Asymptotics for Threshold Regression Under Misspecification*

Ping Yu†
University of Hong Kong

Started: October 2012
First Version: August 2015
This Version: December 2019

Abstract

In this paper, we develop the asymptotic theory for threshold regression under misspecification, which is especially useful in the regression tree analysis of machine learning. First, we provide a thorough characterization of the asymptotic distribution of the least square estimator, which integrates some fragmented asymptotic results of threshold regression in the literature into one unified framework of misspecification. The asymptotic distribution depends on the fitted threshold regression model being discontinuous or continuous and also on the rate of the limit objective function shrinking to zero in the direction of threshold parameter. The partially identified and unidentified models are also discussed. Second, we provide a LR-based inference method for the threshold point, which can be treated as a misspecification-robust extension of the method in Hansen (2000, *Econometrica*, 68, 575-603).

KEYWORDS: Threshold Regression, Regression Tree, Misspecification, Partial Identification, Unidentification, LR Inference
JEL-CLASSIFICATION: C21, C24.

*I want to thank Howell Tong for providing me some early references on approximating the time series by TAR and Bruce Hansen for helpful comments on the early version of the paper.

†Faculty of Business and Economics, The University of Hong Kong, Pokfulam Road, Hong Kong; email: pingyu@hku.hk.

1 Introduction

Misspecification is a popular problem in econometrics. This problem got much attention since White (1982) who examines the consequences and detection of model misspecification when using maximum likelihood techniques for estimation and inference. Actually, the usual OLS estimator is a misspecified estimator since the conditional mean may not take the linear form of the covariates, see, e.g., White (1980, 1981). In quantile regression, Angrist et al. (2006) study the estimation and inference in a misspecified model. In the GMM framework, Hall and Inoue (2003) study the consequences of misspecification and develop the asymptotics for the pseudo-parameters. In treatment effects evaluation, Yu (2016) studies the pseudo-true values of four estimators in the framework of Heckman and Vytlacil (2005) and also check two responses to model misspecification: the local sensitivity analysis and the partial identification analysis. Following these pioneers, this paper intends to develop the asymptotic theory for the least squares estimator (LSE) in misspecified threshold regression (TR).

In fact, the threshold autoregressive regression (TAR) is initiated in misspecified time series models. Tong and Lim (1980, p. 250) motivate the threshold model for approximating the dynamics of some time series. Later, Petruccelli (1992) provides a rigorous argument that threshold autoregressive models can approximate a general class of time series processes (e.g., exponential autoregressive and invertible bilinear processes) almost surely. Tong (1982) gives some Bayesian underpinnings for the threshold approximation. However, asymptotic results for the least squares estimation in the misspecified model are yet to be developed.

This paper is more motivated by the regression tree analysis in machine learning. In this sense, this paper is close to Bühlmann and Yu (2002) (BY hereafter) and Banerjee and McKeague (2007) (BM hereafter) in spirit. *Decision tree learning* is a basic approximation and predictive approach nowadays. Depending on the predicted outcome is discrete or continuous, it is called *classification tree* and *regression tree* respectively, and combined as so-called *Classification And Regression Tree* (CART). See Breiman et al. (1984) for an early summary about CART, and Hastie et al. (2009) and Efron and Hastie (2016) for recent treatments at the textbook level. In TR language, the regression at each step of regression tree is a *discontinuous threshold regression* (DTR), i.e., the fitted model has a discontinuity at the threshold point. We focus on parametric DTR in this paper although nonparametric DTR is also popular in the literature. For example, threshold regression with endogeneity will reduce to a nonparametric DTR as shown in Yu and Phillips (2018a), and *regression discontinuity designs* (RDDs) with unknown discontinuity points studied in Porter and Yu (2015) are also special cases of nonparametric DTR. Besides DTR, *continuous threshold regression* (CTR) introduced by Chan and Tsay (1998) (CT hereafter) in the context of autoregression is also a standard tool in nonlinear econometric modeling; see also Feder (1975a) for early developments with triangular independent samples. CTR imposes restrictions on DTR that guarantee the fitted model is continuous but has a kink (i.e., the slope of the threshold variable has a discontinuity) at the threshold point; consequently, the decisions based on DTR and CTR are often called *hardthresholding* (or *hard*) and *softthresholding* (or *soft*) *decision*, respectively. Many nonparametric techniques are essentially CTR under different guises. For example, *Multivariate Adaptive Regression Spline* (MARS) proposed by Friedman (1991) is basically an extension of CTR. Similar to nonparametric DTR, nonparametric CTR also has important applications in econometrics, e.g., *regression kink designs* (RKD) popularized by Card et al. (2015) are actually nonparametric CTR with the threshold points known. This paper also studies parametric CTR but takes a different view. We do not impose restrictions as in CT; rather, we run DTR but DTR degenerates to CTR, which is close to Hidalgo, Lee and Seo (2019) (HLS hereafter) in spirit but they assume the model is correctly specified. For future references, we label DTR as Case I and CTR as Case II. It turns out the asymptotic theories of LSE in these two cases are dramatically different.

Suppose the true model is

$$y = m(x,q) + \varepsilon,\ \mathbb{E}[\varepsilon|x,q] = 0,$$

but we mistakenly fit the model as

$$y = \mathbf{x}'\beta_1 1\left(q \le \gamma\right) + \mathbf{x}'\beta_2 1\left(q > \gamma\right) + e = \mathbf{x}'\beta_2 + \mathbf{x}'\delta 1\left(q \le \gamma\right) + e, \tag{1}$$

where $\mathbf{x}' = (1, x', q) \in \mathbb{R}^{d+1}$, $1(\cdot)$ is the indicator function, and the parameter of interest is $\theta = \left(\gamma, \beta'\right)'$ with $\beta = \left(\beta_1', \beta_2'\right)'$ or equivalently, $\theta = \left(\gamma, \beta_2', \delta'\right)'$ with $\delta = \beta_1 - \beta_2$ being the threshold effect. To distinguish $m(x,q)$ in the two regimes, we write

$$m(x,q) = m_1\left(x,q\right) 1(q \le \gamma_0) + m_2\left(x,q\right) 1(q > \gamma_0),$$

where we use the subscript 0 to indicate the pseudo-true value of a parameter of interest. Correspondingly, the original error term $\varepsilon = \varepsilon_1 1(q \le \gamma_0) + \varepsilon_2 1(q > \gamma_0)$ with $\varepsilon_\ell = y - m_\ell\left(x,q\right)$ and the pseudo-true error term $e = e_1 1(q \le \gamma_0) + e_2 1(q > \gamma_0)$ with $e_\ell = y - \mathbf{x}'\beta_{\ell 0}$, $\ell = 1, 2$. Besides the most general form of $m(x,q)$ in the true model, we give two specific examples to show the possibility of misspecification. In the first example,

$$m(x,q) = \sum_{\ell=1}^{L} m_\ell(x,q) 1(\gamma_{\ell-1} < z \le \gamma_\ell), \tag{2}$$

where $m_\ell(x,q)$ is a smooth function, z is a threshold variable which may be different from q, the number of regimes $L \ge 2$, and $-\infty = \gamma_0 < \gamma_1 < \cdots < \gamma_L = \infty$. In other words, the true model is different from the fitted model in at least three aspects: (i) the threshold variable may not be q; (ii) there may be more then two regimes; (iii) the conditional mean in each regime may not be a linear function of covariates. The second example is the varying coefficient model (VCM); see Fan and Zhang (2008) for a review on this model. The VCM is specified as

$$m(x,q) = \mathbf{x}'\beta(q),$$

where $\beta(q)$ is a smooth function of q. In the fitted TR model, $\beta(q)$ takes a parametric discontinuous form: $\beta_{10} 1\left(q \le \gamma_0\right) + \beta_{20} 1\left(q > \gamma_0\right)$. In other words, β_{10} is an average of $\beta(q)$ for $q \le \gamma_0$, and β_{20} is an average of $\beta(q)$ for $q > \gamma_0$.

