Inference in the Threshold Model

Yulong Wang*

Princeton University, Department of Economics

Job Market Paper

Jan. 2017

Abstract

This paper studies inference about the values of the parameters in the threshold model in a generalized method of moments (GMM) framework. First, we establish that the extensively studied least squares method leads to substantially oversized tests and confidence intervals when the coefficient change is not large. Second, by re-ordering the data to recast the threshold model as a structural break problem, we construct tests that control size under a large range of empirically relevant moderate coefficient changes and are approximately efficient in a well-defined sense. Finally, we modify our approach to encompass inference problems in a variety of additional widely studied models. The accuracy of the asymptotic approximations is evaluated by Monte Carlo simulations. The empirical applicability is illustrated through two examples: (i) testing if public debt has a threshold effect on economic growth; and (ii) constructing a confidence interval for the tipping point in the segregation problem studied by Card, Mas, and Rothstein (2008).

Keywords: Statistical Inference; Sample Splitting; Threshold Model; Structural Break; Parameter Instability; Limits of Experiments; Re-ordering

*www.princeton.edu/~yulong and yulong@princeton.edu. I am very grateful to Ulrich K. Müller, Bo E. Honoré, and Mark W. Watson for continuous guidance and support. I also thank Kirill Evdokimov, Michal Kolesár, and participants at the Princeton Econometrics seminar for helpful discussion and comments.

1 Introduction

Empirical studies routinely involve splitting the sample and conducting econometric analysis in different subsamples. When the subsample is selected based on a continuous variable such as firm size, three major questions naturally evolve, namely, where to split the sample, whether the parameters are stable across subsamples, and what the parameter value is in a certain subsample. The answer to these questions has been exclusively studied in the linear regression case but not in other, possibly nonlinear setups.

This paper considers a general method of moments (GMM) setup and constructs tests and confidence intervals with attractive size and power properties for the parameters of interest. Such a GMM framework has not been considered in the threshold model literature. It covers a variety of empirically relevant models related to sample splitting, including, for example, the linear threshold model, the binary choice model with a coefficient change, and the panel data threshold model. To be precise, we study the following GMM threshold model:

$$\begin{aligned} 0 &= E\left[m\left(W_i;\beta\right)\right] && \text{if } q_i > \gamma \\ 0 &= E\left[m\left(W_i;\beta+\delta\right)\right] && \text{otherwise} \end{aligned} \tag{1}$$

where $m(\cdot;\cdot)$ is a known moment function, (W_i, q_i) is the observed data, and (β, δ, γ) is the unknown parameter. The scalar variable q_i is defined as the *threshold variable* that drives a coefficient change when it surpasses the unknown *threshold* γ. The parameter δ denotes the magnitude of the coefficient change and β the post-change level.

Our approach is designed to address two drawbacks of the extensively studied least squares method for inference about (β, δ, γ). First, the least squares method has been developed only in the context of linear regressions. See, for example, Hansen (1997, 1999, 2000a) and Caner and Hansen (2004). This is restrictive, since we may encounter nonlinear models in many situations. Second, the asymptotic analysis relies on the important assumption that the coefficient change across subsamples is sufficiently large. Consequently, the coefficient change can be detected by reasonable tests with probability close to 1. Inference for the presence of the coefficient change then becomes straightforward, and the researcher can perform further analysis by simply treating the estimated threshold as the true cutoff and ignoring the estimation error about γ. However, in many applications, assuming large values of δ is questionable, since the magnitude and even the presence of the change are unknown a priori. Moreover, the least squares method can lead to substantially oversized tests and confidence intervals when δ is not large. This is shown in the present paper by both derivation of the analytical limit behavior of the least squares test and by simulation of its finite sample performance. In particular, the confidence interval for γ based on the least squares estimator may produce finite sample coverage that is far below the nominal level.

In comparison, our tests and confidence intervals have two advantages: they control size under a large range of empirically relevant *moderate* coefficient changes and are approximately optimal in a well-defined sense. The moderate threshold effect captures the fact that both the presence and the location of the coefficient change are uncertain, even in a large sample. Analytically, a moderate coefficient change can be represented by an asymptotic analysis where δ is of the order of magnitude $N^{-1/2}$ in a sample of size N (cf. Elliott and Müller (2007)). Such a moderate δ can be thought of as a local change that leads to non-trivial power when testing whether a coefficient change exists. In this situation, the coefficient change provides only a limited amount of information so that γ cannot be consistently estimated. To avoid estimation of δ and γ, this paper directly constructs tests that control size uniformly over moderate values of δ. As demonstrated by Monte Carlo experiments, these tests provide good finite sample performance for a wide range of plausible magnitudes of change.

The main idea of our approach is to re-order the data according to the threshold variable, q_i, so that the threshold model is transformed into a structural break model. By doing so, we can resort to the methods developed in the structural break literature for inference about β, δ, and γ, which are then defined as the post-break level, the break magnitude and the break date, respectively. See, for example, Quandt (1958, 1960), Nyblom (1989), Andrews (1993), Andrews and Ploberger (1994), Sowell (1996), Bai (1994, 1997, 1999), Bai and Perron (1998), Bai, Lumsdaine, and Stock (1998), Qu and Perron (2007), Elliott and Müller (2007), Chen and Hong (2012), Elliott and Müller (2014), Elliott, Müller, and Watson (2015), and Eo and Morley (2015).

The re-ordering approach faces a serious challenge, however. When the data are sorted by the rank of q_i, one typically obtains a non-stationary structural break model. This is because q_i can be correlated with W_i, leading to time-varying moments with the rank of q_i interpreted as time. The above mentioned structural break literature is all based on an assumption of stationarity, though. One exception is Hansen (2000b), who studies testing whether $\delta = 0$ and proposes a bootstrap test. But this bootstrap idea can neither claim any efficiency in general nor can it be easily applied for inference about β and γ, since these problems are characterized by nuisance parameters under the null hypotheses.

To overcome this key challenge of non-stationarity, we develop a suitable time transformation. Such a time transformation recovers the asymptotic distribution of the stationary case by standardizing the time-varying second moments asymptotically. Our time transformation depends on the variance-covariance structure of the model, which is categorized into three cases. In the first one, which we refer to as the *canonical* case, all variables are stationary, and q_i is independent. Then, re-ordering according to q_i reduces to a random ordering, and the resulting structural break model is thus of a form to which existing methods can be applied directly. In the other two noncanonical cases, q_i can be correlated with other variables, and then stationarity is violated. But after our transformation is performed, the tests relying on stationarity can be applied again.

Given our time transformation, we can extend the statistically efficient tests developed in the canonical case to noncanonical cases, which cover both the threshold model and the structural break model with non-stationary data. In the canonical case, the efficiency of these tests is established in the situation where the moment function is the score function of a parametric model. Since our time transformation is a one-to-one mapping, the efficiency is retained in noncanonical cases.

In addition to the baseline GMM threshold model (1), we generalize our approach to encompass inference problems in additional empirically relevant models, such as the threshold model with a preliminary estimator of an infinite dimensional nuisance parameter, and the threshold model with non-smooth moment conditions. Our approach is illustrated through two widely applied models: the binary choice/single index model with a coefficient change and the static panel data threshold model. If an instrument for lagged dependent variables is available, our approach can also be generalized to cover the dynamic panel data threshold model (cf. Seo and Shin (2014)).

In summary, this paper makes four contributions to the literature. First, we point out the equivalence between the threshold model and the structural break problem by re-ordering the data. Second, we develop methods for inference about (β, δ, γ) with attractive size and power properties in a general GMM framework. Third, we illustrate the invalidity of the least squares inference method in the situation where δ is not large, which has not previously been studied in the threshold literature. Finally, we extend the efficient tests developed for the structural break model with stationary data to allow for data with time-varying second moments and generalize them to additional models of empirical importance.

The remainder of this paper is organized as follows. Section 2 performs asymptotic analysis to reduce the finite sample problem into an asymptotic problem that involves a Gaussian process as the observation. Time transformations then directly involve this Gaussian observation. In the end of this section, we also discuss the properties of the least squares method when the coefficient change is moderate. Section 3 introduces the tests as functions of the observation in the asymptotic problem for β, δ, and γ, respectively, and discusses their efficiency by resorting to Le Cam's limits of experiments theory. Section 4 shows how to implement these tests in finite samples. Section 5 extends the main results to additional models, and Section 6 applies the results to the binary choice model and the panel data model. Section 7 examines the finite sample performance of our approach by several Monte Carlo experiments, and Section 8 reports two empirical applications. Unreported details and all proofs are collected in the appendix.

Throughout the paper, we use the linear regression model as the running example. Let '$\xrightarrow{p}$' denote convergence in probability and '$\Rightarrow$' weak convergence of the underlying probability measures as $N \to \infty$. Let $[r]$ denote the biggest integer smaller than r and $\mathbf{1}[A]$ the indicator function of a generic event A. Let $||B||$ denote the Euclidean norm of a vector or matrix B, and let C denote a generic constant.

2 Asymptotic Analysis

This section performs the asymptotic analysis and derives the main convergence result. Specifically, the partial sum of the re-ordered moment functions leads to a Gaussian process in the limit, which is useful in constructing the tests that are introduced later.

Section 2.1 sets up the problem, and Section 2.2 investigates the canonical case. In Section 2.3, we analyze the noncanonical cases where the limiting Gaussian process has a more complicated form. In Section 2.4, we show that, by constructing a time transformation, we can recover the same experiment as in the canonical case. Finally, in Section 2.5, we analyze the asymptotic properties of the least squares test in the linear regression context.

2.1 Setup

Let $m(W_i;\beta)$ be some known $\mathbb{R}^k$-valued moment function indexed by some unknown parameter β, so that $E[m(W_i;\beta_0)] = 0$ when the true parameter for the i-th observation is given by $\beta = \beta_0$. In the following three sections, we study the exact identification case, i.e., $\dim(m) = \dim(\beta) = k$. The over-identification case is postponed to Section 5.1. Write $m_i(\beta)$ for $m(W_i;\beta)$.

We refer to the model with some benchmark value $\bar{\beta}$ as the 'stable' model, where no coefficient change exists. In the unstable model, we assume that for the i-th observation, the parameter β evolves as

$$\beta_i = \bar{\beta} + N^{-1/2}Mb + N^{-1/2}Md\mathbf{1}\left[q_i \leq \gamma\right] \tag{2}$$

where $b, d \in \mathbb{R}^k$ and M is some generic normalizing matrix to be specified later. Such a specification can be treated as a re-parameterization, which captures a local deviation, b, from the stable benchmark value $\bar{\beta}$ and a moderate threshold effect, d. We assume that $\bar{\beta}$ is known in this and the next sections for simplicity of illustration. This is without loss of generality because first, if we are testing about β, $\bar{\beta}$ is then the value under the null hypothesis. Hence, the problem is equivalent to testing the null $b = 0$ with a known $\bar{\beta}$, and $N^{-1/2}Mb$ can be thought as a local alternative. Second, if we consider inference about d or γ, while treating $\bar{\beta}$ as known is unrealistic, such knowledge will not compromise the validity of our approach as we show later, and the tests introduced in the next section can be implemented by replacing $\bar{\beta}$ with some $\sqrt{N}$-consistent estimator whose estimation error is $N^{-1/2}Mb$.

Given the above re-parameterization (2), the main focus of this paper is to construct tests with good finite sample performance for the following three scalar testing problems:

$$H_0 \ : \ \gamma = \gamma_0 \qquad \text{against} \qquad H_1 \ : \ \gamma \neq \gamma_0 \tag{3}$$

$$H_0 \ : \ d_1 = 0 \qquad \text{against} \qquad H_1 \ : \ d_1 \neq 0 \tag{4}$$

$$H_0 \ : \ b_1 = 0 \qquad \text{against} \qquad H_1 \ : \ b_1 \neq 0 \tag{5}$$

where d_1 and b_1 are the first components of d and b, respectively. Notice that the above problems are nonstandard in the sense that they involve nuisance parameters under both the null and the alternative hypotheses. For example, as for problem (3), the whole vectors b and d are nuisance parameters. We will simplify these problems by focusing on a specific combination of the moment functions only and imposing some invariance restriction, as shown in Sections 2 and 3.

Our method also applies to inference about $v'b$ and $v'd$ for any non-zero $k \times 1$ known vector v. Without loss of generality, we focus on $v = e_1$, the first column of the $k \times k$ identity matrix, I_k. The pre-break parameter $b + d$ can be easily covered by flipping the sign of q_i.

Let $Q(\cdot)$ denote the quantile function of the threshold variable q_i and $\hat{Q}(\cdot)$ the corresponding empirical quantile. We assume that q_i is i.i.d. and $Q(\cdot)$ is continuous; so any γ that is not on the boundary of the support of q_i corresponds to a certain quantile, $Q(r)$, for some $0 < r < 1$. Since $Q(\cdot)$ is continuous and strictly increasing, r is uniquely defined by $Q^{-1}(\gamma)$, and problem (3) can be rewritten as

$$H_0 \;:\; r = r_0 \qquad \text{against} \qquad H_1 \;:\; r \neq r_0. \tag{6}$$

It is then equivalent (at least asymptotically) to conduct inference about r instead of γ even if $Q(\cdot)$ is unknown because it can be consistently estimated by $\hat{Q}(\cdot)$. See Lemma E.2 in the appendix for details.

By rearranging the data according to the order of q_i (cf. Tsay (1989, 1998)), we re-order the moment functions evaluated at $\bar{\beta}$ as $m_{(i)}(\bar{\beta})$, where the order of $m_{(i)}(\bar{\beta})$ is induced by the order statistics $q_{(1)} \leq q_{(2)} \leq \ldots \leq q_{(N)}$, i.e., $m_{(i)}(\bar{\beta}) = m_j(\bar{\beta})$ if $q_{(i)} = q_j$. Such a representation allows us to suppress the marginal distribution of q_i asymptotically by considering the indicator function in (2) as $\mathbf{1}[i/N \leq r]$. We trim the boundary thresholds and restrict r to be in $[\underline{r}, \bar{r}]$ for some $0 < \underline{r} < \bar{r} < 1$. Such a restriction is standard in the structural break literature, see, for example, Andrews and Ploberger (1994). For simplicity, we refer to r as the threshold and the rank of q_i as time when we analyze the re-ordered series.

Let $W(\cdot)$ be a k-dimensional standard Wiener process on $[0,1]$. For the asymptotic results, we impose the following regularity condition:

Condition 1

1. $\{W_i, q_i\}$ *is independent over* i.
2. q_i *is identically distributed over* i *with a continuous density function* f_q *such that for all* s, $0 < f_q(s) < C$ *for some* $C < \infty$.
3. *The true threshold* γ_0 *is such that* $\gamma_0 = Q(r_0)$ *for some* $\underline{r} < r_0 < \bar{r}$.
4. *For all* i, $m_i(\beta)$ *is twice continuously differentiable w.r.t.* β *with* $\Lambda_i(\beta) = -\frac{\partial m_i(\beta)}{\partial \beta'}$.
5. *For all* i *and uniformly in* $s \in [0,1]$, $E[\Lambda_i(\beta_i) | q_i = Q(s)]$ *and* $E\left[m_i(\beta_i) m_i(\beta_i)' | q_i = Q(s)\right]$ *are bounded*[1] *and satisfy*

$$N^{-1} \sum_{i=1}^{N} E[\Lambda_i(\beta_i) | q_i = Q(s)] \;\; \rightarrow \;\; \Gamma(s)$$

[1] These conditional expectations are taken with respect to the conditional density $f_{W_i|q_i=Q(s)}(\cdot) \equiv f_{W_i,q_i}(\cdot, Q(s)) / f_q(Q(s))$ where $f_{W_i,q_i}(\cdot,\cdot)$ denotes the joint density of W_i and q_i. The existence of this conditional density is implied by Theorem 10.2.2 of Dudley (2002).

$$N^{-1} \sum_{i=1}^{N} E\left[m_i\left(\beta_i\right) m_i\left(\beta_i\right)' | q_i = Q\left(s\right)\right] \quad \rightarrow \quad H\left(s\right)$$

where $\Gamma(\cdot)$ and $H(\cdot)$ are uniformly bounded matrix-valued functions defined on $[0,1]$ that are piece-wise twice continuously differentiable with at most a finite number of discontinuities, and left and right limits everywhere.

6. *For all i and uniformly in β on some open ball centered at $\bar{\beta}$, $E\left[||m_i(\beta)||^4\right] < C$, $E\left[||\Lambda_i(\beta)||^4\right] < C$, and $E\left[\left|\left|\frac{\partial \Lambda_i(\beta)}{\partial \beta}\right|\right|^2\right] < C$ for some $C < \infty$.*

7. *$E\left[m_i\left(\beta_i\right) | q_i\right] = 0$ for all i.*

Condition 1.1 is usually satisfied in cross-sectional data. Time series with weak dependence are discussed in Section 5.4. Condition 1.2 implies that the quantile function of q_i is continuous and uniquely defined for all i. Condition 1.3 imposes the constraint that the number of observations in both subsamples are in fixed proportions. Condition 1.4 requires $m_i(\cdot)$ to be smooth. Violation of this condition is discussed in Section 5.3 where the estimators with non-smooth moment functions are covered, for example, the least absolute deviation (LAD) estimator and the quantile estimator. Condition 1.5 defines $\Gamma(\cdot)$ and $H(\cdot)$. Such a definition is unique up to a set with Lebesgue measure zero. When $\{W_i, q_i\}$ is i.i.d., $\Gamma(s)$ and $H(s)$ reduce to $E\left[\Lambda_i\left(\beta_i\right) | q_i = Q(s)\right]$ and $E\left[m_i\left(\beta_i\right) m_i\left(\beta_i\right)' | q_i = Q(s)\right]$, respectively. When we allow series correlation, we need another term in addition to $H(\cdot)$ to capture the covariance. This is discussed in Section 5.4. The uniform boundedness at the boundary can be relaxed by trimming. See Appendix A for details. Condition 1.6 bounds the relevant moments in a neighborhood of $\bar{\beta}$. Condition 1.7 is a key assumption that the moment condition holds conditional on the threshold variable. Given Condition 1, we can derive the following result:

Lemma 1 *Suppose Condition 1 holds. Then in the unstable model (2),*

$$N^{-1/2} \sum_{i=1}^{[sN]} m_{(i)}\left(\beta_{(i)}\right) \Rightarrow \int_0^s H^{1/2}(l) dW(l)$$

for $s \in [0,1]$ and

$$\sup_{s \in [0,1]} \sup_{\beta \in B_{\varepsilon_N}(\bar{\beta})} \left|\left| N^{-1} \sum_{i=1}^{[sN]} \Lambda_{(i)}(\beta) - \int_0^s \Gamma(l) dl \right|\right| \xrightarrow{p} 0$$

for any $\varepsilon_N \rightarrow 0$ as $N \rightarrow \infty$ where $B_{\varepsilon_N}\left(\bar{\beta}\right)$ is the open ball centered at $\bar{\beta}$ with radius ε_N.

Table 1: Three Cases of the Variance-Covariance

Case	$E\left[\Lambda_{(i)}\left(\beta_{(i)}\right)\right]$	$E\left[m_{(i)}\left(\beta_{(i)}\right)m_{(i)}\left(\beta_{(i)}\right)'\right]$
Canonical	$\bar{\Gamma}$	$\bar{\Gamma}\bar{\sigma}^2$
Noncanonical I	$\bar{\Gamma}$	$\bar{\Gamma}\sigma^2\left(i/N\right)$
Noncanonical II	$\Gamma\left(i/N\right)$	$H\left(i/N\right)$

Lemma 1 lays the foundation of our analysis. The special case where $\Gamma(\cdot)$ and $H(\cdot)$ are constant matrices is usually treated as a high-level time series condition in the structural break literature (cf. Sowell (1996) and Li and Müller (2009)). In this paper, we prove it as a general result and the method developed later naturally applies to the time series structural break model with non-stationary data.

Running Example: Throughout the paper, we use the linear threshold model as the leading example for illustration. The linear threshold model can be written as

$$y_i = x_i'\left(\bar{\beta} + bN^{-1/2} + dN^{-1/2}\mathbf{1}\left[q_i \leq \gamma\right]\right) + u_i.$$

Suppose $Q(\cdot)$ is known. Then, re-ordering the data according to q_i leads to the structural break model

$$y_{(i)} = x_{(i)}'\left(\bar{\beta} + bN^{-1/2} + dN^{-1/2}\mathbf{1}\left[i/N \leq r\right]\right) + u_{(i)}. \tag{7}$$

We also assume that (u_i, x_i, q_i) is i.i.d., and then we have $\Gamma(l) = E\left[x_i x_i' | q_i = Q(l)\right]$ and $H(l) = E\left[x_i x_i' u_i^2 | q_i = Q(l)\right].$

Our analysis is performed in three scenarios based on the level of the complexity of $\Gamma(\cdot)$ and $H(\cdot)$, as described in Table 1. The first case is referred to as the canonical one, as it is widely studied in the structural break literature, including, for instance, Andrews (1993), Andrews and Ploberger (1994), Elliott and Müller (2014), and Elliott, Müller, and Watson (2015). It arises in the threshold context under the condition that the data are i.i.d. with constant variance of the error term. The noncanonical I case covers heteroskedasticity by allowing the variance of $m_i(\beta_i)$ to depend on q_i but still imposes the stationarity of $\Lambda_{(i)}\left(\beta_{(i)}\right)$. In noncanonical II, we allow for time-varying moments of $\Lambda_{(i)}\left(\beta_{(i)}\right)$ and $m_{(i)}\left(\beta_{(i)}\right)$.

In this and the next sections where we study the asymptotic problems, it is assumed that $\Gamma(\cdot)$, and $H(\cdot)$ are known. The estimation in finite samples is discussed in Section 4.

2.2 Canonical Case

Our test statistics are built on the partial sum of the re-ordered moment functions, which leads to the Gaussian process in the asymptotic problem. Define the function $a:[0,1]\to\mathbb{R}^k$ as $a(\eta)=m_{(i)}(\bar{\beta}+\eta\left(\beta_{(i)}-\bar{\beta}\right))$, and then $a'(\eta)=-\Lambda_{(i)}\left(\bar{\beta}+\eta\left(\beta_{(i)}-\bar{\beta}\right)\right)\left(\beta_{(i)}-\bar{\beta}\right)$. Note that

$$
\begin{aligned}
m_{(i)}\left(\beta_{(i)}\right)-m_{(i)}\left(\bar{\beta}\right) &= a(1)-a(0)\\
&= -\int_0^1 \Lambda_{(i)}\left(\bar{\beta}+\eta\left(\beta_{(i)}-\bar{\beta}\right)\right)d\eta\left(\beta_{(i)}-\bar{\beta}\right).
\end{aligned}
$$

Then, in the canonical case where $\Gamma(\cdot)=\bar{\Gamma}$ and $H(\cdot)=\bar{H}$, we have

$$
\begin{aligned}
&\bar{H}^{-1/2}N^{-1/2}\sum_{i=1}^{[sN]}m_{(i)}(\bar{\beta})\\
=\;&\bar{H}^{-1/2}N^{-1/2}\sum_{i=1}^{[sN]}m_{(i)}\left(\beta_{(i)}\right)+\bar{H}^{-1/2}\int_0^1\left(N^{-1}\sum_{i=1}^{[sN]}\Lambda_{(i)}\left(\bar{\beta}+\eta\left(\beta_{(i)}-\bar{\beta}\right)\right)Mb\right)d\eta\\
&+\bar{H}^{-1/2}\int_0^1\left(N^{-1}\sum_{i=1}^{[sN]}\Lambda_{(i)}\left(\bar{\beta}+\eta\left(\beta_{(i)}-\bar{\beta}\right)\right)Md\mathbf{1}\left[i/N\le r\right]\right)d\eta+o_p(1).
\end{aligned}
$$

By the continuous mapping theorem and Lemma 1 and setting $M=\bar{\Gamma}^{-1}\bar{H}^{1/2}$, we have

$$
\bar{H}^{-1/2}N^{-1/2}\sum_{i=1}^{[sN]}m_{(i)}(\bar{\beta})\Rightarrow G(s)\equiv W(s)+sb+\min(s,r)\,d. \tag{8}
$$

Running Example-cont'd: In the linear threshold model, the moment function is simply $m_i(\beta)=x_i\left(y_i-x_i'\beta\right)$. In the unstable model with $\bar{\beta}=0$ and $M=\bar{H}=\bar{\Gamma}=I_k$, we have

$$
\begin{aligned}
N^{-1/2}\sum_{i=1}^{[sN]}x_{(i)}y_{(i)} &= N^{-1/2}\sum_{i=1}^{[sN]}x_{(i)}\left(u_{(i)}+x_{(i)}'\left(b+d\mathbf{1}\left[i/N\le r\right]\right)N^{-1/2}\right)+o_p(1)\\
&\Rightarrow G(s).
\end{aligned}
$$

For problems (4), (5), and (6), it is convenient to consider the first element of G, that is,

$$
G_1(s) = e_1'G(s)
$$

$$= \quad W_1(s) + sb_1 + \min(s,r)\, d_1 \tag{9}$$

where $W_1 = e_1' W$ is a scalar Wiener process. Our method might suffer some loss of efficiency due to ignoring the last $k-1$ elements of the vector process G. But tests that are functions of G_1 are robust in the sense of providing reliable inference about (r, d_1, b_1) regardless of the other components of d and b. We provide more details in Section 3.4 where the efficiency of these tests is discussed.

By standard arguments (see, for example, Chapter 7 of Liptser and Shiryaev (2013)), the log of the Radon-Nikodym derivative of the distribution of G_1 relative to the distribution of $W_1(\cdot)$, evaluated at G_1, is given by

$$\log f(G_1; r, d_1, b_1) = b_1 G_1(1) + d_1 G_1(r) - \frac{1}{2}\left(b_1^2 + d_1^2 r + 2b_1 d_1 r\right). \tag{10}$$

The observation G_1 is extensively studied in the structural break literature. Given its density (10), we can construct Neyman-Pearson type tests φ_r, φ_b and φ_d for inference about r, b_1, and d_1, respectively. They all rely on G_1 and are constructed to be statistically efficient. Before introducing these tests, we first discuss the noncanonical cases where the limit of the partial sum of the moment functions is more complicated than G_1 (and G). The properties of the tests φ_r, φ_b and φ_d are postponed to the next section.

2.3 Noncanonical Cases

The limiting Gaussian process in the noncanonical cases has a more complicated form. We first relax the constancy of $H(\cdot)$ by assuming $H(\cdot) = \bar{\Gamma}\sigma^2(\cdot)$ where $0 < \varepsilon \leq \sigma(s) \leq C < \infty$ for $s \in [0,1]$ and some $\varepsilon, C > 0$. This is important in the classic structural break problem, as it captures the time-varying variance of the error term. For $s \in [0,1]$, define $\tilde{h}(s) = \int_0^s \sigma^2(l)\, dl$ and $\tilde{g}(s) = \tilde{h}(s)/\tilde{h}(1)$. After normalizing the coefficients by setting $M = \tilde{h}(1)^{1/2}\bar{\Gamma}^{-1/2}$, one can show that, by an argument similar to that in (8),

$$\tilde{h}(1)^{-1/2}\bar{\Gamma}^{-1/2} N^{-1/2} \sum_{i=1}^{[sN]} m_{(i)}(\bar{\beta}) \Rightarrow W(\tilde{g}(s)) + sb + \min(s,r)\, d.$$

Although the non-constancy of the moments is fully captured by $\tilde{g}(\cdot)$, we cannot simply treat $\tilde{g}(\cdot)$ as time to obtain G_1, because doing so will lead to a nonlinear drift term. One possible solution is to apply a GLS-type transformation, that is, to consider the partial sum of $\sigma^{-2}(i/n)\, m_{(i)}(\bar{\beta})$. However, such a transformation is insufficient to obtain G_1 in the more general noncanonical II case.

Consider the noncanonical II case, where we allow both $\Gamma(\cdot)$ and $H(\cdot)$ to be nonlinear matrix-valued functions. Then, Lemma 1 implies that

$$N^{-1/2}\sum_{i=1}^{[sN]} m_{(i)}(\bar{\beta}) \Rightarrow \tilde{G}(s) \tag{11}$$
$$\equiv \int_0^s H^{1/2}(s)\,dW(s) + \left(\int_0^s \Gamma(l)\,dl\right) Mb + \left(\int_0^{\min(s,r)} \Gamma(l)\,dl\right) Md.$$

Non-constant $H(\cdot)$ and $\Gamma(\cdot)$ thus induce the more complicated limiting process, $\tilde{G}$. The stochastic term is not a Wiener process and the drift process is not linear with a kink as in (8). It follows that a single GLS-type transformation cannot retain G_1. But since only the parameters involved in the first coefficient change are of interest, we next construct a more sophisticated transformation and apply it to $\tilde{G}$ whose first element can then be transformed back to G_1.

2.4 Transformation in the Asymptotic Problem

The above three cases demonstrate that the complexity of the asymptotic problem stems from the time-varying moments. They can be standardized if the non-constancy is fully characterized by a scalar time transformation. If the expected Jacobian is also time-varying, a single time transformation is insufficient to simplify the whole vector $\tilde{G}$ due to a lack of degrees of freedom. Alternatively, we could consider the first component of $\tilde{G}$ only and resort to the time transformation introduced in the noncanonical I case. This idea needs more subtle care on the deterministic term $\int_0^s \Gamma(l)\,dl + \int_0^{\min(s,r)} \Gamma(l)\,dl$, which varies across time in a different manner than the stochastic term. We make the following additional assumption:

Condition 2

1. *The matrix $\Gamma(l)$ is invertible for all $l \in [0,1]$.*

2. *For all $l \in [0,1]$, $e_1'\Gamma(l)^{-1} H(l)\Gamma(l)^{-1\prime} e_1 > 0$, $||\Gamma(l)||^2 \leq C$, and $\left|\left|\Gamma(l)^{-1}\right|\right|^2 \leq C$ for some positive constant C.*

Condition 2.1 requires the invertibility of the expectation of the Jacobian of the moment function. Condition 2.2 requires some boundedness of the spot variance-covariance matrices.

Running Example-cont'd: In the linear threshold case with i.i.d. data, Condition 2.2 requires that $E\left[x_i x_i' | q_i = Q(l)\right]$ be invertible for all $l \in [0,1]$. Such an invertibility excludes the situation where q_i is a linear combination of x_i. To see this, suppose x_i includes a constant and $x_{1i} = q_i$. This corresponds to the situation where the regressor with a change in coefficient is equal to the threshold variable. For any $r_1 \in (0,1)$, since $q_{([r_1 N])}$ is an empirical quantile and converges to the true quantile, $Q(r_1)$, then the first column of $\Gamma(r_1)$ is equivalent to $Q(r_1) E\left[x_i | q_i = Q(r_1)\right]$, which is proportional to the column associated with the constant, $E\left[x_i | q_i = Q(r_1)\right]$. See Romano and Wolf (2001) for inference about β based on subsampling in this situation.

Under Condition 2, we can define two continuous and strictly increasing functions,

$$
\begin{aligned}
h(l) &= \int_0^l \frac{1}{e_1' \Gamma(s)^{-1} H(s) \Gamma(s)^{-1\prime} e_1} ds \quad \text{and} \\
g(l) &= h(l) / h(1).
\end{aligned} \tag{12}
$$

Since $g(0) = 0$ and $g(1) = 1$, we can treat $g(\cdot)$ as a transformed time. Then, we define our transformation mapping $\mathcal{T} : C^k_{[0,1]} \to C^k_{[0,1]}$ such that for any $\bar{G} \in C^k_{[0,1]}$,

$$
\mathcal{T}\bar{G}(s) = \sqrt{h(1)} \int_0^{g^{-1}(s)} \frac{\partial g(l)}{\partial l} \Gamma^{-1}(l) \, d\bar{G}(l) \tag{13}
$$

where $\partial g(l) / \partial l = \left(e_1' \Gamma(l)^{-1} H(l) \Gamma(l)^{-1\prime} e_1 h(1)\right)^{-1}$. Note that this mapping is one-to-one, since $g(\cdot)$ is strictly increasing. The following lemma establishes the usefulness of this transformation.

Theorem 1 *Set the normalization matrix M as $h(1)^{-1/2} I_k$. For any $s \in [0,1]$, we have*

$$
e_1' \mathcal{T}\tilde{G}(s) = W_1(s) + b_1 s + d_1 \min(s, g(r)).
$$

With such a transformation, the first component of $\mathcal{T}\tilde{G}$ is exactly G_1 as in (9) with r transformed into $g(r)$. Then, we may consider $e_1' \mathcal{T}\tilde{G}$ as the observation for the tests relying on G_1.

The intuition for constructing $\mathcal{T}$ can be explained as follows. First, we standardize the non-constant variance-covariance matrix $\Gamma(\cdot)$ by pre-multiplying with $\Gamma^{-1}(\cdot)$. This turns the nonlinear function $\int_0^{\cdot} \Gamma(l) \, dl$ into a linear one. Second, to standardize the spot variance, we set the derivative of $g(\cdot)$ as the weighting function that is proportional to the 1,1 element of the inverse local Fisher information w.r.t. β, that is, $e_1'\Gamma(l)^{-1} H(l)\Gamma(l)^{-1\prime} e_1$. Finally, because $\int_0^{g^{-1}(s)} (\partial g(l)/\partial l)\, dl = s$, we transform the stochastic integral back to a standard Wiener process, and the deterministic term to a linear function with a kink.

In summary, we have studied the asymptotic problem in three situations defined by the different variance-covariance matrices of m_i and Λ_i. The canonical case is simple and widely explored in the literature, and the other two are complicated and rarely investigated. Before introducing the asymptotically efficient tests that are functions of G_1, we next analyze the properties of the least squares test for inference about the threshold.

2.5 Properties of the Least Squares Test under a Moderate Threshold Effect

This subsection discusses the least squares method for the testing problem (6). We derive the asymptotic distribution of the least squares test statistics and show that it leads to substantially over-sized tests and confidence intervals when δ is moderate, i.e., $\delta = MdN^{-1/2}$. Such a moderate threshold effect has not been considered in the threshold model literature. This case is of significant empirical importance for two reasons. First, after all, we do not know how large the coefficient change is a priori. Second, a moderate δ only means the change is not large in a statistical sense but not necessarily small in an economic sense. The local parameter d measures the magnitude of the coefficient change in multiples of the standard error (reflected by M). A reasonable d, say 4, could imply a non-negligible change in the data. See Elliott and Müller (2007) for structural break examples as well as Section 7.2 for threshold model examples.

In the running example where $y_i = x_i'(\beta + \delta \mathbf{1}[q_i \leq Q(r)]) + u_i$, the least squares method proceeds as follows: minimize the sum of the squared residuals of the linear regression (7) over all coefficients and thresholds that are not too close to the boundary. Denote the least squares estimators of the threshold by $\hat{r}$. Let $S_N(r)$ be the sum of squared residuals of the regression (7) given r. The least squares test of (6) is then constructed as, for some critical value ξ,

$$\begin{aligned} \varphi_{LS}(r_0) &= \mathbf{1}[R_N(r_0) > \xi] \\ \text{where } R_N(r) &= N\frac{S_N(r) - S_N(\hat{r})}{S_N(\hat{r})}. \end{aligned} \tag{14}$$

To derive the asymptotic distribution of R_N, Hansen (2000a) imposes the condition that $\delta = dN^{-a}$ for some $0 < a < 1/2$. This condition dates back to Picard (1985) and is also applied by Yao (1987), Dümbgen (1997), Bai (1997, 1999), Bai and Perron (1998), Bai, Lumsdaine, and Stock (1998), Qu and Perron (2007), and Eo and Morley (2015). As pointed out by Elliott and Müller (2007), however, this assumption might not provide good approximations in finite samples. When the break magnitude is large (corresponding to $a = 0$), Chan (1993) establishes that $\hat{r}$ converges at the rate N. The limit distribution of $\hat{r}$ can hardly be employed for inference about r since it depends on unknown parameters and the distribution of u_i. When the break magnitude is moderate (corresponding to $a = 1/2$), both the presence and the location of the threshold are uncertain, even asymptotically. In this situation, the estimator $\hat{r}$ is no longer consistent, and the test (14) does not control size. As shown in Table 1 of Hansen (2000a), the test (14) can overreject the null that $r = r_0$ substantially when δ is small.

To illustrate the effect of a moderate threshold, we establish the asymptotic distribution of R_N under $\delta = N^{-1/2}d$ in the canonical case, as summarized in the following lemma.