For a possibly misspecified TR model, we can first conduct specification testing as in Yu et al. (2018) before estimation. However, we will follow the spirit of White (1980, 1981, 1982) – estimate the misspecified model directly but make the asymptotic theory robust to misspecification. The only estimator studied in this paper is the LSE of θ, say $\widehat{\theta}$, which is defined as the minimizer of

$$S_n\left(\theta\right) = \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\left(y - \mathbf{x}'\beta_1 1(q \le \gamma) - \mathbf{x}'\beta_2 1(q > \gamma)\right)^2, \tag{3}$$

where the constant $1/2$ is added to $S_n\left(\theta\right)$ for convenience in expressing the asymptotic distribution of LSE. Besides the distinction of DTR and CTR, the asymptotic theory of LSE also critically depends on the behavior of the probability limit of $S_n\left(\theta\right)$, say $S\left(\theta\right)$, in the neighborhood of γ_0, especially the rate of $S\left(\theta\right)$ shrinking to zero in the direction of γ. We will index this rate by α; in DTR, α is restricted in $[1, 2]$ and in CTR, in $[2, 4]$. Combining with the distinction of DTR and CTR, we will label a case in DTR with $\alpha = 1.5$ as I(1.5) and a case in CTR with $\alpha = 2.7$ as II(2.7), etc. Under this labeling, the correctly specified DTR in Chan (1993) is a special case of I(1), the correctly specified DTR but degenerating to CTR in HLS is a special case of II(3). All other cases can happen only in misspecified TR models. In other words, the two

cases in the correctly specified DTR are rare but become the focal points of research. It may be unexpected that the asymptotic theory for LSE in misspecified TR is much more complicated than that in White (1980, 1981) given that only a threshold point is added in White's linear approximation of the conditional mean.

Besides the correctly specified models, many misspecified TR models studied in the literature also fall in the framework of this paper as special cases. First, the model studied in BY and BM is a special case of I(2) with $\mathbf{x} = 1$. As in this paper, their data are randomly sampled. Later, Seo (2015) extends BY and BM to include nonconstant regressors and cover time series and Koo and Seo (2015) extend to structural break models for forecasting purposes. Both Seo (2015) and Koo and Seo (2015) require $S(\theta)$ to be the second-order differentiable and $\varepsilon_1 = \varepsilon_2$, while we do not impose such restrictions. Second, the structural change model studied in Bai (1997a) (see also Chong (1995)) can be treated as a special case of I(1) with $\mathbf{x} = 1$. He assumes $m(x, q)$ takes the piece-wise constant form. Later, Gonzalo and Pitarakis (2002) (GP hereafter) extend Bai (1997a) to multiple-regime TR. Specifically, their $m(x, q)$ takes the form of (2) with $m_\ell(x, q) = \mathbf{x}'\beta_\ell$ and $L \geq 3$. Both Bai (1997a) and GP show that $\widehat{\gamma}$ must converge to one of the original threshold points, $\gamma_2, \cdots, \gamma_{L-1}$. Bai (1997a) develops the asymptotic theory of LSE in his simple setup, while GP show that $\widehat{\gamma}$ is n-consistent but without any asymptotic distribution; this paper will fill this gap. Third, Perron and Yamamoto (2015) use the LSE to estimate the structural break points when there is endogeneity, so their model can be treated as a special case of (2) with $z = t$ and their estimator is constructed under misspecification; see also Chong (2003) and Bai et al. (2008) for related setups. Actually, their model can be treated as a special case of I(1). Yu (2015b) shows that their consistency proof is flawed but the LSE is indeed consistent to the true structural break point because the threshold variable in the structural change model is the time index t which is independent of the rest components of the model. On the other hand, Yu (2019) shows that their asymptotic distribution of $\widehat{\gamma}$ is correct only if the endogeneity takes the linear form and is incorrect in general; Yu (2019) also provides the correct asymptotic theory for LSE under more general endogeneity forms. This paper extends Yu (2019) to misspecified TR where the threshold variable q need not be independent of the other elements of the model. Note that in misspecified TR, $\widehat{\gamma}$ need not converge to any true threshold point as shown in Yu (2013). To the best of our knowledge, these are the only misspecified TR models studied in the literature. Combining both correctly specified and misspecified TR models in the literature, only some special cases of I(1), I(2) and II(3) are carefully studied until now.

It is well known that when the model is misspecified, the parameters of interest can be partially identified or even unidentified. This also happens in misspecified TR. For example, Bai (1997a) studies the asymptotic theory of $\widehat{\gamma}$ when $L = 3$ and $S(\theta)$ achieves the minimum at the two original break points. For another example, Yu and Phillips (2019) derive the asymptotic distributions of $\widehat{\gamma}$ and $\widehat{\beta}$ in correctly specified TR models with $\delta_0 = 0$, i.e., the model is fully unidentified. Note that in correctly specified TR models, θ is either point identified or fully unidentified, and no intermediate scenario can happen, while in misspecified TR models, partial identification can indeed happen. We are not aware of any developments in the asymptotic theory of $\widehat{\theta}$ when θ is not point identified in general misspecified TR models. This paper will fill this gap.

The asymptotic theory developed in this paper is very useful for prediction purposes as illustrated in BY, Seo (2015) and Koo and Seo (2015), but we will focus our attention on inferences especially on the inference of γ as in Hansen (2000), BM and HLS. In all models regardless of point identified or not, we follow Feder (1975b) and Hansen (2000, 2017) and use the likelihood ratio (LR) statistic to conduct inference on γ; Feder (1975b) and Hansen (2017) consider the LR test in correctly specified CTR and Hansen (2000) considers correctly specified DTR with shrinking threshold effects. Particularly, we solve a long-standing question since Hansen (2000) – how to conduct inference on γ when the TR model is misspecified especially in I(1).

There is also some other literature related to this paper. First, Hansen (2017) develops the asymptotic distribution of the LSE with the CTR restrictions imposed as in CT. His asymptotic distribution is robust to

misspecification, but because the objective function is continuous in parameters, the asymptotic distribution of $\widehat{\theta}$ is jointly normal, which is similar to White (1980, 1981). Second, this paper pays special attentions to how the rates of $S(\theta)$ shrinking to zero in the direction of γ affect the asymptotic distribution of $\widehat{\theta}$, and Yu and Zhao (2013) (YZ hereafter) also study the effects of $S(\theta)$ shrinking to zero in correctly specified DTR but the reasons of shrinking to zero are different. Intuitively, this paper assumes the jump size of $S_n(\theta)$ in the direction of γ shrinks to zero while YZ assume the jump intensity shrinks to zero (technically, they assume the density of q shrinks to zero as γ converge to γ_0). Of course, because YZ consider correctly specified DTR models, it is impossible for the jump size to go to zero. As a result, the asymptotic distributions of $\widehat{\theta}$ are very different. Specifically, for $\alpha \in (1,2]$ in DTR ($\alpha = 1$ is excluded because the jump size is positive when $\alpha = 1$), our asymptotic distribution of $\widehat{\gamma}$ is always related to a Gaussian process while YZ's asymptotic distribution is always related to a compound Poisson process although the Poisson process need not be homogeneous as in the $\alpha = 1$ case; $\widehat{\beta}$ and $\widehat{\gamma}$ in YZ are always asymptotically independent and estimating $\widehat{\gamma}$ will not affect the asymptotic distribution of $\widehat{\beta}$, while $\widehat{\beta}$ and $\widehat{\gamma}$ can be perfectly collinear or partially correlated in this paper.

The rest of this paper is organized as follows. In Section 2, we make the necessary preparations for the developments of our asymptotic theory. In Section 3, we use a simple example with q being the only covariate to illustrate how to derive the convergence rates of $\widehat{\gamma}$ and $\widehat{\beta}$. In Section 4, we discuss the asymptotic theory in I(1) where the threshold effect δ_0 can be either fixed or shrinking to zero. Sections 5 and 6 discuss the asymptotic theories in DTR with other α values (i.e., $\alpha \in (1,2]$), and in CTR, respectively. Section 7 includes the asymptotic theory with identification failure. Section 8 covers some possible extensions of previous sections and some unsolved problems in this paper. Section 9 presents some numerical examples, and Section 9 concludes. All proofs and lemmas are collected in four supplementary appendices.

A word on notation: $\|\cdot\|$ denotes the Euclidean norm. For a matrix A, $A > 0$ means it is positive definite. $U[a,b]$ is the uniform distribution on an interval $[a,b]$ and $N(\mu,\Sigma)$ is the multivariate normal distribution with mean vector μ and variance-covariance matrix Σ. The symbol ℓ is used to indicate the two regimes in (1) and, to simplify notation in what follows, the explicit values "$\ell = 1, 2$" are often omitted. The subscripts "$\leq \gamma$" and "$> \gamma$" signify use of the indicator functions $1(q \leq \gamma)$ and $1(q > \gamma)$, so that $\mathbf{x}_{i,\leq\gamma} = \mathbf{x}_i 1(q_i \leq \gamma)$ and $\mathbf{x}_{i,>\gamma} = \mathbf{x}_i 1(q_i > \gamma)$. For a real number a, $a_\oplus = \max(a,0)$ and $a_\ominus = \min(a,0)$. For two real numbers a and b, $a \wedge b = \min(a,b)$ and $a \vee b = \max(a,b)$. For two sequences of real numbers a_n and b_n, $a_n \prec b_n$ means $a_n = o(b_n)$, $a_n \preceq b_n$ means $a_n = O(b_n)$, $a_n \sim b_n$ means a_n and b_n have the same rate as $n \to \infty$, and if they are random variables, $a_n \prec b_n$ means $a_n = o_p(b_n)$, and $a_n \preceq b_n$ means $a_n = O_p(b_n)$; $a_n \succ b_n$ and $a_n \succeq b_n$ are similarly understood. $\approx$ means higher-order terms are neglected. Subscripts are used to indicate the arguments of differentiation, e.g., $S^+_{\beta\gamma} = \frac{\partial^2 S(\theta_0)}{\partial\beta\partial\gamma_+}$ is the right derivative of $\frac{\partial S(\theta)}{\partial\beta}$ at θ_0, and $S^-_{\gamma^4}$ is the fourth order left derivative of $S(\theta)$ at θ_0. An object without the superscripts or subscripts $\pm$ indicates the common value, e.g., λ is the common value of λ_+ and λ_-, and $S_{\beta\gamma}$ is the common value of $S^+_{\beta\gamma}$ and $S^-_{\beta\gamma}$. Finally, we use CS for an abbreviation of "correctly specified and MS for "misspecified".

2 The Setup

This section includes the setup for the least squares estimator and our asymptotic theory.

2.1 The Least Squares Estimator and Likelihood Ratio Statistic

We re-write the objective function of LSE as

$$S_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} s(w_i|\theta),$$

where

$$s\left(w|\theta\right)=\frac{1}{2}\left(y-\mathbf{x}'\beta_1 1(q\leq\gamma)-\mathbf{x}'\beta_2 1(q>\gamma)\right)^2,$$

and $S\left(\theta\right)=$plim $S_n\left(\theta\right)=\mathbb{E}\left[s\left(w|\theta\right)\right]$. Usually, γ is estimated through the concentrated objection function

$$S_n\left(\gamma\right)=\frac{1}{n}\sum_{i=1}^{n}s\left(w_i|\gamma\right),$$

where

$$s\left(w|\gamma\right)=\frac{1}{2}\left(y-\mathbf{x}'\widehat{\beta}_1\left(\gamma\right)1\left(q\leq\gamma\right)-\mathbf{x}'\widehat{\beta}_2\left(\gamma\right)1\left(q>\gamma\right)\right)^2$$

with

$$\widehat{\beta}\left(\gamma\right):=\left(\begin{array}{c}\widehat{\beta}_1\left(\gamma\right)\\ \widehat{\beta}_2\left(\gamma\right)\end{array}\right)=\arg\min_{\beta_1,\beta_2}S_n\left(\theta\right)=\left(\begin{array}{c}\left(X_{\leq\gamma}'X_{\leq\gamma}\right)^{-1}X_{\leq\gamma}'Y\\ \left(X_{>\gamma}'X_{>\gamma}\right)^{-1}X_{>\gamma}'Y\end{array}\right)$$

and $X_{\leq\gamma}$ and $X_{>\gamma}$ being matrices stacking the vectors $\mathbf{x}_{i,\leq\gamma}'$ and $\mathbf{x}_{i,>\gamma}'$. The probability limit of $S_n\left(\gamma\right)$ is denoted as $S\left(\gamma\right)$. There is an interval of γ, $[\widehat{\gamma}_-,\widehat{\gamma}_+)$, minimizing $S_n\left(\gamma\right)$. Following Yu (2012, 2015a), we therefore take the mid-point of the interval as our estimator of γ because the mid-point $\frac{\widehat{\gamma}_-+\widehat{\gamma}_+}{2}$ is more efficient than the left-endpoint $\widehat{\gamma}_-$ in most cases when the model is CS. This choice of $\widehat{\gamma}$ will affect the asymptotic distribution only in I(1) with fixed threshold effects. Given $\widehat{\gamma}$, $\widehat{\beta}=\widehat{\beta}\left(\widehat{\gamma}\right)=\left(\widehat{\beta}_1',\widehat{\beta}_2'\right)'$.

It can be shown that

$$\begin{aligned}S\left(\theta\right)-S\left(\theta_0\right)=\Phi\left(\beta_1\right)+\overline{\Phi}\left(\beta_2\right)+\left[\Lambda_-\left(\gamma\right)+\Psi_-\left(\beta_1,\gamma\right)+\Psi_-\left(\beta_2,\gamma\right)\right]1(\gamma\leq\gamma_0)\\ +\left[\Lambda_+\left(\gamma\right)+\Psi_-\left(\beta_1,\gamma\right)+\Psi_+\left(\beta_2,\gamma\right)\right]1(\gamma>\gamma_0),\end{aligned}$$

where

$$\begin{aligned}\Phi\left(\beta\right)&=\left(\begin{array}{c}\mathbb{E}\left[\left(m_1(x,q)-\mathbf{x}'\frac{\beta_{10}+\beta_1}{2}\right)\mathbf{x}'\left(\beta_{10}-\beta_1\right)1(q\leq\gamma_0)\right]\\ \mathbb{E}\left[\left(m_2(x,q)-\mathbf{x}'\frac{\beta_{20}+\beta_2}{2}\right)\mathbf{x}'\left(\beta_{20}-\beta_2\right)1(q>\gamma_0)\right]\end{array}\right)=:\left(\begin{array}{c}\Phi\left(\beta_1\right)\\ \overline{\Phi}\left(\beta_2\right)\end{array}\right)\\ \Lambda_-\left(\gamma\right)&=\mathbb{E}\left[\left(m_1(x,q)-\mathbf{x}'\overline{\beta}_0\right)\mathbf{x}'\delta_0 1(\gamma<q\leq\gamma_0)\right]=\mathbb{E}\left[\overline{z}_1 1(\gamma<q\leq\gamma_0)\right],\\ \Lambda_+\left(\gamma\right)&=-\mathbb{E}\left[\left(m_2(x,q)-\mathbf{x}'\overline{\beta}_0\right)\mathbf{x}'\delta_0 1(\gamma_0<q\leq\gamma)\right]=\mathbb{E}\left[\overline{z}_2 1(\gamma_0<q\leq\gamma)\right],\\ \Psi_-\left(\beta,\gamma\right)&=\left(\begin{array}{c}-\mathbb{E}\left[\left(m_1(x,q)-\mathbf{x}'\frac{\beta_{10}+\beta_1}{2}\right)\mathbf{x}'\left(\beta_{10}-\beta_1\right)1(\gamma<q\leq\gamma_0)\right]\\ \mathbb{E}\left[\left(m_1(x,q)-\mathbf{x}'\frac{\beta_{10}+\beta_2}{2}\right)\mathbf{x}'\left(\beta_{10}-\beta_2\right)1(\gamma<q\leq\gamma_0)\right]-\Lambda_-\left(\gamma\right)\end{array}\right)=:\left(\begin{array}{c}\Psi_-\left(\beta_1,\gamma\right)\\ \Psi_-\left(\beta_2,\gamma\right)\end{array}\right),\\ \Psi_+\left(\beta,\gamma\right)&=\left(\begin{array}{c}\mathbb{E}\left[\left(m_2(x,q)-\mathbf{x}'\frac{\beta_{20}+\beta_1}{2}\right)\mathbf{x}'\left(\beta_{20}-\beta_1\right)1(\gamma_0<q\leq\gamma)\right]-\Lambda_+\left(\gamma\right)\\ -\mathbb{E}\left[\left(m_2(x,q)-\mathbf{x}'\frac{\beta_{20}+\beta_2}{2}\right)\mathbf{x}'\left(\beta_{20}-\beta_2\right)1(\gamma_0<q\leq\gamma)\right]\end{array}\right)=:\left(\begin{array}{c}\Psi_+\left(\beta_1,\gamma\right)\\ \Psi_+\left(\beta_2,\gamma\right)\end{array}\right),\end{aligned}$$

and in $\Lambda_\pm\left(\gamma\right)$,

$$\overline{z}_1=\left(m_1(x,q)-\mathbf{x}'\overline{\beta}_0\right)\mathbf{x}'\delta_0+\delta_0'\mathbf{x}\varepsilon_1=\frac{1}{2}\left[2\left(m_1(x,q)-\mathbf{x}'\beta_{10}\right)+\mathbf{x}'\delta_0\right]\mathbf{x}'\delta_0+\delta_0'\mathbf{x}\varepsilon_1=\frac{1}{2}\left(2e_1+\mathbf{x}'\delta_0\right)\mathbf{x}'\delta_0,$$

which equals $\overline{z}:=\left(y-\mathbf{x}'\overline{\beta}_0\right)\left(\delta_0'\mathbf{x}\right)$ with $\overline{\beta}_0=\frac{\beta_{10}+\beta_{20}}{2}$ as $q\leq\gamma_0$ and represents the effect on $S_n\left(\gamma\right)-S_n\left(\gamma_0\right)$ when γ is displaced on the left of γ_0 while β_0 is fixed, and