Lemma 2 *For any $s \in (0,1)$ and $r_0 \in (\underline{r}, \bar{r})$, define*

$$\begin{aligned} U(s) &= \bar{H}^{1/2} W(s) + \min(s, r_0)\, \bar{\Gamma} d, \\ V(s) &= U(s)' \bar{\Gamma}^{-1} U(s)/s + (U(1) - U(s))' \bar{\Gamma}^{-1} (U(1) - U(s))/(1-s). \end{aligned}$$

Then under Condition 1, when $V(s)$ has a unique maximum with probability 1 on $[\underline{r}, \bar{r}]$, $\Gamma(\cdot) = \bar{\Gamma}$, $H(\cdot) = \bar{H}$ and $\delta = N^{-1/2}d$,

$$\hat{r} \Rightarrow r_a$$

where $r_a = \arg\max_{\underline{r} \le s \le \bar{r}} V(s)$. Moreover, when $N^{-1}\sum_{i=1}^{N} u_i^2 \xrightarrow{p} \sigma_u^2$ for some $\sigma_u > 0$,

$$R_N(r_0) \Rightarrow \sigma_u^{-2}(V(r_a) - V(r_0)).$$

Lemma 2 is analogous to Proposition 1 of Elliott and Müller (2007). When δ is of the order of magnitude $N^{-1/2}$, the least squares estimator is no longer consistent, and the likelihood ratio statistic R_N does not converge to the limit tabulated in Hansen (2000a). When $\Gamma(\cdot)$ and $H(\cdot)$ are not constant matrices, the asymptotic distribution of R_N depends on r_0, Γ, and H in a complicated way.

The finite sample performance of the test (14) requires numerical evaluation. Table II of Hansen (2000a) depicts the finite sample rejection probabilities of (14) with data generated from the following simple experiment: $x_i = (x_{1i}, 1)$ where x_{1i}, q_i, and u_i are jointly i.i.d. standard normal, $\delta = (d_1 N^{-1/2}, 0)$ and $r_0 = 0.5$. For $d_1 = 5.6$ (which renders $\delta_1 = 0.25$ with sample size 500), the rejection probability is around 20% for the nominal 10% level.

Table 2: Asymptotic properties of the least squares test

Panel A: Asymptotic Rejection Probabilities				
d_1	$r_0 = 0.4$		$r_0 = 0.2$	
	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 10\%$
4	0.085	0.160	0.081	0.152
8	0.068	0.129	0.052	0.104
12	0.054	0.101	0.043	0.090
16	0.047	0.084	0.041	0.089
Panel B: Percentiles of r_a				
	5% percentile	95% percentile	5% percentile	95% percentile
4	0.169	0.826	0.156	0.831

Note: Data are generated from $U(\cdot)$ and $V(\cdot)$ defined in Lemma 2 with $\bar{H} = \bar{\Gamma} = I_2$. Significance level is α. Continuous processes are approximated by 2,000 steps. Based on 5,000 replications.

Panel A of Table 8 here provides the asymptotic rejection probabilities of the test (14) with data generated from $U(\cdot)$ and $V(\cdot)$ with $\bar{H} = \bar{\Gamma} = I_2$ and $[\underline{r}, \bar{r}] = [0.15, 0.85]$. It is clear that under a moderate threshold effect, say, $d_1 = 4$, the test (14) overrejects asymptotically. See also Table 7 in Section 7 for finite sample rejection probabilities.

Panel B of Table 2 lists the 5% and 95% percentiles of r_a, which are substantially different from r_0. This suggests that the least squares confidence interval obtained by inverting (14) may be centered around the boundary when the break magnitude is moderate.

3 Asymptotically Efficient Tests

The previous section performs the asymptotic analysis and constructs the time transformation that recovers the canonical observation from its noncanonical counterpart. In principle, we can choose any tests that are functions of the canonical limiting Gaussian process for problems (4), (5), and (6). In this section, we introduce the statistically efficient one for each of them. The efficiency is established by resorting to the limits of experiments theory pioneered by Le Cam in the situation where the moment function is the score function of some parametric model.

3.1 Review of the Efficient Tests Developed in the Canonical Case

This subsection reviews the efficient tests for inference about d, b, and r, which are developed in the canonical case by Andrews and Ploberger (1994), Elliott and Müller (2014), and Elliott, Müller, and Watson (2015), respectively. For conciseness, we present the size and power properties of these tests only and leave their expressions and more details to Appendix B.

By treating G_1 as the only observation in the limiting problem, problem (5) is simplified as

$$H_0 : b_1 = 0, d_1 \in \mathbb{R}, r \in (\underline{r}, \bar{r}) \quad \text{against} \quad H_1 : b_1 \neq 0, d_1 \in \mathbb{R}, r \in (\underline{r}, \bar{r}) \, .$$

The efficient test φ_b, as a function of G_1, is constructed as a generalized likelihood ratio-type test, which utilizes the knowledge that the density of G_1 is $f(G_1; r, b_1, d_1)$ as specified in (10). It controls size uniformly regardless the value of $d_1 \in \mathbb{R}$ and $r \in (\underline{r}, \bar{r})$.

For inference about r and d_1, the problems (4) and (6) with the observation G_1 has a natural invariance structure and hence it makes sense to impose the corresponding invariance. Consider the group of transformations

$$G_1(s) \to G_1(s) + b_1 s, \;\; s \in [0,1] \;\text{ and } b_1 \in \mathbb{R}. \tag{15}$$

It can be easily shown that

$$\begin{aligned} G_1^*(s) &\equiv G_1(s) - sG_1(1) \\ &= W_1(s) - sW_1(1) + d_1(\min(s,r) - sr) \end{aligned} \tag{16}$$

is a maximal invariant to (15), and hence by Theorem 6.2.1 of Lehmann and Romano (2005), invariant tests to (15) can be constructed as functions of $G_1^*(\cdot)$ only. The testing problems (4) and (6) are then simplified as

$$\begin{array}{lll} H_0 : r = r_0, d_1 \in \mathbb{R} & \text{against} & H_1 : r \neq r_0, d_1 \in \mathbb{R} \\ H_0 : d_1 = 0 & \text{against} & H_1 : d_1 \neq 0, r \in (\underline{r}, \bar{r}) \, . \end{array}$$

By standard arguments, the Radon-Nikodym derivative of the distribution of G_1^* relative to the distribution of a standard one-dimensional Brownian bridge, $W_1(s) - sW_1(1)$, evaluated at G_1^*, is given by

$$f^*(G_1^*; r, d_1) = \exp\left(d_1 G_1^*(r) - \frac{d_1^2}{2} r(1-r)\right). \tag{17}$$

Then, the efficient tests φ_r and φ_d, as functions of G_1^*, are constructed as generalized likelihood ratio-type tests, which rely on the knowledge that the above expression is the density of G_1^*. In particular, the test φ_r controls size uniformly for d_1 in Δ, some compact subset of $\mathbb{R}$ centered at 0.[2] See Appendix B for more details.

The reason we pick these tests is that they are asymptotically efficient in the sense that they maximize some weighted average power (WAP), as specified in the original papers. In particular, the test φ_r maximizes the WAP specified in Elliott, Müller, and Watson (2015) among all tests that control size uniformly over d_1; the test φ_d maximizes the WAP introduced in Andrews and Ploberger (1994); and the test φ_d maximizes the WAP specified in Elliott and Müller (2014) among all tests that control size uniformly over $d_1 \in \mathbb{R}$ and $r \in [\underline{r}, \bar{r}]$.

The tests φ_r, φ_d, and φ_b are all functions of G_1 or G_1^*, and hence are directly applicable in the canonical case. In noncanonical cases where $\tilde{G}$ is the limiting Gaussian process, we can apply the transformation (13) to $\tilde{G}^*(s) \equiv \tilde{G}(s) - s\tilde{G}(1)$ to obtain

$$e_1'\mathcal{T}\tilde{G}^*(s) = W_1(s) - sW_1(1) + d_1(\min(s, g(r)) - sg(r)).$$

Then, these tests are again applicable with G_1^* and G_1 replaced with $e_1'\mathcal{T}\tilde{G}^*$ and $e_1'\mathcal{T}\tilde{G}$, respectively. Note that our transformation is not limited to these tests, but can be applied to any test designed in the canonical case.

3.2 Asymptotic Efficiency

In addition to satisfying the size constraint, the tests φ_r, φ_d, and φ_b are statistically efficient in a well-defined sense when $m_i(\cdot)$ is the normalized score function. In this subsection, we discuss the efficiency properties of these tests by applying Le Cam's limits of experiments theory.

Suppose the data $(W_1, ..., W_N)$ are generated conditional on $\{q_1, ...q_N\}$ from the density $\prod_{i=1}^N f_{W_i}(W_i; \beta_i)$ relative to some σ-finite measure μ_N when the parameter β takes the value β_i for the i-th observation. Define

$$\Gamma(l) \quad = \quad \lim_{N\to\infty} N^{-1} \sum_{i=1}^{N} E\left[\frac{\partial}{\partial\beta} \log f_{W_i}(W_i; \beta) \,|q_i = Q(l)\right]$$

[2]This set is $[-20, 20]$, which is large enough for empirical studies with a reasonable sample size, say, $N = 500$, as it covers the coefficient change that is less than 20 times the standard error. If the change size is even larger, which is not quite empirically relevant, then the threshold can be estimated so sharply that we can simply ignore the uncertainty and treat the estimated threshold as the true value.

$$H(l) = \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N} E\left[-\frac{\partial^2}{\partial\beta\partial\beta'}\log f_{W_i}(W_i;\beta)\,|q_i = Q(l)\right]$$

whose validity is implied by Condition 1.5, and assume the moment function is the normalized score function, that is,

$$m_i(\beta) = H(R_i/N)\,\Gamma(R_i/N)^{-1}\frac{\partial}{\partial\beta}\log f_{W_i}(W_i;\beta) \tag{18}$$

where R_i is the rank of q_i. The score function is the essential element while $\Gamma(\cdot)$ and $H(\cdot)$ are for normalization purposes only. After re-ordering according to q_i, we obtain

$$\begin{aligned} m_{(i)}(\beta) &= H(i/N)\,\Gamma(i/N)^{-1}\frac{\partial}{\partial\beta}\log f_{W_{(i)}}\left(W_{(i)};\beta\right) \\ \Lambda_{(i)}(\beta) &= -H(i/N)\,\Gamma(i/N)^{-1}\frac{\partial^2}{\partial\beta\partial\beta'}\log f_{W_{(i)}}\left(W_{(i)};\beta\right) \end{aligned}$$

where $\{\log f_{W_{(i)}}\left(W_{(i)};\beta\right)\}$ is understood as the log of the original density $\{\log f_{W_i}(W_i;\beta)\}$ after being re-ordered. Then, the log-likelihood ratio statistic between the unstable model and the stable model, with data generated from the unstable model, can be written as

$$\begin{aligned} \mathrm{LR}_N &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(b + d\mathbf{1}[i/N \le r])'\,M'\Gamma(i/N)\,H(i/N)^{-1}\,m_{(i)}\left(\bar\beta\right) \\ &\quad -\frac{1}{2N}\sum_{i=1}^{N}(b + d\mathbf{1}[i/N \le r])'\,M'\Gamma(i/N)\,H(i/N)^{-1}\,\Lambda_{(i)}\left(\dot\beta_{(i)}\right)M\,(b + d\mathbf{1}[i/N \le r]) + o_p(1) \end{aligned}$$

where $\dot\beta_{(i)}$ lies on the segment between $\bar\beta$ and $\beta_{(i)}$. Under Condition 1, Lemma 1 implies that

$$\begin{aligned} \mathrm{LR}_N \Rightarrow\; & b'M'\int_0^1 \Gamma(l)\,H(l)^{-1}\,d\tilde G(l) + d'M'\int_0^r \Gamma(l)\,H(l)^{-1}\,d\tilde G(l) \\ & -\frac{1}{2}\int_0^1 (b + d\mathbf{1}[l \le r])'\,M'\Gamma(l)\,H(l)^{-1}\,\Gamma(l)\,M\,(b + d\mathbf{1}[l \le r])\,dl \end{aligned} \tag{19}$$

where $\tilde G(\cdot)$ is defined in (11). See the proof of the following theorem for more details.

By standard arguments, the log of the Radon-Nikodym derivative of the measure of $\tilde G$ with parameter $(r, b, d) \in [\underline{r}, \bar r] \times \mathbb{R}^k \times \mathbb{R}^k$ relative to the measure of $\int_0^{\cdot} H^{1/2}(l)\,dW(l)$, evaluated at $\tilde G$, is given by exactly the same expression as (19). By Le Cam's limits of experiments theory (see, for instance, Chapter 9 of van der Vaart (1998)), we conclude that the information contained in $\{W_i, q_i\}_{i=1}^N$ is asymptotically the same as that obtained by observing $\tilde G$, as formalized by the following theorem.

Theorem 2 *Suppose m_i is defined as in (18). Then under Condition 1 with known* $\Gamma(\cdot)$ *and* $H(\cdot)$*, the statistical experiment of observing* $\{W_i, q_i\}_{i=1}^N$ *with parameter space* $[\underline{r}, \bar{r}] \times \mathbb{R}^k \times \mathbb{R}^k$ *converges weakly to the limit experiment of observing* $\tilde{G}$ *as in (11).*

Theorem 2 adapts Proposition 1 of Elliott and Müller (2014) to the threshold context, and generalizes it to noncanonical cases. By the convergence of experiments argument, the asymptotic problem is about the change in the drift component of a Gaussian process on the unit interval.

The asymptotic representation theorem (see, for example, Theorem 9.3 of van der Vaart (1998)) and Prohorov's theorem ensure that the asymptotic properties of any test in the finite sample problem with a converging power function can be matched by some test in the limiting problem with the single observation $\tilde{G}$. Thus, the optimal test in the limiting problem provides an upper bound of the asymptotic power within all tests in the finite sample problem that have a converging power function and satisfy the size constraint asymptotically.

In the special situation where $k = 1$, the following corollary is sufficient to establish that the statistical efficiency of the tests φ_r, φ_d, and φ_b holds in both the canonical and the noncanonical cases.

Corollary 1 *The experiment of observing* $\tilde{G}$ with parameter space $[\underline{r}, \bar{r}] \times \mathbb{R}^k \times \mathbb{R}^k$ is the same as that of observing $\mathcal{T}\tilde{G}$.

A straightforward implication of Corollary 1 is that if $k = 1$, then with $M = \left(\int_0^1 \Gamma^2(l)/H(l)\,dl\right)^{-1/2}$, the limit experiment is G_1 with $(r, b_1, d_1) \in [g(\underline{r}), g(\bar{r})] \times \mathbb{R} \times \mathbb{R}$ where $g(s) = \left(\int_0^s \Gamma^2(l)/H(l)\,dl\right) / \left(\int_0^1 \Gamma^2(l)/H(l)\,dl\right)$ for $s \in [0,1]$. In the canonical case where $\Gamma(\cdot)$ and $H(\cdot)$ are constants, the limit of the likelihood ratio statistic (19) reduces to (10). In this situation, the tests φ_r, φ_d, and φ_b are constructed to be efficient in the sense that they maximize some weighted average powers. In noncanonical cases, the limit experiment can be written as $G_1(g(s))$, that is, the canonical limiting Gaussian process with a time transformation. Since $g(\cdot)$ is one-to-one, the efficiency of the tests φ_r, φ_d, and φ_b is retained. In the following subsection, we illustrate such efficiency by comparing the asymptotic power between the test φ_d and a widely used bootstrap test for testing if $d_1 = 0$.

It is worth mentioning that the efficiency is difficult to establish in a multivariate setup. Our method might suffer efficiency loss due to ignoring the last $k-1$ elements of the vector process $\mathcal{T}\tilde{G}$, even if it is known that only the first coefficient involves a break. Intuitively, observing the last $k-1$ components of $\mathcal{T}\tilde{G}$, denoted by $\mathcal{T}\tilde{G}_{-1}$, provides information about

$W_1(s)$ that is useful for learning about (r, d_1, b_1). In the extreme case of perfect correlation, one can even back out these parameters by simply comparing $\mathcal{T}\tilde{G}_1$ and $\mathcal{T}\tilde{G}_{-1}$. However, tests utilizing all components of $\mathcal{T}\tilde{G}$ may violate the size constraint if more than one component of b and d are non-zero. Moreover, when we allow different components of β to change values at different thresholds, the moderate magnitude for the whole vector δ further complicates the limit of the likelihood ratio statistic, which then involves all local break magnitudes $d_1, ..., d_k$ as well as the thresholds $r_1, ..., r_k$.

3.3 Comparison of Asymptotic Power

In this section, we compare the asymptotic power of the test φ_d and the widely used bootstrap test proposed by Hansen (2000b) for problem (4) in the linear regression context. Both tests are based on the $AveF$ test proposed by Andrews and Ploberger (1994).

Consider the model

$$y_i = x_i\beta + x_i d_1 N^{-1/2}\mathbf{1}\left[q_i \leq Q(r)\right] + u_i$$

where (x_i, u_i, q_i) is i.i.d. such that q_i is standard normal and $(x_i, u_i)|q_i \sim N\left(0, diag\left(1, \sigma_i^2\right)\right)$ with $\sigma_i^2 = \rho^2 q_i^2 + 1 - \rho^2$. The parameter $\rho \in [0, 1]$ governs the time transformation g in our method. When $\rho = 0$, the limiting problem corresponds to the canonical case, and the $AveF$ test is directly applicable and performed as follows. Let S_N^0 be the sum of square residuals of regressing y_i on x_i and $S_N(r)$ be that of regressing y_i on x_i and $x_i\left[q_i \leq \hat{Q}(r)\right]$. Then, let $F_N(s) = \left(S_N^0 - S_N(s)\right)/S_N(s)$ and the test statistic (with uniform weight on $[\underline{r}, \bar{r}]$) can be written as

$$AveF_N = \frac{N}{\bar{r} - \underline{r}}\int_{\underline{r}}^{\bar{r}} F_N(s)\, ds \tag{20}$$

which converges to the limit tabulated by Andrews and Ploberger (1994) when $\rho = 0$. However, the limit distribution is different when $\rho \neq 0$. In our setup, one can show that

$$AveF_N \Rightarrow \frac{1}{\bar{r} - \underline{r}}\int_{\underline{r}}^{\bar{r}} \frac{\left(J(s) - rJ(1) + d_1\left(\min(s, r_0) - sr_0\right)\right)^2}{s - s^2} ds$$

where $J(r)$ is a mean-zero Gaussian process with covariance kernel $E\left[J(r)J(s)\right] = \int_0^{\min(s,r)}\left(1 - \rho^2 + \rho^2 Q(t)^2\right)dt$. Thus, the $AveF$ test cannot be directly applied if $\rho \neq 0$.

One solution is developed by Hansen (2000b) who suggests adjusting the critical value by the bootstrap. However, the efficiency of this bootstrap test can not be claimed unless

$\rho = 0$. To maintain efficiency, we can apply the transformation to recover the canonical limit experiment, that is,

$$G_1^*(s) = W_1(s) - sW_1(1) + d_1\sqrt{h(1)}\left(\min(s, g(r_0)) - sg(r_0)\right), \; s \in [0,1]$$

where $g(s) = h(s)/h(1)$ and $h(s) = \int_0^s \left(1 - \rho^2 + \rho^2 Q(s)\right)^{-1} ds$. Our transformation is one-to-one, and hence the test φ_d remains efficient as claimed by Corollary 1. To compare the weighted average power, we set a uniform weight on $g(r) \in [0.15, 0.85]$, so that the original $AveF$ test rejects large values of $\int_{g^{-1}(0.15)}^{g^{-1}(0.85)} (\partial g(s)/\partial s) F_N(g(s))\, ds$.

Figure 1 depicts the asymptotic powers of the test φ_d (transform) and Hansen's (2000b) test (bootstrap) with data generated from J and G_1^*, for $\rho = 0, 0.3, 0.6$, and 0.9 at $d_1 = 3$. The continuous processes J and G_1^* are approximated with 2,000 steps, and the results are based on 5,000 simulations. As expected, these two methods are asymptotically the same in the canonical case where $\rho = 0$. When ρ is not zero, our method outperforms the bootstrap test. In the case where $\rho = 0.9$, our method has an approximately 20% larger weighted average power.

4 Finite Sample Implementation

This section provides finite sample analogues of the tests introduced in the asymptotic problem. We end the section with a summary of implementation steps for convenient reference.

We first assume $\bar{\beta}$, $\Gamma(\cdot)$, and $H(\cdot)$ are known. The finite sample analogue of $\tilde{G}$ is constructed as the partial sum process $N^{-1/2}\sum_{i=1}^{[sN]} m_{(i)}\left(\bar{\beta}\right)$. To obtain G_1, we replace the integral in the mapping $\mathcal{T}$ with a partial sum and evaluate $\Gamma(\cdot)$, $H(\cdot)$, and $g(\cdot)$ at $\{i/N\}_{i=1}^N$. Then, we construct the following finite sample analogues of $\mathcal{T}\tilde{G}$ and $\mathcal{T}\tilde{G}^*$, that is, for $s \in [0,1]$,

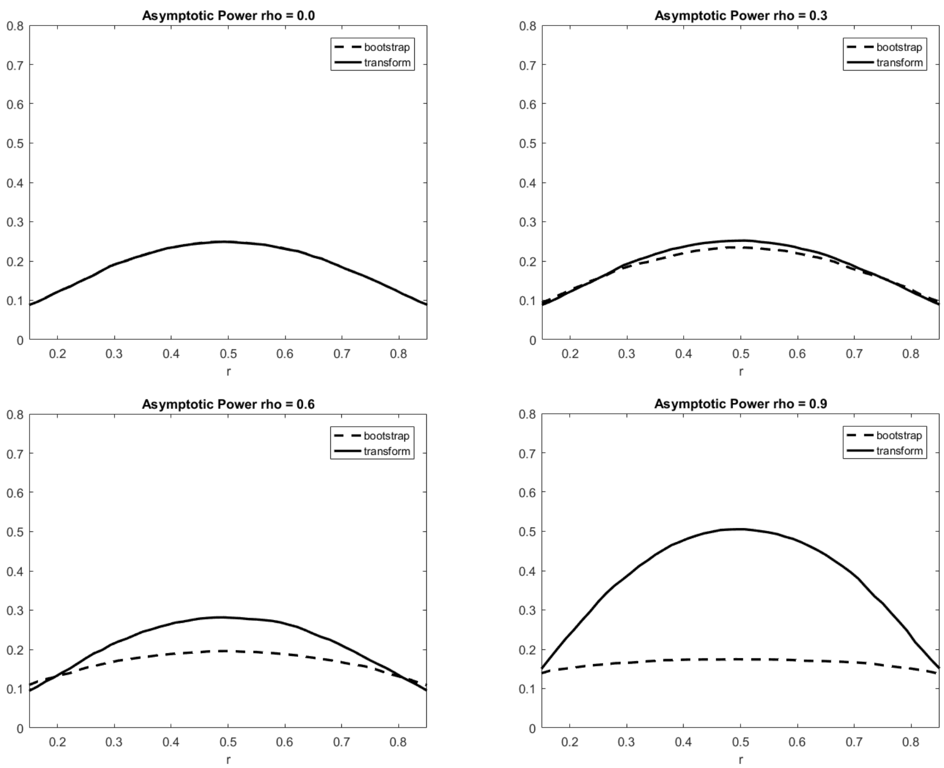
$$\mathcal{T}\tilde{G}_N(s) \equiv \sqrt{h(1)}N^{-1/2}\sum_{i=1}^{\left[g^{-1}(s)N\right]} \frac{\partial g(i/N)}{\partial l}\Gamma(i/N)^{-1} m_{(i)}\left(\bar{\beta}\right) \quad (21)$$

$$\mathcal{T}\tilde{G}_N^*(s) = \mathcal{T}\tilde{G}_N(s) - s\mathcal{T}\tilde{G}_N(1) \quad (22)$$

where $\partial g(i/N)/\partial l$ is understood as $\partial g(l)/\partial l|_{l=i/N}$.

Condition 2.1 implies that $\Gamma(s)$ and $H(s)$ are well defined at $s = 0$ and 1. When it is not the case, we might need to truncate the boundary and modify $h(\cdot)$ and $g(\cdot)$. See Appendix A for more details. The following theorem establishes that the first elements of $\mathcal{T}\tilde{G}_N$ and $\mathcal{T}\tilde{G}_N^*$ converge weakly to G_1 and G_1^*, respectively.

Figure 1: Power Comparison in Testing for Structural Break



Note: Significance level is 5%. Data are generated from the continuous processes J and G_1^* with $d_1 = 3$ that are approximated with 2,000 steps. Based on 5,000 replications.

Lemma 3 *Suppose Conditions 1 and 2 hold. Then in the unstable model (2) with a known* $\bar{\beta}$ *and* $M = h(1)^{-1/2} I_k$*, we have, for* $s \in [0,1]$*,*

$$e_1' \mathcal{T} \tilde{G}_N(s) \Rightarrow W_1(s) + b_1 s + d_1 \min(s, g(r))$$

and

$$e_1' \mathcal{T} \tilde{G}_N^*(s) \Rightarrow W_1(s) - sW_1(1) + d_1 (\min(s, g(r)) - sg(r)).$$

Lemma 3 implies that we can implement the transformation (13) by adding a weighting function and defining a transformed time for the partial sum of $m_{(i)}(\bar{\beta})$. The construction of (21) shares the same intuition as the limit version (13). By multiplying $\Gamma^{-1}(i/N)$, we recover that $N^{-1} \sum_{i=1}^{[sN]} \Gamma(i/N)^{-1} \Lambda_{(i)}(\beta_{(i)}) \xrightarrow{p} sI_k$. Regarding the variance, we observe that the variance of $e_1' \Gamma(i/N)^{-1} m_{(i)}(\beta_{(i)})$ is approximately $e_1' \Gamma(l)^{-1} H(l) \Gamma(l)^{-1\prime} e_1$ when i/N is close to l. Then, we choose $\partial g(l)/\partial l|_{l=i/N}$ to be the weighting function that is proportional to the inverse of the Fisher information at that percentile. Therefore, together with summation along the transformed time $g^{-1}(\cdot)$, the weighting function standardizes the spot variance.

Given Lemma 3, we next deal with the unknown parameter $\bar{\beta}$. For the inference about r and d_1, we can replace $\bar{\beta}$ with its estimator, denoted as $\hat{\beta}$. For example, we can use the moment estimator that satisfies $\sum_{i=1}^{N} m_i(\hat{\beta}) = 0$. Since δ is of the order of magnitude $N^{-1/2}$, this moment estimator is still $\sqrt{N}$-consistent even if δ is ignored, and the error in estimating $\bar{\beta}$ is asymptotically eliminated by imposing the invariance to (15). For inference about $e_1'\beta$, under the null hypothesis $e_1'\beta = \beta_{10}$, we can use the vector $\bar{\beta} \equiv (\beta_{10}, \hat{\beta}_2, ..., \hat{\beta}_k)'$ where $(\hat{\beta}_2, ..., \hat{\beta}_k)'$ denotes the vector of the last $k-1$ components of $\hat{\beta}$.

Finally, we discuss the estimation of the variance-covariance functions $\Gamma(\cdot)$ and $H(\cdot)$. Many nonparametric estimators can be applied, for example, kernel estimators and sieve estimators. We propose a simple Nadaraya-Watson kernel estimator, that is, for any $l \in (0,1)$,

$$\begin{aligned} \hat{\Gamma}(l) &= \frac{\sum_{i=1, i\neq[lN]}^{N} \Lambda_{(i)}(\hat{\beta}) K\left(\frac{i/N-l}{b_N}\right)}{\sum_{i=1, i\neq[lN]}^{N} K\left(\frac{i/N-l}{b_N}\right)} \\ \hat{H}(l) &= \frac{\sum_{i=1, i\neq[lN]}^{N} m_{(i)}(\hat{\beta}) m_{(i)}(\hat{\beta})' K\left(\frac{i/N-l}{b_N}\right)}{\sum_{i=1, i\neq[lN]}^{N} K\left(\frac{i/N-l}{b_N}\right)} \end{aligned} \tag{23}$$

where we use the leave-one-out kernel.

Running Example-cont'd: In the linear regression case with i.i.d. data, we have $\Gamma(s) = E[x_i x_i' | q_i = Q(s)]$ and $H(s) = E[x_i x_i' u_i^2 | q_i = Q(s)]$. Then the nonparametric estimators can be constructed as $\hat{\Gamma}(l) = \sum_{i=1, i\neq[lN]}^{N} x_{(i)} x_{(i)}' K\left(\frac{s/N-l}{b_N}\right) / \sum_{i=1, i\neq[lN]}^{N} K\left(\frac{i/N-l}{b_N}\right)$ and $\hat{H}(l) = \sum_{i=1, i\neq[lN]}^{N} x_{(i)} x_{(i)}' \hat{u}_{(i)}^2 K\left(\frac{i/N-l}{b_N}\right) / \sum_{i=1, i\neq[lN]}^{N} K\left(\frac{i/N-l}{b_N}\right)$ where $\hat{u}_{(i)}$ is the reordered residual of regressing y_i on x_i. If we are testing the null hypothesis $r = r_0$, we can include $x_i \mathbf{1}\left[q_i \leq \hat{Q}(r_0)\right]$ as a regressor for $\hat{u}_{(i)}$.[3]

To obtain uniform consistency, we impose the following additional assumptions:

Condition 3

1. *For all i, $E[\Lambda_i(\beta_i) | q_i = Q(s)]$ and $E\left[m_i(\beta_i) m_i(\beta_i)' | q_i = Q(s)\right]$ are twice continuously differentiable w.r.t. s with uniformly bounded first and second derivatives, except on a subset of $[0,1]$ including only up to a finite number of points.*

2. *The kernel function K satisfies $\int K(s)\, ds = 1$, $\int sK(s)\, dv = 0$, $0 < K(s) < \infty$ for all s, $\lim_{|s|\to\infty} sK(s) = 0$, and $\lim_{|s|\to\infty} s^2 K'(s) = 0$.*

3. *$b_N \to 0$ and $b_N^2 N \to \infty$.*

4. *$\hat{\beta}$ has bounded 4th moment and satisfies that, for all $s \in [0,1]$, $N^{-1/2} \sum_{i=1}^{N} m_i(\beta_i) \mathbf{1}[q_i \leq Q(s)]$ and $\sqrt{N}\left(\hat{\beta} - \bar{\beta}\right)$ jointly weakly converge.*

Condition 3.1 imposes some smoothness and boundedness on the conditional moments. The boundedness at the boundary ($s = 0$ or 1) can be relaxed. See Appendix A for details. Conditions 3.2 and 3.3 impose some conventional restrictions on the kernel function and the bandwidth. The restrictions $\lim_{|s|\to\infty} sK(s) = 0$, $\lim_{|s|\to\infty} s^2 K'(s) = 0$, and $b_N^2 N \to 0$ are imposed to simplify the proof and can be relaxed. Condition 3.4 is readily satisfied for the moment estimator. The boundedness of the fourth moment is also imposed to simplify the proof and can be relaxed. Together with Conditions 1 and 2, we derive the following uniform consistency.

[3]Including $x_i 1\left[q_i \leq \hat{Q}(r_0)\right]$ may lead to a worse estimation of $H(\cdot)$ when the sample size is small, since the estimation error in $\hat{Q}$ is large in this situation. We may also include $x_i \mathbf{1}\left[q_i \leq \hat{Q}(\hat{r})\right]$ for $\hat{u}_i$ when conducting inference about δ and β where $\hat{r}$ is the least squares estimator of r. Adding this regressor may improve the estimation of Γ and H when δ is large. However, it may lead to a worse estimate of Γ and H when δ is small, since $\hat{r}$ might be substantially different from r, as we show in Section 2.5.

Lemma 4 *Suppose Conditions 1-3 hold, then*

(i) $\sup_{l\in S}\left|\left|\hat{\Gamma}(l)-\Gamma(l)\right|\right| \xrightarrow{p} 0$ *for any compact set* $S\subset(0,1)$ *on which* $\Gamma(\cdot)$ *is continuous.*

(ii) $\sup_{l\in S}\left|\left|\hat{H}(l)-H(l)\right|\right| \xrightarrow{p} 0$ *for any compact set* $S\subset(0,1)$ *on which* $H(\cdot)$ *is continuous.*

(iii) For $l\in[0,1]$, *let* $\hat{h}(l)=N^{-1}\sum_{i=1}^{[lN]}\left(e_1'\hat{\Gamma}(i/N)^{-1}\hat{H}(i/N)'\hat{\Gamma}(i/N)^{-1\prime}e_1\right)^{-1}$, *and* $\hat{g}(l)=\hat{h}(l)/\hat{h}(1)$, *then* $\hat{h}(1)\xrightarrow{p}h(1)$, *and* $\sup_{0\le l\le 1}|\hat{g}(l)-g(l)|\xrightarrow{p}0$.

Lemma 4 shows that the unknown conditional moments and the transformed time can be uniformly consistently estimated.[4] Regarding the choice of the bandwidth, many suggestions have been made in the kernel estimation literature, for instance, cross-validation and plug-in. Monte Carlo experiments demonstrate that our approach is not sensitive to the bandwidth; so we choose the simplest rule-of-thumb bandwidth, that is, $b_N^*=\sqrt{1/12}N^{-1/5}$ where $\sqrt{1/12}$ is the standard deviation of a uniform distribution on $(0,1)$.

Replacing all unknown items in (21) with their consistent estimators, we obtain the feasible analogues of $\mathcal{T}\tilde{G}_N(s)$ and $\mathcal{T}\tilde{G}_N^*(s)$, that is,

$$\mathcal{T}\hat{G}_N(s)=\sqrt{\hat{h}(1)}N^{-1/2}\sum_{i=1}^{\left[\hat{g}^{-1}(s)N\right]}\frac{\widehat{\partial g}(i/N)}{\partial l}\hat{\Gamma}(i/N)^{-1}m_{(i)}\left(\bar{\beta}\right). \tag{24}$$

and

$$\begin{aligned}\mathcal{T}\hat{G}_N^*(s) &= \sqrt{\hat{h}(1)}N^{-1/2}\sum_{i=1}^{\left[\hat{g}^{-1}(s)N\right]}\frac{\widehat{\partial g}(i/N)}{\partial l}\hat{\Gamma}(i/N)^{-1}m_{(i)}\left(\hat{\beta}\right)\\ &\quad -s\sqrt{\hat{h}(1)}N^{-1/2}\sum_{i=1}^{N}\frac{\widehat{\partial g}(i/N)}{\partial l}\hat{\Gamma}(i/N)^{-1}m_{(i)}\left(\hat{\beta}\right). \end{aligned}\tag{25}$$

By Theorem 2 and the uniform consistency of the kernel estimators, we establish that the first elements of $\mathcal{T}\hat{G}_N$ and $\mathcal{T}\hat{G}_N^*$ converge weakly to the desired limits G_1 and G_1^*, respectively, as summarized in the following theorem.

[4]The model (2) can also be relaxed by allowing the last $k-1$ components of β to be time-varying. In this situation, we may employ a kernel-smoothed average estimator as studied in Robinson (1989, 1991). This estimator is $\sqrt{N}$-consistent under the assumption that $\beta_2(s),...,\beta_k(s)$ are continuous in $s\in[0,1]$ (and other standard regularity conditions). Pronounced instabilities are hence allowed; so choosing a time-varying coefficient might improve approximation accuracy even in the single break model.

Theorem 3 *Suppose Conditions 1-3 hold, then in the unstable model (2) with* $M = h(1)^{-1/2} I_k$*, we have, for* $s \in [0,1]$*, (i) under the null hypothesis* $e_1'\beta = \beta_{10}$*,*

$$e_1' \mathcal{T} \hat{G}_N(s) \Rightarrow W_1(s) + b_1 s + d_1 \min(s, g(r)),$$

and (ii)

$$e_1' \mathcal{T} \hat{G}_N^*(s) \Rightarrow W_1(s) - sW_1(1) + d_1(\min(s, g(r)) - sg(r)).$$

Theorem 3 suggests that the tests φ_r, φ_d, and φ_b are implementable with G_1 and G_1^* replaced with $e_1'\mathcal{T}\hat{G}_N$ and $e_1'\mathcal{T}\hat{G}_N^*$, respectively.

Running Example–cont'd: The effect of b in the asymptotic problem is eliminated by imposing invariance to (15). In finite samples, we may also impose the following invariance constraint in the linear regression context. Let X denote the $N \times k$ matrix with the i-th row x_i' and Y the vector of $\{y_i\}_{i=1}^N$. Consider the group of transformations

$$\{Y, X\} \to \{Y + X\tilde{b}, X\} \text{ for all } \tilde{b} \in \mathbb{R}^k. \tag{26}$$

One maximal invariant to (26) is the vector of the least squares residual, $\{\hat{u}_i\}$, of regressing y_i on x_i only. Then, the second convergence in Theorem 3 holds with $m_{(i)}\left(\hat{\beta}\right)$ understood as $x_{(i)}\hat{u}_{(i)}$. Regarding inference about b, we can eliminate the effect of the last $k-1$ components of β by imposing the additional invariance to the group of transformations

$$\{Y, X\} \to \{Y + X(0, \tilde{b}_{-1}')', X\} \text{ for all } \tilde{b}_{-1} \in \mathbb{R}^{k-1}. \tag{27}$$

Accordingly, we can replace $y_{(i)}$ with $y_{(i)} - x_{(i)}'\left(0, \hat{\beta}_{-1}'\right)'$ in constructing $e_1'\mathcal{T}\hat{G}_N$ where $\hat{\beta}_{-1}$ is the coefficient of regressing y_i on $(x_{i2}, ..., x_{ik})$.

As a summary, we have provided details about how to construct finite sample analogues of the efficient tests developed in the asymptotic problems. We conclude this section by listing the steps to implementing our approach:

Step 1 Construct a preliminary estimator $\hat{\beta}$, for example, the moment estimator satisfying $0 = \sum_{i=1}^N m_i\left(\hat{\beta}\right)$, and re-order the moment functions according to q_i.

Step 2 Estimate the spot variance-covariance matrices $\hat{\Gamma}(l)$ and $\hat{H}(l)$ for $l \in \{i/N\}_{i=1}^N$ as in (23) and construct $\hat{h}(\cdot)$ and $\hat{g}(\cdot)$ as

$$\hat{h}(l) = N^{-1} \sum_{i=1}^{[lN]} \left(e_1'\hat{\Gamma}(i/N)^{-1} \hat{H}(i/N) \hat{\Gamma}(i/N)^{-1\prime} e_1\right)^{-1}$$

$$\hat{g}(l) = \hat{h}(l)/\hat{h}(1).$$

See Appendix A for truncating boundary observations when necessary.

Step 3 Numerically invert the function $\hat{g}(\cdot)$.

Step 4 Construct the partial sum $e_1'\mathcal{T}\hat{G}_N(\cdot)$ as in (24) or $e_1'\mathcal{T}\hat{G}_N^*(\cdot)$ as in (25).

Step 5 Apply the tests φ_r, φ_d, and φ_b for inference about r, d_1, and b_1, respectively, and invert them for the corresponding confidence intervals. See Appendix B for details of these tests.

5 Extensions

This section first extends our approach to three additional models: the over-identified GMM threshold model, the threshold model with a preliminary estimator of an infinite dimensional nuisance parameter, and the threshold model with a non-smooth moment function. Then in the last subsection, we discuss time series applications with weak dependence conditions.

5.1 Over-identification

The preceding analysis focuses on exact identification. In this subsection, we extend our approach to cover over-identification. Suppose $\dim(\beta) < \dim(m_i)$ and define $L_N = (\sum_{i=1}^N \Lambda_i\left(\hat{\beta}_0\right))'(\sum_{i=1}^N m_i\left(\hat{\beta}_0\right)m_i\left(\hat{\beta}_0\right)')^{-1}$ where $\hat{\beta}_0$ is some preliminary estimator of β, say, the GMM estimator with an identity weighting matrix. Under Conditions 1-3, it is easily shown that $L_N \xrightarrow{p} L \equiv \lim_{N\to\infty}(N^{-1}\sum_{i=1}^N E\left[\Lambda_i\left(\beta_i\right)\right])'(N^{-1}\sum_{i=1}^N E\left[m_i\left(\beta_i\right)m_i\left(\beta_i\right)'\right])^{-1}$. Then, we can prove a similar version of Lemma 1, that is,

$$\sup_{s\in[0,1]}\sup_{\beta\in B_{\varepsilon_N}\left(\bar{\beta}\right)}\left\|N^{-1}\sum_{i=1}^{[sN]}L_N\Lambda_{(i)}(\beta) - \int_0^s L\Gamma(l)\,dl\right\| \xrightarrow{p} 0 \text{ and} \tag{28}$$

$$N^{-1/2}\sum_{i=1}^{[sN]}Lm_{(i)}\left(\beta_{(i)}\right) \Rightarrow \int_0^s LH^{1/2}(l)\,dW_k(l)\text{, for } s\in[0,1] \tag{29}$$

where $B_{\varepsilon_N}\left(\bar{\beta}\right)$ is the open ball centered around $\bar{\beta}$ with radius $\varepsilon_N \to 0$. Assume that $L\Gamma(l)$ and $LH(l)L'$ are invertible at all $l\in[0,1]$ except for up to a finite number of points.[5] Then,

[5]This is can be relaxed as discussed in Appendix A.

we can adjust the time transformation $g(\cdot)$ as

$$\begin{aligned} h(l) &= \int_0^l \frac{1}{e_1'(L\Gamma(s))^{-1} LH(s) L' (L\Gamma(s))^{-1'} e_1} ds \text{ and} \\ g(l) &= h(l)/h(1), \end{aligned}$$

and hence the asymptotic analysis in Sections 2 and 3 remains valid. Regarding the finite sample implementation, we assume that $L_N \hat{\Gamma}(i/N)$ is also invertible for all i. Then, we can construct the over-identification version of $\mathcal{T}\hat{G}_N$ and $\mathcal{T}\hat{G}_N^*$, that is, for $s \in [0,1]$,

$$\begin{aligned} \mathcal{T}\hat{G}_N^{OI}(s) &= \sqrt{\hat{h}(1)} N^{-1/2} \sum_{i=1}^{[\hat{g}^{-1}(s)N]} \frac{\widehat{\partial g}(i/N)}{\partial l} \left(L_N \hat{\Gamma}(i/N) \right)^{-1} L m_{(i)}(\bar{\beta}) \\ \mathcal{T}\hat{G}_N^{*OI}(s) &= \sqrt{\hat{h}(1)} N^{-1/2} \sum_{i=1}^{[\hat{g}^{-1}(s)N]} \frac{\widehat{\partial g}(i/N)}{\partial l} \left(L_N \hat{\Gamma}(i/N) \right)^{-1} L m_{(i)}\left(\hat{\beta}\right) \\ &\quad - s\sqrt{\hat{h}(1)} N^{-1/2} \sum_{i=1}^{N} \frac{\widehat{\partial g}(i/N)}{\partial l} \left(L_N \hat{\Gamma}(i/N) \right)^{-1} L m_{(i)}\left(\hat{\beta}\right) \end{aligned}$$

where $\hat{\beta}$ can be the GMM estimator of $\bar{\beta}$ with the optimal weighting matrix, and $\hat{\Gamma}(\cdot), \hat{g}(\cdot)$, and $\hat{h}(\cdot)$ are kernel estimators of $\Gamma(\cdot), g(\cdot)$, and $h(\cdot)$, respectively, as defined in Section 4. Under Conditions 1-3, we can establish the desired convergence of $e_1'\mathcal{T}\hat{G}_N^{OI}$ and $e_1'\mathcal{T}\hat{G}_N^{*OI}$ by the exact argument of Theorem 3. Therefore, the tests φ_r, φ_d, and φ_b are readily applied with $e_1'\mathcal{T}\hat{G}_N$ and $e_1'\mathcal{T}\hat{G}_N^*$ replaced with $e_1'\mathcal{T}\hat{G}_N^{OI}$ and $e_1'\mathcal{T}\hat{G}_N^{*OI}$, respectively.

Note that since our transformation g standardizes the time-varying variance-covariance matrices, the asymptotic distributions of $e_1'\mathcal{T}\hat{G}_N^{OI}$ and $e_1'\mathcal{T}\hat{G}_N^{*OI}$ remain unchanged regardless of the choice of L_N.

Running Example–cont'd: In the i.i.d. case where $y_i = x_i'\beta_i + u_i$, suppose x_i is correlated with u_i, and there exists a set of instrumental variables z_i that satisfies $E[z_i u_i | q_i] = 0$. Then, we may consider the moment function as $z_i(y_i - x_i'\beta)$. Assume $\dim(z_i) > \dim(x_i)$; so $L\Gamma(s)$ and $LH(s)L'$ can be understood as $LE[z_i x_i' | q_i = Q(s)]$ and $H(s) = LE[z_i z_i' u_i^2 | q_i = Q(s)] L'$. Then, $e_1'\mathcal{T}\hat{G}_N^{*OI}$ can be constructed as the partial sum of $L_N x_{(i)} \hat{u}_{(i)}$. The GMM estimator $\hat{\beta}_{GMM}$ is equivalent to the two-stage least squares estimator when L is $(E[x_i z_i'])(E[z_i z_i'])^{-1}$, whose finite sample analogue, L_N, is $\left(\sum_{i=1}^N x_i z_i'\right)\left(\sum_{i=1}^N z_i z_i'\right)^{-1}$.

5.2 Nonparametric Component

In many econometric problems, the moment function involves a nonparametric component, for example, the partially linear model with an unknown nonlinear function (see, for example, Robinson (1988)) and the binary choice model with an unknown CDF of the error term (see, for example, Klein and Spady (1993) and Ichimura (1993)). Estimation of such models usually relies on a preliminary estimator of this nonparametric component (cf. Andrews (1994)). In this subsection, we show that under some additional weak regularity conditions, the estimation error from introducing the nonparametric estimator is negligible, and hence, our previous analysis goes through.

Let τ be some unknown $\mathbb{R}^{\tau}$-valued function and $\hat{\tau}$ be its nonparametric estimator. Denote the argument of τ as X_i, and assume that $\hat{\tau}$ is consistent with respect to some metric ρ_τ, for example, $\rho_\tau(\tau, \tau_0) = \left(E\left[||\tau(X_i) - \tau_0(X_i)||^2\right]\right)^{1/2}$. We consider the moment function as $m_i(\beta, \tau)$, which satisfies $E[m_i(\beta_i, \tau_0)|q_i] = 0$ when the data are generated from the model with β_i as in (2) and with the true function of τ as τ_0. In addition, we assume that X_i has a bounded support $\mathcal{X}_N^*$, which is an increasing sequence of compact subsets of $\mathbb{R}^x$. Such a restriction can be enforced by trimming, which is implemented in many semiparametric models. For notational ease, we suppress X_i in $\tau(X_i)$ in the following derivation. Denote $\xi_i = \mathbf{1}[X_i \in \mathcal{X}_N^*]$ as the trimming function, and we assume that the moment condition with trimming satisfies $N^{-1/2}\sum_{i=1}^N E[\xi_i m_i(\beta_i, \tau_0)|q_i = Q(s)] = o(1)$ for all $s \in [0,1]$.

Modify $\Lambda_i(\beta, \tau)$ as $-\partial m_i(\beta, \tau)/\partial\beta'$, and define the function $a : [0,1] \to \mathbb{R}^k$ as $a(\eta) = m_{(i)}(\bar{\beta} + \eta\left(\beta_{(i)} - \bar{\beta}\right), \hat{\tau})$. We have $a'(\eta) = -\Lambda_{(i)}\left(\bar{\beta} + \eta\left(\beta_{(i)} - \bar{\beta}\right), \hat{\tau}\right)\left(\beta_{(i)} - \bar{\beta}\right)$ and

$$\begin{aligned} m_{(i)}\left(\beta_{(i)}, \hat{\tau}\right) - m_{(i)}\left(\bar{\beta}, \hat{\tau}\right) &= a(1) - a(0) \\ &= \int_0^1 -\Lambda_{(i)}\left(\bar{\beta} + \eta\left(\beta_{(i)} - \bar{\beta}\right), \hat{\tau}\right) d\eta \left(\beta_{(i)} - \bar{\beta}\right). \end{aligned}$$

In the unstable model (2) with a known $\bar{\beta}$, we then obtain

$$\begin{aligned} & \bar{H}^{-1/2} N^{-1/2} \sum_{i=1}^{[sN]} m_{(i)}(\bar{\beta}, \hat{\tau}) \\ = \; & \bar{H}^{-1/2} N^{-1/2} \sum_{i=1}^{[sN]} m_{(i)}\left(\beta_{(i)}, \hat{\tau}\right) \\ & + \bar{H}^{-1/2} \int_0^1 \left(N^{-1} \sum_{i=1}^{[sN]} \Lambda_{(i)}\left(\bar{\beta} + \eta\left(\beta_{(i)} - \bar{\beta}\right), \hat{\tau}\right) Mb \right) d\eta \end{aligned}$$

$$+\bar{H}^{-1/2}\int_0^1\left(N^{-1}\sum_{i=1}^{[sN]}\Lambda_{(i)}\left(\bar{\beta}+\eta\left(\beta_{(i)}-\bar{\beta}\right),\hat{\tau}\right)Md\mathbf{1}\left[i/N\leq r\right]\right)d\eta+o_p\left(1\right).$$

Assume Condition 1 holds with $m_i(\beta)$ and $\Lambda_i(\beta)$ replaced with $m_i(\beta,\tau_0)$ and $\Lambda_i(\beta,\tau_0)$, respectively, and with $\Gamma(\cdot)$ and $H(\cdot)$ adjusted accordingly. Our previously suggested approach is valid if we establish

$$N^{-1/2}\sum_{i=1}^{[sN]}\xi_{(i)}m_{(i)}\left(\beta_{(i)},\hat{\tau}\right)\Rightarrow\int_0^s H^{1/2}\left(l\right)dW\left(l\right),\ s\in\left[0,1\right],\tag{30}$$

and

$$\sup_{s\in[0,1]}\sup_{\beta\in B_{\varepsilon_N}\left(\bar{\beta}\right)}\left|\left|N^{-1}\sum_{i=1}^{[sN]}\Lambda_{(i)}\left(\beta,\hat{\tau}\right)-\int_0^s\Gamma(l)dl\right|\right|\xrightarrow{p}0\tag{31}$$

for any $\varepsilon_N\to 0$ as $N\to\infty$ where B_{ε_N} is the open ball centered at $\bar{\beta}$ with radius ε_N.

The convergence of (30) can be established by resorting to the stochastic equicontinuity condition. Define the following empirical processes indexed by τ that for any fixed $s\in(0,1)$,

$$v_N\left(s,\tau\right)=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\xi_i m_i\left(\beta_i,\tau\right)\mathbf{1}\left[q_i\leq Q\left(s\right)\right]-E\left[\xi_i m_i\left(\beta_i,\tau\right)\mathbf{1}\left[q_i\leq Q\left(s\right)\right]\right]\right).$$

Then, $v_N(s,\cdot)$ is stochastic equicontinuous at τ_0 if for all $\varepsilon>0$ and $\varepsilon_1>0$, there exists $\eta>0$ such that

$$\limsup_{N\to\infty}P\left(\sup_{\rho_\tau(\tau,\tau_0)<\eta}\left|\left|v_N\left(s,\tau\right)-v_N\left(s,\tau\right)\right|\right|>\varepsilon\right)<\varepsilon_1.$$

To prove (30) and (31), we make the following additional assumptions:

Condition 4:

1. $N^{-1/2}\sum_{i=1}^{N}E\left[\xi_i m_i\left(\beta_i,\tau\right)\mathbf{1}\left[q_i\leq Q\left(s\right)\right]\right]|_{\tau=\hat{\tau}}=o_p\left(1\right)$ *uniformly in* $s\in\left[0,1\right].$
2. $v_N(s,\tau)$ *is stochastic equicontinuous in* τ *for any fixed* $s\in[0,1]$.
3. $\rho_\tau(\hat{\tau},\tau_0)\xrightarrow{p}0$, *and for all* i *and uniformly for* β *and* τ *in some open ball centered at* $\left(\bar{\beta},\tau_0\right)$, $m_i(\beta,\tau)$ *and* $\Lambda_i(\beta,\tau)$ *have well-defined partial derivatives w.r.t.* τ *and satisfy* $E\left[||m_i(\beta,\tau)||^4\right]<C$, $E\left[||\Lambda_i(\beta,\tau)||^4\right]<C$, $E\left[||\partial m_i(\beta,\tau)/\partial\tau'||^4\right]<C$, *and* $E\left[||\partial\Lambda_i(\beta,\tau)/\partial\tau'||^4\right]<C$ *for some* $C<\infty$.

4. $N^{-1/2}\sum_{i=1}^{N} E\left[\xi_i m_i\left(\beta_i,\tau_0\right)\mathbf{1}\left[q_i\leq Q\left(s\right)\right]\right]=o\left(1\right)$ *for all* $s\in[0,1]$.

Condition 4 is extensively discussed by Andrews (1994), who also provides primitive assumptions. This condition is satisfied in many semiparametric models, including, for example, the single index model and the partial linear model. In particular, Condition 4.1 is an asymptotic orthogonality condition between $\hat{\beta}$ and $\hat{\tau}$, requiring that the estimator error of $\hat{\tau}$ has an effect on the moment conditions of an order of magnitude no larger than $N^{-1/2}$. It can be directly verified in the partial linear model. It also holds if $E\left[\xi_i \frac{\partial}{\partial\tau} m_i\left(\beta_i,\tau_0\right)\mathbf{1}\left[q_i\leq Q\left(s\right)\right]|X_i=x\right]=0$ for any s, x, and i, and $\hat{\tau}$ converges to τ_0 at a rate faster than $N^{-1/4}$. Condition 4.2 is satisfied under Condition 1 and some smoothness of m_i in β. Condition 4.3 requires the consistency of $\hat{\tau}$ and reinforces Condition 1.6 by requiring additional boundedness of the moment function in a local neighborhood of τ_0. Condition 4.4 requires that the trimming makes no first-order effect on the asymptotic result. Under Conditions 1 and 4, we have the following lemma.

Lemma 5 *Suppose Conditions 1 and 4 hold with* $m_i\left(\beta\right)$ *and* $\Lambda_i\left(\beta\right)$ *replaced with* $m_i\left(\beta,\tau_0\right)$ *and* $\Lambda_i\left(\beta,\tau_0\right)$*, respectively. Then, (30) and (31) hold. Moreover, in the unstable model (2) with a known* $\bar{\beta}$*,*

$$N^{-1/2}\sum_{i=1}^{[sN]}\xi_{(i)}m_{(i)}\left(\bar{\beta},\hat{\tau}\right)\Rightarrow\tilde{G}\left(s\right)$$

where $\tilde{G}\left(\cdot\right)$ *is defined as in (11).*

With $h\left(\cdot\right)$ and $g\left(\cdot\right)$ modified accordingly, the asymptotic analysis conducted in Section 2 is again applicable. Regarding the finite sample implementation, we need uniformly consistent estimators of $\Gamma\left(\cdot\right)$ and $H\left(\cdot\right)$. The kernel estimator (23) can be adjusted as

$$\begin{aligned}\hat{\Gamma}\left(l\right) &= \frac{\sum_{i=1,i\neq[lN]}^{N}\Lambda_{(i)}\left(\hat{\beta},\hat{\tau}\right)K\left(\frac{i/N-l}{b_N}\right)}{\sum_{i=1,i\neq[lN]}^{N}K\left(\frac{i/N-l}{b_N}\right)}\\ \hat{H}\left(l\right) &= \frac{\sum_{i=1,i\neq[lN]}^{N}m_{(i)}\left(\hat{\beta},\hat{\tau}\right)m_{(i)}\left(\hat{\beta},\hat{\tau}\right)'K\left(\frac{i/N-l}{b_N}\right)}{\sum_{i=1,i\neq[lN]}^{N}K\left(\frac{i/N-l}{b_N}\right)}.\end{aligned}\tag{32}$$

Then, the analogues of $\mathcal{T}\hat{G}_N$ as in (24) and $\mathcal{T}\hat{G}_N^{*}$ as in (25) can be constructed by plugging in the above estimators, that is,

$$\mathcal{T}\hat{G}_N^{NC}\left(s\right) = \sqrt{\hat{h}\left(1\right)}N^{-1/2}\sum_{i=1}^{\left[\hat{g}^{-1}(s)N\right]}\frac{\widehat{\partial g}\left(i/N\right)}{\partial l}\hat{\Gamma}\left(i/N,\hat{\tau}\right)^{-1}m_{(i)}\left(\bar{\beta},\hat{\tau}\right)\tag{33}$$

$$
\begin{aligned}
\mathcal{T}\hat{G}_N^{*NC}(s) &= \sqrt{\hat{h}(1)}N^{-1/2}\sum_{i=1}^{[\hat{g}^{-1}(s)N]}\frac{\widehat{\partial g}(i/N)}{\partial l}\hat{\Gamma}(i/N,\hat{\tau})^{-1}m_{(i)}\left(\hat{\beta},\hat{\tau}\right) \\
&\quad -s\sqrt{\hat{h}(1)}N^{-1/2}\sum_{i=1}^{N}\frac{\widehat{\partial g}(i/N)}{\partial l}\hat{\Gamma}(i/N,\hat{\tau})^{-1}m_{(i)}\left(\hat{\beta},\hat{\tau}\right). \qquad (34)
\end{aligned}
$$

The following theorem establishes the desired convergence.

Theorem 4 *Suppose Conditions 1-4 hold with $m_i(\beta)$ and $\Lambda_i(\beta)$ replaced with $m_i(\beta,\tau_0)$ and $\Lambda_i(\beta,\tau_0)$, respectively. Then in the unstable model (2) with $M = h(1)^{-1/2} I_k$, we have, for $s \in [0,1]$, (i) under the null hypothesis $e_1'\beta = \beta_{10}$,*

$$
e_1'\mathcal{T}\hat{G}_N^{NC}(s) \Rightarrow W_1(s) + b_1 s + d_1 \min(s, g(r)),
$$

and (ii)

$$
e_1'\hat{G}_N^{*NC}(s) \Rightarrow W_1(s) - sW_1(1) + d_1(\min(s,g(r)) - sg(r)).
$$

Based on the above theorem, the tests φ_r, φ_d, and φ_b are again implementable with G_1 and G_1^* replaced with $e_1'\mathcal{T}\hat{G}_N^{NC}$ and $e_1'\mathcal{T}\hat{G}_N^{*NC}$, respectively.

5.3 Non-smooth Moment Function

In the preceding analysis, the definition of $\Gamma(\cdot)$ requires the differentiability of m_i. A leading example where the differentiability is violated is the least absolute deviation estimator, as defined by $\hat{\beta}_{LAD} = \arg\min_\beta N^{-1}\sum_{i=1}^N |y_i - x_i'\beta|$. The corresponding moment condition in the unstable model can be specified as $E[x_i(-\mathbf{1}[y_i > x_i'\beta_i] + \mathbf{1}[y_i < x_i'\beta_i])] = 0$. In this situation, the moment condition is non-differentiable, and hence Λ_i is not defined.

As an alternative, we consider the partial derivative of the expectation of the moment function and make the following modifications to Conditions 1.4-1.6:

Condition 1.4' *For all i and $s \in [0,1]$, $E[m_i(\beta)\mathbf{1}[q_i \le Q(s)]]$ is twice continuously differentiable w.r.t. β.*

Condition 1.5' *For all i and uniformly in $s \in [0,1]$, $\partial E[m_i(\beta)|q_i = Q(s)]/\partial\beta'|_{\beta=\bar{\beta}}$ and $E\left[m_i(\beta_i)m_i(\beta_i)'|q_i = Q(s)\right]$ are bounded and satisfy*

$$
-N^{-1}\sum_{i=1}^N \frac{\partial}{\partial\beta'}E[m_i(\beta)|q_i = Q(s)]\bigg|_{\beta=\bar{\beta}} \quad\rightarrow\quad \Gamma(s)
$$

$$N^{-1}\sum_{i=1}^{N} E\left[m_i\left(\beta_i\right) m_i\left(\beta_i\right)' | q_i = Q\left(s\right)\right] \quad \rightarrow \quad H\left(s\right)$$

where $\Gamma(\cdot)$ *and* $H(\cdot)$ *are uniformly bounded matrix-valued functions defined on* $[0,1]$ *that are piece-wise twice continuously differentiable with at most a finite number of discontinuities, and left and right limits everywhere.*

Condition 1.6' *For all* i *and uniformly for* β *in some open ball centered at* $\bar{\beta}$, $E\left[||m_i(\beta)||^4\right] < C$, $\left|\left|\frac{\partial}{\partial \beta'} E\left[m_i(\beta)\right]\right|\right|^4 < C$, $\left|\left|\frac{\partial^2 E[m_i(\beta)]}{\partial \beta' \partial \beta}\right|\right|^2 < C$ *for some* $C < \infty$.

Condition 1.4' imposes restriction on the smoothness of $E\left[m_i(\beta)\mathbf{1}\left[q_i \leq Q(s)\right]\right]$ instead of $m_i(\beta)$. Condition 1.5' reinterprets $\Gamma(\cdot)$, and Condition 1.6' requires the corresponding boundedness. Define

$$v_N^0(s) = N^{-1/2}\sum_{i=1}^{N}\left(m_i\left(\bar{\beta}\right)\mathbf{1}\left[q_i \leq Q(s)\right] - E\left[m_i\left(\bar{\beta}\right)\mathbf{1}\left[q_i \leq Q(s)\right]\right]\right)$$

and

$$v_N^1(s) = N^{-1/2}\sum_{i=1}^{N}\left(m_i\left(\beta_i\right)\mathbf{1}\left[q_i \leq Q(s)\right] - E\left[m_i\left(\beta_i\right)\mathbf{1}\left[q_i \leq Q(s)\right]\right]\right).$$

Under Condition 1 with the above modifications, we can resort to an argument similar to that in the previous subsection to establish

$$\sup_{s\in[0,1]}\left|\left|v_N^0(s) - v_N^1(s)\right|\right| = o_p(1). \tag{35}$$

Moreover, by a similar derivation as in Lemma 1, we can show that under Conditions 1.4'-1.6',

$$\sup_{s\in[0,1]}\sup_{\beta\in B_{\varepsilon_N}(\bar{\beta})}\left|\left|N^{-1}\sum_{i=1}^{N}\left(E\left[m_i(\beta)\mathbf{1}\left[q_i \leq Q(s)\right]\right] - E\left[m_i(\beta)\mathbf{1}\left[q_i \leq \hat{Q}(s)\right]\right]\right)\right|\right| \xrightarrow{p} 0 \tag{36}$$

and

$$\sup_{s\in[0,1]}\sup_{\beta\in B_{\varepsilon_N}(\bar{\beta})}\left|\left|-N^{-1}\sum_{i=1}^{N}\frac{\partial}{\partial\beta'}E\left[m_i(\beta)\mathbf{1}\left[q_i \leq Q(s)\right]\right] - \int_0^s \Gamma(l)\,dl\right|\right| \xrightarrow{p} 0 \tag{37}$$

where $B_{\varepsilon_N}\left(\bar{\beta}\right)$ is the open ball centered at $\bar{\beta}$ with radius $\varepsilon_N \rightarrow 0$. The following Lemma formally establishes (35)-(37).

Lemma 6 *Suppose Condition 1 holds with 1.4-1.6 replaced with 1.4'-1.6'. Then, (35)-(37) hold. Moreover, in the unstable model (2) with a known* $\bar{\beta}$*,*

$$N^{-1/2} \sum_{i=1}^{[sN]} m_{(i)} \left(\bar{\beta}\right) \Rightarrow \tilde{G}(s)$$

where $\tilde{G}(\cdot)$ *is defined as in (11).*

By the above lemma, the asymptotic analysis in Section 2 is again applicable, and then we can adjust the transformation accordingly to obtain the canonical limiting Gaussian processes, G_1 and G_1^*, as before.

Regarding the finite sample implementation, the kernel estimator (23) is no longer applicable in general since Condition 3.1 is violated. Alternatively, we can rely on the explicit expression of the moment function to construct estimators for $\Gamma(\cdot)$ and $H(\cdot)$. This is possible in many situations, including, for example, the LAD estimator, whose moment function can be written as $m_i(\beta) = x_i(-\mathbf{1}[y_i > x_i'\beta] + \mathbf{1}[y_i < x_i'\beta])/2$. Suppose (u_i, q_i, x_i) is i.i.d. and u_i is independent of (x_i, q_i) with a continuous PDF f_u and CDF F_u. Then, $\Gamma(s)$ is equal to $f_u(0) E[x_i x_i | q_i = Q(s)]$, and $H(s)$ is equal to $F_u(0) E[x_i x_i | q_i = Q(s)]/2$. Therefore, given some consistent estimators of $f_u(0)$ and $F_u(0)$, the tests φ_r, φ_d, and φ_b can be again implemented.

5.4 Time Series

In the previous discussion, we assume that the data are independent, and hence, $H(\cdot)$ fully captures the second moment of $m_i(\beta_i)$. In this subsection, we extend the preceding analysis to allow for weak dependence across data. Accordingly, the (conditional) long-run variance of $m_i(\beta_i)$ depends not only on the variance but also on the covariance. To modify our approach, we define

$$\begin{aligned}
L_{1N}(s) &= \frac{1}{N} \sum_{i=2}^{N} \sum_{j=1}^{i-1} E\left[m_i(\beta_i)\, m_j\left(\beta_j\right)' \mathbf{1}\left[q_j \le Q(s)\right] | q_i = Q(s)\right] \\
L_{2N}(s) &= \frac{1}{N} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} E\left[m_i(\beta_i)\, m_j\left(\beta_j\right)' \mathbf{1}\left[q_j \le Q(s)\right] | q_i = Q(s)\right]
\end{aligned}$$

and make the following adjustment to Condition 1:

Condition 1.1" *$\{W_i, q_i\}$ is ϕ-mixing, with ϕ-mixing coefficients satisfying* $\sum_{s=1}^{\infty} \phi_s^{1/2} < \infty$.

Condition 1.2" *q_i is strictly stationary with a continuous density function f_q such that for all s,* $0 < f_q(s) < C$ for some $C < \infty$.

Condition 1.5" *For all i and N and uniformly in $s \in [0,1]$, $E[\Lambda_i(\beta_i)|q_i = Q(s)]$, $E\left[m_i(\beta_i) m_i(\beta_i)' |q_i = Q(s)\right]$, $L_{1N}(s)$, and $L_{2N}(s)$ are bounded and satisfy*

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N} E[\Lambda_i(\beta_i)|q_i = Q(s)] &\rightarrow \Gamma(s)\\ \frac{1}{N}\sum_{i=1}^{N} E\left[m_i(\beta_i) m_i(\beta_i)' |q_i = Q(s)\right] &\rightarrow H(s)\\ L_{1N}(s) &\rightarrow L_1(s)\\ L_{2N}(s) &\rightarrow L_2(s)\end{aligned}$$

where $\Gamma(\cdot)$, $H(\cdot)$, $L_1(\cdot)$, and $L_2(\cdot)$ *are uniformly bounded matrix-valued processes that are piece-wise twice continuously differentiable on $(0,1)$ with at most a finite number of discontinuities, and left and right limits everywhere.*

Conditions 1.1" and 1.2" replace the independence assumption with the weak dependence. Note that we do not require W_i to be stationary unless it includes q_i. The first two convergences in Condition 1.5" are the same as in Section 2. The last two require additionally that the covariance term satisfies a certain summability. Under these assumptions, Lemma 7 in the following derives the time series version of the limiting Gaussian process. Define

$$\tilde{H}(s) = H(s) + 2(L_1(s) + L_2(s)).$$

Lemma 7 *Suppose Condition 1 holds with 1.1, 1.2 and 1.5 replaced with 1.1", 1.2", and 1.5". Then in the unstable model (2),*

$$N^{-1/2}\sum_{i=1}^{[sN]} m_{(i)}\left(\beta_{(i)}\right) \Rightarrow \int_0^s \tilde{H}^{1/2}(l)dW(l)$$

for $s \in [0,1]$ and

$$\sup_{s\in[0,1]}\sup_{\beta\in B_{\varepsilon_N}(\bar{\beta})}\left\|N^{-1}\sum_{i=1}^{[sN]}\Lambda_{(i)}(\beta) - \int_0^s \Gamma(l)dl\right\| \xrightarrow{p} 0$$

for any $\varepsilon_N \to 0$ as $N \to \infty$ where B_{ε_N} is the open ball centered at $\bar{\beta}$ with radius ε_N. Moreover, in the unstable model (2) with a known $\bar{\beta}$,

$$N^{-1/2} \sum_{i=1}^{[sN]} \xi_{(i)} m_{(i)} \left(\bar{\beta}\right) \Rightarrow \tilde{G}(s)$$

where $\tilde{G}(\cdot)$ is defined as in (11).

Based on Lemma 7 and Condition 2, the asymptotic analysis introduced in Section 2 can be again performed with the time transformation (12) constructed as a function of $\Gamma(\cdot)$ and $\tilde{H}(\cdot)$.

To implement the efficient tests φ_r, φ_d, and φ_b, we need to consistently estimate the variance-covariance matrix functions. The kernel estimator (23) for $\Gamma(\cdot)$ and $H(\cdot)$ are still applicable and consistent by an argument similar to that in Lemma 4 (see also the proof of the following theorem). To estimate $L_1(s)$, we can rewrite $L_{1N}(s)$ as

$$L_{1N}(s) = \sum_{k=1}^{N-1} \frac{1}{N} \sum_{i=k+1}^{N} E\left[m_i(\beta_i) m_{i-k}\left(\beta_{i-k}\right) \mathbf{1}\left[q_{i-k} \leq Q(s)\right] | q_i = Q(s)\right]$$

and define, for $1 \leq p_N < N$,

$$L_{1p_N}(s) = \sum_{k=1}^{p_N} \frac{1}{N} \sum_{i=k+1}^{N} E\left[m_i(\beta_i) m_{i-k}\left(\beta_{i-k}\right) \mathbf{1}\left[q_{i-k} \leq Q(s)\right] | q_i = Q(s)\right].$$

Then, the estimator of $L_1(s)$ can be constructed as

$$\hat{L}_{1p_N}(s) = \sum_{k=1}^{p_N} \frac{1}{N b_N} \sum_{i=k+1}^{N} m_i\left(\hat{\beta}\right) m_{i-k}\left(\hat{\beta}\right) \mathbf{1}\left[q_{i-k} \leq \hat{Q}(s)\right] K\left(\frac{R_i/N - s}{b_N}\right)$$

where R_i is the rank of q_i, and $\hat{L}_{2p_N}(\cdot)$ as well in a similar fashion. The desired consistency of $\hat{L}_{1p_N}(s)$ and $\hat{L}_{2p_N}(s)$ is established in the following theorem.

Lemma 8 *Suppose Conditions 1-3 hold with 1.1, 1.2 and 1.5 replaced with 1.1", 1.2" and 1.5", $p_N \to \infty$, and $p_N^4\left(N^{-1}b_N^{-1} + b_N^4\right) \to 0$, then for any compact subset $S \subset (0,1)$ on which $L_i(\cdot)$ is continuous,*

$$\sup_{s \in S} \left\| \hat{L}_{ip_N}(s) - L_i(s) \right\| = o_p(1)$$

for $i = 1$ and 2.

Lemma 8 proves that $\hat{L}_{1p_N}(\cdot)$ and $\hat{L}_{2p_N}(\cdot)$ are uniformly consistent. Then, the time series adjusted analogues of $\mathcal{T}\hat{G}_N$ and $\mathcal{T}\hat{G}_N^*$, denoted by $\mathcal{T}\hat{G}_N^{TS}$ and $\mathcal{T}\hat{G}_N^{*TS}$, can be constructed by replacing $\hat{H}(\cdot)$ with $\hat{H}(\cdot)+2\left(\hat{L}_{1p_N}(\cdot)+\hat{L}_{2p_N}(\cdot)\right)$. Thus, the previously suggested tests are implemented again, as summarized by the following theorem.

Theorem 5 *Suppose Conditions 1-3 hold with 1.1, 1.2 and 1.5 replaced with 1.1", 1.2" and 1.5". Then in the unstable model (2) with $M = h(1)^{-1/2} I_k$, we have, for $s \in [0,1]$, (i) under the null hypothesis $e_1'\beta = \beta_{10}$,*

$$e_1'\mathcal{T}\hat{G}_N^{TS}(s) \Rightarrow W_1(s) + b_1 s + d_1 \min(s, g(r))$$

and

$$e_1'\mathcal{T}\hat{G}_N^{*TS}(s) \Rightarrow W_1(s) - sW_1(1) + d_1(\min(s, g(r)) - sg(r)).$$

6 Applications

In this section, we illustrate the previous established results in two empirically relevant models: the binary choice model and the panel data model with fixed effects. The first one is covered by Section 5.2, and the second one is shown to fit the baseline GMM threshold model (1) after taking a subsample. In both models, we focus on the testing problem (6), i.e., $H_0: r = r_0$ against $H_1: r \neq r_0$ and demonstrate how the canonical observation G_1^* can be obtained by considering the partial sum of the re-ordered moment functions.