$$\overline{z}_2=-\left(m_2(x,q)-\mathbf{x}'\overline{\beta}_0\right)\mathbf{x}'\delta_0-\delta_0'\mathbf{x}\varepsilon_2=-\frac{1}{2}\left[2\left(m_2(x,q)-\mathbf{x}'\beta_{20}\right)-\mathbf{x}'\delta_0\right]\mathbf{x}'\delta_0-\delta_0'\mathbf{x}\varepsilon_2=\frac{1}{2}\left(\mathbf{x}'\delta_0-2e_2\right)\mathbf{x}'\delta_0,$$

which equals $-\overline{z}=-\left(y-\mathbf{x}'\overline{\beta}_0\right)\left(\delta_0'\mathbf{x}\right)$ as $q>\gamma_0$ and represents the converse case. Among the various terms

in $S(\theta) - S(\theta_0)$, we use $\Phi(\beta)$ to represent the variation in the direction of β, $\Lambda_-(\gamma)$ ($\Lambda_+(\gamma)$) to represent the variation in the direction of γ and $\Psi_-(\beta,\gamma)$ ($\Psi_+(\beta,\gamma)$) to represent the covariation of β and γ in the left (right) neighborhood of γ_0 in the limit objective function. Note that $\Lambda_\pm(\gamma_0) = 0$ and are positive when $\gamma \in \mathcal{N} \setminus \{\gamma_0\}$ by the point identification of γ_0, where $\mathcal{N}$ is a neighborhood of γ_0.

Following Hansen (2000), we conduct inference on γ always based on the LR-like statistic. In all cases, the LR statistic takes the following form:

$$LR_n(\gamma) = \frac{\tau_n\left(S_n(\gamma) - S_n(\widehat{\gamma})\right)}{\widehat{b}},$$

where τ_n is the normalization rate and $\widehat{b}$ is a consistent estimator of the normalization constant. The test statistic is a by-product of estimation and can be used for hypotheses concerning γ such as $H_0 : \gamma = \gamma_0$.

2.2 Distinction Between DTR and CTR

To develop the asymptotic distribution of $\widehat{\gamma}$, we first distinguish DTR and CTR which are labeled as case I and II in this paper. Recall that in CTR, $\delta_{x0} = \mathbf{0}$ and $\delta_{c0} + \delta_{q0}\gamma_0 = 0$ so that $\mathbf{x}'\delta_0 = (q - \gamma_0)\delta_{q0}$. In Hansen (2000), the assumption $\delta_0'\mathbb{E}[\mathbf{x}\mathbf{x}'|q = \gamma_0]\delta_0 > 0$ is used to exclude the CTR. A natural question is whether there are other cases besides CTR where $\delta_0'\mathbb{E}[\mathbf{x}\mathbf{x}'|q = \gamma_0]\delta_0 = 0$. The following proposition shows that under a regularity condition, no intermediate cases between CTR and DTR can happen.

Proposition 1 *If* $Var(x|q = \gamma_0) > 0$, *then* $\delta_0 \neq \mathbf{0}$ *but* $\delta_0'\mathbb{E}[\mathbf{x}\mathbf{x}'|q = \gamma_0]\delta_0 = 0$ *if and only if* $\delta_{x0} = \mathbf{0}$ *and* $\delta_{c0} + \delta_{q0}\gamma_0 = 0$.

When q is not a regressor, i.e., $\delta_{q0} = 0$, we have either DTR ($\delta_0 \neq \mathbf{0}$) or unidentification ($\delta_0 = \mathbf{0}$). Quite often, we normalize $\gamma_0 = 0$ (which can be achieved by a location shift on q, i.e., replacing $q - \gamma_0$ for q), then the CTR is equivalent to $\delta_{x0} = \mathbf{0}$ and $\delta_{c0} = 0$, and $\mathbf{x}'\delta_0 = q\delta_{q0} = O_p(q)$. In DTR, since $\delta_{x0} \neq \mathbf{0}$ and/or $\delta_{c0} \neq 0$, $\mathbf{x}'\delta_0 = O_p(1)$.

For a simple setup, we check the partition of the parameter space for the three cases - DTR, CTR and unidentification. Suppose $y = (\delta_{c0} + \delta_{q0}q)\,1(q \leq \gamma_0) + \varepsilon$, where the parameter spaces of δ_c, δ_q and γ are all $[-1, 1]$. Then Figure 1 shows the areas of $(\delta_c, \delta_q, \gamma)$ where the three cases happen.

2.3 Rates of $\Lambda_\pm(\gamma)$ Shrinking to Zero

Another critical factor that affects the asymptotic distribution, especially the convergence rate, of $\widehat{\gamma}$ is the rate of $\Lambda_\pm(\gamma)$ shrinking to zero, where $\Lambda_\pm(\gamma)$ is defined in Section 2.1. This rate indicates the information to identify γ_0 with a smaller rate indicating more information. In Chan (1993), $\Lambda_\pm(\gamma)$ is linear in γ, in BY and BM, $\Lambda_\pm(\gamma)$ is quadratic in γ, and in HLS, $\Lambda_\pm(\gamma)$ is cubic in γ. In this paper, we allow the rate to be in the interval $[1, 2]$ in DTR and in the interval $[2, 4]$ in CTR; in other words, the existing literature considers only the rates 1 and 2 in DTR and the rate 3 in CTR so can be treated as special cases of this paper.

To extend $\Lambda_\pm(\gamma)$ to functions like $|\gamma|^\alpha |\log(1/|\gamma|)|$, we introduce the regularly varying functions at zero. A positive locally integrable function $\Lambda : (0,\infty) \to (0,\infty)$ is called *slowly varying* at zero if $\lim_{\gamma \downarrow 0} \frac{\Lambda(c\gamma)}{\Lambda(\gamma)} = 1$, for any $c > 0$, denoted as $\Lambda \in RV_0$ as $\gamma \to 0$. If this limit is finite but nonzero for any $c > 0$, then Λ is called *regularly varying* at zero. Typical examples of slowly varying functions are the constant function and the logarithm; other examples are the powers and the iterations of the logarithm, e.g., $\ln^\alpha$, $\alpha \in \mathbb{R}$ and $\ln\ln$. The function $\Lambda(\gamma) = \gamma$ is not slowly varying, neither is $\Lambda(\gamma) = \gamma^\alpha$ for any real $\alpha \neq 0$. They are regularly varying functions. By Karamata's characterization theorem, any regularly varying function Λ is of the form $\gamma^\alpha L(\gamma)$

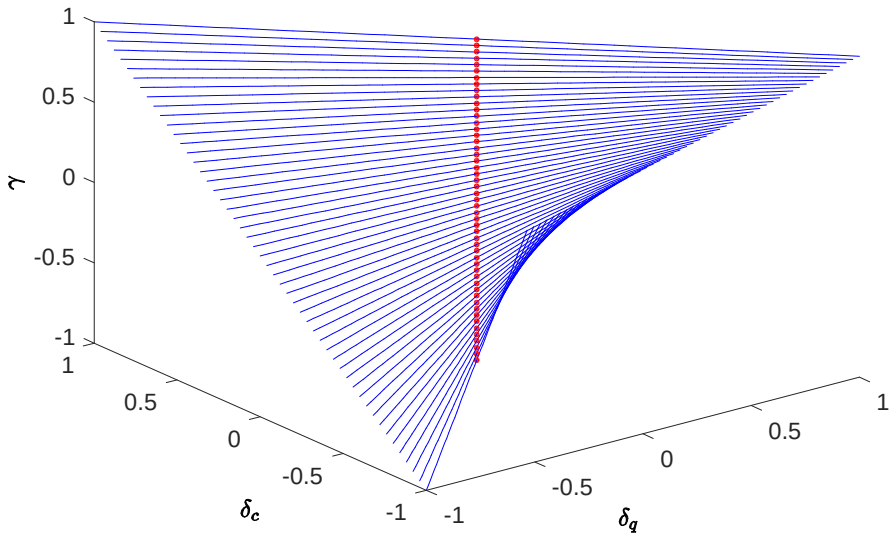

Figure 1: Partition of the Parameter Space of $(\delta_c, \delta_q, \gamma)$ for the Three Cases: Blue Suface For CTR, Red Line for Linear Regression and Other Area for DTR

where $\alpha \in \mathbb{R}$ and $L \in RV_0$. Intuitively, $L(\gamma)$ can be treated as a small disturbance on the main function γ^α. We denote Λ as $\Lambda \in RV_\alpha$ as $\gamma \to 0$ and call α as the *exponent of variation*. See Section 0.4 of Resnick (1987) and Seneta (1976) for more details on this type of functions.

Because the level information in $L(\gamma)$ is important in this paper, we extract this part of information in $L(\gamma)$ out and make $L(\gamma)$ include only the rate information. Specifically, we assume $\Lambda_\pm(\gamma) = \lambda_\pm |\gamma|^{\alpha_\pm} L_\pm(\gamma)$ for γ in a neighborhood of $\gamma_0 = 0$, where $\lambda_\pm > 0$ is the level information, $\alpha_\pm \in [1, 2]$ in DTR and $\alpha_\pm \in [2, 4]$ in CTR. Here, we implicitly extend the domain of regularly varying functions and slowly varying functions to $(-\infty, 0)$ but maintain the range as $(0, \infty)$. We start α_+ from 2 in CTR because

$$\Lambda_+(\gamma) = -\mathbb{E}\left[\left(m_2(x,q) - \mathbf{x}'\overline{\beta}_0\right)\mathbf{x}'\delta_0 1(\gamma_0 < q \leq \gamma)\right] = -\mathbb{E}\left[\left(m_2(x,q) - \mathbf{x}'\overline{\beta}_0\right) q\delta_{q0} 1(\gamma_0 < q \leq \gamma)\right]$$

so that even if $m_2(x,q) - \mathbf{x}'\overline{\beta}_0$ were a constant, $\Lambda_+(\gamma) \in RV_2$; similar arguments apply to α_-. As to why α is bounded above by 2 in DTR and by 4 in CTR, we will explain in the next section. In most parts of the paper, we assume $\alpha_+ = \alpha_-$ and $L_+(\cdot) = L_-(\cdot)$, i.e., $\Lambda_+(\gamma)$ and $\Lambda_-(\gamma)$ are different only in the level information; we will extend $\Lambda_+(\gamma)$ and $\Lambda_-(\gamma)$ to have different rates in Section 8. As a result, we can write $\Lambda_\pm(\gamma) = \lambda_\pm |\gamma|^\alpha L(\gamma)$ for simplicity. Often, we just use $\Lambda(|\gamma|)$ to represent this common rate information of $\Lambda_\pm(\gamma)$. Note that $\Lambda(|\gamma|)$ need not be monotone in $|\gamma|$, but we assume it so for ease of analysis since we need define in our proofs the inverse function of $\Lambda(\cdot)$, $\Lambda^{\leftarrow}(t) = \inf\{s|\Lambda(s) \geq t\}$. We do not assume the $[\alpha]$th-order differentiability of $\Lambda(|\gamma|)$ at γ_0, where $[\alpha]$ is the largest integer not greater than α.[1] When $\Lambda(|\gamma|)$ is indeed the $[\alpha]$th-order differentiable at γ_0, $\lambda_\pm = \frac{S^\pm_{\gamma\gamma}}{2}$when $\alpha = 2$ and $\lambda_\pm = \frac{S^\pm_{\gamma^4}}{4!}$ when $\alpha = 4$. Sometimes, we abuse notations and define $S^\pm_{\gamma\gamma} = 2\lambda_\pm$ even if $\Lambda(|\gamma|)$ is not second-order differentiable.

[1] Although a monotone function is differentiable almost everywhere, it need not be higher-order differentiable.

In the future discussions, some values of α receive special treatments; these values include $\alpha = 1, 1.5$ and 2 in DTR and $\alpha = 2, 2.5, 3, 3.5$ and 4 in CTR. For these values of α, we assume $\lim_{|\gamma|\downarrow 0} L(\gamma) = 1$, i.e., all rate information is included in $|\gamma|^\alpha$ and all level information is included in $\lambda_\pm$. When $\lim_{|\gamma|\downarrow 0} L(\gamma) = 0$ or ∞, the corresponding $\Lambda_\pm(\gamma)$ functions with index α are absorbed in the contiguous α interval. For example, the function $\lambda_\pm |\gamma|^3 |\log(1/|\gamma|)|$ is included in the α interval $(3, 3.5)$, and $\lambda_\pm |\gamma|^3 |\log(1/|\gamma|)|^{-1}$ in the α interval $(2.5, 3)$. As mentioned in the Introduction, we will index each $\Lambda_\pm(\gamma)$ function by its α value.

We illustrate how to obtain the level constants $\lambda_\pm$ at the end of this subsection. In DTR, suppose $\mathbf{x} = (1, q)'$, $\gamma_0 = 0$ and $m_2(q) - \mathbf{x}'\overline{\beta}_0 = Aq^{\alpha-1}$, $1 \le \alpha \le 2$. Then

$$\Lambda_+(\gamma) = -\int_0^\gamma A\nu^{\alpha-1}(1, \nu)\,\delta_0 f(\nu)\,d\nu \approx -\frac{A\delta_{c0} f_0}{\alpha}\gamma^\alpha,$$

so $\lambda_+ = -\frac{Af_0\delta_{c0}}{\alpha}$, whose positiveness implies $A\delta_{c0} < 0$. This also implies $L(\gamma) = \frac{\Lambda_+(\gamma)}{\lambda_+\gamma^\alpha}$ in the right neighborhood of 0 and satisfies $\lim_{\gamma\downarrow 0} L(\gamma) = 1$. $\Lambda_-(\gamma)$ can be similarly discussed. In HLS, the model is the CS CTR, so $m_\ell(x, q) - \mathbf{x}'\beta_{\ell 0} = 0$ such that $m_1(x, q) - \mathbf{x}'\overline{\beta}_0 = \mathbf{x}'\delta_0/2 = \delta_{q0}q/2$ and $m_2(x, q) - \mathbf{x}'\overline{\beta}_0 = -\mathbf{x}'\delta_0/2 = -\delta_{q0}q/2$. As a result,

$$\begin{array}{l} \Lambda_-(\gamma) = \mathbb{E}\left[\left(m_1(x, q) - \mathbf{x}'\overline{\beta}_0\right)\mathbf{x}'\delta_0 1(\gamma < q \le \gamma_0)\right] = \int_\gamma^0 \frac{(\delta_{q0}\nu)^2}{2} f(\nu)\,d\nu \approx \frac{1}{6} f_0 \delta_{q0}^2 |\gamma|^3, \\ \Lambda_+(\gamma) = -\mathbb{E}\left[\left(m_2(x, q) - \mathbf{x}'\overline{\beta}_0\right)\mathbf{x}'\delta_0 1(\gamma_0 < q \le \gamma)\right] = \int_0^\gamma \frac{(\delta_{q0}\nu)^2}{2} f(\nu)\,d\nu \approx \frac{1}{6} f_0 \delta_{q0}^2 \gamma^3, \end{array} \tag{4}$$

so $\lambda_+ = \lambda_+ = \frac{1}{6} f_0 \delta_{q0}^2 =: \lambda$, which is the constant appearing in HLS's Theorem 1 (after dividing by 2 since our $S_n(\theta)$ is their $S_n(\theta)/2$). This also implies $L(\gamma) = \frac{\Lambda_+(\gamma)}{\lambda\gamma^3}$ in the right neighborhood of 0 and $L(\gamma) = \frac{\Lambda_-(\gamma)}{\lambda\gamma^3}$ in the left neighborhood of 0, and $\lim_{|\gamma|\downarrow 0} L(\gamma) = 1$.

2.4 Maximizer and Maximum of A Class of Stochastic Processes

In developing the asymptotic theory for $\widehat{\gamma}$, we often need the distributions of the maximizer and maximum of the following stochastic process:

$$\begin{cases} -\frac{1}{2}\mu_- |v|^\tau + \sqrt{\varpi_-} B_1(-v), & \text{if } v \le 0, \\ -\frac{1}{2}\mu_+ v^\tau + \sqrt{\varpi_+} B_2(v), & \text{if } v > 0, \end{cases}$$

where $B_1(v)$ and $B_2(v)$ are two independent standard Wiener Processes on $[0, \infty)$. Given $B_1(v)$ and $B_2(v)$, we can define a Wiener Processes on $\mathbb{R}$ as $B(v) = B_1(-v)1(v \le 0) + B_2(v)1(v > 0)$.

In the following proposition, we simplify these targets to some basic objects.

Proposition 2 *For $\mu_\pm > 0$, $\varpi_\pm > 0$ and $\tau > 1/2$, we have the following results: (i)*

$$\arg\max_v \begin{cases} -\frac{1}{2}\mu_- |v|^\tau + \sqrt{\varpi_-} B_1(-v), & \text{if } v \le 0, \\ -\frac{1}{2}\mu_+ v^\tau + \sqrt{\varpi_+} B_2(v), & \text{if } v > 0. \end{cases} = \omega^{\frac{1}{2\tau-1}} \zeta(\varphi, \phi; \tau),$$

where $\omega = \frac{\varpi_-}{\mu_-^2}$, and $\zeta(\varphi, \phi; \tau) := \arg\max_r \begin{cases} -\frac{1}{2}|r|^\tau + B_1(-r), & \text{if } r \le 0, \\ -\frac{1}{2}\varphi r^\tau + \sqrt{\phi} B_2(r), & \text{if } r > 0, \end{cases}$ with $\varphi = \frac{\mu_+}{\mu_-}$ and $\phi = \frac{\varpi_0^+}{\varpi_0^-}$.