6.1 Binary Choice Model

The binary choice/single index model is extensively studied in the econometric literature, for example, Han (1987), Ichimura (1993), and Klein and Spady (1993). Estimation and inference are sometimes conducted in subsamples selected based on a continuous random variable (cf. Lee, Rosenzweig, and Pitt (1997)). However, a systematic approach about how to split the sample has not been developed. In this section, we show that the test φ_r is applicable.

For expositional ease, we assume that $\{x_i, q_i, u_i\}$ is i.i.d. and satisfies

$$\begin{aligned} y_i &= \mathbf{1}[x_i'\beta_i + u_i \geq 0] \\ \beta_i &= \bar{\beta} + bN^{-1/2} + dN^{-1/2}\mathbf{1}[q_i \leq Q(r)]. \end{aligned}$$

We also assume that u_i is independent of x_i and q_i, and the CDF of $-u_i$, denoted by $F_u(\cdot)$, is twice continuously differentiable with derivative $f_u(\cdot)$. Other primitive assumptions are collected in Condition C in Appendix C. We show that these primitive assumptions are sufficient for Conditions 1-4 so that Theorem 4 in Section 5.2 holds with a suitably chosen moment function.

Since F_u is unspecified, we need some normalization on β for point identification. We assume that x_i has no intercept, and $\beta_k = 1$ for $k > 2$. Following Andrews (1994), we trim the data by restricting $x_i'\bar{\beta} \in \mathcal{X}_N^*$, an increasing sequence of compact subsets of $\mathbb{R}$. Denote the trimming function as $\xi_i = \mathbf{1}\left[x_i'\bar{\beta} \in \mathcal{X}_N^*\right]$.

To fit in the analysis in Section 5.2, the unknown infinite dimensional nuisance parameter can be defined as $\tau_1(z) = F_u(z)$ and $\tau_2(z) = E\left[X_i | X_i'\bar{\beta} = z\right]$. If $\tau_1(\cdot)$ and $\tau_2(\cdot)$ were known, the moment function can be chosen as $\xi_i(y_i - \tau_1(x_i'\beta))(x_i - \tau_2(x_i'\beta))$. Then after re-ordering according to q_i, we have, in the unstable model (2) with a known $\bar{\beta}$,

$$
\begin{aligned}
G_N^{BC}(s) &\equiv N^{-1/2}\sum_{i=1}^{[sN]} \xi_{(i)}\left(y_{(i)} - \tau_1\left(x_{(i)}'\bar{\beta}\right)\right)\left(x_{(i)} - \tau_2\left(x_{(i)}'\bar{\beta}\right)\right) \\
&= N^{-1/2}\sum_{i=1}^{[sN]} \xi_{(i)}\left(y_{(i)} - \tau_1\left(x_{(i)}'\beta_{(i)}\right)\right)\left(x_{(i)} - \tau_2\left(x_{(i)}'\bar{\beta}\right)\right) \\
&\quad + N^{-1/2}\sum_{i=1}^{[sN]} \xi_{(i)} f_u\left(x_{(i)}'\dot{\beta}_{(i)}\right)\left(x_{(i)} - \tau_2\left(x_{(i)}'\bar{\beta}\right)\right) x_{(i)}'\left(\beta_{(i)} - \bar{\beta}\right)
\end{aligned}
$$

where $\dot{\beta}_{(i)}$ lies between $\beta_{(i)}$ and $\bar{\beta}$. Then, Lemma 5 implies that, with $M = I_k$,

$$
\begin{aligned}
& N^{-1/2}\sum_{i=1}^{[sN]} \xi_{(i)}\left(y_{(i)} - \tau_1\left(x_{(i)}'\bar{\beta}\right)\right)\left(x_{(i)} - \tau_2\left(x_{(i)}'\bar{\beta}\right)\right) \\
\Rightarrow & \int_0^s H^{1/2}(l)\, dW(l) + \left(\int_0^s \Gamma(l)\, dl\right) b + \left(\int_0^{\min(s,r)} \Gamma(l)\, dl\right)
\end{aligned}
$$

where

$$
\begin{aligned}
H(l) &= E\left[\xi_i\left(\tau_1(x_i'\beta_i) - \tau_1(x_i'\beta_i)^2\right) | q_i = Q(l)\right] \\
\Gamma(l) &= E\left[\xi_i f_u(x_i'\beta_i)(x_i - \tau_2(x_i'\beta_i)) x_i' | q_i = Q(l)\right].
\end{aligned}
$$

Thus, our transformation can be implemented as before.

In practice, the unknown ingredients need to be estimated. Let $\hat{\beta}$ be some preliminary$\sqrt{N}$-consistent estimator of β. In the following analysis, we use the estimator proposed by Klein and Spady (1993), denoted by $\hat{\beta}_{KS}$. The kernel estimator of $\tau_1(\cdot)$, $\tau_2(\cdot)$, and $f_u(\cdot)$ can then be constructed, for $x_i'\hat{\beta}_{KS} \in \mathcal{X}_N^*$, as

$$\begin{aligned}
\hat{\tau}_1\left(x_i'\hat{\beta}_{KS}\right) &= \frac{\sum_{j=1,j\neq i}^{N} y_j K_1\left(\frac{x_j'\hat{\beta}_{KS}-x_i'\hat{\beta}_{KS}}{b_{N1}}\right)}{\sum_{j=1,j\neq i}^{N} K_1\left(\frac{x_j'\hat{\beta}_{KS}-x_i'\hat{\beta}_{KS}}{b_{N1}}\right)} \\
\hat{\tau}_2\left(x_i'\hat{\beta}_{KS}\right) &= \frac{\sum_{j=1,j\neq i}^{N} x_i K_1\left(\frac{x_j'\hat{\beta}_{KS}-x_i'\hat{\beta}_{KS}}{b_{N1}}\right)}{\sum_{j=1,j\neq i}^{N} K_1\left(\frac{x_j'\hat{\beta}_{KS}-x_i'\hat{\beta}_{KS}}{b_{N1}}\right)}
\end{aligned}$$

and

$$\hat{f}_u\left(x_i'\hat{\beta}_{KS}\right) = \frac{1}{Nb_{N1}} \sum_{j=1,j\neq i}^{N} K_1\left(\frac{x_j'\hat{\beta}_{KS}-x_i'\hat{\beta}_{KS}}{b_{N1}}\right).$$

Then, $\Gamma(\cdot)$ can be estimated by

$$\hat{\Gamma}(s) = \frac{\sum_{i=1,i\neq[sN]}^{N} \hat{\xi}_{(i)} \hat{f}\left(x_{(i)}'\hat{\beta}_{KS}\right)\left(x_{(i)} - \hat{\tau}_2\left(x_{(i)}'\hat{\beta}_{KS}\right)\right) x_{(i)}' K_2\left(\frac{i/N-s}{b_{N2}}\right)}{\sum_{i=1,i\neq[sN]}^{N} K_2\left(\frac{i/N-s}{b_{N2}}\right)}$$

where $\hat{\xi}_i = \mathbf{1}\left[x_i'\hat{\beta} \in \mathcal{X}_N^*\right]$, and $\hat{H}(\cdot), \hat{g}(\cdot)$, and $\hat{h}(\cdot)$ can be constructed in a similar fashion. The kernel functions and the bandwidths are different for estimating the nuisance parameters and $\Gamma(\cdot)$. See Condition C for more details. Plug in the above estimators, we arrive at the binary choice analogue of (34), that is,

$$\begin{aligned}
\mathcal{T}\hat{G}_N^{*BC}(s) &= \sqrt{\hat{h}(1)} N^{-1/2} \times \\
&\left\{ \sum_{i=1}^{\left[\hat{g}^{-1}(s)N\right]} \hat{\xi}_{(i)} \frac{\widehat{\partial g}(i/N)}{\partial l} \hat{\Gamma}(i/N)^{-1} \left(y_{(i)} - \hat{\tau}_1\left(x_{(i)}'\hat{\beta}_{KS}\right)\right)\left(x_{(i)} - \hat{\tau}_2\left(x_{(i)}'\hat{\beta}_{KS}\right)\right) \right. \\
&\left. -s \sum_{i=1}^{N} \hat{\xi}_{(i)} \frac{\widehat{\partial g}(i/N)}{\partial l} \hat{\Gamma}(i/N)^{-1} \left(y_{(i)} - \hat{\tau}_1\left(x_{(i)}'\hat{\beta}_{KS}\right)\right)\left(x_{(i)} - \hat{\tau}_2\left(x_{(i)}'\hat{\beta}_{KS}\right)\right) \right\},
\end{aligned}$$

and then Theorem 4 leads to the following corollary.

Corollary 2 *Suppose Condition C in Appendix C is satisfied. Then, Conditions 1-4 are satisfied. Moreover, in the unstable model (2) with* $M = h(1)^{-1/2} I_k$*, we have, for* $s \in [0,1]$*,*

$$e_1' \mathcal{T} \hat{G}_N^{*BC}(s) \Rightarrow G_1^*(s).$$

Given Corollary 2, the test φ_r can be implemented with G_1^* replaced with $e_1' \mathcal{T} \hat{G}_N^{*BC}$. The above choice of the moment function leads to a relatively simple implementation, which, however, may suffer some loss of efficiency. As discussed in Section 3.4, the efficiency of the test φ_r can be established when the moment function is the score function. Given our parametric assumption, the log-likelihood conditional on $\{x_i, q_i\}_{i=1}^N$ is written as

$$\begin{aligned} \mathcal{L}_N(b,d,r) &= \sum_{i=1}^N y_i \log \tau_1 \left(x_i' \left(\bar{\beta} + bN^{-1/2} + dN^{-1/2} \mathbf{1} \left[q_i \le \hat{Q}(r) \right] \right) \right) \\ &\quad + \sum_{i=1}^N (1-y_i) \log \left(1 - \tau_1 \left(x_i' \left(\bar{\beta} + bN^{-1/2} + dN^{-1/2} \mathbf{1} \left[q_i \le \hat{Q}(r) \right] \right) \right) \right). \end{aligned}$$

Then, by choosing the moment function as the partial derivation of $\mathcal{L}_N$ with respect to b and d, we can construct, with known $\bar{\beta}$, $f(\cdot)$, $\tau_1(\cdot)$, and $\tau_2(\cdot)$,

$$G_N^{BC_2}(s) \equiv N^{-1/2} \sum_{i=1}^{[sN]} \xi_{(i)} \frac{\left(y_{(i)} - \tau_1 \left(x_{(i)}' \bar{\beta} \right) \right) f_u \left(x_{(i)}' \bar{\beta} \right)}{\tau_1 \left(x_{(i)}' \bar{\beta} \right) \left(1 - \tau_1 \left(x_{(i)}' \bar{\beta} \right) \right)} \left(x_{(i)} - \tau_2 \left(x_{(i)}' \bar{\beta} \right) \right).$$

Replacing the unknown elements with their estimators, we can obtain the implementable version of $G_N^{BC_2}$, whose finite sample transformed analogue $\mathcal{T} \hat{G}_N^{*BC_2}$ shares the same convergence as in Corollary 2.

6.2 Linear Panel Data Model with Fixed Effects

In addition to cross-sectional data, the panel data threshold model with fixed effects is also widely applied in empirical studies. Recent examples can be found in Fazzari et al. (1988), Girma (2005), and Giuliano and Ruiz-Arranz (2009). The state-of-the-art method is developed by Hansen (1999), who assumes that the true threshold remains unchanged in different time periods. It is controversial that the marginal distribution of the threshold variable q_{it} remains unchanged across time t. For instance, in Fazzari et al. (1988), q_{it} is the dividend-income ratio of firms whose distribution is heavily affected by macroeconomics. It is then more convenient to model the threshold as $Q_t(r)$ for a fixed r and a time-varying quantile function of q. For expositional ease and to reduce notational burden, we illustrate only the case where $T = 2$ and leave the general T case to Appendix D.

We focus on the linear regression context and start with the following static threshold model

$$y_{it} = \mu_i + x_{it}'\beta + x_{it}'\delta\mathbf{1}\left[q_{it} \leq Q_t(r)\right] + u_{it}, \quad t = 1, 2$$

where μ_i is the unobserved fixed effect. We allow x_{it} to include a time dummy, i.e., $\mathbf{1}[t = 2]$. Suppose we observe a balanced panel $\{y_{it}, x_{it}, q_{it}\}$ and remain interested in testing

$$H_0 : r = r_0 \text{ against } H_1 : r \neq r_0.$$

For brevity, we assume (q_{it}, x_{it}, u_{it}) is i.i.d. across i but not across t, and build our asymptotic analysis on $N \to \infty$.[6] Other primitive assumptions are collected in Condition D in Appendix D, which is sufficient for Conditions 1-3 to hold.

Due to the existence the fixed effects, we cannot directly apply the previously established results. As an alternative, we take the subsample such that under the null hypothesis $r = r_0$, the threshold variables in two periods are either both above or both below the empirical quantile functions evaluated at r_0. To be precise, include the i-th observation if $\mathbf{1}\left[q_{i1} \leq \hat{Q}_1(r_0)\right] = \mathbf{1}\left[q_{i2} \leq \hat{Q}_2(r_0)\right]$ where $\hat{Q}_t(\cdot)$ denotes the empirical quantile function of q_{it} for $t = 1$ and 2. In such a subsample, taking time difference eliminates μ_i.

Define $\Delta x_i = x_{i1} - x_{i2}$ and similarly for Δy_i and Δu_i. Also define $N(r_0) = \sum_{i=1}^{N} \mathbf{1}\left[\hat{A}_i\right]$ for $\hat{A}_i = \{\mathbf{1}\left[q_{i1} \leq \hat{Q}_1(r_0)\right] = \mathbf{1}\left[q_{i2} \leq \hat{Q}_2(r_0)\right]\}$, and assume

$$N(r_0)/N \to P(\mathbf{1}[A_i]) > 0$$

where $A_i = \{\mathbf{1}[q_{i1} \leq Q_1(r_0)] = \mathbf{1}[q_{i2} \leq Q_2(r_0)]\}$. Then, we obtain a subsample model under the null hypothesis, that is, for $t = 1$ and 2,

$$\Delta\dot{y}_i = \Delta\dot{x}_i'\beta + \Delta\dot{x}_i'\delta\mathbf{1}\left[\dot{q}_{it} \leq \hat{Q}_t(r_0)\right] + \Delta\dot{u}_i \tag{38}$$

where the variables with a 'dot' denote the subsample.

To perform the previous analysis, we order this subsample ascendingly in $\dot{q}_{i1}$.[7] Consider the moment function as

$$m_i(\beta) = \Delta\dot{x}_i\left(\Delta y_i - \Delta\dot{x}_i'\beta\right).$$

[6] Hansen (1999) assumes that u_{it} is i.i.d. across both i and t and independent of $\{x_{it}, q_{it}\}_{t=1}^{N}$ to construct confidence intervals for r.

[7] We may also re-order the data according to $\dot{q}_{i2}$ and arrive at another confidence interval.

The following lemma establishes a conditional moment version of Lemma 1. Let $Q_1^c(\cdot)$ denote the quantile function of q_{i1} in the subpopulation where $\mathbf{1}[q_{i1} \leq Q_1(r_0)] = \mathbf{1}[q_{i2} \leq Q_2(r_0)]$ holds, and redefine

$$\begin{aligned} \Gamma(l) &= \Gamma_1(l)\,\mathbf{1}[Q_1^c(l) \leq Q_1(r_0)] + \Gamma_2(l)\,\mathbf{1}[Q_1^c(l) > Q_1(r_0)] \\ H(l) &= H_1(l)\,\mathbf{1}[Q_1^c(l) \leq Q_1(r_0)] + H_2(l)\,\mathbf{1}[Q_1^c(l) > Q_1(r_0)] \end{aligned}$$

where for $l \leq r_0^c$ that satisfies $Q_1^c(r_0^c) = Q_1(r_0)$,

$$\begin{aligned} \Gamma_1(l) &= E[\Delta x_i \Delta x_i' | q_{i2} \leq Q_2(r_0), q_{i1} = Q_1^c(l)] \\ H_1(l) &= E\left[\Delta x_i \Delta x_i' \Delta u_i^2 | q_{i2} \leq Q_2(r_0), q_{i1} = Q_1^c(l)\right] \end{aligned}$$

and for $l > r_0^c$

$$\begin{aligned} \Gamma_2(l) &= E[\Delta x_i \Delta x_i' | q_{i2} > Q_2(r_0), q_{i1} = Q_1^c(l)] \\ H_2(l) &= E\left[\Delta x_i \Delta x_i' \Delta u_i^2 | q_{i2} > Q_2(r_0), q_{i1} = Q_1^c(l)\right]. \end{aligned}$$

Lemma 9 *Suppose Condition D in Appendix D holds. Then, Condition 1 holds and hence*

$$N(r_0)^{-1/2} \sum_{i=1}^{[sN(r_0)]} \Delta \dot{x}_{(i)} \Delta \dot{u}_{(i)} \Rightarrow \int_0^s H^{1/2}(l)\, dW(l)$$

for $s \in [0,1]$ and

$$N(r_0)^{-1} \sum_{i=1}^{[sN(r_0)]} \Delta \dot{x}_{(i)} \Delta \dot{x}_{(i)}' \xrightarrow{p} \int_0^s \Gamma(l)\, dl$$

uniformly in $s \in [0,1]$. Moreover, in the unstable model (2) with a known $\bar{\beta}$,

$$N^{-1/2} \sum_{i=1}^{[sN]} \Delta \dot{x}_{(i)} \left(\Delta \dot{x}_{(i)} - \Delta \dot{y}_{(i)} \hat{\beta} \right) \Rightarrow \tilde{G}(s)$$

where $\hat{\beta}$ is the coefficient of regressing $\Delta \dot{y}_i$ on $\Delta \dot{x}_i$, and $\tilde{G}(\cdot)$ is defined as in (11).

Given the above lemma, we can perform the same asymptotic analysis as in Section 2 to the time-differenced model (38) with $\Gamma(\cdot)$ and $H(\cdot)$ redefined above. For implementation, although $\Gamma(\cdot)$ and $H(\cdot)$ are discontinuous at r_0^c, we can use the kernel estimators (23) as before since our time transformation allows for a finite number of discontinuities.[8] Adjust

[8] To improve finite sample performance, we may resort to the local linear kernel estimator for $\hat{\Gamma}, \hat{H}, \hat{h}$, and $\hat{g}$ (see, for example, Fan and Gijbels (1996)).

the time transformation accordingly and plug in the corresponding estimators, we obtain an implementable panel data version of $\mathcal{T}\hat{G}_N^*$, that is,

$$\begin{aligned}\mathcal{T}\hat{G}_N^{*Panel}(s) &= \sqrt{\hat{h}(1)}N(r_0)^{-1/2}\sum_{i=1}^{\left[\hat{g}^{-1}(s)N(r_0)\right]}\frac{\widehat{\partial g}(i/N(r_0))}{\partial l}\hat{\Gamma}(i/N(r_0))^{-1}\Delta\dot{x}_{(i)}\widehat{\Delta\dot{u}}_{(i)} \\ &\quad -s\sqrt{\hat{h}(1)}N(r_0)^{-1/2}\sum_{i=1}^{N(r_0)}\frac{\widehat{\partial g}(i/N(r_0))}{\partial l}\hat{\Gamma}(i/N(r_0))^{-1}\Delta\dot{x}_{(i)}\widehat{\Delta\dot{u}}_{(i)}\end{aligned}$$

where $\widehat{\Delta\dot{u}}_i$ is the residual of regressing $\Delta\dot{y}_i$ on $\Delta\dot{x}_i$. The desired convergence of $e_1'\mathcal{T}\hat{G}_N^{*Panel}$ is established in the following corollary.

Corollary 3 *Suppose Condition D is satisfied, then Conditions 1-3 are satisfied. Moreover, under the null hypothesis $r = r_0$, in the unstable model (2) with $M = h(1)^{-1/2} I_k$, we have, for $s \in [0,1]$,*

$$e_1'\mathcal{T}\hat{G}_N^{*Panel}(s) \Rightarrow W_1(s) - sW_1(1) + d_1\sqrt{h(1)}(\min(s, g(r_0)) - sg(r_0)).$$

By Corollary 3, the test φ_r is readily applicable by replacing $e_1'\mathcal{T}\hat{G}_N^*$ with $e_1'\mathcal{T}\hat{G}_N^{*Panel}$. Our approach can be generalized to cover the dynamic threshold model if an instrument variable exists. Suppose we also observe the data at $t = 0$ and x_{i0} is independent of u_{it} for $t = 1$ and 2. If x_{it} involves a lagged dependent variable, we can use x_{i0} as the instrument for $\Delta\dot{x}_i$. Then, the test φ_r is again implementable by considering $x_{i0}(\Delta\dot{y}_i - \Delta\dot{x}_i'\beta)$ as the moment function.

7 Monte Carlo Experiments

In this section, we implement the tests discussed before in several Monte Carlo experiments. Through all experiments, we use the standard normal kernel.

Consider the following DGP

$$y_i = \bar{\beta} + x_i b_1 N^{-1/2} + d_1 N^{-1/2} x_i \mathbf{1}[q_i \leq Q(r)] + u_i \tag{39}$$

where x_i, u_i and q_i are jointly normal with the following covariance structures: denote $\Phi(\cdot)$ as the CDF of a standard normal random variable,

S1 $(x_i, q_i, u_i) \sim^{iid} \mathcal{N}(0, I_3)$.

S2 (x_i, q_i, u_i) is i.i.d. multivariate normal such that $(q_i, u_i) \sim \mathcal{N}(0, I_2)$, x_i is independent of u_i, and $x_i|q_i \sim \mathcal{N}(0, (|\Phi(q_i)/N - 0.5| + 1)^2)$.

S3 (x_i, q_i, v_i, u_i) is i.i.d. multivariate normal such that $v_i|q_i \sim \mathcal{N}(0, (|\Phi(q_i)/N - 0.5| + 1)^2)$, $x_i = 0.5q_i + \sqrt{1 - 0.5^2}v_i$, and $u_i|x_i \sim \mathcal{N}(0, (1 + |x_i|)^2)$.

The first example corresponds to the canonical case, where (x_i, u_i) is i.i.d. and independent of q_i. In S2, we introduce the non-constant $\Gamma(\cdot)$, and in S3, we include time-varying $\Gamma(\cdot)$ and $H(\cdot)$.

We consider the testing problems (4), (5), and (6), that is, (i) $H_0 : r = r_0$ against $H_1 : r \neq r_0$; (ii) $H_0 : d_1 = 0$ against $H_1 : d_1 \neq 0$; and (iii) $H_0 : b_1 = 0$ against $H_1 : b_1 \neq 0$. In the first two problems, we impose invariance to (26), and in the third, we impose invariance to (27).

Tables 3-5 depict the finite sample rejection probabilities under the null hypothesis of the tests φ_r,[9], φ_d, and φ_b, respectively. These results show that our approach performs well in finite samples in controlling size. Table 6 lists the rejection probabilities of the test φ_r with different bandwidths. Together with the third to fifth columns of Table 3 for Model S3, these results imply that our tests are not sensitive to the choice of bandwidth. Some unreported results indicate that our method is also robust to the choice of kernel function.

After demonstrating the rejection probabilities under the null hypothesis, we next examine the finite sample power of our approach. We focus on problem (6) and consider the DGP model (39) with variance structure S1. Table 7 depicts the rejection probabilities of the test φ_r under the alternative hypothesis $r = r_1$ with $d_1 = 8$ and $r_0 = 0.4$. For comparison, we also report the power of the least squares test (14) introduced in Section 2.5, denoted as φ_{LS}. We show that although the uniform size constraint results in some power loss compared with φ_{LS}, the test φ_r still has good finite sample powers. Note that the weighting function w puts more weight on small breaks and less weight on large ones. When the sample size is small, i.e., $N = 100$, the break magnitude is relatively large, and hence, our test is expected to have a small power against the particular alternative with $d_1 = 8$.

[9] For comparison with the finite sample power in Table 7, we do not include $x_i\mathbf{1}\left[q_i \leq \hat{Q}(r_0)\right]$ as a regressor to estimate $H(\cdot)$. In practice, we do recommend to include it. Unreported Monte Carlo results show that including $x_i\mathbf{1}\left[q_i \leq \hat{Q}(r_0)\right]$ leads to a better estimate of $H(\cdot)$ and a better finite sample size control when the sample size is small (relative to the standard deviation of the error term), i.e., $N = 100$. When N is 500 or larger, this improvement is quite small.

Table 3: Small sample rejection probabilities of the test φ_r

DGP	d_1	$r_0 = 0.4$			$r_0 = 0.2$		
		N=100	N=500	N=1000	N=100	N=500	N=1000
S1	4	0.026	0.040	0.050	0.034	0.045	0.041
	8	0.023	0.032	0.043	0.031	0.044	0.041
	12	0.021	0.028	0.040	0.024	0.040	0.033
	16	0.021	0.027	0.039	0.018	0.029	0.032
S2	4	0.024	0.038	0.045	0.036	0.045	0.045
	8	0.024	0.033	0.043	0.029	0.039	0.038
	12	0.020	0.029	0.043	0.017	0.031	0.032
	16	0.017	0.025	0.039	0.011	0.022	0.025
S3	4	0.018	0.037	0.042	0.025	0.041	0.040
	8	0.018	0.035	0.044	0.028	0.040	0.042
	12	0.019	0.033	0.044	0.028	0.039	0.039
	16	0.017	0.031	0.041	0.026	0.038	0.037

Note: Significance level is 5%. Data are generated from (39) with the variance structures S1-S3. Based on the rule-of-thumb bandwidth and 5,000 replications.

Table 4: Small sample rejection probabilities of the test φ_d

DGP	N=100	N=500	N=1000
S1	0.024	0.043	0.045
S2	0.023	0.045	0.044
S3	0.017	0.041	0.043

Note: Significance level is 5%. Data are generated from (39) with $d_1 = 0$ and the variance structures S1-S3. Based on the rule-of-thumb bandwidth and 5,000 replications.

Table 5: Small sample rejection probabilities of the test φ_b

DGP	d_1	$r_0 = 0.5$			$r_0 = 0.25$		
		N=100	N=500	N=1000	N=100	N=500	N=1000
S1	1	0.039	0.036	0.039	0.040	0.036	0.039
	4	0.035	0.040	0.037	0.039	0.041	0.043
	8	0.028	0.034	0.036	0.032	0.040	0.040
	16	0.012	0.026	0.029	0.018	0.036	0.031
S2	1	0.041	0.041	0.038	0.042	0.042	0.040
	4	0.037	0.042	0.038	0.040	0.045	0.042
	8	0.032	0.040	0.038	0.029	0.043	0.038
	16	0.008	0.018	0.022	0.010	0.027	0.022
S3	1	0.044	0.038	0.036	0.043	0.039	0.038
	4	0.044	0.036	0.039	0.043	0.038	0.039
	8	0.046	0.038	0.040	0.045	0.039	0.041
	16	0.049	0.039	0.040	0.044	0.037	0.038

Note: Significance level is 5%. Data are generated from (39) with $b_1 = 0$ and the variance structures S1-S3. Based on the rule-of-thumb bandwidth and 5,000 replications.

Table 6: Small sample rejection probabilities of the test φ_r with different bandwidths

d_1	$b_N = b_N^*/2$			$b_N = 2b_N^*$		
	N=100	N=500	N=1000	N=100	N=500	N=1000
4	0.020	0.035	0.042	0.027	0.040	0.046
8	0.020	0.031	0.042	0.024	0.039	0.045
12	0.019	0.030	0.042	0.020	0.036	0.045
16	0.021	0.030	0.040	0.018	0.034	0.043

Note: Significance level is 5%. Data are generated from (39) with the variance structure S3 and $r_0 = 0.4$. The second to fourth columns use half of the rule-of-thumb bandwidth b_N^*, and the last three columns use twice the rule-of-thumb bandwidth. Based on 5,000 replications.

Table 7: Small sample powers of the test φ_r and the least squares test φ_{LS}

r_1	φ_r			φ_{LS}		
	N=100	N=500	N=1000	N=100	N=500	N=1000
0.2	0.063	0.380	0.463	0.729	0.778	0.772
0.4	0.023	0.032	0.043	0.047	0.049	0.051
0.6	0.547	0.712	0.733	0.668	0.730	0.732
0.8	0.297	0.621	0.688	0.839	0.881	0.874

Note: Significance level is 5%. Data are generated from (39) with the variance structure S1, $r_0 = 0.4$, $d_1 = 8$, and different values of r_1. Based on the rule-of-thumb bandwidth and 5,000 replications.

8 Empirical Examples

In this section, we apply the tests suggested in Section 4 to a Macroeconomic and a Microeconomic example. In both examples, we find that our approach provides substantially different results than the least squares method.

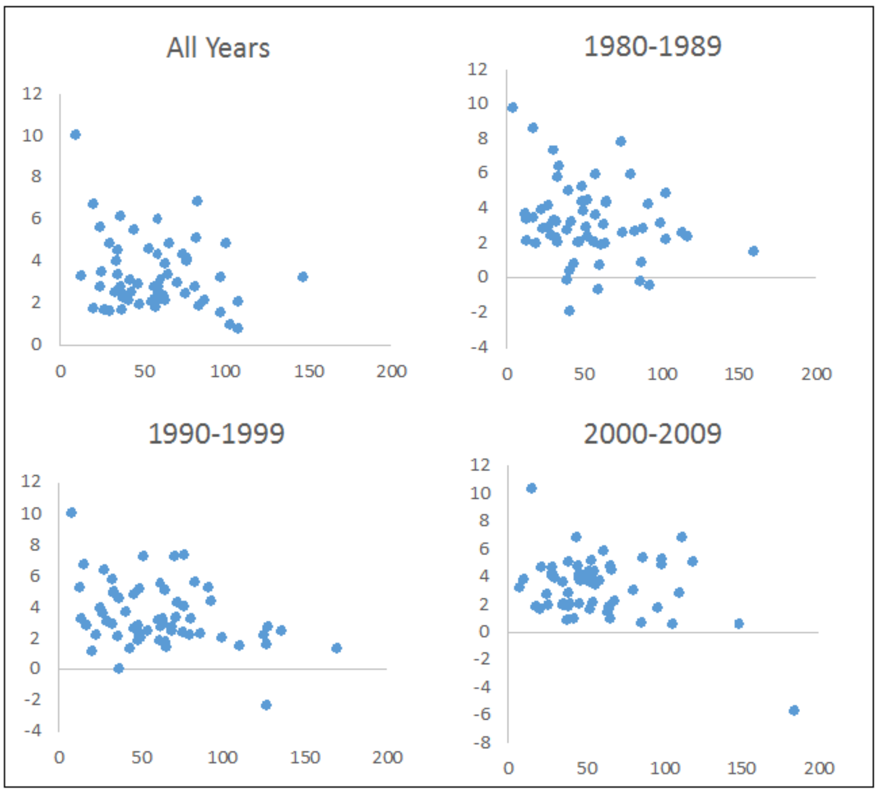
8.1 Public Debt and GDP Growth

Reinhart and Rogoff (2010) study the relationship between economic growth and public debt. They divide the data into four subsamples in terms of debt-to-GDP ratios: ≤30%, 30%-60%, 60%-90%, and >90% and find that the average annual GDP growth in the last group is substantially lower than in the other three groups. This method is criticized by Herndon, Ash, and Pollin (2014), who argue that (i) the relationship between public debt and GDP growth varies significantly by period and country; and (ii) the ad hoc sample splitting suffers from significant errors. To provide some evidence on whether such a relationship exists across countries, we collect the debt-to-GDP ratio data from another paper by the same authors and the GDP growth rate data from the World Bank's website to estimate the following simple regression model

$$y_i = \beta + \delta \mathbf{1}\left[q_i > \gamma\right] + u_i$$

where y_i denotes the average annual GDP growth rate over certain years and q_i the average total public debt-to-GDP ratio over the same period. All data are measured in percentage. We use the data from four time periods: 1980-1989, 1990-1999, 2000-2009, and all three combined.

Figure 2: Annual GDP Growth Rate Ordered by Debt-to-GDP Ratio



Note: The vertical axis is average annual GDP growth rate measured in percentage, and the horizontal is the public debt-to-GDP ratio measured in percentage. Data are from Reinhart and Rogoff (2011) and the website of the World Bank: http://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG.

Figure 2 depicts the scatter plot of the average GDP growth against the average debt-to-GDP ratio. The upper-left panel pools all the data and the others are based on three separate periods. The estimated transformed time $g(\cdot)$ is highly nonlinear for all four groups of data, suggesting that the $AveF$ test is not directly applicable due to the non-stationarity along the rank of q_i. Following the steps listed in Section 4, we apply the test φ_d to test if $\delta = 0$. At 5% level, our test does not reject the null that there is no threshold effect in all four groups of data. Hence, we favor the idea that a high public debt has no systematic effect on economic growth. This conclusion is also supported by several recent papers. See, for example, Panizza and Presbitero (2014), Eberhardt and Presbitero (2015), and Egert (2015).[10]

8.2 Tipping Point and Segregation

The second application is about segregation that arises from social interactions within a neighborhood. Card, Mas, and Rothstein (2008) empirically examine the theory proposed by Schelling (1971) that the white population decreases substantially once the minority share in a neighborhood exceeds a "tipping point." In particular, they specify the following threshold model

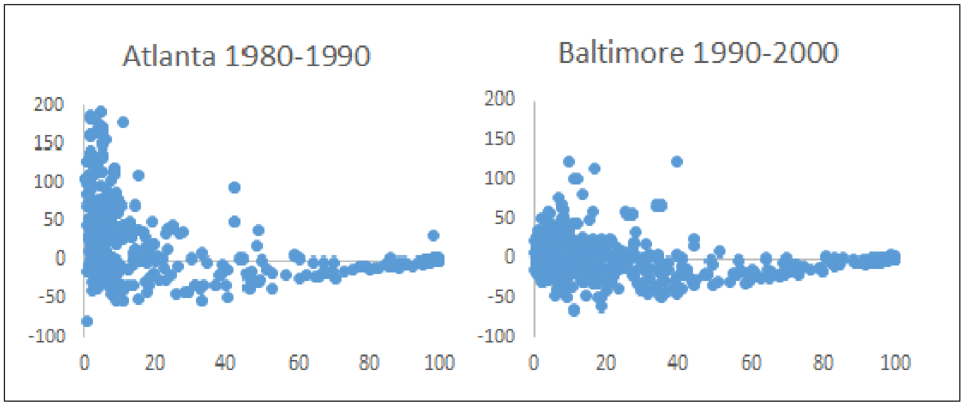
$$y_i = \beta + \delta \mathbf{1}[q_i > \gamma] + u_i$$

where for each neighborhood i (called tract in their paper) in a certain city, q_i denotes the minority share in percentage at the beginning of a certain time period and y_i the normalized white population change in percentage within this period. The data are collected from a variety of cities in three time periods: 1970-80, 1980-90, and 1990-2000. They apply the least squares method to estimate the tipping point γ. For most cities and all three periods, they find that white population flows exhibit the tipping-like behavior, with the estimated tipping points ranging from 5% to 20% across cities.

In their data, the number of neighborhoods varies a lot across cities. For illustration purposes, we apply our techniques to two cities with moderately large sample sizes: Atlanta, GA in the period 1980-90 (with sample size 579) and Baltimore, MD in the period 1990-2000 (with sample size 611). The scatter plot of the white population change against the minority share is depicted in Figure 3. The sample standard deviations of y_i are roughly 60 and 23

[10] The $AveF$ test with the bootstrap critical value proposed by Hansen (2000b) does not reject the null hypothesis either in all four groups. But the test with the original critical value rejects the null hypothesis for the data from 1980-1989 and 1990-1999.

Figure 3: White Population Change Ordered by Minority Share



Note: The vertical axis is normalized white population change measured in percent, and the horizontal is the minority share measured in percent. Data are from http://eml.berkeley.edu/~jrothst/.

in two areas, suggesting that δ has to be larger than approximately 12 and 8 percent for the least squares method to work reliably. However, the estimated threshold effect by least squares is roughly of the same magnitude so that the least squares confidence interval might deliver a coverage that is much lower than the nominal level, as we show in Section 2.5.

To provide a reliable analysis robust to a moderate δ, we apply our techniques to recast the problem into dating the timing of a structural break. To avoid outliers, we exclude the neighborhoods where the white population growth rate exceeds the sample mean plus five times the sample standard deviation and where the minority share at the starting point of that period is either 0 or 100%. We also restrict the tipping point to be between the 15% and 85% empirical quantiles. Table 7 presents the 95% confidence intervals of the tipping point produced by inverting the test φ_r (CI_φ_r) and the least squares test (14) introduced in Section 2.5 (CI_φ_{LS}). Our confidence intervals are much longer than those produced by the least squares method in both samples, indicating considerable uncertainty about the tipping point.

9 Conclusion

Threshold models have a wide variety of applications in economics. The least squares method for inference about the values of the parameters is limited to the linear regression case and

Table 8: Confidence Intervals of the Tipping Point

	Atlanta	Baltimore
CI_φ_r	$[5.43, 9.12]$	$[4.78, 17.38]$
CI_φ_{LS}	$[5.44, 5.87]$	$[7.69, 15.35]$
N	579	611

Note: Significance level is 5%. Data are from http://eml.berkeley.edu/~jrothst/.

relies on a key assumption that the coefficient change is large. This paper studies a general GMM threshold model and constructs tests that control size for a large range of moderate coefficient changes that are empirically relevant. These tests are statistically efficient in a well-defined sense. By inverting these tests, we can obtain the corresponding confidence intervals.

Our approach relies on the key idea of re-ordering. After re-ordering the data according to the values of the threshold variable, the threshold model shares a core similarity with the time series structural break model. We show that these two models are actually asymptotically equivalent in the sense that they induce the same limiting Gaussian process. Then, we can resort to the methods developed in the structural break literature. However, these methods are usually developed for stationary data only, but our rearranged data exhibit non-stationarity. This is because the threshold variable can be correlated with other variables, leading to time-varying moments with the rank of the threshold variable treated as time. To overcome this difficulty, we standardize the rearranged data by defining a time transformation so that we recover the same limiting distribution as in the stationary case. By doing so, all methods relying on stationarity can be implemented again.

Among all methods developed under stationarity, we choose an optimal test for inference about the threshold, the magnitude of the change, and the post-change level, respectively. When the moment function is the score function of some parametric model, we show that these tests with our time transformation are statistically efficient in a well-defined sense by applying Le Cam's limits of experiments theory. As demonstrated in Monte Carlo experiments, our approach outperforms the least squares method by a substantial margin.

Our approach is generalized to cover inference problems in many additional empirically relevant models. In particular, we consider the threshold models with a preliminary estimator of infinite dimensional nuisance parameters, with non-smooth moment functions, and with weakly dependent data. Our approach is illustrated through two widely studied models: the binary choice threshold model and the panel data threshold model with fixed effects. We also apply our techniques to two empirical questions: testing for the threshold effect of public debt on economic growth, and inference about the tipping point in segregation. In the first one, our method finds no threshold effect; and in the second one, our approach provides much longer confidence intervals than the least squares method.

Appendices

Appendix A: Details of Boundary Truncation

In the main text, Condition 1.5 imposes that $\Gamma(\cdot)$ and $H(\cdot)$ are uniformly bounded on $[0,1]$. This can be relaxed to allow $\Gamma(l)$ and $H(l)$ to be infinite at $l = 0$ and 1. One example for such unboundedness is the linear regression model (7), where x_i is a linear function of q_i and q_i has unbounded support. To avoid the unboundedness, we may trim the smallest and largest η percents, say, 1%, of the re-ordered data to avoid the boundary observations. Such a trimming ensures a good finite sample performance of the kernel estimator. In particular, we can pick a small $\eta \in (0, 1/2)$, and modify $h(l)$ and $g(l)$ into $h^{\eta}(l) = \int_{\eta}^{l}\left(e_1'\Gamma(s)^{-1}H(s)\Gamma(s)^{-1\prime}e_1\right)^{-1}ds$ and $g^{\eta}(l) = h^{\eta}(l)/h^{\eta}(1-\eta)$ for $l \in [\eta, 1-\eta]$. Then, the estimators of $\Gamma(\cdot)$ and $H(\cdot)$ can be adjusted as

$$
\begin{aligned}
\hat{\Gamma}(l) &= \frac{\sum_{i=[\eta N/2]+1, i\neq[lN]}^{[(1-\eta/2)N]}\Lambda_{(i)}\left(\hat{\beta}\right)K\left(\frac{i/N-l}{b_N}\right)}{\sum_{i=[\eta N/2]+1, i\neq[lN]}^{[(1-\eta/2)N]}K\left(\frac{i/N-l}{b_N}\right)} \\
\hat{H}(l) &= \frac{\sum_{i=[\eta N/2]+1, i\neq[lN]}^{[(1-\eta/2)N]}m_{(i)}\left(\hat{\beta}\right)m_{(i)}\left(\hat{\beta}\right)'K\left(\frac{i/N-l}{b_N}\right)}{\sum_{i=[\eta N/2]+1, i\neq[lN]}^{[(1-\eta/2)N]}K\left(\frac{i/N-l}{b_N}\right)},
\end{aligned}
$$

and we arrive at the truncated version of $\mathcal{T}\hat{G}_N$ and $\mathcal{T}\hat{G}_N^{*}$, that is,

$$
\begin{aligned}
\mathcal{T}\hat{G}_N^{\eta}(s) &= \sqrt{\hat{h}^{\eta}(1-\eta)}N^{-1/2}\sum_{i=[\eta N]+1}^{\left[\hat{g}^{\eta-1}(s)N\right]}\frac{\widehat{\partial g^{\eta}}(i/N)}{\partial l}\hat{\Gamma}(i/N)^{-1}m_{(i)}\left(\bar{\beta}\right) \\
\mathcal{T}\hat{G}_N^{\eta}(s) &= \sqrt{\hat{h}^{\eta}(1-\eta)}N^{-1/2}\sum_{i=[\eta N]+1}^{\left[\hat{g}^{\eta-1}(s)N\right]}\frac{\widehat{\partial g^{\eta}}(i/N)}{\partial l}\hat{\Gamma}(i/N)^{-1}m_{(i)}\left(\hat{\beta}\right) \\
&\quad -s\sqrt{\hat{h}^{\eta}(1-\eta)}N^{-1/2}\sum_{i=[\eta N]+1}^{\left[\hat{g}^{\eta-1}(1)N\right]}\frac{\widehat{\partial g^{\eta}}(i/N)}{\partial l}\hat{\Gamma}(i/N)^{-1}m_{(i)}\left(\hat{\beta}\right)
\end{aligned}
$$

where $\widehat{\partial g^\eta}(i/N)/\partial l$ denotes $\widehat{\partial g^\eta}(l)/\partial l|_{l=i/N}$, and $\hat{\beta}$ is constructed with the truncated data.

By adjusting the boundedness and smoothness imposed in Conditions 1.5, 2, and 3.1 accordingly, one can establish a similar version of Theorem 3 for $e_1'\mathcal{T}\hat{G}_N^\eta$ and $e_1'\mathcal{T}\hat{G}_N^{\eta *}$. Then, the tests φ_r, φ_d, and φ_b are again implementable.

Appendix B: Details of the Efficient Tests

This appendix reviews more details of the tests φ_r, φ_d, and φ_b as introduced in Section 2. As functions of G_1^*, the tests φ_r and φ_d are written as

$$\varphi_r(G_1^*) = \mathbf{1}\left[\frac{\int_{\Theta_1} f^*(G_1^*; r, c)\, w(c, r)\, dr dc}{\int_\Delta \lambda(c; r_0)\, f^*(G_1^*; r_0, c)\, dc} > 1\right] \tag{40}$$

and

$$\varphi_d(G_1^*) = \mathbf{1}\left[\int_{\underline{r}}^{\bar{r}} \frac{G_1^*(r)^2}{r(1-r)} dr > \xi\right] \tag{41}$$

where $w(\cdot,\cdot)$ and $\lambda(\cdot, r_0)$ are weighting functions defined on $\Theta_1 = \Delta \times [\underline{r}, \bar{r}]$ and Δ, respectively, for some compact set $\Delta \subset \mathbb{R}$. Regarding the problem (5), the efficient test φ_b involves a switching scheme and reads

$$\begin{aligned} \varphi_b(G_1) &= \mathbf{1}[\sup F_0(G_1) > \kappa_1] \times \mathbf{1}[|t_0(G_1)| > \xi_1] \\ &\quad + \mathbf{1}[\sup F_0(G_1) \leq \kappa_1] \times \mathbf{1}[R_0(G_1) > \xi_2] \end{aligned} \tag{42}$$

where κ_1 is the switching cutoff, ξ_1 and ξ_2 are some critical values, and $\sup F_0(G_1)$, $t_0(G_1)$, and $R_0(G_1)$ are functions of G_1. This test rejects H_0 if (i) $\sup F_0(G_1) > \kappa_1$ and $|t_0(G_1)| > \xi_1$, or (ii) $\sup F_0(G_1) \leq \kappa_1$ and the likelihood ratio statistic $R_0(G_1)$ is larger than ξ_2. Before introducing the values of κ_1, ξ_1, and ξ_2 and the expressions of these statistics, we first provide some comments on the test (40), which also apply to the other two tests.

First, the function λ is defined as the Approximately Least Favorable Distribution (ALFD), which plays a similar role as the Least Favorable Distribution (cf. chapter 3.8 of Lehmann and Romano (2005)). It is constructed to maintain the size constraint for all d in Δ. Elliott, Müller, and Watson (2015) provides an algorithm to numerically compute the ALFD, which we employ in the present paper.

Second, The weighting function w is chosen by the econometrician. In principle, any weighting function can be chosen, which leads to a different set of the ALFD λ. Monte Carlo results show that our tests are not sensitive to the choice of w. So for computational ease, given the exponential structure in the density (17), the test (40) is constructed with a normal density with mean zero and standard deviation $1/\sigma$ as the weight for d_1, so that the numerator can be substantially simplified.

The parameter σ measures the level of concentration of the weighting function. A smaller σ indicates that the weight on d_1 is more spread out and closer to uniform. In our experiments, we follow Elliott, Müller, and Watson (2015) to set $\sigma = 0.1$ and $\Delta = [-20, 20]$

Finally, we restrict $g(r) \in [0.15, 0.85]$. The truncation on the break point is standard in the literature of testing for structural breaks (cf. Andrews (1993) and Andrews and Ploberger (1994)). In the classic structural break literature where stationarity is assumed, the truncation imposed on r is appropriate since the information is accumulated linearly. In the non-stationary case, the information is accumulated linearly along the transformed time $g(\cdot)$. Hence, truncation of boundary points should be imposed on $g(r)$ instead of r.

For computational purpose, the integrals in (40), (41), and (42) are approximated numerically, and the parameter spaces Δ and Θ_1 are replaced with grids. In particular, the intervals $[0.15, 0.85]$ and $[-20, 20]$ are discretized into the grids $[0.15, 0.16, ..., 0.85]$ and $[-20, -19..., 20]$, respectively. Accordingly λ is tabulated as a 71×41 matrix, with each row corresponding to the ALFD for that particular r.

The implementable version of (40) is written as, for $\hat{g}(r_0) \in [0.15, 0.16, ..., 0.85]$,

$$\varphi_r\left(e_1'\mathcal{T}\hat{G}_N^*\right) = \mathbf{1}\left[\frac{\sum_{i=1}^{71}\frac{\left(v\left(e_1'\mathcal{T}\hat{G}_N^*, r_i, -20\right) - v\left(e_1'\mathcal{T}\hat{G}_N^*, r_i, 20\right)\right)}{10\sqrt{\left(r_i - r_i^2 + 0.1^2\right)}}\exp\left(\frac{e_1'\mathcal{T}\hat{G}_N^{*2}(r_i)}{2\left(r_i - r_i^2 + 0.1^2\right)}\right)}{\sum_{j=1}^{41}\lambda\left(c_j; \hat{g}(r_0)\right) f^*\left(e_1'\mathcal{T}\hat{G}_N^*; \hat{g}(r_0), c_j\right)} > 1\right]$$

where $v(G_1^*, r_i, c) = \Phi\left(\sqrt{(r - r^2 + 0.1^2)}\left(c - \frac{G_1^*(r)}{(r - r^2 + 0.1^2)}\right)\right)$ and Φ is the CDF of a standard normal. If $\hat{g}(r_0) \notin [0.15, 0.16, ..., 0.85]$, we may adjust λ by linear interpolation. The function v results from completing the squares, and the constant 0.1^2 and ± 20 correspond to σ^2 and the boundary of Δ, respectively.

The implementable version of (41) reads

$$\varphi_d\left(e_1'\mathcal{T}\hat{G}_N^*\right) = \mathbf{1}\left[\frac{1}{0.7N}\sum_{i=[0.15N]+1}^{i=[0.85N]}\frac{\left(e_1'\mathcal{T}\hat{G}_N^*(i/N)\right)^2}{(i/N)(1 - i/N)} > \xi\right].$$

The critical value ξ is 2.16, 2.88, and 4.72 for $\alpha = 1\%$, 5%, and 10%, respectively.

For the test (42), the exact expression of $\sup F_0, t_0$, and R_0 are as follows. For implementation, simply replace G_1 with $e_1'\mathcal{T}\hat{G}_N$.

$$\begin{aligned}
\sup F_0(G_1) &= \max_{16 \le l \le 85}\frac{100}{99}\frac{\left((100 - l)G_1\left(\frac{l-1}{100}\right) - (l - 1)\left(G_1(1) - G_1\left(\frac{l}{100}\right)\right)\right)^2}{(100 - l)(l - 1)} \\
S_0(l) &= \frac{100\left(100G_1\left(\frac{l}{100}\right) - lG(1)\right)^2}{l(100 - l)} + 100^2\frac{\left(lG_1\left(\frac{l-1}{100}\right) - (l - 1)G_1\left(\frac{l}{100}\right)\right)^2}{l(l - 1)}
\end{aligned}$$

Table 9: Approximate least favorable distribution for the test φ_b

j	1	2	3	4	5	6	7	8	9
p_j	0.588	0.123	0.067	0.057	0.038	0.032	0.026	0.020	0.009
a_j	15	85	85	20	75	20	20	75	45
b_j	85	85	85	74	85	74	74	82	59
$\sigma^2_{\delta,j}$	100	10	4	300	200	10	3	10	10
μ_j	20	5	3	16	28	9	6	7	11
j	10	11	12	13	14	15	16	17	18
p_j	0.009	0.008	0.006	0.005	0.004	0.004	0.002	0.001	0.001
a_j	70	15	15	60	80	60	83	85	75
b_j	74	19	24	69	82	69	84	85	82
$\sigma^2_{\delta,j}$	10	10	200	10	10	3	10	3	3
μ_j	9	5	28	12	11	8	13	15.5	13

$$\begin{aligned}
\hat{l}_0 &= \arg\max_{16\le l\le 85} S_0(l) \\
t_0(G_1) &= \frac{G_1(1) - G_1\left(\min\left(\hat{l}_0+1,85\right)/100\right)}{\sqrt{1-\min\left(\hat{l}_0+1,85\right)/100}} \\
R_0(G_1) &= \frac{\sum_{l=15}^{85} v\left(l,\sigma^2_{pre}\right)^{-1/2} v\left(100-l,\sigma^2_\beta\right)^{-1/2} \exp\left(\frac{1}{2}\frac{\sigma^2_{pre}G_1(l/100)^2}{v\left(l,\sigma^2_{pre}\right)} + \frac{1}{2}\frac{\sigma^2_\beta\left(G_1(1)-G_1(l/100)\right)^2}{v\left(100-l,\sigma^2_\beta\right)}\right)}{\sum_{j=1}^{18}\sum_{l=a_j}^{b_j} \frac{p_j}{b_j-a_j+1} v\left(l,\sigma^2_{\delta,j}\right)^{-1/2} \exp\left(-\frac{1}{2}\frac{\mu_j^2 l}{v\left(l,\sigma^2_{\delta,j}\right)}\right)\cosh\left(\frac{\mu_j G_1(l/100)}{v\left(l,\sigma^2_{\delta,j}\right)}\right)}
\end{aligned}$$

where $v\left(l,\sigma^2\right) = 1+\sigma^2 l/100$, $\sigma^2_{pre} = 378$, $\sigma^2_\beta = 22$, and $p_j, a_j, b_j, \sigma^2_{\delta,j}$ and μ_j are defined in Table 9.

Appendix C: Details of the Binary Choice Model

This subsection provides the primitive conditions for the binary choice model discussed in Section 6.1. The following conditions are imposed for Corollary 2.

Condition C: *1.* (u_i, x_i, q_i) *is i.i.d., and* u_i *is independent of* x_i *and* q_i with zero mean*; 2.* $F(\cdot)$ *is twice continuously differentiable; 3. For all* s, $\varepsilon \le F(s) \le 1-\varepsilon$ *and* $\varepsilon < f(s) \le C$ *for some constants* $C > \varepsilon > 0$*; 4.* x_i *has a compact support and includes at least one continuous random variable,* $E\left[||x_i||^4\right] < C$ *for some constant* C*; 5.* $\bar{\beta}$ *lies in the interior of a compact parameter space* Θ_β*; 6. Condition C.6 of Klein and Spady (1993) is satisfied; 7.* $K_1(s)$ *is uniformly bounded and satisfies* $\int K_1(s) = 1$, $\int K_1(s)\, s = 0$, $\int s^2 K_1(s) = 0$, $\lim_{|s|\to\infty} sK_1(s) = 0$, *and*

$\lim_{|s|\to\infty} s^2 K_1'(s) = 0$, *and* $b_{N1} = O(N^{-a})$ for $a \in [-1/6, -1/8]$; *8.* $K_2(\cdot)$ *and* b_{N2} *satisfy Conditions 3.2-3.3;*

Condition C.1 can be relaxed to allow weak convergence and heteroskedasticity. Conditions C.2-3 impose some smoothness and boundedness on the CDF and PDF of u_i. Condition C.4 restricts x_i to be on a compact support, which can be implemented by trimming. We can also relax this condition by imposing similar trimming conditions as in Klein and Spady (1993) at the price of a more complicated proof. Conditions C.5-7 are imposed for the consistency of $\hat{\beta}_{KS}$. In particular, C.6 imposes additional smoothness on the density of x_i and can be replaced with other conditions if we use other estimators of β. Condition C.7 requires a higher order kernel and an under-smoothing bandwidth. Condition C.8 is the same as in the main text.

Regarding the statistically efficient version $\mathcal{T}\hat{G}_N^{BC_2}$, since $F(\cdot)$, $f(\cdot)$, $\hat{F}(\cdot)$, and $\hat{f}(\cdot)$ are all uniformly bounded from both below and above, and x_i has bounded 4th moment, the convergence of $\mathcal{T}\hat{G}_N^{BC_2}$ follows from a similar (but much more tedious) argument as that for $\mathcal{T}\hat{G}_N^{BC}$. The proof of Corollary 2 is provided in Appendix E.

Appendix D: Details for the Panel Data Threshold Model

In this section, we discuss some details of Section 6.2 for $T = 2$ and extend the results to a general T. First, the kernel estimators of Γ and H can be written as

$$\hat{\Gamma}(l) = \frac{\sum_{s=1, s\neq[lN(r_0)]}^{N(r_0)} \Delta\dot{x}_{(s)} \Delta\dot{x}_{(s)}' K\left(\frac{s/N(r_0)-l}{b_N}\right)}{\sum_{s=1, s\neq[lN(r_0)]}^{N(r_0)} K\left(\frac{s/N(r_0)-l}{b_N}\right)}$$

$$\hat{H}(l) = \frac{\sum_{s=1, s\neq[lN(r_0)]}^{N(r_0)} \Delta\dot{x}_{(s)} \Delta\dot{x}_{(s)}' \widehat{\Delta\dot{u}}_{(s)}^2 K\left(\frac{s/N(r_0)-l}{b_N}\right)}{\sum_{s=1, s\neq[lN(r_0)]}^{N(r_0)} K\left(\frac{s/N(r_0)-l}{b_N}\right)}$$

where $\widehat{\Delta\dot{u}}_i$ is the residual of regressing $\Delta\dot{y}_i$ and $\Delta\dot{x}_i$. To prove Lemma 9 and Corollary 3, we make the following condition:

Condition D: *1.* $\{q_{i1}, q_{i2}, x_{i1}, x_{i2}, u_{i1}, u_{i2}\}$ *is i.i.d. across* i *with continuous (conditional) quantile functions of* q_{it}, $Q_t(\cdot)$ *and* $Q_t^c(\cdot)$, *for all* t; *2.* $\gamma_0 = Q_1^c(r_0)$ *for some* $\underline{r} < r_0 < \bar{r}$; *3.* $\Gamma(l)$ *is invertible for any* $l \in (0,1)\backslash\{r_0^c\}$ *and* $e_1'\Gamma(l)^{-1} H(l) \Gamma(l)^{-1\prime} e_1 > 0$ *for any* $l \in (0,1)\backslash\{r_0^c\}$; *4. for all* i, $E\left[u_{it}|\{x_{it}, q_{it}\}_{i=1}^N\right] = 0$; *5. For all* i *and* t, $E\left[||x_{it}||^8\right] < C$ *and* $E\left[||x_{it}u_{it}||^4\right] < C$ *for some* $C < \infty$; *6.* $E\left[\Delta x_i \Delta x_i' | q_{i2} \leq Q_2(s), q_{i1} = Q_1^c(l)\right]$ *and* $E\left[\Delta x_i \Delta x_i' \Delta u_i^2 | q_{i2} \leq Q_2(s), q_{i1} = Q_1^c(l)\right]$ *are twice continuously differentiable w.r.t.* s *and* l *on* $(0,1) \times (0, r_0^c)$ *except for up to a finite number of points, with left and right limits everywhere and* $E\left[\Delta x_i \Delta x_i' | q_{i2} > Q_2(s), q_{i1} = Q_2^c(l)\right]$ *and* $E\left[\Delta x_i \Delta x_i' \Delta u_i^2 | q_{i2} > Q_2(s), q_{i1} = Q_1^c(l)\right]$ *are piece-wise twice continuously differentiable w.r.t.* s *and* l *on* $(0,1) \times (0,1)$ *except for up to a finite number of points, with left and right limits everywhere; 7. Conditions 3.2-3.3.*

Condition D is sufficient for Conditions 1-3. Hence, Lemma 4 implies that $\hat{\Gamma}$ and $\hat{H}$ are uniformly consistent on any compact subset of $(0,1)$ without discontinuities. Then, the desired convergence of $\mathcal{T}\hat{G}_{N1}^{*Panel}$ follows from Theorem 3. To avoid the unboundedness on the boundary, we may trim away η percent of the data at each boundary.

Regarding the panel data threshold model with a general $T > 2$, that is,

$$y_{it} = \mu_i + x_{it}'\beta + x_{it}'\delta\mathbf{1}\left[q_{it} \leq Q_t(r)\right] + u_{it}, \quad t = 1, ..., T,$$

we may perform the previous analysis for any two consecutive periods to obtain $T-1$ confidence intervals. If we are willing to believe that the threshold r remains unchanged across all periods, we can construct a Bonferroni-type confidence interval by considering the union of all $T-1$ intervals. Otherwise, each confidence interval is asymptotically valid for those particular (two consecutive) periods.

If we impose an additional assumption that q_{it} is i.i.d. across both i and t, and the threshold remains unchanged at r for all t, we can then pool all data to perform a similar analysis. Denote $\hat{Q}_{Pool}(\cdot)$ as the empirical quantile function of q_i in the pooled sample. Under the null hypothesis $r = r_0$, we define

$$\begin{aligned}
\bar{y}_i^+ &= \left(\sum_{t=1}^{T} y_{it}\mathbf{1}\left[q_{it} \leq \hat{Q}_{Pool}(r_0)\right]\right) \Bigg/ \left(\sum_{t=1}^{T} \mathbf{1}\left[q_{it} \leq \hat{Q}_{Pool}(r_0)\right]\right) \\
\bar{y}_i^- &= \left(\sum_{t=1}^{T} y_{it}\mathbf{1}\left[q_{it} > \hat{Q}_{Pool}(r_0)\right]\right) \Bigg/ \left(\sum_{t=1}^{T} \mathbf{1}\left[q_{it} > \hat{Q}_{Pool}(r_0)\right]\right)
\end{aligned}$$

and

$$\tilde{y}_{it} = y_{it} - \bar{y}_i^+\mathbf{1}\left[q_{it} \leq \hat{Q}_{Pool}(r_0)\right] - \bar{y}_i^-\mathbf{1}\left[q_{it} > \hat{Q}_{Pool}(r_0)\right]$$

where $\bar{y}_i^+$ and $\bar{y}_i^-$ are understood as 0 if $\sum_{t=1}^{T}\mathbf{1}\left[q_{it} \leq \hat{Q}^{Pool}(r_0)\right] = 0$ and $\sum_{t=1}^{T}\mathbf{1}\left[q_{it} > \hat{Q}^{Pool}(r_0)\right] = 0$, respectively. Define $\tilde{x}_{it}$ and $\tilde{u}_{it}$ in a similar fashion so that we can exclude the fixed effects to arrive at a subsample threshold model, that is,

$$\tilde{y}_{it} = \tilde{x}_{it}'\beta + \tilde{x}_{it}'\delta\mathbf{1}\left[q_{it} \leq \hat{Q}_{Pool}(r_0)\right] + \tilde{u}_{it}.$$

Following the same argument as in the $T = 2$ case, we consider a subpopulation where we aim for a generalized version of Lemma 9. Define $A_{it}(r_0)$ as the stochastic event

$$\begin{aligned}
\{q_{it} &\leq Q(r_0) \text{ and } \sum_{t=1}^{T}\mathbf{1}\left[q_{it} \leq Q(r_0)\right] > 1\} \cup \\
\{q_{it} &> Q(r_0) \text{ and } \sum_{t=1}^{T}\mathbf{1}\left[q_{it} > Q(r_0)\right] > 1\},
\end{aligned}$$

and $\hat{A}_{it}(r_0)$ its empirical counterpart with Q replaced with $\hat{Q}_{Pool}$. Also define $N(r_0) = \sum_{t=1}^{T}\sum_{i=1}^{N}\mathbf{1}\left[\hat{A}_{it}(r_0)\right]$. Let $Q^c_{Pool}(\cdot)$ denote the quantile function of q_{it} in the subpopulation where $\mathbf{1}[A_{it}(r_0)] = 1$ holds and $\hat{Q}^c_{Pool}(\cdot)$ is its finite sample analogue. Then, one can show that under a set of regularity conditions similar to those in Condition D, in the pooled and re-ordered subsample,

$$N(r_0)^{-1}\sum_{i=1}^{[sN(r_0)]}\tilde{x}_{(i)}\tilde{x}'_{(i)} \xrightarrow{p} \int_0^s \Gamma(l)\,dl$$

uniformly in $s \in [0,1]$ and

$$N(r_0)^{-1/2}\sum_{i=1}^{[sN(r_0)]}\tilde{x}_{(i)}\tilde{u}_{(i)} \Rightarrow \int_0^s H^{1/2}(l)\,dW_k(l)$$

for $s \in [0,1]$ where the 'dot' denotes the observation in the subsample, and

$$\begin{aligned}
\Gamma(l) &= \Gamma_1(l)\,\mathbf{1}[Q^c_{Pool}(l) \le Q(r_0)] + \Gamma_2(l)\,\mathbf{1}[Q^c_{Pool}(l) > Q(r_0)] \\
H(l) &= H_1(l)\,\mathbf{1}[Q^c_{Pool}(l) \le Q(r_0)] + H_2(l)\,\mathbf{1}[Q^c_{Pool}(l) > Q(r_0)]
\end{aligned}$$

for some matrix processes $\Gamma_1(\cdot), \Gamma_2(\cdot), H_1(\cdot)$, and $H_2(\cdot)$. When (x_{it}, u_{it}) is i.i.d. across both i and t, we have

$$\begin{aligned}
\Gamma_1(l) &= E\left[\tilde{x}_{it}\tilde{x}'_{it}\left|\sum_{s\neq t}\mathbf{1}[q_{is} \le Q(r_0)] > 1, q_{it} = Q^c_1(l)\right.\right] \\
\Gamma_2(l) &= E\left[\tilde{x}_{it}\tilde{x}'_{it}\left|\sum_{s\neq t}\mathbf{1}[q_{is} \le Q(r_0)] > 1, q_{it} = Q^c_1(l)\right.\right] \\
H_1(l) &= E\left[\tilde{x}_{it}\tilde{x}'_{it}\tilde{u}^2_{it}\left|\sum_{s\neq t}\mathbf{1}[q_{is} > Q(r_0)] > 1, q_{i1} = Q^c_1(l)\right.\right] \\
H_2(l) &= E\left[\tilde{x}_{it}\tilde{x}'_{it}\tilde{u}^2_{it}\left|\sum_{s\neq t}\mathbf{1}[q_{is} > Q(r_0)] > 1, q_{i1} = Q^c_1(l)\right.\right].
\end{aligned}$$

The rest of our approach follows similar to that as in the main text.

Appendix E: Proofs

The following two lemmas establish that the estimation error from replacing the true quantile $Q(\cdot)$ with the empirical quantile $\hat{Q}(\cdot)$ is asymptotically negligible.

Lemma E.1 *Under Conditions 1.1-1.2 or 1.1"-1.2", for any* $0 < \eta < 1/2$, $\sup_{\eta \le l \le 1-\eta}\left|\hat{Q}(l) - Q(l)\right| = O_p\left(N^{-1/2}\right)$.

Proof of Lemma E.1. This result shows that the empirical quantile is a uniformly consistent estimator of the true quantile. The proof can be easily derived from the uniform Bahadur presentation with stationary and mixing data. For details and more comments, see, for example, Sen (1972), Babu and Singh (1978), and Wu (2005). ■

Lemma E.2 *Under Conditions 1.1-1.2,*

(i) for any $\varepsilon_N \to 0$ as $N \to \infty$,

$$\sup_{s\in[0,1]} \sup_{\beta\in B_{\varepsilon_N}(\bar{\beta})} \left\| N^{-1} \sum_{i=1}^{[sN]} \Lambda_{(i)}(\beta) - N^{-1} \sum_{i=1}^{N} \Lambda_i(\beta_i) \mathbf{1}[q_i \le Q(s)] \right\| = o_p(1)$$

where $B_{\varepsilon_N}(\bar{\beta})$ *is the open ball centered at* $\bar{\beta}$ *with radius* ε_N.

(ii)

$$\sup_{s\in[0,1]} \left\| N^{-1/2} \sum_{i=1}^{[sN]} m_{(i)}\left(\beta_{(i)}\right) - N^{-1/2} \sum_{i=1}^{N} m_i(\beta_i) \mathbf{1}[q_i \le Q(s)] \right\| = o_p(1).$$

Proof of Lemma E.2. Denote

$$\dot{v}_N(\beta, s) = N^{-1} \sum_{i=1}^{N} \Lambda_i(\beta) \mathbf{1}\left[q_i \le \hat{Q}(s)\right] - N^{-1} \sum_{i=1}^{N} \Lambda_i(\beta_i) \mathbf{1}[q_i \le Q(s)]$$

and

$$\ddot{v}_N(s) = N^{-1/2} \sum_{i=1}^{[sN]} m_{(i)}\left(\beta_{(i)}\right) - N^{-1/2} \sum_{i=1}^{N} m_i(\beta_i) \mathbf{1}[q_i \le Q(s)].$$

When $l = 0$ and 1, the results hold trivially. For $l \in (0,1)$, by Lemma E.1, for any small positive ε, there exists a positive integer N^* and a positive constant v such that for any $N > N^*$

$$P\left(\left|\hat{Q}(l) - Q(l)\right| \le vN^{-1/2}\right) > 1 - \varepsilon. \tag{43}$$

Since ε is arbitrary, it suffices to establish (i) and (ii) on the set $\{\left|\hat{Q}(l) - Q(l)\right| \le vN^{-1/2}\}$.

To prove (i), we first prove $\sup_{\beta\in B_{\varepsilon_N}(\bar{\beta})} |\dot{v}_N(\beta, s)|$ is $o_p(1)$ for each $s \in (0,1)$. Then by Theorem 15.5 of Billingsley (1968), the uniform convergence in s can be established by the pointwise convergence and the uniform tightness of $\sup_{\beta\in B_{\varepsilon_N}(\bar{\beta})} |\dot{v}_N(\beta, s)|$ in s. For conciseness, we prove the scalar case only, and the proof of the multivariate case follows from a similar argument. For each $s \in (0,1)$, we have, on the set $\{0 \le \hat{Q}(s) - Q(s) \le vN^{-1/2}\}$

$$\begin{aligned} |\dot{v}_N(\beta, s)| \le\ & \left| N^{-1} \sum_{i=1}^{N} \Lambda_i(\beta_i) \mathbf{1}\left[\hat{Q}(s) < q_i \le Q(s)\right] \right| \\ & + \left| N^{-1} \sum_{i=1}^{N} \frac{\partial \Lambda_i\left(\dot{\beta}_i\right)}{\partial \beta} \mathbf{1}[q_i \le Q(s)](\beta - \beta_i) \right| \end{aligned}$$

$$
\begin{aligned}
\leq & \left(N^{-1}\sum_{i=1}^{N}\left|\Lambda_i\left(\beta_i\right)\right|^2\right)^{1/2}\left(N^{-1}\sum_{i=1}^{N}\mathbf{1}\left[\hat{Q}\left(s\right)<q_i\leq Q\left(s\right)\right]\right)^{1/2} \\
& +\left(N^{-1}\sum_{i=1}^{N}\left|\frac{\partial\Lambda_i\left(\dot{\beta}_i\right)}{\partial\beta}\right|^2\right)^{1/2}\left(N^{-1}\sum_{i=1}^{N}\left|\beta-\beta_i\right|^2\right)^{1/2}
\end{aligned}
$$

where $\dot{\beta}_i$ lies between β and β_i. By Condition 1.6 and Markov's inequality, the above item is $o_p(1)$. This also holds on the set $\{0 \leq Q(s) - \hat{Q}(s) \leq vN^{-1/2}\}$. Thus, $\sup_{\beta\in B_{\varepsilon_N}(\bar{\beta})}|\dot{v}_N(\beta,s)| = o_p(1)$ for any fixed s.

Next, we establish the uniform tightness of $\sup_{\beta\in B_{\varepsilon_N}(\bar{\beta})}|\dot{v}_N(\beta,s)|$ in s. For $l<s$, we have

$$
\left|\sup_{\beta\in B_{\varepsilon_N}(\bar{\beta})}\left|\dot{v}_N(\beta,l)\right|-\sup_{\beta\in B_{\varepsilon_N}(\bar{\beta})}\left|\dot{v}_N(\beta,s)\right|\right|\leq\left|\sup_{\beta\in B_{\varepsilon_N}(\bar{\beta})}\left|\dot{v}_N(\beta,l)-\dot{v}_N(\beta,s)\right|\right|
$$

$$
\begin{aligned}
\leq & \sup_{\beta\in B_{\varepsilon_N}(\bar{\beta})}\left|N^{-1}\sum_{i=1}^{N}\Lambda_i(\beta)\,\mathbf{1}\left[Q(l)<q_i\leq Q(s)\right]\right| \\
& +\sup_{\beta\in B_{\varepsilon_N}(\bar{\beta})}\left|N^{-1}\sum_{i=1}^{N}\Lambda_i(\beta)\,\mathbf{1}\left[\hat{Q}(l)<q_i\leq\hat{Q}(s)\right]\right| \\
\leq & \sup_{\beta\in B_{\varepsilon_N}(\bar{\beta})}\left(N^{-1}\sum_{i=1}^{N}\left|\Lambda_i(\beta)\right|^2\right)^{1/2}\times \\
& \left(\left(N^{-1}\sum_{i=1}^{N}\mathbf{1}\left[Q(l)<q_i\leq Q(s)\right]+\right)^{1/2}+\left(\frac{[sN]-[lN]}{N}\right)^{1/2}\right).
\end{aligned}
$$

By Lemma E.1, Condition 1.6, and Markov's inequality, the above term is bounded in probability by $C|l-s|$ for some constant $C<\infty$. By the exact argument for $l>s$, the uniform tightness of $\dot{v}_N(\cdot)$ is established.