(ii)

$$\max_v \begin{cases} -\frac{1}{2}\mu_- |v|^\tau + \sqrt{\varpi_-} B_1(-v), & \text{if } v \le 0, \\ -\frac{1}{2}\mu_+ v^\tau + \sqrt{\varpi_+} B_2(v), & \text{if } v > 0. \end{cases} = \eta^{\frac{2}{2\tau-1}} \xi(\varphi, \phi; \tau),$$

where $\eta^2 = \frac{\varpi^\tau}{\mu_-}$, *and* $\xi(\varphi,\phi;\tau) := \max_r \begin{cases} -\frac{1}{2}|r|^\tau + B_1(-r), & \text{if } r \le 0, \\ -\frac{1}{2}\varphi r^\tau + \sqrt{\phi}B_2(r), & \text{if } r > 0, \end{cases}$ *with* φ *and* ϕ *defined above. (iii)* $\xi(\varphi,\phi;1)$ *has the distribution* $P(\xi(\varphi,\phi;1) \le x) = (1-e^{-x})(1-e^{-x\varphi/\phi})$.

We restrict $\tau > 1/2$ because by the law of the iterated logarithms for Brownian motion, $B(v) \le \sqrt{2v\log\log v}$ as $|v| \to \infty$, and $\tau > 1/2$ guarantees the $|v|^\tau$ term dominate the $B(v)$ term and the maximizer be $O_p(1)$. The distribution of $\zeta(\varphi,\phi;1)$ is developed in Appendix B of Bai (1997b), but for other τ values, it is unknown whether $\zeta(\varphi,\phi;\tau)$ has a closed-form distribution. A special case of $\xi(\varphi,\phi;2)$ attracts much attention in the literature. Define

$$\zeta_c = \arg\max_r \left\{-cr^2 + B(r)\right\},$$

and then $\zeta(1,1;2) = \zeta_{1/2}$. Groeneboom (1989) derives the distribution of ζ_1 and shows that it is related to the so-called *Airy function* so has no explicit form; Groeneboom and Wellner (2001) then show how to compute it. Since Brownian scaling implies that $\zeta_c = c^{-2/3}\zeta_1$, the case $c = 1$ is considered without loss of generality. The distribution of ζ_1 is referred to as *Chernoff's distribution* in the literature. Note that the distribution of $\zeta(\varphi,\phi;\tau)$ is not required in this paper because the inference on γ is based on the LR statistic so $\xi(\varphi,\phi;\tau)$ rather than $\zeta(\varphi,\phi;\tau)$ is relevant. Except $\xi(\varphi,\phi;1)$, we guess $\xi(\varphi,\phi;\tau)$ with other τ values does not have a closed-form distribution. Parallel to ζ_c, we can define

$$\xi_c = \max_r \left\{-cr^2 + B(r)\right\},$$

and then $\xi(1,1;2) = \xi_{1/2}$; Brownian scaling implies $\xi_c = c^{-1/3}\xi_1$.

2.5 Maintained Assumptions

We collect some maintained assumptions here for future references. These maintained assumptions will not be repeated and only adjustments will be stated in the discussions of each case. In this way, we can focus our attention on the assumptions that are different and critical in each case. First, define

$$\begin{aligned}
M &= \mathbb{E}[\mathbf{x}\mathbf{x}'], N = \mathbb{E}[\mathbf{x}y], M_\gamma = \mathbb{E}[\mathbf{x}\mathbf{x}'1(q \le \gamma)], \overline{M}_\gamma = M - M_\gamma = \mathbb{E}[\mathbf{x}\mathbf{x}'1(q > \gamma)], \\
M_0 &= M_{\gamma_0} = S_{\beta_1\beta_1}, \overline{M}_0 = \overline{M}_{\gamma_0} = S_{\beta_2\beta_2}, S_{\beta\beta} = \operatorname{diag}\left\{M_0, \overline{M}_0\right\}, \\
N_\gamma &= \mathbb{E}[\mathbf{x}y1(q \le \gamma)], \overline{N}_\gamma = N - N_\gamma = \mathbb{E}[\mathbf{x}y1(q > \gamma)], N_0 = N_{\gamma_0}, \overline{N}_0 = \overline{N}_{\gamma_0}, \\
\beta_{1\gamma} &= M_\gamma^{-1}N_\gamma, \beta_{2\gamma} = \overline{M}_\gamma^{-1}\overline{N}_\gamma, \beta_{10} = \beta_{1\gamma_0} = M_0^{-1}N_0, \beta_{20} = \beta_{2\gamma_0} = \overline{M}_0^{-1}\overline{N}_0, \\
\beta_\gamma &= \left(\beta_{1\gamma}', \beta_{2\gamma}'\right)', \delta_\gamma = \beta_{1\gamma} - \beta_{2\gamma}, \beta_0 = \beta_{\gamma_0}, \delta_0 = \delta_{\gamma_0}, \\
\Sigma_\gamma &= \mathbb{E}\left[\mathbf{x}\mathbf{x}'\left(e_\gamma^-\right)^2 1(q \le \gamma)\right], \overline{\Sigma}_\gamma = \mathbb{E}\left[\mathbf{x}\mathbf{x}'\left(e_\gamma^+\right)^2 1(q > \gamma)\right], \Sigma_0 = \Sigma_{\gamma_0}, \overline{\Sigma}_0 = \overline{\Sigma}_{\gamma_0},
\end{aligned}$$

where $e_\gamma^- = y - \mathbf{x}'\beta_{1\gamma}$ and $e_\gamma^+ = y - \mathbf{x}'\beta_{2\gamma}$, $e_1 = e_{\gamma_0}^- = y - \mathbf{x}'\beta_{10} = \varepsilon_1 + m_1(x,q) - \mathbf{x}'\beta_{10}$ and $e_2 = e_{\gamma_0}^+ = y - \mathbf{x}'\beta_{20} = \varepsilon_2 + m_2(x,q) - \mathbf{x}'\beta_{20}$. Let $f(q)$ denote the density of q, $f_0 = f(\gamma_0)$ and $\mathcal{N}$ be a neighborhood of γ_0 when γ_0 is point identified.

Assumption MA:

(i) The data $\{w_i\}_{i=1}^n$ are randomly sampled, $w_i = (y_i, x_i', q_i)' \in \mathbb{W} \equiv \mathbb{R}\times\mathbb{X}\times\mathbb{Q} \subset \mathbb{R}^{d+1}$, $\beta_\ell \in B_\ell \subset \mathbb{R}^{d+1}$, and $\gamma \in \Gamma = [\underline{\gamma}, \overline{\gamma}] \subsetneqq \mathbb{Q}$ is compact.

(ii) When $q \le \gamma_0$, $m_1(x,q) := \mathbb{E}[y|x,q]$ is left continuous at $q = \gamma_0$ for all $x \in \mathbb{X}$; when $q > \gamma_0$, $m_2(x,q) := \mathbb{E}[y|x,q]$ is right continuous at $q = \gamma_0$ for all $x \in \mathbb{X}$.

(iii) The conditional distribution $f_{(x,\varepsilon_1)|q}(x,\varepsilon_1|q)$ is left continuous at $q=\gamma_0$ and $f_{(x,\varepsilon_2)|q}(x,\varepsilon_2|q)$ is right continuous at $q=\gamma_0$.

(iv) (a) $\mathbb{E}[\varepsilon_\ell^4]<\infty$, $\mathbb{E}[\|\mathbf{x}\|^4]<\infty$ and $\mathbb{E}[y^4]<\infty$; (b) $\sup_{\gamma\in\mathcal{N}}\mathbb{E}[\varepsilon_\ell^4|q=\gamma]<\infty$, $\sup_{\gamma\in\mathcal{N}}\mathbb{E}[\|\mathbf{x}\|^4|q=\gamma]<\infty$ and $\sup_{\gamma\in\mathcal{N}}\mathbb{E}[y^4|q=\gamma]<\infty$.

(v) $M>M_\gamma>0$ for $\gamma\in\mathcal{N}$.

(vi) $\Sigma_\gamma>0$ and $\overline{\Sigma}_\gamma>0$ for $\gamma\in\mathcal{N}$.

(vii) $f(\gamma)$ is continuous at γ_0, and $0<\underline{f}\leq f(\gamma)\leq\overline{f}<\infty$ for $\gamma\in\Gamma$.

(viii) $\arg\min_{\gamma\in\Gamma}S(\gamma)=\gamma_0$ is unique.

(ix) θ_0 satisfies $\delta_{\alpha 0}+\delta_{q0}\gamma_0\neq 0$ and/or $\delta_{x0}\neq\mathbf{0}$.