Regarding (ii), following the same logic as in (i), we prove $\ddot{v}_N(s) = o_p(1)$ for fixed s and $\ddot{v}_N(s)$ is uniform tight. Conditions 1.1 and 1.7 imply that for any $i \neq j$,

$$
E\left[m_i\left(\beta_i\right)m_j\left(\beta_j\right)|\{q_i\}_{i=1}^{N}\right]=0.
$$

Then on the set $\{0 < \hat{Q}(s) - Q(s) \leq vN^{-1/2}\}$ where v is chosen as in (43) for an arbitrary ε, we have

$$
\left|\left|\ddot{v}_N(s)\right|\right|^2 \leq N^{-1}\sum_{i=1}^{N}\left|\left|m_i\left(\beta_i\right)\right|\right|^2\mathbf{1}\left[\hat{Q}(s)<q_i\leq Q(s)\right]
$$

$$\leq \left(N^{-1}\sum_{i=1}^{N}||m_i(\beta_i)||^4\right)^{1/2}\left(N^{-1}\sum_{i=1}^{N}\mathbf{1}\left[\hat{Q}(s)<q_i\leq Q(s)\right]\right)^{1/2}.$$

By Lemma E.1, Condition 1.6, and Markov's inequality, the above item converges to zero in probability for any fixed s. This also holds on the set $\{0\leq Q(s)-\hat{Q}(s)\leq vN^{-1/2}\}$ following the exact argument. Thus, $\ddot{v}_N(s)=o_p(1)$ for fixed s. Finally, the uniform tightness of $\ddot{v}_N(s)$ follows from a similar argument for $\dot{v}_N(s)$, and then the proof is complete. ■

Proof of Lemma 1. By Lemma E.2, it suffices to prove that, for $s\in[0,1]$

$$N^{-1/2}\sum_{i=1}^{N}m_i(\beta_i)\mathbf{1}[q_i\leq Q(s)]\Rightarrow\int_0^s H^{1/2}(l)\,dW(l) \tag{44}$$

and uniformly in $s\in[0,1]$,

$$N^{-1}\sum_{i=1}^{N}\Lambda_i(\beta_i)\mathbf{1}[q_i\leq Q(s)]\xrightarrow{p}\int_0^s\Gamma(l)\,dl. \tag{45}$$

We first establish the pointwise convergence and then the uniform tightness. For $s=0$, (44) and (45) are trivially satisfied. Consider any fixed $s\in(0,1]$. Define

$$M_1(s)=\int_0^s\Gamma(l)\,dl\ \text{ and } M_2(s)=\int_0^s H(l)\,dl.$$

The conditions for Proposition 1 of Andrews (1991) are satisfied for the partial sum $N^{-1/2}\sum_{i=1}^{N}M_2(s)^{-1/2}m_i(\beta_i)\mathbf{1}[q_i\leq Q(s)]$, which then converges pointwisely to $W(s)$. By Corollary 19.3 of Davidson (1994), $N^{-1}\sum_{i=1}^{N}(\Lambda_i(\beta_i)\mathbf{1}[q_i\leq Q(s)]-E[\Lambda_i(\beta_i)\mathbf{1}[q_i\leq Q(s)]])$ converges to 0 in probability pointwisely. Thus, by Condition 1.5 and the continuous mapping theorem, we establish (44) and (45) for any fixed s.

For the uniform tightness of (44), consider any $s,s_1\in(0,1)$. WLOG, let $s_1-s=\eta>0$. Define $v_N(s)=N^{-1/2}\sum_{i=1}^{N}m_i(\beta_i)\mathbf{1}[q_i\leq Q(s)]$. Then, Conditions 1.1 and 1.7 imply $E[v_N(s)-v_N(s_1)]=0$ and $E\left[m_i(\beta_i)m_j(\beta_j)'|q_i,q_j\right]=0$ for any $i\neq j$. Then for some constant C and a large enough N, we have

$$\begin{aligned}
&E\left[||v_N(s)-v_N(s_1)||^2\right]\\
\leq\ & N^{-1}\sum_{i=1}^{N}E\left[m_i(\beta_i)m_i(\beta_i)'\mathbf{1}[Q(s)<q_i\leq Q(s_1)]\right]\\
\leq\ & N^{-1}\sum_{i=1}^{N}(E[\mathbf{1}[Q(s)<q_i\leq Q(s_1)]])^{1/2}\left(\sup_i E\left[||m_i(\beta_i)||^4\right]\right)^{1/2}\\
\leq\ & \eta C.
\end{aligned}$$

The uniform tightness of (45) follows from Markov's inequality and

$$\begin{aligned}
& E\left[\left|\left|N^{-1}\sum_{i=1}^{N}\Lambda_i\left(\beta_i\right)\mathbf{1}\left[q_i\leq Q\left(s\right)\right]-N^{-1}\sum_{i=1}^{N}\Lambda_i\left(\beta_i\right)\mathbf{1}\left[q_i\leq Q\left(s_1\right)\right]\right|\right|\right] \\
= \; & E\left[\left|\left|N^{-1}\sum_{i=1}^{N}\Lambda_i\left(\beta_i\right)\mathbf{1}\left[Q\left(s\right)<q_i\leq Q\left(s_1\right)\right]\right|\right|\right] \\
\leq \; & N^{-1}\sum_{i=1}^{N}E\left[\left|\left|\Lambda_i\left(\beta_i\right)\mathbf{1}\left[Q\left(s\right)<q_i\leq Q\left(s_1\right)\right]\right|\right|\right] \\
\leq \; & \sup_i E\left[\left|\left|\Lambda_i\left(\beta_i\right)\right|\right|^2\right]^{1/2}\left(E\left[\mathbf{1}\left[Q\left(s\right)<q_i\leq Q\left(s_1\right)\right]\right]\right)^{1/2} \\
\leq \; & C\eta \text{ for a large enough } N.
\end{aligned}$$

Hence, the proof is complete by Theorem 15.5 of Billingsley (1968). ■

Proof of Theorem 1. The proof follows readily from

$$\begin{aligned}
e_1'\mathcal{T}\tilde{G}\left(s\right) &= \sqrt{h\left(1\right)}\int_0^{g^{-1}(s)}\frac{\partial g\left(l\right)}{\partial l}e_1'\Gamma^{-1}\left(l\right)H^{1/2}\left(l\right)dW\left(l\right) \\
&+ d_1\int_0^{\min(g^{-1}(s),r)}\frac{\partial g\left(l\right)}{\partial l}dl+b_1\int_0^1\frac{\partial g\left(l\right)}{\partial l}dl
\end{aligned}$$

and the facts that

$$h\left(1\right)\int_0^{g^{-1}(s)}\frac{\partial g\left(l\right)}{\partial l}e_1'\Gamma\left(l\right)^{-1}H^{1/2}\left(l\right)dW\left(l\right)=W_1\left(s\right),$$

and $\int_0^{g^{-1}(s)}\frac{\partial g(l)}{\partial l}dl=s$ for $s\in[0,1]$. ■

Proof of Lemma 2. In the situation where $\Gamma\left(\cdot\right)=\bar{\Gamma}$ and $H\left(\cdot\right)=\bar{H}$, Lemma 1 reduces to $N^{-1}\sum_{i=1}^{[sN]}x_{(i)}x_{(i)}'\xrightarrow{p}s\bar{\Gamma}$ and $N^{-1}\sum_{i=1}^{[sN]}x_{(i)}u_{(i)}\Rightarrow\bar{H}^{1/2}W\left(s\right)$. Thus, the limit behavior of $\hat{r}$ readily follows from Proposition 1 in Elliott and Müller (2007) and the continuous mapping theorem. To derive the limit of the likelihood ratio statistic, define $z_{(i)}=u_{(i)}+N^{-1/2}x_{(i)}'d\mathbf{1}\left[q_i\leq Q\left(r_0\right)\right].$ By Lemma E.2, we have, for any $r\in[\underline{r},\bar{r}]$ and $\gamma=Q\left(r\right),$

$$\begin{aligned}
& \arg\min_{\underline{r}\leq r\leq\bar{r}}S_N\left(r\right) \\
= \; & \arg\max_{\underline{r}\leq r\leq\bar{r}}\left\{\left(\sum_{i=1}^{[rN]}x_{(i)}z_{(i)}\right)'\left(\sum_{i=1}^{[rN]}x_{(i)}x_{(i)}'\right)^{-1}\left(\sum_{i=1}^{[rN]}x_{(i)}z_{(i)}\right)\right. \\
& \left.+\left(\sum_{i=[rN]+1}^{N}x_{(i)}z_{(i)}\right)'\left(\sum_{i=[rN]+1}^{N}x_{(i)}x_{(i)}'\right)^{-1}\left(\sum_{i=[rN]+1}^{N}x_{(i)}z_{(i)}\right)\right\}+o_p\left(1\right) \\
\Rightarrow \; & \arg\max_{s\in[\underline{r},\bar{r}]}V\left(s\right).
\end{aligned}$$

Then the proof follows from the continuous mapping theorem and the fact that $N^{-1}S_N(r_0)$ converges to σ_u^2. ■

Proof of Theorem 2 and Corollary 1. The proof follows closely to that of Proposition 1 of Elliott and Müller (2014). Let θ denote (r, d, b) and $\Theta^l = (\theta_1, ..., \theta_l) \subset [\underline{r}, \bar{r}] \times \mathbb{R}^k \times \mathbb{R}^k$ be an arbitrary set of $l < \infty$ values of θ. With some abuse of notation, let $\mathrm{LR}_N(\theta)$ denote the log-likelihood ratio statistic between the unstable model with parameter θ and the stable model. By Definition 9.1 in van der Vaart (1998), it suffices to establish $\{\mathrm{LR}_N(\theta)\}_{\theta\in\Theta^l} \Rightarrow \{\log f\left(\tilde{G};\theta\right)\}_{\theta\in\Theta^l}$ where $\{W_i\}_{i=1}^N$ and $\tilde{G}$ are generated under any $\theta_0 \in [\underline{r}, \bar{r}] \times \mathbb{R}^k \times \mathbb{R}^k$. We first establish the convergence with data generated from the stable model ($b = d = 0$). By Lemma 1, the continuous mapping theorem, and a similar version of Lemma E.2, we have, for any fixed M,

$$\begin{aligned}
& \mathrm{LR}_N(\theta) \\
= & \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(b + d\mathbf{1}\left[q_i \le Q(r)\right]\right)' M'\Gamma(R_i/N) H(R_i/N)^{-1} m_i\left(\bar{\beta}\right) \\
& -\frac{1}{2N}\sum_{i=1}^{N}\left(b + d\mathbf{1}\left[q_i \le Q(r)\right]\right)' M'\Gamma(R_i/N) H(R_i/N)^{-1} \Lambda_i\left(\dot{\beta}_i\right) M\left(b + d\mathbf{1}\left[q_i \le Q(r)\right]\right) \\
= & \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(b + d\mathbf{1}\left[i/N \le r\right]\right)' M'\Gamma(i/N) H(i/N)^{-1} m_{(i)}\left(\bar{\beta}\right) \\
& -\frac{1}{2N}\sum_{i=1}^{N}\left(b + d\mathbf{1}\left[i/N \le r\right]\right)' M'\Gamma(i/N) H(i/N)^{-1} \Lambda_{(i)}\left(\dot{\beta}_{(i)}\right) M\left(b + d\mathbf{1}\left[i/N \le r\right]\right) + o_p(1) \\
\Rightarrow & b'M'\int_0^1 \Gamma(l) H(l)^{-1/2} dW(l) + d'M'\int_0^r \Gamma(l) H(l)^{-1/2} dW(l) \\
& -\frac{1}{2}\int_0^1 \left(b + d\mathbf{1}\left[l \le r\right]\right)' M'\Gamma(l) H(l)^{-1} \Gamma(l) M\left(b + d\mathbf{1}\left[l \le r\right]\right) dl \\
= & \log f\left(\tilde{G};\theta\right).
\end{aligned}$$

To prove this convergence with data generated from any fixed $\theta \in [\underline{r}, \bar{r}] \times \mathbb{R}^k \times \mathbb{R}^k$, by the exact argument as above, we have the joint convergence $\{\mathrm{LR}_N(\theta_0), \{\mathrm{LR}_N(\theta_0)\}_{\theta\in\Theta^l}\} \Rightarrow \{\log f\left(\tilde{G};\theta_0\right), \{\log f\left(\tilde{G};\theta\right)\}_{\theta\in\Theta^l}\}$ in the stable model. Since $E\left[f\left(\int_0^{\cdot}\Gamma(l) H(l)^{-1/2} dW(l);\theta_0\right)\right] = 1$, by Le Cam's First Lemma, the unstable model with parameter θ_0 is contiguous to the stable model. Thus, by a general version of Le Cam's Third Lemma, $\{\mathrm{LR}_N(\theta)\}_{\theta\in\Theta^l}$ also converges weakly when data are generated in the model with θ, and the limiting distribution is given by $\{\log f\left(\tilde{G};\theta\right)\}_{\theta\in\Theta^l}$, with $\tilde{G}$ generated under $\theta = \theta_0$. The rest of the proof follows readily from the continuous mapping theorem. ■

Proof of Lemma 3. Denote $\partial g(i/N)/\partial l$ for $\partial g(l)/\partial l|_{l=i/N}$. Plugging in model (2), we obtain that by a similar version of Lemma E.2,

$$\begin{aligned}
e_1'\mathcal{T}\tilde{G}_N(s) &= \sqrt{h(1)}N^{-1/2}\sum_{i=1}^{[g^{-1}(s)N]}\frac{\partial g(i/N)}{\partial l}e_1'\Gamma(i/N)^{-1}m_{(i)}(\bar{\beta}) \\
&= \sqrt{h(1)}N^{-1/2}\sum_{i=1}^{[g^{-1}(s)N]}\frac{\partial g(i/N)}{\partial l}e_1'\Gamma(i/N)^{-1}m_{(i)}\left(\beta_{(i)}\right) \\
&\quad+\sqrt{h(1)}N^{-1}\sum_{i=1}^{[g^{-1}(s)N]}\frac{\partial g(i/N)}{\partial l}e_1'\Gamma(i/N)^{-1}\Lambda_{(i)}\left(\dot{\beta}_{(i)}\right)b \\
&\quad+\sqrt{h(1)}N^{-1}\sum_{i=1}^{[g^{-1}(s)N]}\frac{\partial g(i/N)}{\partial l}e_1'\Gamma(i/N)^{-1}\Lambda_{(i)}\left(\dot{\beta}_{(i)}\right)d\mathbf{1}[i/N\le r]+o_p(1) \\
&= A_{1N}(s)+A_{2N}(s)+A_{3N}(s)+o_p(1)
\end{aligned}$$

where $\dot{\beta}_{(i)}$ is between $\beta_{(i)}$ and $\bar{\beta}$. Then by $g(0)=0$ and $g(1)=1$, we have,

$$\begin{aligned}
A_{1N}(s) &= \sqrt{h(1)}N^{-1/2}\sum_{i=1}^{[g^{-1}(s)N]}\frac{\partial g(i/N)}{\partial l}e_1'\Gamma(i/N)^{-1}m_{(i)}\left(\beta_{(i)}\right) \\
&\Rightarrow \sqrt{h(1)}\int_{g^{-1}(0)}^{g^{-1}(s)}\frac{\partial g(l)}{\partial l}e_1'\Gamma(l)^{-1}H(l)^{1/2}dW(l) \qquad (46) \\
&= W_1(s)
\end{aligned}$$

where the weak convergence (46) is established as follows. Denote $\Psi(s)$ as $\frac{\partial g(s)}{\partial l}\Gamma(s)^{-1}$. We first assume $\Gamma(\cdot)$ and $H(\cdot)$ are continuous and differentiable on $[0,1]$, implying that $\Psi(s)$ is also continuous. Denote $\Delta\Psi(i/N)$ for $\Psi(i/N)-\Psi((i-1)/N)$ and $J(s)=\int_0^s H(l)^{1/2}dW(l)$. By summation by parts, Lemma 1, the continuous mapping theorem, and Ito's lemma, we have

$$\begin{aligned}
N^{-1/2}\sum_{i=1}^{[g^{-1}(s)N]}\Psi(i/N)m_{(i)}\left(\beta_{(i)}\right) &= \Psi\left(\frac{[g^{-1}(s)N]}{N}\right)N^{-1/2}\sum_{i=1}^{[g^{-1}(s)N]}m_{(i)}\left(\beta_{(i)}\right) \\
&\quad-N^{-1}\sum_{i=1}^{[g^{-1}(s)N]}\left((N\Delta\Psi(i/N))N^{-1/2}\sum_{j=1}^{i}m_{(i)}\left(\beta_{(i)}\right)\right) \\
&\Rightarrow \Psi\left(g^{-1}(s)\right)J\left(g^{-1}(s)\right)-\int_0^{g^{-1}(s)}\frac{\partial\Psi(t)}{\partial t}J(t)\,dt \\
&= \int_0^{g^{-1}(s)}\Psi(l)\,dJ(l)
\end{aligned}$$

Then, the proof of (46) follows from Theorem 1.

When $\Gamma(\cdot)$ and $H(\cdot)$ have finite number of discontinuities, denoted as $r_1, ..., r_M$, we can approximate Ψ by a differentiable function $\tilde{\Psi}$ such that $\tilde{\Psi}(s) = \Psi(s)$ for $s \notin \cup_{j=1}^{M} B_{\varepsilon}(r_j)$ where $B_{\varepsilon/N}(r_j)$ is an open interval centered around r_j with length ε/N. By the boundedness of Γ and H, $\sup_{s\in[0,1]} \left|\tilde{\Psi}(s) - \Psi(s)\right| < C$ and $\int_0^1 \left|\tilde{\Psi}(s) - \Psi(s)\right| ds \leq C/N$ for some $C < \infty$. It is then easy to show that $N^{-1/2} \sum_{i=1}^{[g^{-1}(s)N]} \Psi(i/N)\, m_{(i)}\left(\beta_{(i)}\right) - N^{-1/2} \sum_{i=1}^{[g^{-1}(s)N]} \tilde{\Psi}(i/N)\, m_{(i)}\left(\beta_{(i)}\right) = o_p(1)$. Hence, the convergence (46) can be established by replacing Ψ with $\tilde{\Psi}$.

For A_{2N} and A_{3N}, we have, by Lemma 1

$$\begin{aligned}
A_{2N}(s) &= N^{-1} \sum_{i=1}^{[g^{-1}(s)N]} \frac{\partial g(i/N)}{\partial l} e_1' \Gamma(i/N)^{-1} \Lambda_{(i)}\left(\dot{\beta}_{(i)}\right) d\mathbf{1}[i/N \leq r] \\
&\xrightarrow{p} d_1 \int_{g^{-1}(0)}^{\min\left(r, g^{-1}(s)\right)} \frac{\partial g(l)}{\partial l} e_1' \Gamma(l)^{-1} \Gamma(l)\, e_1 dl \\
&= d_1 \min(s, g(r)),
\end{aligned}$$

and

$$\begin{aligned}
A_{3N}(s) &= N^{-1/2} \sum_{i=1}^{[g^{-1}(s)N]} \frac{\partial g(i/N)}{\partial l} e_1' \Gamma(i/N)^{-1} \Lambda_{(i)}\left(\dot{\beta}\right) bN^{-1/2} \\
&= b_1 \int_{g^{-1}(0)}^{g^{-1}(s)} \frac{\partial g(l)}{\partial l} e_1' \Gamma(l)^{-1} \Gamma(l)\, e_1 dl \\
&\xrightarrow{p} b_1 s.
\end{aligned}$$

Then, the rest of the proof follows readily from the Slutsky's theorem and the continuous mapping theorem. ■

Lemma E.3 *Under Conditions 1 and 3, for any* $0 < \eta < 1/2$,

$$\sup_{\eta \leq l \leq 1-\eta} \left\| \hat{H}(l) - \frac{\sum_{i=1, i\neq [lN]}^{N} m_{(i)}\left(\beta_{(i)}\right) m_{(i)}\left(\beta_{(i)}\right)' K\left(\frac{i/N - l}{b_N}\right)}{\sum_{i=1, i\neq [lN]}^{N} K\left(\frac{i/N - l}{b_N}\right)} \right\| = o_p(1).$$

Proof of Lemma E.3. To reduce notational burden, we prove the scalar case, and the multivariate case follows from a similar argument. Let $\hat{H}^0(l) = \frac{\sum_{i=1, i\neq [lN]}^{N} m_{(i)}\left(\beta_{(i)}\right)^2 K\left(\frac{i/N - l}{b_N}\right)}{\sum_{i=1, i\neq [lN]}^{N} K\left(\frac{i/N - l}{b_N}\right)}$, then

$$\hat{H}(l) - \hat{H}^0(l) = \frac{\frac{1}{Nb_N} \sum_{i=1, i\neq [lN]}^{N} \left(m_{(i)}\left(\beta_{(i)}\right)^2 - m_{(i)}\left(\hat{\beta}\right)^2 \right) K\left(\frac{i/N - l}{b_N}\right)}{\frac{1}{Nb_N} \sum_{i=1, i\neq [lN]}^{N} K\left(\frac{i/N - l}{b_N}\right)}.$$

The denominator is $O(1)$ by standard arguments, and the numerator can be written as

$$\begin{aligned}
& \frac{1}{Nb_N}\sum_{i=1,i\neq[lN]}^{N}\left(m_{(i)}\left(\beta_{(i)}\right)+m_{(i)}\left(\hat{\beta}\right)\right)\left(m_{(i)}\left(\beta_{(i)}\right)-m_{(i)}\left(\hat{\beta}\right)\right)K\left(\frac{i/N-l}{b_N}\right) \\
= & \frac{1}{Nb_N}\sum_{i=1,i\neq[lN]}^{N}\left(m_{(i)}\left(\beta_{(i)}\right)+m_{(i)}\left(\hat{\beta}\right)\right)\left(\Lambda_{(i)}\left(\dot{\beta}_{(i)}\right)\left(\beta_{(i)}-\hat{\beta}\right)\right)K\left(\frac{i/N-l}{b_N}\right) \\
= & \frac{1}{Nb_N}\sum_{i=1,R_i\neq lN}^{N}\left(m_i\left(\beta_i\right)+m_i\left(\hat{\beta}\right)\right)\left(\Lambda_i\left(\dot{\beta}_i\right)\left(\beta_i-\hat{\beta}\right)\right)K\left(\frac{R_i/N-l}{b_N}\right) \\
\equiv & B_N
\end{aligned}$$

where $\dot{\beta}_{(i)}$ lies on the segment line joining $\hat{\beta}$ and $\beta_{(i)}$ and R_i is the rank of q_i. Then, we have

$$\begin{aligned}
E\left[\|B_N\|\right] \leq & \frac{1}{Nb_N}\sum_{i=1}^{N}E\left[\left\|\left(m_i\left(\beta_i\right)+m_i\left(\hat{\beta}\right)\right)\left(\Lambda_i\left(\dot{\beta}_i\right)\left(\beta_i-\hat{\beta}\right)\right)K\left(\frac{R_i/N-l}{b_N}\right)\right\|\right] \\
\leq & \frac{1}{N}\sum_{i=1}^{N}\left(E\left[\left\|\left(m_i\left(\beta_i\right)+m_i\left(\hat{\beta}\right)\right)\Lambda_i\left(\dot{\beta}_i\right)\right\|^2\right]\right)^{1/2}\times \\
& b_N^{-1}\sup_s K\left(\frac{s-l}{b_N}\right)\max_i\left(E\left\|\beta_i-\hat{\beta}\right\|^2\right)^{1/2}.
\end{aligned}$$

The proof follows from Markov's inequality and Conditions 1.6 and 3.2-3.4. ■

Proof of Lemma 4. To reduce notational burden, we present the proof for the scalar case, which can be easily generalized for the multivariate case.

For (i), we first prove the pointwise convergence that $\hat{\Gamma}(l)-\Gamma(l)=o_p(1)$ for any fixed l at which $\Gamma(\cdot)$ is continuous. Regarding the bias term, we have,

$$\begin{aligned}
E\left[\hat{\Gamma}(l)\right]-\Gamma(l) = & \frac{1}{Nb_N}\sum_{i=1}^{N}E\left[\Lambda_{(i)}\left(\hat{\beta}\right)\right]K\left(\frac{i/N-l}{b_N}\right)-\Gamma(l)+O\left(N^{-1}\right) \\
= & \frac{1}{Nb_N}\sum_{i=1}^{N}E\left[\Lambda_{(i)}\left(\beta_{(i)}\right)+\frac{\partial}{\partial\beta}\Lambda_{(i)}\left(\dot{\beta}\right)\left(\hat{\beta}-\beta_{(i)}\right)\right]K\left(\frac{i/N-l}{b_N}\right) \\
= & \frac{1}{Nb_N}\sum_{i=1}^{N}E\left[\Lambda_i\left(\beta_i\right)K\left(\frac{R_i/N-l}{b_N}\right)\right]-\Gamma(l) \\
& +\frac{1}{Nb_N}\sum_{i=1}^{N}E\left[\frac{\partial}{\partial\beta}\Lambda_{(i)}\left(\dot{\beta}\right)\left(\hat{\beta}-\beta_{(i)}\right)\right]+O\left(N^{-1}\right).
\end{aligned}$$

Note that by Conditions 1.6 and 3.4,

$$\frac{1}{Nb_N}\sum_{i=1}^{N}E\left[\frac{\partial}{\partial\beta}\Lambda_{(i)}\left(\dot{\beta}\right)\left(\hat{\beta}-\beta_{(i)}\right)\right]$$

$$
\begin{aligned}
&\leq \frac{1}{b_N}\left(N^{-1}\sum_{i=1}^{N}E\left[\left|\frac{\partial}{\partial\beta}\Lambda_i\left(\dot{\beta}_i\right)\right|^2\right]\right)^{1/2}\left(N^{-1}\sum_{i=1}^{N}E\left[\left|\hat{\beta}-\beta_i\right|^2\right]\right)^{1/2} \\
&\leq \frac{1}{b_N}\left(\sup_i E\left[\left|\frac{\partial}{\partial\beta}\Lambda_i\left(\dot{\beta}\right)\right|^2\right]\sup_i E\left[\left|\hat{\beta}-\beta_i\right|^2\right]\right)^{1/2} \\
&= o\left(1\right).
\end{aligned}
$$

Conditions 3.1 leads to

$$
\begin{aligned}
& b_N^{-1}E\left[\Lambda_i\left(\beta_i\right)K\left(\frac{R_i/N-l}{b_N}\right)\right] \\
=\;& b_N^{-1}\sum_{j=1}^{N}E\left[\Lambda_i\left(\beta_i\right)K\left(\frac{j/N-l}{b_N}\right)|R_i=j\right]P\left(R_i=j\right) \\
=\;& \int_0^1 E\left[\Lambda_i\left(\beta_i\right)|Q\left(l-1/N+vb_N\right)\leq q_i\leq Q\left(l+vb_N\right)\right]K\left(v\right)dv+O\left(N^{-1}\right) \\
=\;& E\left[\Lambda_i\left(\beta_i\right)|q_i=Q\left(l\right)\right]+O\left(N^{-1}+b_N^{-2}\right),
\end{aligned}
$$

and then

$$
\begin{aligned}
& \frac{1}{Nb_N}\sum_{i=1}^{N}E\left[\Lambda_i\left(\beta_i\right)K\left(\frac{R_i/N-l}{b_N}\right)\right]-\Gamma\left(l\right) \\
=\;& \frac{1}{N}\sum_{i=1}^{N}E\left[\Lambda_i\left(\beta_i\right)|q_i=Q\left(l\right)\right]+O\left(N^{-1}+b_N^{-2}\right)-\Gamma\left(l\right) \\
=\;& o\left(1\right).
\end{aligned}
$$

Thus, the bias is $o\left(1\right)$. For the variance, we have

$$
\begin{aligned}
Var\left[\hat{\Gamma}\left(l\right)\right] &= \frac{1}{N^2b_N^2}Var\left[\sum_{i=1}^{N}\Lambda_{(i)}\left(\hat{\beta}\right)K\left(\frac{i/N-l}{b_N}\right)\right] \\
&= \frac{1}{N^2b_N^2}Var\left[\sum_{i=1}^{N}\Lambda_i\left(\hat{\beta}\right)K\left(\frac{R_i/N-l}{b_N}\right)\right] \\
&= \frac{1}{N^2b_N^2}\sum_{i=1}^{N}Var\left[\Lambda_i\left(\hat{\beta}\right)K\left(\frac{R_i/N-l}{b_N}\right)\right] \\
&\quad +\frac{2}{N^2b_N^2}\sum_{1\leq i<j\leq N}Cov\left[\Lambda_i\left(\hat{\beta}\right)K\left(\frac{R_i/N-l}{b_N}\right),\Lambda_j\left(\hat{\beta}\right)K\left(\frac{R_j/N-l}{b_N}\right)\right].
\end{aligned}
$$

where R_i is the rank of q_i. To bound the first term, we have

$$
\frac{1}{N^2b_N^2}\sum_{i=1}^{N}Var\left[\Lambda_i\left(\hat{\beta}\right)K\left(\frac{R_i/N-l}{b_N}\right)\right]
$$

$$
\begin{aligned}
&\leq \frac{1}{N^2 b_N^2}\sum_{i=1}^{N} E\left[\left|\Lambda_i\left(\hat{\beta}\right)\right|^2\right]\sup_{s\in[0,1]} K\left(\frac{s-l}{b_N}\right) \\
&= O\left(\frac{1}{Nb_N}\right) = o(1).
\end{aligned}
$$

To bound the covariance term, by a Taylor expansion of $\Lambda_i\left(\hat{\beta}\right)$ at β_i and Conditions 1.1 and 3.2, we have

$$
\begin{aligned}
& Cov\left[\Lambda_i\left(\hat{\beta}\right)K\left(\frac{R_i/N-l}{b_N}\right), \Lambda_j\left(\hat{\beta}\right)K\left(\frac{R_j/N-l}{b_N}\right)\right] \\
= \; & Cov\left[\frac{\partial}{\partial\beta}\Lambda_i\left(\dot{\beta}_i\right)\left(\hat{\beta}-\beta_i\right)K\left(\frac{R_i/N-l}{b_N}\right), \Lambda_j\left(\beta_j\right)K\left(\frac{R_j/N-l}{b_N}\right)\right] \\
& + Cov\left[\Lambda_i\left(\beta_i\right)K\left(\frac{R_i/N-l}{b_N}\right), \frac{\partial}{\partial\beta}\Lambda_j\left(\dot{\beta}_j\right)\left(\hat{\beta}-\beta_i\right)K\left(\frac{R_j/N-l}{b_N}\right)\right] \\
& + Cov\left[\frac{\partial}{\partial\beta}\Lambda_i\left(\dot{\beta}_i\right)\left(\hat{\beta}-\beta_i\right)K\left(\frac{R_i/N-l}{b_N}\right), \frac{\partial}{\partial\beta}\Lambda_j\left(\dot{\beta}_j\right)\left(\hat{\beta}-\beta_i\right)K\left(\frac{R_j/N-l}{b_N}\right)\right] \\
\leq \; & C\left(\max_i E\left[\left|\hat{\beta}-\beta_i\right|^4\right]\right)^{1/4}\left(E\left[K\left(\frac{R_j/N-l}{b_N}\right)^2\right]\right)^{1/2} \\
& + C\left(\max_j E\left[\left|\hat{\beta}-\beta_j\right|^4\right]\right)^{1/4}\left(E\left[K\left(\frac{R_i/N-l}{b_N}\right)^2\right]\right)^{1/2} + C\left(\max_j E\left[\left|\hat{\beta}-\beta_j\right|^4\right]\right)^{1/2}
\end{aligned}
$$

Then for some finite constants C' and C'', substitute the above in to obtain

$$
\begin{aligned}
& \frac{2}{N^2 b_N^2}\sum_{1\leq i<j\leq N} Cov\left[\Lambda_i\left(\hat{\beta}\right)K\left(\frac{R_i/N-l}{b_N}\right), \Lambda_j\left(\hat{\beta}\right)K\left(\frac{R_j/N-l}{b_N}\right)\right] \\
\leq \; & \frac{C'}{Nb_N^2}\sum_{i=1}^{N} K\left(\frac{i/N-l}{b_N}\right)\left(\max_i E\left[\left|\hat{\beta}-\beta_i\right|^4\right]\right)^{1/4} \\
\leq \; & C'' N^{-1/2} b_N^{-1} = o(1)
\end{aligned}
$$

where the last inequality is implied by Conditions 3.2-3.4. Hence, by Markov's inequality, we establish that for any fixed l, $\hat{\Gamma}(l) - \Gamma(l) = o_p(1)$.

To prove the uniform convergence on S, any compact subset of $(0,1)$ on which $\Gamma(\cdot)$ is continuous, by Theorem 15.5 of Billingsley (1968), it suffices to prove the uniform tightness of $\hat{\Gamma}(\cdot)$. For any $s, s_1 \in S$,

$$
\begin{aligned}
\left|\hat{\Gamma}(s)-\hat{\Gamma}(s_1)\right| &\leq \frac{1}{Nb_N}\sum_{i=1}^{N}\left|\Lambda_{(i)}\left(\hat{\beta}\right)\right|\left|K\left(\frac{i/N-s}{b_N}\right)-K\left(\frac{i/N-s_1}{b_N}\right)\right| + O\left(N^{-1}\right) \\
&\leq \left(\frac{1}{N}\sum_{i=1}^{N}\left|\Lambda_i\left(\hat{\beta}\right)\right|\right)\sup_{l\in S}\frac{1}{b_N}\left|K\left(\frac{l-s}{b_N}\right)-K\left(\frac{l-s_1}{b_N}\right)\right| + O\left(N^{-1}\right)
\end{aligned}
$$

$$\leq \left(\frac{1}{N}\sum_{i=1}^{N}\left|\Lambda_i\left(\hat{\beta}\right)\right|\right)\sup_{l\in S}\left|K'\left(\frac{l}{b_N}\right)b_N^{-2}\right||s-s_1|+O\left(N^{-1}\right).$$

Then by Conditions 1.6 and 3.2 and Markov's inequality,

$$\begin{aligned} P\left(\sup_{\substack{s_1\leq s\leq s_1+v\\ s_1,s_1+v\in S}}\left|\hat{\Gamma}(s)-\hat{\Gamma}(s_1)\right|>\varepsilon_1\right) &\leq P\left(\left(\frac{1}{N}\sum_{i=1}^{N}\left|\Lambda_i\left(\hat{\beta}\right)\right|\right)\sup_{l\in S}\left|K'\left(\frac{l}{b_N}\right)b_N^{-2}\right|v>\varepsilon_1\right) \\ &\leq \max_i\left(E\left[\left|\Lambda_i\left(\hat{\beta}\right)\right|^2\right]\right)^{1/2}\sup_{l\in S}\left|K'\left(\frac{l}{b_N}\right)b_N^{-2}\right|v/\varepsilon_1 \\ &\leq \varepsilon \text{ for a sufficiently large } N. \end{aligned}$$

Part (ii) follows from Lemma E.3 and a similar derivation as in part (i).

For Part (iii), we first establish the pointwise convergence and the tightness under the assumption that $\Gamma(\cdot)$ and $H(\cdot)$ are twice continuously differentiable on $(0,1)$. The case where $\Gamma(\cdot)$ and $H(\cdot)$ have a finite number of discontinuities is studied at the end of the proof. When $l=0$, the proof is trivial.

For any fixed $l\in(0,1]$,

$$\begin{aligned} \hat{h}(l)-h(l) &= N^{-1}\sum_{i=1}^{[lN]}\frac{1}{e_1'\hat{\Gamma}^{-1}(i/N)\hat{H}(i/N)'\hat{\Gamma}^{-1}(i/N)e_1}-\int_0^l\frac{1}{e_1'\Gamma^{-1}(s)H(s)'\Gamma^{-1}(s)e_1}ds \\ &= N^{-1}\sum_{i=1}^{[lN]}\frac{1}{e_1'\hat{\Gamma}^{-1}(i/N)\hat{H}(i/N)'\hat{\Gamma}^{-1}(i/N)e_1}-N^{-1}\sum_{i=1}^{[lN]}\frac{1}{e_1'\Gamma^{-1}(i/N)H(i/N)'\Gamma^{-1}(i/N)e_1} \\ &\quad +N^{-1}\sum_{i=1}^{[lN]}\frac{1}{e_1'\Gamma^{-1}(i/N)H(i/N)'\Gamma^{-1}(i/N)e_1}-\int_0^l\frac{1}{e_1'\Gamma^{-1}(s)H(s)'\Gamma^{-1}(s)e_1}ds \\ &\equiv A_{N1}(l)+A_{N2}(l). \end{aligned}$$

By the continuity of $\Gamma(\cdot)$ and $H(\cdot)$, A_{N2} is $o(1)$. For A_{N1}, denote $\hat{\Omega}(s)=\left(e_1'\hat{\Gamma}^{-1}(s)\hat{H}(s)'\hat{\Gamma}^{-1}(s)e_1\right)^{-1}$ and $\Omega(s)=\left(e_1'\Gamma^{-1}(s)H(s)'\Gamma^{-1}(s)e_1\right)^{-1}$. Define $Z(s)=\hat{\Omega}(s)-\Omega(s)$; so $A_{N1}(l)=N^{-1}\sum_{i=1}^{[lN]}Z(i/N)$. Then for an arbitrarily small $\varepsilon>0$ and $l\in(\varepsilon,1-\varepsilon)$, we have $N^{-1}\sum_{i=[\varepsilon N]}^{[lN]}Z_{i/N}=o_p(1)$ by parts (i) and (ii), and for any fixed $v>0$, by Markov's inequality,

$$P\left(\left|N^{-1}\sum_{i=1}^{[\varepsilon N]}Z(i/N)\right|>v\right)\leq N^{-1}\sum_{i=1}^{[\varepsilon N]}E\left[|Z(i/N)|\right]/v$$

$$\leq \frac{1}{v}\int_0^{\varepsilon}E\left[|Z(s)|\right]ds$$

$$
\begin{aligned}
&\leq \frac{\varepsilon}{v} \sup_{s\in[0,\varepsilon]} E\left[|Z(s)|\right] \\
&\leq \frac{\varepsilon}{v}\left(\sup_{s\in[0,\varepsilon]} E\left[\left|\hat{\Omega}(s)\right|\right] + \sup_{s\in[0,\varepsilon]} E\left[|\Omega(s)|\right]\right)
\end{aligned}
$$

By Conditions 1.6 and 3.2, it is easy to establish that the item in the bracket is bounded. A similar bound can be found for $P\left(N^{-1}\sum_{i=[(1-\varepsilon)N]}^{N} Z(i/N) > v\right)$, and then the pointwise convergence of $\hat{h}(\cdot) - h(\cdot)$ to 0 in probability follows from an arbitrary ε. Since $\hat{h}(1) \xrightarrow{p} h(1)$, the uniform convergences of $\hat{h}(\cdot)$ and $\hat{g}(\cdot)$ are equivalent by Slutsky's theorem. WLOG, we focus on $\hat{g}(\cdot)$.

Next, we prove the tightness of $\hat{g}(\cdot)$, from which the uniform convergence follows. The proof is based on the argument in Lemma A.3 of Hansen (2000a). For any $s_1 \in [0,1), \varepsilon > 0, \varepsilon_1 > 0$, and $v \geq N^{-1}$, let n be an integer satisfying $Nv/2 \leq n \leq Nv$. Set $v_n = v/n$. For $k = 1, ..., n+1$, let $s_k = s_1 + v_n(k-1)$. Note that for $s_k \leq s \leq s_{k+1}$,

$$
\begin{aligned}
\left|\hat{g}(s) - \hat{g}(s_k)\right| &\leq \hat{g}(s_{k+1}) - \hat{g}(s_k) \\
&\leq \left|\hat{g}(s_{k+1}) - \hat{g}(s_k) - E\left[\hat{g}(s_{k+1}) - \hat{g}(s_k)\right]\right| + E\left[\hat{g}(s_{k+1}) - \hat{g}(s_k)\right].
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\sup_{s_1\leq s\leq s_1+v} \left|\hat{g}(s) - \hat{g}(s_1)\right| \leq\ & \max_{2\leq k\leq n+1} \left|\hat{g}(s_1) - \hat{g}(s_k)\right| \\
&+ \max_{1\leq k\leq n} \left|\hat{g}(s_{k+1}) - \hat{g}(s_k) - E\left[\hat{g}(s_{k+1}) - \hat{g}(s_k)\right]\right| \\
&+ \max_{1\leq k\leq n} E\left[\hat{g}(s_{k+1}) - \hat{g}(s_k)\right].
\end{aligned} \tag{47}
$$

Note that by parts (i) and (ii), it is easy to establish that $E\left[\left|\hat{\Omega}(s) - \Omega(s)\right|^2\right] = O\left(b_N^4 + (Nb_N)^{-1}\right)$ for any $s \in (0,1)$. Then to bound the first term, since $\hat{\Omega}(s) \geq 0$, we have, for any $\eta > 0$

$$
\begin{aligned}
P\left(\max_{2\leq k\leq n+1} \left|\hat{g}(s_1) - \hat{g}(s_k)\right| > \eta\right) &\leq P\left(\hat{g}(s_n) - \hat{g}(s_1) > \eta\right) \\
&\leq E\left[\hat{g}(s_n) - \hat{g}(s_1)\right]/\eta \\
&= \left(g(s_n) - g(s_1) + O\left(b_N^2\right)\right)/\eta \\
&\leq \varepsilon/(3\eta)
\end{aligned}
$$

for a sufficiently large N. For the second term of (47), we have

$$
\begin{aligned}
& E\left[\left|\hat{g}(s_{k+1}) - \hat{g}(s_k) - E\left[\hat{g}(s_{k+1}) - \hat{g}(s_k)\right]\right|\right] \\
\leq\ & E\left[\left|\hat{g}(s_{k+1}) - E\left[\hat{g}(s_{k+1})\right]\right|\right] \\
& + E\left[\left|\hat{g}(s_k) - E\left[\hat{g}(s_k)\right]\right|\right]
\end{aligned}
$$

$$
\begin{aligned}
&\leq E\left[\left|\hat{g}\left(s_{k+1}\right)-E\left[\hat{g}\left(s_{k+1}\right)\right]\right|^{2}\right]^{1/2} \\
&\quad +E\left[\left|\hat{g}\left(s_{k}\right)-E\left[\hat{g}\left(s_{k}\right)\right]\right|^{2}\right]^{1/2} \\
&= O\left(b_{N}^{2}+\left(Nb_{N}\right)^{-1/2}\right). \qquad (48)
\end{aligned}
$$

Since the bound is independent of s, we have

$$
P\left(\max_{1\leq k\leq n}\left|\hat{g}\left(s_{k+1}\right)-\hat{g}\left(s_{k}\right)-E\left[\hat{g}\left(s_{k+1}\right)-\hat{g}\left(s_{k}\right)\right]\right|>\eta\right)\leq\frac{C}{\eta}\left(b_{N}^{2}+\left(Nb_{N}\right)^{-1/2}\right)\leq\varepsilon/\left(3\eta\right) \quad (49)
$$

for a sufficiently large N where C is some finite positive constant.

Finally, the last term of (47) is bounded by

$$
\begin{aligned}
E\left[\hat{g}\left(s_{k+1}\right)-\hat{g}\left(s_{k}\right)\right] &= E\left[g\left(s_{k+1}\right)-g\left(s_{k}\right)\right]+O\left(b_{N}^{2}\right) \\
&= \sup_{l\in[0,1]}\Omega\left(l\right)v_{n}+O\left(b_{N}^{2}\right) \\
&< \varepsilon/3 \qquad (50)
\end{aligned}
$$

for a sufficiently large N. Thus the tightness of $\hat{g}\left(s\right)$ follows from combining (48)-(50).

Now suppose there exist a finite number of points $s_{1}^{*},...,s_{p}^{*}$ at which $\Gamma\left(\cdot\right)$ and $H\left(\cdot\right)$ are discontinuous. WLOG, assume $p=1$ and denote the discontinuity as s^{*}. For any small positive η, Theorem 3 implies that $\hat{\Gamma}\left(s\right)$ and $\hat{H}\left(s\right)$ are uniformly consistent on $[\eta,s^{*}-\eta]$ and $[s^{*}+\eta,1-\eta]$. Thus, we have $\sup_{s\in[\eta,s^{*}-\eta]}\left|\hat{g}\left(s\right)-g\left(s\right)\right|\xrightarrow{p}0$ and $\sup_{s\in[s^{*}+\eta,1-\eta]}\left|\int_{s^{*}+\eta}^{s}Z\left(s\right)ds\right|\xrightarrow{p}0$. Finally, by a similar argument for bounding $N^{-1}\sum_{i=1}^{[\varepsilon N]}Z\left(i/N\right)$ as presented above, it is easy to show that $\int_{s^{*}-\eta}^{s^{*}+\eta}Z\left(s\right)ds$ is $o_{p}\left(1\right)$ since η is arbitrary. Then, the proof is complete. ■

Proof of Theorem 3. This theorem holds trivially for $s=0$. For any fixed $s\in(0,1]$, WLOG, we assume $\hat{g}^{-1}\left(s\right)\leq g\left(s\right)$. Given $\hat{h}\left(1\right)\xrightarrow{p}h\left(1\right)$, we have

$$
\begin{aligned}
& e_{1}'\mathcal{T}\hat{G}_{N}\left(s\right)-e_{1}'\mathcal{T}\tilde{G}_{N}\left(s\right) \\
={}& \sqrt{h\left(1\right)}N^{-1/2}\sum_{i=1}^{\left[\hat{g}^{-1}(s)N\right]}e_{1}'\frac{\widehat{\partial g}\left(i/N\right)}{\partial l}\hat{\Gamma}\left(i/N\right)^{-1}m_{(i)}\left(\hat{\beta}\right) \\
& -\sqrt{h\left(1\right)}N^{-1/2}\sum_{i=1}^{\left[g^{-1}(s)N\right]}e_{1}'\frac{\partial g\left(i/N\right)}{\partial l}\Gamma\left(i/N\right)^{-1}m_{(i)}\left(\hat{\beta}\right)+o_{p}\left(1\right) \\
={}& \sqrt{h\left(1\right)}N^{-1/2}\sum_{i=1}^{[\varepsilon N]}e_{1}'\left(\frac{\widehat{\partial g}\left(i/N\right)}{\partial l}\hat{\Gamma}\left(i/N\right)^{-1}-\frac{\partial g\left(i/N\right)}{\partial l}\Gamma\left(i/N\right)^{-1}\right)m_{(i)}\left(\hat{\beta}\right) \\
& +\sqrt{h\left(1\right)}N^{-1/2}\sum_{i=[\varepsilon N]+1}^{\left[\hat{g}^{-1}(s)N\right]}e_{1}'\left(\frac{\widehat{\partial g}\left(i/N\right)}{\partial l}\hat{\Gamma}\left(i/N\right)^{-1}-\frac{\partial g\left(i/N\right)}{\partial l}\Gamma\left(i/N\right)^{-1}\right)m_{(i)}\left(\hat{\beta}\right)
\end{aligned}
$$

$$
\begin{aligned}
& -\sqrt{h(1)}N^{-1/2}\sum_{i=\left[\hat{g}^{-1}(s)N\right]+1}^{\left[g^{-1}(s)N\right]} e_1'\frac{\partial g(i/N)}{\partial l}\Gamma(i/N)^{-1} m_{(i)}\left(\hat{\beta}\right)+o_p(1) \\
= \quad & A_{1N}(s)+A_{2N}(s)-A_{3N}(s)+o_p(1).
\end{aligned}
$$

We now show that A_{jN} is $o_p(1)$ for $j=1,2,3$. First, by Theorem 3 and the continuous mapping theorem, $A_{2N}(s)=o_p(1)$. Second, denote

$$
\begin{aligned}
z_{(i)} &= e_1'\frac{\partial g(i/N)}{\partial l}\Gamma(i/N)^{-1} m_{(i)}\left(\hat{\beta}\right) \\
z_i &= e_1'\frac{\partial g(R_i/N)}{\partial l}\Gamma(R_i/N)^{-1} m_i\left(\hat{\beta}\right) \\
z_i^0 &= e_1'\frac{\partial g(R_i/N)}{\partial l}\Gamma(R_i/N)^{-1} m_i(\beta_i)
\end{aligned}
$$

and $\mathbf{1}_s(R_i)$ for $\mathbf{1}\left[\left[\hat{g}^{-1}(s)N\right]+1\leq R_i\leq\left[g^{-1}(s)N\right]\right]$. Then, $A_{3N}(s)$ can be decomposed into $\sqrt{h(1)}\left(B_{1N}(s)+B_{2N}(s)\right)$ where

$$
\begin{aligned}
B_{1N}(s) &= N^{-1/2}\sum_{i=1}^{N} z_i^0\mathbf{1}_s(R_i) \\
B_{2N}(s) &= N^{-1/2}\sum_{i=1}^{N} e_1'\frac{\partial g(R_i/N)}{\partial l}\Gamma(R_i/N)^{-1}\Lambda_i\left(\dot{\beta}_i\right)\left(\hat{\beta}-\beta_i\right)\mathbf{1}_s(R_i)
\end{aligned}
$$

for some $\dot{\beta}_i$ between β_i and $\hat{\beta}$. By Conditions 1.1 and 1.7, we have $E\left[B_{1N}(s)\right]=0$ and $E\left[z_i^0\mathbf{1}_s(R_i)\,z_j^0\mathbf{1}_s(R_j)\right]=0$, thus $B_{1N}(s)=o_p(1)$ follows from Markov's inequality and Condition 1.6. To bound $B_{2N}(s)$, by Cauchy Schwarz, we have

$$
\left|\left|B_{2N}(s)\right|\right|\leq CN^{-1}\sum_{i=1}^{N}\left|\left|\Lambda_i\left(\dot{\beta}_i\right)\right|\right|\mathbf{1}_s(R_i)\max_i\sqrt{N}\left|\left|\hat{\beta}-\beta_i\right|\right|.
$$

Since

$$
\begin{aligned}
N^{-1}\sum_{i=1}^{N}E\left[\left|\left|\Lambda_i\left(\dot{\beta}_i\right)\right|\right|\mathbf{1}_s(R_i)\right] &\leq N^{-1}\sum_{i=1}^{N}\left(E\left[\left|\left|\Lambda_i\left(\dot{\beta}_i\right)\right|\right|^2\right]\right)^{1/2}\max_i\left(E\left[\mathbf{1}_s(R_i)\right]\right)^{1/2} \\
&= o(1),
\end{aligned}
$$

the claim that $B_{2N}(s)=o_p(1)$ follows from Condition 3.4 and $\hat{g}^{-1}(s)\xrightarrow{p}g^{-1}(s)$ uniformly on $s\in[0,1]$. Thus, by combining B_{1N} and B_{2N}, we have established $P\left(\left|A_{3N}(s)\right|>\varepsilon\right)\to 0$.

Finally, for $A_{1N}(s)$, denote $\mathbf{1}_\varepsilon(R_i)$ for $\mathbf{1}\left[R_i\leq[\varepsilon N]\right]$ and rewrite $A_{1N}(s)$ as $\sqrt{h(1)}\left(B_{3N}(s)+B_{4N}(s)\right)$ where

$$
B_{3N}(s) = N^{-1/2}\sum_{i=1}^{N} e_1'\left(\frac{\widehat{\partial g}(R_i/N)}{\partial l}\hat{\Gamma}(R_i/N)^{-1}-\frac{\partial g(R_i/N)}{\partial l}\Gamma(R_i/N)^{-1}\right)m_i(\beta_i)\mathbf{1}_\varepsilon(R_i)
$$

$$B_{4N}(s) = N^{-1/2}\sum_{i=1}^{N} e_1'\left(\frac{\widehat{\partial g}(R_i/N)}{\partial l}\hat{\Gamma}(R_i/N)^{-1} - \frac{\partial g(i/N)}{\partial l}\Gamma(i/N)^{-1}\right)\Lambda_i\left(\dot{\beta}_i\right)\left(\hat{\beta}-\beta_i\right)\mathbf{1}_{\varepsilon}(R_i).$$

By a similar argument for B_{2N}, it is readily shown that $B_{4N}(s) = o_p(1)$. For $B_{3N}(s)$, since we use the leave-one-out kernel to estimate $\Gamma(\cdot)$ and $H(\cdot)$, Conditions 1.1 and 1.7 imply that $E[B_{3N}(s)] = 0$. Moreover, by independence and the construction of $\hat{\Gamma}$ and $\hat{H}$, some tedious calculation shows that for any $i \neq j$,

$$\left|E\left[e_1'\frac{\widehat{\partial g}(R_i/N)}{\partial l}\hat{\Gamma}(R_i/N)^{-1} m_i(\beta_i)\mathbf{1}_{\varepsilon}(R_i)\, e_1'\frac{\widehat{\partial g}(R_j/N)}{\partial l}\hat{\Gamma}(R_j/N)^{-1} m_j\left(\beta_j\right)\mathbf{1}_{\varepsilon}(R_j)\right]\right| \leq CN^{-2}b_N^{-2}. \tag{51}$$

For simplicity, we show (51) without the error from $\widehat{\partial g}(R_i/N)/\partial l$ and in the scalar case. Define $\tilde{\Gamma}(R_i/N)$ as the estimator with both the i-th and j-th observations excluded. Thus, $\hat{\Gamma}(R_i/N) - \tilde{\Gamma}(R_i/N) = \gamma_{1j} + \gamma_{2j}$ where $\gamma_{1j} = \Lambda_j\left(\beta_j\right)K\left(\frac{R_j/N - R_i/N}{b_N}\right)/\sum_{s=1,s\neq i,s\neq j}^{N} K\left(\frac{s/N - i/N}{b_N}\right)$ and $\gamma_{2j} = \partial\Lambda_j\left(\dot{\beta}_j\right)/\partial\beta\left(\hat{\beta}-\beta_j\right)K\left(\frac{R_j/N - R_i/N}{b_N}\right)/\sum_{s=1,s\neq i,s\neq j}^{N} K\left(\frac{s/N - i/N}{b_N}\right)$. Following a similar argument for bounding B_{2N}, γ_{2j} makes no asymptotic effect on the first order behavior of B_{3N}. Thus, it suffices to consider $\hat{\Gamma}(R_i/N) - \tilde{\Gamma}(R_i/N) = \gamma_{1j}$. Since $\Lambda_j\left(\beta_j\right)$ has uniformly bounded second moments, $\gamma_{1j} = O\left(N^{-1}b_N^{-1}\right)$. Then by Conditions 1.1 and 1.6, we have

$$\begin{aligned}
& E\left[e_1'\hat{\Gamma}(R_i/N)^{-1} m_i(\beta_i)\, e_1'\hat{\Gamma}(R_j/N)^{-1} m_j\left(\beta_j\right)\right] \\
= \; & E\left[e_1'\left(\tilde{\Gamma}(R_i/N) + \gamma_{1j}\right)^{-1} m_i(\beta_i)\, e_1'\left(\tilde{\Gamma}(R_j/N)^{-1} + \gamma_{1i}\right) m_j\left(\beta_j\right)\right] \\
= \; & E\left[e_1'\left(\tilde{\Gamma}(R_i/N)^{-1} - \left(I + \tilde{\Gamma}(R_i/N)\gamma_{1j}\right)^{-1}\tilde{\Gamma}(R_i/N)^{-1}\gamma_{1j}\tilde{\Gamma}(R_i/N)\right) m_i(\beta_i)\right. \\
& \left.\times e_1'\left(\tilde{\Gamma}(R_j/N)^{-1} - \left(I + \tilde{\Gamma}(R_j/N)\gamma_{1i}\right)^{-1}\tilde{\Gamma}(R_j/N)^{-1}\gamma_{1i}\tilde{\Gamma}(R_j/N)\right) m_j\left(\beta_j\right)\right] \\
= \; & E\left[e_1'\left(\tilde{\Gamma}(R_i/N)^{-1} - \left(I + O_p\left(N^{-1}b_N^{-1}\right)\right)\tilde{\Gamma}(R_i/N)^{-1}\gamma_{1j}\tilde{\Gamma}(R_i/N)\right) m_i(\beta_i)\right. \\
& \left.\times e_1'\left(\tilde{\Gamma}(R_j/N)^{-1} - \left(I + O_p\left(N^{-1}b_N^{-1}\right)\right)\tilde{\Gamma}(R_j/N)^{-1}\gamma_{1i}\tilde{\Gamma}(R_j/N)\right) m_j\left(\beta_j\right)\right] \\
= \; & O\left(N^{-2}b_N^{-2}\right).
\end{aligned}$$

Following a similar but more tedious argument, the above derivation holds with $\partial g(R_i/N)/\partial l$ and its estimator included. It follows from (51) that

$$E\left[\|B_{3N}(s)\|^2\right] \leq CN^{-1}\sum_{i=1}^{N} E\left[\|m_i(\beta_i)\|^2\right]\max_i E\left[\mathbf{1}_{\varepsilon}(R_i)\right] + O\left(N^{-1}b_N^{-2}\right).$$

Since ε is arbitrary, the above item is arbitrarily close to 0. Then, the proof follows Markov's inequality.

We have established the pointwise convergence that $\mathcal{T}\hat{G}_N(s) - \mathcal{T}\tilde{G}_N(s) \xrightarrow{p} 0$. The uniform tightness of $\mathcal{T}\hat{G}_N(s) - \mathcal{T}\tilde{G}_N(s)$ follows from an argument similar to that in Lemma 1. Thus, $\mathcal{T}\hat{G}_N(s) - \mathcal{T}\tilde{G}_N(s)$ uniformly converges to 0 in probability, and the rest of the proof follows from Theorem 2.

Proof of Lemma 5. It suffices to establish for $s \in [0,1]$,

$$N^{-1/2} \sum_{i=1}^{[sN]} \xi_{(i)} m_{(i)} \left(\beta_{(i)}, \hat{\tau}\right) \Rightarrow \int_0^s H^{1/2}(l)\, dW(l) \tag{52}$$

and

$$\sup_{s\in[0,1]} \sup_{\beta \in B_{\varepsilon_N}(\bar{\beta})} \left|\left| N^{-1} \sum_{i=1}^{[sN]} \Lambda_{(i)}(\beta, \hat{\tau}) - \int_0^s \Gamma(l)\, dl \right|\right| \xrightarrow{p} 0 \tag{53}$$

where $B_{\varepsilon_N}\left(\bar{\beta}\right)$ is the open ball centered at $\bar{\beta}$ with any positive $\varepsilon_N \to 0$. ■

To prove (52), we first establish an important intermediate result

$$\sup_{s\in[0,1]} ||v_N(s;\hat{\tau}) - v_N(s;\tau_0)|| = o_p(1) \tag{54}$$

where $v_N(s,\tau) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \Delta m_i(s,\tau)$ for

$$\Delta m_i(s,\tau) = \xi_i m_i(\beta_i,\tau) \mathbf{1}[q_i \le Q(s)] - E[\xi_i m_i(\beta_i,\tau) \mathbf{1}[q_i \le Q(s)]].$$

By Conditions 4.2 and 4.3, we obtain that for any fixed s, $v_N(s,\hat{\tau}) - v_N(s,\tau_0) \xrightarrow{p} 0$. If we show that $v_N(s,\tau)$ for any fixed τ is uniformly tight in s over $[0,1]$, then $\sup_{s\in[0,1]} ||\Delta v_N(s,\hat{\tau})|| \le \sup_{s\in[0,1]} ||v_N(s,\hat{\tau})|| + \sup_{s\in[0,1]} ||v_N(s,\tau_0)|| \xrightarrow{p} 0$ follows from Theorem 15.5 of Billingsley (1968). For notational simplicity, we establish the uniform tightness in the scalar case. The proof for the multivariate case follows the same logic. For any fixed τ, any $s \in (0,1)$, and $s_1 = s + \eta$ for any positive η, define

$$\Delta \dot{m}_i(\eta,\tau) = \xi_i m_i(\beta_i,\tau) \mathbf{1}[Q(s) < q_i \le Q(s_1)] - E[\xi_i m_i(\beta_i,\tau) \mathbf{1}[Q(s) < q_i \le Q(s)]].$$

Condition 4.3, $E[\Delta \dot{m}_i(\eta,\tau) = 0]$ and the independence of m_i across i imply that[11]

$$E\left[|v_N(s,\tau) - v_N(s_1,\tau)|^2\right] = E\left[\left| N^{-1/2} \sum_{i=1}^N \Delta \dot{m}_i(\eta,\tau) \right|^2\right]$$

$$= \frac{1}{N} \sum_{i=1}^N E\left[\Delta \dot{m}_i(\eta,\tau)^2\right] \le C\eta \text{ for some finite constant } C.$$

[11] This bound also holds under the conditions specified in Section 5.4 by resorting to Corollary 14.5 of Davidson (1994).

Since the above constant C is uniform for all τ that is in a neighborhood of τ_0 by Condition 4.3, the proof of (54) follows from Markov's inequality and the consistency of $\hat{\tau}$.

Then, by a similar argument as above, (54), Proposition 1 of Andrews (1991), and Conditions 4.1 and 4.4, (52) is obtained as follows.

$$
\begin{aligned}
& N^{-1/2}\sum_{i=1}^{[sN]}\xi_{(i)}m_{(i)}\left(\beta_{(i)},\hat{\tau}\right) \\
= \; & N^{-1/2}\sum_{i=1}^{N}\left(\xi_i m_i\left(\beta_i,\tau_0\right)\mathbf{1}\left[q_i\leq Q\left(s\right)\right]-E\left[\xi_i m_i\left(\beta_i,\tau_0\right)\mathbf{1}\left[q_i\leq Q\left(s\right)\right]\right]\right) \\
& +N^{-1/2}\sum_{i=1}^{N}E\left[\xi_i m_i\left(\beta_i,\tau_0\right)\mathbf{1}\left[q_i\leq Q\left(s\right)\right]\right] \\
& +N^{-1/2}\sum_{i=1}^{N}E\left[\xi_i m_i\left(\beta_i,\tau\right)\mathbf{1}\left[q_i\leq Q\left(s\right)\right]\right]|_{\tau=\hat{\tau}} \\
& +v_N\left(s,\hat{\tau}\right)-v_N\left(s,\tau_0\right)+o_p\left(1\right) \\
= \; & N^{-1/2}\sum_{i=1}^{N}\xi_i m_i\left(\beta_i,\tau_0\right)\mathbf{1}\left[q_i\leq Q\left(s\right)\right]+o_p\left(1\right) \\
\Rightarrow \; & \int_0^s H^{1/2}\left(l\right)dW\left(l\right).
\end{aligned}
$$

To prove (53), it suffices to establish

$$
\sup_{s\in[0,1]}\sup_{\beta\in B_{\varepsilon_N}\left(\bar{\beta}\right)}\left\|N^{-1}\sum_{i=1}^{[sN]}\Lambda_{(i)}\left(\beta,\hat{\tau}\right)-N^{-1}\sum_{i=1}^{[sN]}\Lambda_{(i)}\left(\beta,\tau_0\right)\right\|\xrightarrow{p}0.
$$

This follows from Conditions 4.3-4.4 and a similar argument as Lemma 1. Therefore, the proof is complete. ■

Proof of Theorem 4. Following the exact argument of Lemma 4 and Theorem 3, it suffices to establish the consistency of $\hat{\Gamma}\left(\cdot\right)$ and $\hat{H}\left(\cdot\right)$. We show the consistency of $\hat{H}\left(\cdot\right)$ and the same argument applies to $\hat{\Gamma}\left(\cdot\right)$. By an argument similar to that in Lemma E.3, it suffices to establish, for any $\eta\in(0,1/2)$,

$$
\sup_{\eta\leq l\leq 1-\eta}\left\|\hat{H}\left(l\right)-\frac{\sum_{i=1,i\neq[lN]}^{N}m_{(i)}\left(\hat{\beta},\tau_0\right)m_{(i)}\left(\hat{\beta},\tau_0\right)'K\left(\frac{i/N-l}{b_N}\right)}{\sum_{i=1,i\neq[lN]}^{N}K\left(\frac{i/N-l}{b_N}\right)}\right\|=o_p\left(1\right). \tag{55}
$$

For notational ease, we establish (55) only for the scalar case. By standard arguments, $(Nb_N)^{-1}\sum_{i=1,i\neq[lN]}^{N}K\left(\frac{i/N-l}{b_N}\right)$ is $O\left(1\right)$, and then the item in (55) can be written as

$$
\frac{1}{Nb_N}\sum_{i=1,i\neq[lN]}^{N}\left(m_{(i)}\left(\hat{\beta},\hat{\tau}\right)+m_{(i)}\left(\hat{\beta},\tau_0\right)\right)\left(m_{(i)}\left(\hat{\beta},\hat{\tau}\right)-m_{(i)}\left(\hat{\beta},\tau_0\right)\right)K\left(\frac{i/N-l}{b_N}\right)+o_p\left(1\right)
$$

$$= \frac{1}{Nb_N}\sum_{i=1,R_i\neq[lN]}^{N}\left(m_i\left(\hat{\beta},\hat{\tau}\right)+m_i\left(\hat{\beta},\tau_0\right)\right)\frac{\partial m_i\left(\hat{\beta},\dot{\tau}_i\right)}{\partial\tau}\left(\hat{\tau}\left(X_i\right)-\tau_0\left(X_i\right)\right)K\left(\frac{R_i/N-l}{b_N}\right)+o_p\left(1\right)$$

where $\dot{\tau}_i$ is between $\hat{\tau}\left(X_i\right)$ and $\tau_0\left(X_i\right)$. By Condition 4.3 and Cauchy Shwartz, the first item in the above expression is $o_p\left(1\right)$ and hence (55) holds. ■

Proof of Lemma 6. We first establish (35)-(37). For (37), the exact argument in Lemma 1 applies. For (35), note that by Condition 1.6' and Cauchy Shwartz, for any fixed s

$$\begin{aligned} E\left[\left|\left|v_N^0\left(s\right)-v_N^1\left(s\right)\right|\right|^2\right] &= N^{-1}\sum_{i=1}^{N}E\left[\left|\left|\Delta m_i\mathbf{1}\left[q_i\leq Q\left(s\right)\right]-E\left[\Delta m_i\mathbf{1}\left[q_i\leq Q\left(s\right)\right]\right]\right|\right|^2\right] \\ &\to 0 \end{aligned}$$

where $\Delta m_i = m_i\left(\bar{\beta}\right)-m_i\left(\beta_i\right)$. Then $v_N^0\left(s\right)-v_N^1\left(s\right)=o_p\left(1\right)$ holds pointwisely. The uniform tightness of $v_N^0\left(s\right)-v_N^1\left(s\right)$ follows from a similar argument as in Lemma E.2, and hence (35) follows from Theorem 15.5 of Billingsley (1968).

For (36), note that for any fixed $\in B_{\varepsilon_N}\left(\bar{\beta}\right)$ and $s\in[0,1]$, Lemma E.1, Condition 1.6', and Cauchy Shwartz imply that, for a large enough N,

$$N^{-1}\sum_{i=1}^{N}E\left[m_i\left(\beta\right)\left(\mathbf{1}\left[q_i\leq Q\left(s\right)\right]-\mathbf{1}\left[q_i\leq\hat{Q}\left(s\right)\right]\right)\right]\leq CN^{-1/2}.$$

Since the above constant C is uniform for all β, a similar argument of Lemma 1 leads to (36).

Next, we establish the main convergence result of Lemma 6. Define the function $a:[0,1]\to\mathbb{R}^k$ as $a\left(\eta\right)=E\left[m_i\left(\bar{\beta}+\eta\left(\beta_{(i)}-\bar{\beta}\right)\right)\mathbf{1}\left[q_i\leq Q\left(s\right)\right]\right]$, and then $a'\left(\eta\right)=\left(\partial E\left[m_i\left(\bar{\beta}+\eta\left(\beta_{(i)}-\bar{\beta}\right)\right)\mathbf{1}\left[q_i\leq Q\left(s\right)\right]\right]/\partial\eta\right)\left(\beta_{(i)}-\bar{\beta}\right)$. Note that

$$\begin{aligned} & E\left[m_i\left(\beta_i\right)\mathbf{1}\left[q_i\leq Q\left(s\right)\right]\right]-E\left[m_i\left(\bar{\beta}\right)\mathbf{1}\left[q_i\leq Q\left(s\right)\right]\right] \\ =\; & a\left(1\right)-a\left(0\right) \\ =\; & \int_0^1\left(\frac{\partial}{\partial\eta}E\left[m_i\left(\bar{\beta}+\eta\left(\beta_i-\bar{\beta}\right)\right)\mathbf{1}\left[q_i\leq Q\left(s\right)\right]\right]\right)d\eta\left(\beta_i-\bar{\beta}\right). \end{aligned}$$

Then given (35)-(37), we have, in the unstable model (2),

$$\begin{aligned} & N^{-1/2}\sum_{i=1}^{[sN]}m_{(i)}(\bar{\beta}) \\ =\; & N^{-1/2}\sum_{i=1}^{N}m_i(\beta_i)\mathbf{1}\left[q_i\leq Q\left(s\right)\right]+N^{-1/2}\sum_{i=1}^{N}E\left[m_i\left(\bar{\beta}\right)\mathbf{1}\left[q_i\leq Q\left(s\right)\right]\right] \\ & +v_N^0\left(s\right)-v_N^1\left(s\right)+o_p\left(1\right) \end{aligned}$$

$$
\begin{aligned}
= \quad & N^{-1/2}\sum_{i=1}^{N} m_i(\beta_i)\mathbf{1}\left[q_i \le Q(s)\right] + v_N^0(s) - v_N^1(s) \\
& -\int_0^1 \left(N^{-1}\sum_{i=1}^{N} \frac{\partial}{\partial \eta} E\left[m_i\left(\bar{\beta} + \eta\left(\beta_i - \bar{\beta}\right)\right)\mathbf{1}\left[q_i \le Q(s)\right]\right] Mb \right) d\eta \\
& -\int_0^1 \left(N^{-1}\sum_{i=1}^{N} \frac{\partial}{\partial \eta} E\left[m_i\left(\bar{\beta} + \eta\left(\beta_i - \bar{\beta}\right)\right)\mathbf{1}\left[q_i \le Q(s)\right]\right] Md\mathbf{1}\left[i/N \le r\right] \right) d\eta + o_p(1) \\
\Rightarrow \quad & \tilde{G}(s).
\end{aligned}
$$

■

Proof of Lemma 7. For the first convergence, by the central limit theorem with weak dependent data (see, for example, Andrews (1991)), it suffices to establish that uniformly in $s \in [0,1]$

$$
N^{-1/2}\sum_{i=1}^{N} m_i(\beta_i)\mathbf{1}\left[q_i \le \hat{Q}(s)\right] - N^{-1/2}\sum_{i=1}^{N} m_i(\beta_i)\mathbf{1}\left[q_i \le Q(s)\right] = o_p(1). \tag{56}
$$

We first establish the pointwise convergence. For notational simplicity, we denote γ_0 and $\hat{\gamma}$ for $Q(s)$ and $\hat{Q}(s)$, and restrict attention to the scalar case. The proof of the multivariate case follows from the same logic. When $s = 0$ and 1, (56) holds trivially. For any $s \in (0,1)$, we can choose a constant v such that $P\left(\left|\hat{Q}(s) - Q(s)\right| \le vN^{-1/2}\right) \ge 1 - \varepsilon$ for any ε. Consider $\{\gamma_j\}_{j=1}^{m_\gamma}$ to be an equally-spaced grid on $[-vN^{-1/2}, vN^{-1/2}]$ such that $\gamma_{j+1} - \gamma_j = 2vN^{-1/2}/m_\gamma$. So for any η, we have

$$
\begin{aligned}
& P\left(\left| N^{-1/2}\sum_{i=1}^{N} m_i(\beta_i)\left(\mathbf{1}\left[q_i \le \hat{\gamma}\right] - \mathbf{1}\left[q_i \le \gamma_0\right]\right)\right|\right) \\
\le \quad & P\left(\sup_{|\gamma-\gamma_0| \le vN^{-1/2}}\left| N^{-1/2}\sum_{i=1}^{N} m_i(\beta_i)\left(\mathbf{1}\left[q_i \le \gamma\right] - \mathbf{1}\left[q_i \le \gamma_0\right]\right)\right|\right) + \varepsilon \\
\le \quad & 2P\left(\max_j\left| N^{-1/2}\sum_{i=1}^{N} m_i(\beta_i)\left(\mathbf{1}\left[q_i \le \gamma_j\right] - \mathbf{1}\left[q_i \le \gamma_0\right]\right)\right| > \eta\right) \\
& +2P\left(\max_j \sup_{\gamma_j \le \gamma \le \gamma_{j+1}}\left| N^{-1/2}\sum_{i=1}^{N} m_i(\beta_i)\left(\mathbf{1}\left[q_i \le \gamma\right] - \mathbf{1}\left[q_i \le \gamma_j\right]\right)\right| > \eta\right).
\end{aligned}
$$