As mentioned in the Introduction, we consider only random samples which are explicitly stated in Assumption (i). As usual, we restrict the parameter space of γ to be a perfect subset of the support of q. Assumption (ii) imposes some regularity conditions on $m_\ell(x,q)$; under Assumption (ii), we can write $m_1(x,\gamma_0)$ for $m_1(x,\gamma_0-)$ and $m_2(x,\gamma_0)$ for $m_2(x,\gamma_0+)$.[2] Similarly, Assumption (iii) imposes some regularity conditions on $f_{(x,\varepsilon\ell)|q}$; this assumption guarantees that $\mathbb{E}[g(x,\varepsilon_1)|q=\gamma]$ is left continuous and $\mathbb{E}[g(x,\varepsilon_2)|q=\gamma]$ is right continuous at γ_0 for any function of g as long as the conditional means are well defined, so we can write $\mathbb{E}[g(x,\varepsilon_1)|q=\gamma_0-]$ as $\mathbb{E}[g(x,\varepsilon_1)|q=\gamma_0]$ and $\mathbb{E}[g(x,\varepsilon_2)|q=\gamma_0+]$ as $\mathbb{E}[g(x,\varepsilon_2)|q=\gamma_0]$.[3] These two assumptions make sure that we can focus on the (dis-)continuity property of $m(x,q)$ and $q=\gamma_0$ because all other components of the model are continuous at γ_0. Assumption (iv) implies $\mathbb{E}\left[m_\ell(x,q)^4\right]<\infty$, $\mathbb{E}\left[\left(m_\ell(x,q)-\mathbf{x}'\beta_{\ell 0}\right)^4\right]<\infty$ and $\mathbb{E}\left[e_\ell^4\right]<\infty$; Assumption (iv)(b) implies $\sup_{\gamma\in\mathcal{N}}\mathbb{E}\left[m_\ell(x,q)^4|q=\gamma\right]<\infty$, $\sup_{\gamma\in\mathcal{N}}\mathbb{E}\left[\left(m_\ell(x,q)-\mathbf{x}'\beta_{\ell 0}\right)^4|q=\gamma\right]<\infty$ and $\sup_{\gamma\in\mathcal{N}}\mathbb{E}\left[e_\ell^4|q=\gamma\right]<\infty$, where $\mathcal{N}$ is understood as the left neighborhood of γ_0 when $m_1(x,q),\varepsilon_1$ and e_1 are involved and the right neighborhood when $m_2(x,q),\varepsilon_2$ and e_2 are involved. Note that $\mathbb{E}[\varepsilon_\ell^4]<\infty$ implies $\mathbb{E}[\varepsilon_\ell^4|q]<\infty$ for q almost everywhere if $0<\underline{f}\leq f(q)\leq\overline{f}<\infty$ for all $q\in\mathbb{Q}$, so combined with Assumption (iii) and Assumption (vii) below, $\mathbb{E}[\varepsilon_\ell^4]<\infty$ indeed implies that there is a $\mathcal{N}$ such that $\sup_{\gamma\in\mathcal{N}}\mathbb{E}[\varepsilon_\ell^4|q=\gamma]<\infty$ (otherwise, $\mathbb{E}[\varepsilon_\ell^4]$ cannot be finite); similarly for $\mathbb{E}[\|\mathbf{x}\|^4]<\infty$ and $\mathbb{E}[y^4]<\infty$. We explicitly state these implications here because they will be used in some of our proofs; sometimes, we need to strengthen Assumption (iv)(b) to apply a Donsker's theorem. In Assumption (v), we only require $M_\gamma>0$ and $\overline{M}_\gamma>0$ for $\gamma\in\mathcal{N}$ while Hansen (2000) requires $M_\gamma>0$ and $\overline{M}_\gamma>0$ for $\gamma\in\Gamma$. This is because Hansen (2000) needs his assumption to prove the consistency of $\widehat{\gamma}$ while we assume the consistency of $\widehat{\gamma}$ in the following Assumption (viii). Note that $M>M_\gamma>0$ for $\gamma\in\Gamma$ implies Γ must be a proper subset of the support of q; since we only restrict $\gamma\in\mathcal{N}$ here, we explicitly specify Γ in Assumption (i). As usual, Σ_γ and $\overline{\Sigma}_\gamma$ in Assumption (vi) will be used in some asymptotic distributions of $\widehat{\beta}$. By the continuity of M_γ, $\overline{M}_\gamma$, Σ_γ and $\overline{\Sigma}_\gamma$ at γ_0, we can actually state Assumption (v) as $M_0>0$ and $\overline{M}_0>0$ and Assumption (vi) as $\Sigma_0>0$ and $\overline{\Sigma}_0>0$.

The following three assumptions may change with the cases, but we still state them here and explain how to adjust them in some cases to save space. In II(α), $3<\alpha\leq 4$, we need to strengthen the continuity of $f(\gamma)$ at γ_0 in Assumption (vii) to the differentiability of $f(\gamma)$ at γ_0. Because $S(\gamma)$ is continuous in γ, Assumption (viii) implies the consistency of $\widehat{\gamma}$ as mentioned above. Given the consistency of $\widehat{\gamma}$, $\widehat{\beta}$ is consistent to β_0. This is also why we did not restrict B_ℓ to be compact in Assumption (i). In Section 7, γ_0 is not point identified (e.g., when $\delta_0=\mathbf{0}$, $\beta_{10}=\beta_{20}=M^{-1}N$, and γ_0 is not identified), so we need to adjust Assumption (viii) and also Assumption (vi) correspondingly.[4] Assumption (ix) restricts the model to be DTR, and in CTR, we

[2] Rigorously, $m_2(x,\gamma_0)$ is not defined, so we define it as $\lim_{\gamma\downarrow\gamma_0}m_2(x,\gamma)=:m_2(x,\gamma_0+)$ to guarantee $m_2(x,q)$ to be right continuous at $q=\gamma_0$. On the other hand, $m_1(x,\gamma_0)=\lim_{\gamma\uparrow\gamma_0}m_1(x,\gamma)=:m_1(x,\gamma_0-)$ since $m_1(x,\gamma_0)$ is well defined. This convention applies to assumption (iii) in defining $f_{(x,\varepsilon_2)|q}(x,\varepsilon_2|q=\gamma)$ at $\gamma=\gamma_0$.

[3] Because $f_{x|q}(x|q)$ is continuous at $q=\gamma_0$, $E[g(x)|q=\gamma]$ is continuous at γ_0. This is why we can define $D_0=E[\mathbf{x}\mathbf{x}'|q=\gamma_0]$ in the future.

[4] Note that γ_0 and $\mathcal{N}$ are meaningful only in point identified models, so all assumptions involving γ_0 and $\mathcal{N}$ need to be

need to adjust $\delta_{\alpha 0}+\delta_{q0}\gamma_0 \neq 0$ and/or $\delta_{x0} \neq \mathbf{0}$ according to Proposition 1. Specially, we replace Assumption (ix) by

(ix)$'$ $\delta_0 \neq \mathbf{0}$ but $\delta_{\alpha 0}+\delta_{q0}\gamma_0 = 0$ and $\delta_{x0} = \mathbf{0}$.

in CTR, and we still call the collected assumptions as Assumption MA. Note that in CTR, $(\delta_{c0}, \delta_{q0}) \neq \mathbf{0}$ implies $\delta_{q0} \neq 0$ because if $\delta_{q0} = 0$ then δ_{c0} must be zero given that $\delta_{c0}+\delta_{q0}\gamma_0 = 0$; on the other hand, when $\gamma_0 = 0$, δ_{c0} can be zero since δ_{q0} need not be zero.

The most important assumption is Assumption (x) which will be stated in each case instead of here because this assumption needs to be changed for each case (i.e., this assumption marks each case). Assumption MA is definitely not the weakest assumption required, but we find the current form of Assumption MA is convenient and intuitive.

3 An Example for Illustration

We use a simple example to illustrate the main results of this paper, especially, the various convergence rates in different cases. With this example in mind, the proofs for the general cases are more accessible. Readers who are only interested in general results can skip this section. Suppose q is the only covariate, and $y = m_1(q)1(q \leq \gamma_0) + m_2(q)1(q > \gamma_0) + \varepsilon$. WLOG, suppose $\gamma_0 = 0$; then for γ is in the neighborhood of 0, $(1,q)\delta_0 = q\delta_{q0} \sim \gamma$ in CTR, and $(1,q)\delta_0 = \delta_{c0} + q\delta_{q0} \sim 1$ in DTR given that $\delta_{c0} \neq 0$. Now, we check the local behavior of $S_n(\theta) - S_n(\theta_0)$.