Markov's inequality, Cauchy Schwarz, and Conditions 1.6 and 1.1" imply that

$$
\begin{aligned}
& P\left(\max_j\left| N^{-1/2}\sum_{i=1}^{N} m_i(\beta_i)\left(\mathbf{1}\left[q_i \le \gamma_j\right] - \mathbf{1}\left[q_i \le \gamma_0\right]\right)\right| > \eta\right) \\
\le \quad & \frac{1}{\eta^2 N}\sum_{j=1}^{m_\gamma}\sum_{i=1}^{N} E\left[m_i(\beta_i)^2\left(\mathbf{1}\left[q_i \le \gamma_j\right] - \mathbf{1}\left[q_i \le \gamma_0\right]\right)^2\right]
\end{aligned}
$$

$$
\begin{aligned}
&\leq \frac{1}{\eta^2}\sup_i E\left[m_i(\beta_i)^4\right]^{1/2}\sum_{j=1}^{m_\gamma} E\left[\left(\mathbf{1}\left[q_i\leq\gamma_j\right]-\mathbf{1}\left[q_i\leq\gamma_0\right]\right)^4\right]^{1/2}\\
&\leq \frac{C}{\eta^2\sqrt{N}} \text{ for some finite constant } C. \qquad (57)
\end{aligned}
$$

Then, by a similar argument as in Lemma A.3 of Hansen (2000a), we have

$$
\begin{aligned}
& P\left(\max_j \sup_{\gamma_j\leq\gamma\leq\gamma_{j+1}}\left|N^{-1/2}\sum_{i=1}^N m_i(\beta_i)\left(\mathbf{1}\left[q_i\leq\gamma\right]-\mathbf{1}\left[q_i\leq\gamma_j\right]\right)\right|>\eta\right)\\
\leq & \sum_{j=1}^{m_\gamma} P\left(\sup_{\gamma_j\leq\gamma\leq\gamma_{j+1}}\left|N^{-1/2}\sum_{i=1}^N m_i(\beta_i)\left(\mathbf{1}\left[q_i\leq\gamma\right]-\mathbf{1}\left[q_i\leq\gamma_j\right]\right)\right|>\eta\right)\\
\leq & \frac{C}{\eta^4}\sum_{j=1}^{m_\gamma}\frac{v^2}{Nm_\gamma^2} \text{ for some finite constant } C. \qquad (58)
\end{aligned}
$$

Hence, (56) holds by combining (57) and (58). The uniform tightness follows from a similar argument as in Lemma E.2. Then, the uniform convergence follows from Theorem 15.5 of Billingsley (1968). For the second convergence, the proof follows from a similar argument as in Lemma 1 and applying Corollary 14.5 of Davidson (1994), and hence is omitted. The last convergence follows readily from the continuous mapping theorem and Slutsky's theorem. ■

Proof of Lemma 8. We present the proof for $\hat{L}_{1N}(s)$, which covers $\hat{L}_{2N}(s)$ by the exact argument. For notational simplicity, we restrict attention to the scalar case. The proof of the multivariate case follows from the same logic. We first establish the pointwise convergence, that is, $\hat{L}_{1p_N}(s)-L_1(s)=o_p(1)$ for any fixed s. When $L_1(s)$ is not well-defined for $s=0$ and 1, we can employ the truncation argument discussed in Appendix A. In the following proof, we assume $L_1(s)$ is well-defined on $[0,1]$. Denote

$$
\begin{aligned}
\hat{z}_i &= m_i\left(\hat{\beta}\right)m_{i-k}\left(\hat{\beta}\right)\mathbf{1}\left[q_{i-k}\leq\hat{Q}(s)\right]\\
z_i &= m_i(\beta_i)\,m_{i-k}\left(\beta_{i-k}\right)\mathbf{1}\left[q_{i-k}\leq Q(s)\right]
\end{aligned}
$$

so that by Condition 1.1" and Corollary 14.5 of Davidson (1994), we have

$$
\begin{aligned}
\left|L_{1N}(s)-L_{1p_N}(s)\right| &= \left|\sum_{k=p_N+1}^N\frac{1}{N}\sum_{i=p_N+1}^N E\left[m_i(\beta_i)\,m_{i-k}\left(\beta_{i-k}\right)\mathbf{1}\left[q_{i-k}\leq Q(s)\right]|q_i=Q(s)\right]\right|\\
&\leq \left(\sum_{k=p_N+1}^N\phi_k^{1/2}\right)\sup_i E\left[m_i(\beta_i)^2|q_i=Q(s)\right]\sup_i E\left[m_i(\beta_i)^2\right]\\
&\to 0 \text{ as } p_N\to\infty.
\end{aligned}
$$

Then it suffices to establish the following item is $o_p(1)$:

$$\left|L_{1p_N}(s) - \hat{L}_{1p_N}(s)\right| = \left|\sum_{k=1}^{p_N} \frac{1}{Nb_N} \sum_{i=k+1}^{N} \left(\hat{z}_i - E\left[z_i | q_i = Q(s)\right]\right) K\left(\frac{R_i/N - s}{b_N}\right)\right|.$$

For fixed k,

$$\begin{aligned}
& \frac{1}{Nb_N} \sum_{i=k+1}^{N} \left(\hat{z}_i - E\left[z_i | q_i = Q(s)\right]\right) K\left(\frac{R_i/N - s}{b_N}\right) \\
= \; & \frac{1}{Nb_N} \sum_{i=k+1}^{N} \left(\hat{z}_i - z_i\right) K\left(\frac{R_i/N - s}{b_N}\right) \\
& + \frac{1}{Nb_N} \sum_{i=k+1}^{N} \left(z_i - E\left[z_i | q_i = Q(s)\right]\right) K\left(\frac{R_i/N - s}{b_N}\right) \\
= \; & \frac{1}{Nb_N} \sum_{i=k+1}^{N} \left(m_i\left(\hat{\beta}\right) m_{i-k}\left(\hat{\beta}\right)\left(\mathbf{1}\left[q_{i-k} \leq \hat{Q}(s)\right] - \mathbf{1}\left[q_{i-k} \leq Q(s)\right]\right)\right) K\left(\frac{R_i/N - s}{b_N}\right) \\
& + \frac{1}{Nb_N} \sum_{i=k+1}^{N} \left(m_i\left(\hat{\beta}\right) m_{i-k}\left(\hat{\beta}\right) - m_i\left(\beta_i\right) m_{i-k}\left(\beta_{i-k}\right) \mathbf{1}\left[q_{i-k} \leq Q(s)\right]\right) K\left(\frac{R_i/N - s}{b_N}\right) \\
& + \frac{1}{Nb_N} \sum_{i=k+1}^{N} \left(z_i - E\left[z_i | q_i = Q(s)\right]\right) K\left(\frac{R_i/N - s}{b_N}\right) \\
\equiv \; & A_{1N}(s) + A_{2N}(s) + A_{3N}(s).
\end{aligned}$$

To bound $A_{1N}(s)$, note that when $s = 0$ and 1, $\mathbf{1}\left[q_i \leq \hat{Q}(s)\right] = \mathbf{1}\left[q_i \leq Q(s)\right]$ for all i, and hence $A_{1N}(s) = 0$. For any $s \in (0,1)$, by Lemma E.1, $\hat{Q}(s) - Q(s) = O_p\left(N^{-1/2}\right)$. Then by Cauchy Schwarz and Conditions 1.6 and 3.2, we find, for some finite constant C

$$E\left[\left|A_{1N}(s)\right|\right] \leq C \sup_{l \in [0,1]} b_N^{-1} K\left(\frac{l - s}{b_N}\right) \frac{1}{N} \sum_{i=k+1}^{N} E\left[\left|\mathbf{1}\left[q_{i-k} \leq \hat{Q}(s)\right] - \mathbf{1}\left[q_{i-k} \leq Q(s)\right]\right|\right].$$

By the argument in proving Lemma 7, $\frac{1}{N}\sum_{i=k+1}^{N} E\left[\left|\mathbf{1}\left[q_{i-k} \leq \hat{Q}(s)\right] - \mathbf{1}\left[q_{i-k} \leq Q(s)\right]\right|\right] = O\left(N^{-1/2}\right)$. Thus, $\left|A_{1N}(s)\right| = O_p\left(N^{-1/2}\right)$.

For $A_{2N}(s)$, we have

$$\begin{aligned}
\left|A_{2N}(s)\right| \leq \; & \left|\frac{1}{Nb_N} \sum_{i=k+1}^{N} \left(m_i\left(\hat{\beta}\right)\left(m_{i-k}\left(\hat{\beta}\right) - m_{i-k}\left(\beta_{i-k}\right)\right) \mathbf{1}\left[q_{i-k} \leq Q(s)\right]\right) K\left(\frac{R_i/N - s}{b_N}\right)\right| \\
& + \left|\frac{1}{Nb_N} \sum_{i=k+1}^{N} \left(\left(m_i\left(\hat{\beta}\right) - m_i\left(\beta_i\right)\right) m_{i-k}\left(\beta_{i-k}\right) \mathbf{1}\left[q_{i-k} \leq Q(s)\right]\right) K\left(\frac{R_i/N - s}{b_N}\right)\right|
\end{aligned}$$

$$\begin{aligned}
&\leq \left|\frac{1}{Nb_N}\sum_{i=k+1}^{N}\left(m_i\left(\hat{\beta}\right)\Lambda_{i-k}\left(\dot{\beta}_{i-k}\right)\left(\hat{\beta}-\beta_{i-k}\right)\mathbf{1}\left[q_{i-k}\leq Q\left(s\right)\right]\right)K\left(\frac{R_i/N-s}{b_N}\right)\right| \\
&\quad +\left|\frac{1}{Nb_N}\sum_{i=k+1}^{N}\Lambda_i\left(\dot{\beta}_i\right)\left(\hat{\beta}-\beta_i\right)m_{i-k}\left(\beta_{i-k}\right)\mathbf{1}\left[q_{i-k}\leq Q\left(s\right)\right]K\left(\frac{R_i/N-s}{b_N}\right)\right| \\
&\equiv B_{1N}\left(s\right)+B_{2N}\left(s\right)
\end{aligned}$$

By Conditions 3.2-3.3, we have

$$\begin{aligned}
\left|B_{1N}\left(s\right)\right| &\leq \frac{1}{Nb_N}\sum_{i=k+1}^{N}\left|m_i\left(\hat{\beta}\right)\Lambda_{i-k}\left(\dot{\beta}_{i-k}\right)\right|K\left(\frac{R_i/N-s}{b_N}\right)\max_i\left|\hat{\beta}-\beta_i\right| \\
&\leq \frac{1}{N}\sum_{i=k+1}^{N}\left|m_i\left(\hat{\beta}\right)\Lambda_{i-k}\left(\dot{\beta}_{i-k}\right)\right|\sup_{l\in[0,1]}b_N^{-1}K\left(\frac{l-s}{b_N}\right)\max_i\left|\hat{\beta}-\beta_i\right| \\
&= O_p\left(N^{-1/2}\right).
\end{aligned}$$

Following the exact argument, $\left|B_{2N}\right|=O_p\left(N^{-1/2}\right)$ and hence $A_{2N}\left(s\right)=O_p\left(N^{-1/2}\right)$.

For $A_{3N}\left(s\right)$, by standard argument, we have $E\left[A_{3N}\left(s\right)\right]=O\left(b_N^2\right)$. By a similar argument of eq. (10) of Newey and West (1987) and Corollary 14.5 of Davidson (1994), we have, for a constant C,

$$\begin{aligned}
Var\left[A_{3N}\left(s\right)\right] &= \frac{1}{N^2b_N^2}\sum_{i=k+1}^{N}E\left[z_i^2K^2\left(\frac{R_i/N-s}{b_N}\right)\right] \\
&\quad +\frac{2}{N^2b_N^2}\sum_{l=k+1}^{N}Cov\left[m_lm_{l-k}K\left(\frac{R_i/N-s}{b_N}\right),m_{l-k}m_{l-2k}K\left(\frac{R_i/N-s}{b_N}\right)\right] \\
&\quad +\frac{2}{N^2b_N^2}\sum_{k+1\leq i\neq l}^{N}Cov\left[m_im_{i-k}K\left(\frac{R_i/N-s}{b_N}\right),m_lm_{l-k}K\left(\frac{R_l/N-s}{b_N}\right)\right] \\
&\leq C(k+1)N^{-1}b_N^{-1}.
\end{aligned}$$

Thus, $E\left[A_{3N}\left(s\right)^2\right]\leq C'\left(b_N^4+(k+1)N^{-1}b_N^{-1}\right)$ for some finite constant C'.

Finally, by Markov's inequality, the bounds of $A_{1N}\left(s\right),A_{2N}\left(s\right)$, and $A_{3N}\left(s\right)$, and $p_N^4\left(N^{-1}b_N^{-1}+b_N^{-4}\right)\to 0$, we find for some finite constant C

$$\begin{aligned}
P\left(\left|L_{1p_N}\left(s\right)-\hat{L}_{1p_N}\left(s\right)\right|>\varepsilon\right) &\leq \sum_{k=1}^{p_N}P\left(\left|\frac{1}{N}\sum_{i=k+1}^{N}\left(\hat{z}_i-E\left[z_i|q_i=Q\left(s\right)\right]\right)K_{b_N}\left(\frac{R_i/N-s}{b_N}\right)\right|>\varepsilon/p_N\right) \\
&\leq \sum_{k=1}^{p_N}\sum_{j=1}^{3}P\left(\left|A_{jN}\left(s\right)\right|>\varepsilon/\left(3p_N\right)\right)
\end{aligned}$$

$$\leq \quad CN^{-1/2}+\sum_{k=1}^{p_N} 9\varepsilon^{-2}\left(k+1\right)p_N^2\left(N^{-1}b_N^{-1}+b_N^{-4}\right)+CN^{-1/2}\rightarrow 0.$$

The uniform tightness of $\hat{L}_{1p_N}(\cdot)$ is established by a similar argument as in Theorem 3, and then the proof is complete by Theorem 15.5 of Billingsley (1968). ■

Proof of Theorem 5. By an argument similar to that in Lemma 8, $\hat{\Gamma}(\cdot)$ and $\hat{H}(\cdot)$ are consistent. Then, the proof follows from the exact argument in Lemma 4 and Theorem 3. ■

Proof of Corollary 2. First, Condition C is sufficient Conditions 1, 2, and 3.1-3.3. Since the assumptions made by Klein and Spady (1993) are also implied by Condition C, Condition 3.4 is also satisfied. Next, Condition 4.2 is satisfied since Assumption SE of Andrews (1994) is implied by Conditions C.1-4. For Condition 4.3, the consistency of $\hat{\tau}$ directly follows from Klein and Spady (1993), and the boundedness restrictions are implied by Conditions C.3-4. Condition 4.4 is also satisfied by Condition C.1. Finally, Condition 4.1 holds if we establish equations (4.12) and (4.13) of Andrews (1994). The first is directly implied by Condition C.1. For the second, by standard arguments (see, for example, chapter 2 of Li and Racine (2007)) and Klein and Spady (1993), $\sup_{z\in\mathcal{X}^*}|\hat{\tau}_1(z)-\tau_1(z)|$ and $\sup_{z\in\mathcal{X}^*}||\hat{\tau}_2(z)-\tau_2(z)||$ are both $o_p\left(N^{-1/4}\right)$. Hence, eq. (4.13) is satisfied, and so is Condition 4.1.

Given Conditions 1-4, we next prove $e_1'\hat{G}_N^{*BC}(s)\Rightarrow G_1^*(s)$. Denote $\hat{\beta}$ for $\hat{\beta}_{KS}$, ξ_i for $\mathbf{1}\left[x_i'\bar{\beta}\in\mathcal{X}_N^*\right]$, and $\hat{\xi}_i$ for $\mathbf{1}\left[x_i'\hat{\beta}\in\mathcal{X}_N^*\right]$. Note that the truncation in $\hat{G}_N^{*BC}$ is based on $\hat{\xi}_i$ instead of ξ_i. To apply Theorem 4, we first prove $\sup_{s\in[0,1]}\left|e_1'\hat{G}_N^{BC}(s)-e_1'\bar{G}_N^{BC}(s)\right|=o_p(1)$ where

$$\begin{aligned}
\hat{G}_N^{BC}(s) &= N^{-1/2}\sum_{i=1}^{[sN]}\hat{\xi}_{(i)}\left(y_{(i)}-\hat{\tau}_1\left(x_{(i)}'\hat{\beta}\right)\right)\left(x_i-\hat{\tau}_2\left(x_i'\hat{\beta}\right)\right)\\
\bar{G}_N^{BC}(s) &= N^{-1/2}\sum_{i=1}^{[sN]}\xi_{(i)}\left(y_{(i)}-\hat{\tau}_1\left(x_{(i)}'\hat{\beta}\right)\right)\left(x_i-\hat{\tau}_2\left(x_i'\hat{\beta}\right)\right).
\end{aligned}$$

We do this by establishing the pointwise convergence and the uniform tightness.

For any ε, we can choose the truncation set $\mathcal{X}_N^*$ such that $P\left(x_i'\bar{\beta}\in\mathcal{X}^*\right)\geq 1-\varepsilon$ for a large enough N. Denote $\mathbf{1}_i^+=\mathbf{1}\left(x_i'\hat{\beta}\in\mathcal{X}_N^*,\, x_i'\bar{\beta}\notin\mathcal{X}_N^*\right)$ and $\mathbf{1}_i^-=\mathbf{1}\left(x_i'\hat{\beta}\notin\mathcal{X}_N^*,\, x_i'\bar{\beta}\in\mathcal{X}_N^*\right)$. Then

$$\begin{aligned}
&\hat{G}_N^{BC}(s)-\bar{G}_N^{BC}(s)\\
=\ & N^{-1/2}\sum_{i=1}^{N}\mathbf{1}_i^+\left(y_i-\hat{\tau}_1\left(x_i'\hat{\beta}\right)\right)\left(x_i-\hat{\tau}_2\left(x_i'\hat{\beta}\right)\right)\mathbf{1}\left[q_i\leq\hat{Q}(s)\right]\\
&+N^{-1/2}\sum_{i=1}^{N}\mathbf{1}_i^-\left(y_i-\hat{\tau}_1\left(x_i'\hat{\beta}\right)\right)\left(x_i-\hat{\tau}_2\left(x_i'\hat{\beta}\right)\right)\mathbf{1}\left[q_i\leq\hat{Q}(s)\right]\\
\equiv\ & B_{N1}(s)+B_{N2}(s).
\end{aligned}$$

WLOG, we prove B_{N1} is $o_p(1)$, and B_{N2} follows from the exact argument. Since $\hat{\tau}_1$ is uniformly bounded, Markov's inequality and Conditions C.4 and C.7 imply, for any $\eta > 0$, there exists some positive constants C and $\tilde{v}$ such that for any i,

$$\begin{aligned}
P\left(\|B_{N1}(s)\| > \eta\right) &\leq N^{1/2} E\left[\mathbf{1}_i^{+}\left|y_i - \hat{\tau}_1\left(x_i'\hat{\beta}\right)\right|\left\|x_i - \hat{\tau}_2\left(x_i'\hat{\beta}\right)\right\|\mathbf{1}\left[q_i \leq \hat{Q}(s)\right]\right]/\eta \\
&\leq 2N^{1/2} E\left[\mathbf{1}_i^{+}\right]\left(E\left[\left\|x_i - \hat{\tau}_2\left(x_i'\hat{\beta}\right)\right\|^2\right]\right)^{1/2}/\eta \\
&= N^{1/2} P\left(x_i'\hat{\beta} \in \mathcal{X}_N^*,\, x_i'\bar{\beta} \notin \mathcal{X}_N^*\right) C/\eta \\
&\leq N^{1/2} P\left(\left\|\hat{\beta} - \bar{\beta}\right\| > \|V_\beta\|\,\tilde{v},\, x_i'\bar{\beta} \notin \mathcal{X}_N^*\right) C/\eta
\end{aligned}$$

where V_β is the asymptotic variance of the Klein-Spady estimator. Following the same argument as in eq. (43) of Klein and Spady (1993), we have

$$\sqrt{N}\left(\hat{\beta} - \bar{\beta}\right) = \sqrt{N}\left(\tilde{\beta} - \bar{\beta}\right) + o_p(1)$$

where $\tilde{\beta}$ can be expressed as $N^{-1}\sum_{j=1}^{N} z_j$ for some i.i.d. mean-zero random variable z_i that satisfies $E\left[|z_i|^3\right] < \infty$. Since z_j has bounded second moments, we can further replace $\tilde{\beta}$ by $\tilde{\beta}_{-i} = (N-1)^{-1}\sum_{j \neq i}^{N} z_j$ which is independent of x_i and satisfies $\sqrt{N}\left(\tilde{\beta} - \tilde{\beta}_{-i}\right) = o_p(1)$. Then by Berry-Esseen theorem, we have, for any i,

$$N^{1/2}\sup_{s>0}\left|P\left(\sqrt{N}\left\|\tilde{\beta}_{-i} - \bar{\beta}\right\| > s\,\|V_\beta\|\right) - (1 - 2\Phi(s))\right| \leq C_1$$

for some constant $C_1 < \infty$ where $\Phi(\cdot)$ is the CDF of a standard normal. Since $N^{1/2}\Phi\left(N^{1/2}s\right) \to 0$ for any fixed $s > 0$, we obtain that for any i and a sufficiently large N,

$$\begin{aligned}
& N^{1/2} P\left(\left\|\hat{\beta} - \bar{\beta}\right\| > \tilde{v},\, x_i'\bar{\beta} \notin \mathcal{X}^*\right) \\
= \;& N^{1/2} P\left(\sqrt{N}\left\|\tilde{\beta}_{-i} - \bar{\beta}\right\| > \sqrt{N}\,\|V_\beta\|\left(\tilde{v} + o_p\left(N^{-1/2}\right)\right),\, x_i'\bar{\beta} \notin \mathcal{X}^*\right) \\
\leq \;& \Big\{2N^{1/2}\left(1 - \Phi\left(N^{1/2}\,\|V_\beta\|\left(\tilde{v} + o_p\left(N^{-1/2}\right)\right)\right)\right) \\
& + N^{1/2}\sup_{s>0}\left|P\left(\sqrt{N}\left\|\tilde{\beta}_{-i} - \bar{\beta}\right\| > z\,\|V_\beta\|\right) - 2\left(1 - \Phi(s)\right)\right|\Big\}\varepsilon \\
\leq \;& C_1\varepsilon.
\end{aligned}$$

Hence, the $B_{N11}(s)$ is $o(1)$ since ε is arbitrary. The uniform tightness of $\hat{G}_N^{BC}(s) - \bar{G}_N^{BC}(s)$ follows from a similar argument as in Theorem 3. Thus, we have established $\sup_{s\in[0,1]}\left|e_1'\hat{G}_N^{BC}(s) - e_1'\bar{G}_N^{BC}(s)\right| = o_p(1)$. By a similar argument as in Lemma 4, $\hat{\tau}_2(z)$ is uniformly consistent on any compact set. The consistency of $\hat{\tau}_1$ and $\hat{f}$ follows from Klein and Spady

(1993). Then, the rest of the proof follows from the continuous mapping theorem, Lemma 5, and Theorem 4. ■

Proof of Lemma 9. First, Conditions 1-3 are directly implied by Condition D. Then for the first convergence, by Lemma E.2, replacing the true quantiles with their empirical counterparts makes no asymptotic effect. Then,

$$\begin{aligned} N(r_0)^{-1/2} \sum_{i=1}^{[sN(r_0)]} \Delta\dot{x}_{(i)} \Delta\dot{u}_{(i)} &= N(r_0)^{-1/2} \sum_{i=1}^{N} \Delta x_{i)} \Delta u_i \mathbf{1}[A_i(r_0)] \mathbf{1}[q_{i1} \leq Q_1^c(s)] + o_p(1) \\ &= J_{1N}(s) + o_p(1) \end{aligned}$$

where $\mathbf{1}[A_i(r_0)] = \mathbf{1}[q_{i1} \leq Q_1(r_0), q_{i2} \leq Q_2(r_0)] + \mathbf{1}[q_{i1} > Q_1(r_0), q_{i2} > Q_2(r_0)]$. For each $s \in (0, r_0^c)$, $\Delta x_i \Delta u_i \mathbf{1}[A_i(r_0)] \mathbf{1}[q_{i1} \leq Q_1^c(s)]$ is a square integrable mean zero random variable, so $J_{1N}(s)$ converges pointwisely to a Gaussian distributed random variable. The uniform tightness can be proved along the lines in the proof of Lemma 1. Then, the theorem follows from Theorem 15.5 of Billingsley (1968) and a similar argument for the part $s > r_0^c$.

For the second convergence, we have for s such that $\hat{Q}_1^c(s) \leq \hat{Q}_1(r_0)$,

$$\begin{aligned} N(r_0)^{-1} \sum_{i=1}^{[sN(r_0)]} \Delta\dot{x}_{(i)} \Delta\dot{x}_{(i)}' &= N(r_0)^{-1} \sum_{i=1}^{N(r_0)} \Delta\dot{x}_{(i)} \Delta\dot{x}_{(i)}' \mathbf{1}\left[\dot{q}_{i1} \leq \hat{Q}_1^c(s)\right] \\ &= N(r_0)^{-1} \sum_{i=1}^{N} \Delta x_i \Delta x_i' \mathbf{1}\left[q_{i1} \leq \hat{Q}_1^c(s)\right] \mathbf{1}\left[\hat{A}_i(r_0)\right] \\ &= N(r_0)^{-1} \sum_{i=1}^{N} \Delta x_i \Delta x_i' \mathbf{1}[q_{i1} \leq Q_1^c(s)] \mathbf{1}[A_i(r_0)] + o_p(1) \\ &\xrightarrow{p} \frac{N}{N(r_0)} E\left[\Delta x_i \Delta x_i' \mathbf{1}[q_{i1} \leq Q_1^c(s)] \mathbf{1}[A_i(r_0)]\right] \\ &= E\left[\Delta x_i \Delta x_i' \mathbf{1}[q_{i1} \leq Q_1^c(s)] | \mathbf{1}[A_i(r_0)]\right] \\ &= \int_0^s E\left[\Delta x_i \Delta x_i' | \mathbf{1}[A_i(r_0)] = 1, q_{i1} = Q_1^c(s)\right] ds \\ &= \int_0^s \Gamma_1(s)\, ds \end{aligned}$$

where the convergence in probability follows from a similar argument in Lemma 1. The same result can be derived for s such that $\hat{Q}_1^c(s) > \hat{Q}_1(r_0)$. The second part is then established. The last convergence readily follows from the continuous mapping theorem and Slutsky's theorem. ■

Proof of Corollary 3. The proof is covered by the argument of Theorem 3 and hence is omitted. ■

References

ANDREWS, D. W. K. (1991): "An Empirical Process Central Limit Theorem for Dependent Non-identically Distributed Random Variables," *Journal of Multivariate Analysis*, 38, 187–203.

——— (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821–856.

——— (1994): "Asymptotics for Semiparametric Econometric Models via Stochastic Equicontinuity," *Econometrica*, 62(1), 43–72.

ANDREWS, D. W. K., AND X. CHENG (2012): "Estimation and Inference with Weak, Semi-strong, and Strong Identification," *Econometrica*, 80(5), 2153–2211.

——— (2013): "Maximum Likelihood Estimation and Uniform Inference with Sporadic Identification Failure," *Journal of Econometrics*, 173, 36–56.

ANDREWS, D. W. K., AND W. PLOBERGER (1994): "Optimal Tests When a Nuisance Parameter Is Present Only under the Alternative," *Econometrica*, 62, 1383–1414.

BABU, G. J., AND K. SINGH (1978): "On Deviations between Empirical and Quantile Processe for Mixing Random Variables," *Journal of Multivariate Analysis*, 8, 532–549.

BAI, J. (1994): "Least Squares Estimation of a Shift in Linear Processes," *Journal of Time Series Analysis*, 15, 453–470.

——— (1997): "Estimating Multiple Breaks One at a Time," *Econometric Theory*, 13, 315–352.

——— (1999): "Likelihood Ratio Tests for Multiple Structural Changes," *Journal of Econometrics*, 91, 299–323.

BAI, J., R. L. LUMSDAINE, AND J. H. STOCK (1998): "Testing for and Dating Common Breaks in Multivariate Time Series," *Review of Economic Studies*, 65, 395–432.

BAI, J., AND P. PERRON (1998): "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 66, 47–78.

BILLINGSLEY, P. (1968): *Convergence of Probability Measure*. Wiley, New York.

CANER, M., AND B. E. HANSEN (2004): "Instrumental Variable Estimation of a Threshold Model," *Econometric Theory*, 20, 813–843.

Card, D., A. Mas, and J. Rothstein (2008): "Tipping and the Dynamics of Segregation," *Quarterly Journal of Economics*, 123(1), 177–218.

Chan, K. S. (1993): "Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model," *Annals of Statistics*, 21, 520–533.

Chen, B., and Y. Hong (2012): "Testing for Smooth Structural Changes in Time Series Models via Nonparametric Regression," *Econometrica*, 80(3), 1157–1183.

Davidson, J. (1994): *Stochastic Limit Theory*. Oxford University Press, New York.

Dudley, R. M. (2002): *Real Analysis and Probability*. Cambridge University Press, Camebridge, UK.

Dümbgen, L. (1997): "The Asymptotic Behavior of Some Nonparametric Change Point Estimators," *Annals of Statistics*, 19, 1471–1495.

Eberhardt, M., and A. Presbitero (2015): "Public Debt and Growth: Heterogeneity and Non-linearity," *Journal of International Economics*, 97(1), 45–58.

Egert, B. (2015): "Public Debt, Economic Growth and Nonlinear Effects: Myth or Reality?," *Journal of Macroeconomics*, 43, 226–238.

Elliott, G., and U. K. Müller (2007): "Confidence Sets for the Date of a Single Break in Linear Time Series Regressions," *Journal of Econometrics*, 141, 1196–1218.

——— (2014): "Pre and Post Break Parameter Inference," *Journal of Econometrics*, 180, 141–157.

Elliott, G., U. K. Müller, and M. W. Watson (2015): "Nearly Optimal Tests When a Nuisance Parameter is Present Under the Null Hypothesis," *Econometrica*, 83, 771–811.

Eo, Y., and J. Morley (2015): "Likelihood-ratio-based Confidence Sets for the Timing of Structural Breaks," *Quantitative Economics*, 6, 463–497.

Fan, J., and I. Gijbels (1996): *Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability*, vol. 66. CRC Press.

Fazzari, S. M., R. G. Hubbard, B. C. Peterson, A. S. Blinder, and J. M. Poterb (1988): "Financing Constraints and Corporate Inverstment," *Brookings Papers on Economic Activity*, 1, 141–206.

Girma, S. (2005): "Absorptive Capacity and Productivity Spillovers from FDI: a Threshold Regression Analyis," *Oxford Bulletin of Economics and Statistics*, 67(3), 281–306.

Giuliano, P., and M. Ruiz-Arranz (2009): "Remittances, Financial Development, and Growth," *Journal of Development Economics*, 90, 144–152.

Han, A. K. (1987): "Nonparametric Analysis of a Generalized Regression Model: The Maximum Rank Correlation Estimator," *Journal of Econometrics*, 35(2-3), 303–316.

Hansen, B. E. (1997): "Inference in TAR Models," *Studies in Nonlinear Dynamics & Econometrics*, 2(1), 1–14.

——— (1999): "Threshold Effects in Non-dynamic Panel: Estimation, Testing and Inference," *Journal of Econometrics*, 93, 345–368.

——— (2000a): "Sample Splitting and Threshold Estimation," *Econometrica*, 68, 575–603.

——— (2000b): "Testing for Structural Change in Conditional Models," *Journal of Econometrics*, 97, 93–115.

Herndon, T., M. Ash, and R. Pollin (2014): "Does High Debt Consistently Stifle Economic Growth? A Critique of Reinhard and Rogoff," *Cambridge Journal of Economics*, 38(2), 257–279.

Ichimura, H. (1993): "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-index Models," *Journal of Econometrics*, 58(1), 71–120.

Klein, R. W., and R. H. Spady (1993): "An Efficienct Semiparametric Estimator for Binary Response Models," *Econometrica*, 61(2), 387–421.

Lee, L., M. R. Rosenzweig, and M. M. Pitt (1997): "The Effects of Improved Nutrition, Sanitation, and Water Quality on Child Health in High-mortality Populations," *Journal of Econometrics*, 77, 209–235.

Lehmann, E. L., and J. P. Romano (2005): *Testing Statistical Hypothesis*. Springer, New York.

Li, H., and U. K. Müller (2009): "Valid Inference in Partially Unstable General Method of Moment Models," *Review of Economic Studies*, 76, 343–365.

Li, Q., and J. S. Racine (2007): *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

Liptser, R., and N. Shiryaev (2013): *Statistics of Random Processes: I. General Theory*, vol. 5. Springer Science & Business Media.

Newey, W. K., and K. West (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Nyblom, J. (1989): "Testing for the Constancy of Parameters Over Time," *Journal of the American Statistical Association*, 84, 223–230.

Panizza, U., and A. Presbitero (2014): "Public Debt and Economic Growth: Is There a Causal Effect?," *Journal of Macroeconomics*, 41, 21–41.

Picard, D. (1985): "Testing and Estimating Change Points in Time Series," *Journal of Applied Probability*, 17, 841–867.

Qu, Z., and P. Perron (2007): "Estimating and Testing Structural Changes in Multivariante Regressions," *Econometrica*, 75(2), 459–502.

Quandt, R. E. (1958): "The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes," *Journal of the American Statistical Association*, 53, 873–880.

——— (1960): "Tests of the Hypotheses That a Linear Regression System Obeys Two Separate Regrimes," *Journal of the American Statistical Association*, 55, 324.

Reinhart, C. M., and K. Rogoff (2010): "Growth in a Time of Debt," *American Economic Review: Paper and Proceedings*, 100, 573–578.

Reinhart, C. M., and K. S. Rogoff (2011): "From Financial Crash to Debt Crisis," *American Economic Review*, 101(5), 1676–1706.

Robinson, P. M. (1987): "Asymptotically Efficient Estimation in the Presence of Heteroscedasticity of Unknown Form," *Ecomometrica*, 55(4), 875–891.

——— (1988): "Root-N-Consistent Semiparametric Regression," *Ecomometrica*, pp. 931–954.

——— (1989): "Nonparametric Estimation of Time-Varying Parameters," in *Statistical Analysis and Forecasting of Economic Structural Change*, ed. by P. Hackl, pp. 253–264. Springer, Berlin.

——— (1991): "Time-Varying Nonlinear Regression," in *Economic Structural Change. Analysis and Forecasting*, ed. by P. Hackl, and A. H. Westlund, pp. 179–190, Berlin. Springer.

Romano, J. P., and M. Wolf (2001): "Subsampling Intervals in Autoregressive Models with Linear Time Trend," *Ecomometrica*, 69, 1283–1314.

Schelling, T. C. (1971): "Dynamic Models of Segregation," *Journal of Mathematical Sociology*, 1(2), 143–186.

Sen, P. K. (1972): "On the Bahadur Representation of Sample Quantiles for Sequences of ϕ-Mixing Random Variables," *Journal of Multivariate Analysis*, 2, 77–95.

Seo, M. H., and Y. Shin (2014): "Dynamic Panels with Threshold Effect and Endogeneity," *Working Paper*.

Sowell, F. (1996): "Optimal Tests for Parameter Instability in the Generalized Method of Moments Framework," *Econometrica*, 64, 1085–1107.

Tong, H. (1978): *On a Threshold Model*, no. 29. Sijthoff & Noordhoff.

——— (1990): *Non-Linear Time Series: A Dynamic System Approach*. Oxford University Press.

——— (2012): *Threshold Models in Nonlinear Time Series Analysis*, vol. 21. Springer.

Tsay, R. S. (1989): "Testing and Modeling Threshold Autoregressive Processes," *Journal of the American Statistical Association*, 84(405), 231–240.

——— (1998): "Testing and Modeling Multivariante Threshold Models," *Journal of the American Statistical Association*, 93(443), 1188–1202.

van der Vaart, A. W. (1998): *Asymptotic Statistics*. Cambridge University Press, Cambridge, UK.

Wu, W. B. (2005): "On the Bahadur Representation of Sample Quantiles for Dependent Sequences," *Annuals of Statistics*, 33(4), 1934–1963.

Yao, Y. C. (1987): "Approximating the Distribution of the Maximum Likelihood Estimate of the Change-point in a Sequence of Independent Random Variables," *Annals of Statistics*, 3, 1321–1328.