3.1 Deterministic and Random Parts of $S_n(\theta) - S_n(\theta_0)$

First, check the deterministic part of $S_n(\theta) - S_n(\theta_0)$, which is $S(\theta) - S(\theta_0)$. From the decomposition in Section 2.1 and noticing that $\mathbf{x} = (1,q)'$, we have

$$
\begin{aligned}
S_{\beta\beta} &= \operatorname{diag}\left\{\tfrac{d^2\Phi(\beta_{10})}{d\beta_1 d\beta_1'}, \tfrac{d^2\overline{\Phi}(\beta_{20})}{d\beta_2 d\beta_2'}\right\} = \operatorname{diag}\left\{\mathbb{E}\left[(1,q)'(1,q)\,1(q\leq\gamma_0)\right], \mathbb{E}\left[(1,q)'(1,q)\,1(q>\gamma_0)\right]\right\} > 0,\\
S^-_{\beta\gamma} &= \begin{pmatrix} \frac{\partial^2\Psi_-(\beta_{10},\gamma_0)}{\partial\beta_1\partial\gamma_-} \\ \frac{\partial^2\Psi_-(\beta_{20},\gamma_0)}{\partial\beta_2\partial\gamma_-} \end{pmatrix} = f_0 \begin{pmatrix} -\left[m_1(\gamma_0) - (1,\gamma_0)\,\beta_{10}\right] \\ m_1(\gamma_0) - (1,\gamma_0)\,\beta_{20} \end{pmatrix} \otimes \begin{pmatrix} 1 \\ \gamma_0 \end{pmatrix},\\
&= f_0 \begin{pmatrix} -\left[m_1(\gamma_0) - (1,\gamma_0)\,\overline{\beta}_0 - \frac{1}{2}(1,\gamma_0)\,\delta_0\right] \\ m_1(\gamma_0) - (1,\gamma_0)\,\overline{\beta}_0 + \frac{1}{2}(1,\gamma_0)\,\delta_0 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ \gamma_0 \end{pmatrix}\\
S^+_{\beta\gamma} &= \begin{pmatrix} \frac{\partial^2\Psi_+(\beta_{10},\gamma_0)}{\partial\beta_1\partial\gamma_+} \\ \frac{\partial^2\Psi_+(\beta_{20},\gamma_0)}{\partial\beta_2\partial\gamma_+} \end{pmatrix} = f_0 \begin{pmatrix} -\left[m_2(\gamma_0) - (1,\gamma_0)\,\beta_{10}\right] \\ m_2(\gamma_0) - (1,\gamma_0)\,\beta_{20} \end{pmatrix} \otimes \begin{pmatrix} 1 \\ \gamma_0 \end{pmatrix},\\
&= f_0 \begin{pmatrix} -\left[m_2(\gamma_0) - (1,\gamma_0)\,\overline{\beta}_0 - \frac{1}{2}(1,\gamma_0)\,\delta_0\right] \\ m_2(\gamma_0) - (1,\gamma_0)\,\overline{\beta}_0 + \frac{1}{2}(1,\gamma_0)\,\delta_0 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ \gamma_0 \end{pmatrix}\\
S^-_{\gamma} &= \tfrac{\partial\Lambda_-(\gamma_0)}{\partial\gamma_-} = -f_0(1,\gamma_0)\,\delta_0\left[m_1(\gamma_0) - (1,\gamma_0)\,\overline{\beta}_0\right],\\
S^+_{\gamma} &= \tfrac{\partial\Lambda_+(\gamma_0)}{\partial\gamma_+} = -f_0(1,\gamma_0)\,\delta_0\left[m_2(\gamma_0) - (1,\gamma_0)\,\overline{\beta}_0\right],
\end{aligned}
$$

regardless of in DTR or CTR, where $\otimes$ is the Kronecker product.

It is obvious that whether $S^{\pm}_{\gamma} = 0$ depends on whether the fitted model is DTR or CTR and the values of $m_1(\gamma_0)$ and $m_2(\gamma_0)$. Only in DTR $S^{\pm}_{\gamma} \neq 0$ can happen. Because $\left(\beta_0', \gamma_0\right)'$ is the minimizer of $S(\theta) - S(\theta_0)$, we have $S^-_{\gamma} < 0$ and $S^+_{\gamma} > 0$ in this case which is our case I(1). If $(1,\gamma_0)\,\delta_0 > 0$, this implies $m_1(\gamma_0) > (1,\gamma_0)\,\overline{\beta}_0$ and $m_2(\gamma_0) < (1,\gamma_0)\,\overline{\beta}_0$, i.e., $m(q)$ is discontinuous at γ_0, and $m_1(\gamma_0)$ and $(1,\gamma_0)\,\beta_{10}$

deleted or adjusted when the model loses point identification.

are both above $(1,\gamma_0)\overline{\beta}_0$ and $m_2(\gamma_0)$ and $(1,\gamma_0)\beta_{20}$ are both below $(1,\gamma_0)\overline{\beta}_0$.[5] In this case, $S(\theta)-S(\theta_0)$ is not differentiable in γ at γ_0.[6] In all other cases, $S_\gamma^{\pm}=0$, and $S(\theta)-S(\theta_0)$ is indeed differentiable in γ at γ_0. In DTR, since $(1,\gamma_0)\delta_0\neq 0$, this means $m_1(\gamma_0)=m_2(\gamma_0)=(1,\gamma_0)\overline{\beta}_0$, which is satisfied in our case I(α), $1<\alpha\leq 2$.[7] In CTR, $(1,\gamma_0)\delta_0=0$, so it must be the case that $S_\gamma^{\pm}=0$. In case II(2), $m_1(\gamma_0)\neq(1,\gamma_0)\overline{\beta}_0\neq m_2(\gamma_0)$, and in case II($\alpha$), $2<\alpha\leq 4$, $m_1(\gamma_0)=m_2(\gamma_0)=(1,\gamma_0)\overline{\beta}_0$. The value of $\alpha-1$ in DTR and $\alpha-2$ in CTR indicate the speed of $m_1(\gamma)-(1,\gamma)\overline{\beta}_0$ and $m_2(\gamma)-(1,\gamma)\overline{\beta}_0$ shrinking to 0 as γ converges to 0. To simplify our discussion, assume here both $m_1(\gamma)-(1,\gamma)\overline{\beta}_0$ and $m_2(\gamma)-(1,\gamma)\overline{\beta}_0$ are $O\left(|\gamma|^{\alpha-1}\right)$ in DTR and $O\left(|\gamma|^{\alpha-2}\right)$ in CTR, and then $\Lambda_{\pm}(\gamma)=O(|\gamma|^{\alpha})$.

Next, we study the covariation of β and γ. In case I(1),

$$S_{\beta\gamma}^{-}\neq\mathbf{0}\neq S_{\beta\gamma}^{+}$$

in general, in case I(α), $1<\alpha\leq 2$,

$$S_{\beta\gamma}^{-}=\frac{1}{2}f_0\begin{pmatrix}(1,\gamma_0)\delta_0\\(1,\gamma_0)\delta_0\end{pmatrix}\otimes\begin{pmatrix}1\\\gamma_0\end{pmatrix}=S_{\beta\gamma}^{+}\neq\mathbf{0},$$

in case II(2),

$$\begin{aligned}S_{\beta\gamma}^{-}&=f_0\begin{pmatrix}-\left[m_1(\gamma_0)-(1,\gamma_0)\overline{\beta}_0\right]\\m_1(\gamma_0)-(1,\gamma_0)\overline{\beta}_0\end{pmatrix}\otimes\begin{pmatrix}1\\\gamma_0\end{pmatrix}\neq\mathbf{0},\\S_{\beta\gamma}^{+}&=f_0\begin{pmatrix}-\left[m_2(\gamma_0)-(1,\gamma_0)\overline{\beta}_0\right]\\m_2(\gamma_0)-(1,\gamma_0)\overline{\beta}_0\end{pmatrix}\otimes\begin{pmatrix}1\\\gamma_0\end{pmatrix}\neq\mathbf{0},\end{aligned}$$

and in case II(α), $2<\alpha\leq 4$,

$$S_{\beta\gamma_-}=S_{\beta\gamma_+}=\mathbf{0}.$$

Given $S_{\beta\gamma_-}=S_{\beta\gamma_+}=\mathbf{0}$, we can further study $S_{\beta\beta\gamma}^{\pm}$ and the behavior of $S_{\beta}^{\pm}(\gamma)$. Notice that

$$S_{\beta\beta\gamma}^{-}=f_0\begin{pmatrix}(1,\gamma_0)'(1,\gamma_0)&\mathbf{0}\\\mathbf{0}&-(1,\gamma_0)'(1,\gamma_0)\end{pmatrix}=S_{\beta\beta\gamma}^{+},$$

and

$$\begin{aligned}S_{\beta}^{-}:=\tfrac{\partial\Psi_-(\beta_0,\gamma)}{\partial\beta}&=\begin{pmatrix}\int_\gamma^0\begin{pmatrix}1\\\nu\end{pmatrix}\left[m_1(\nu)-(1,\nu)\overline{\beta}_0-\frac{\nu\delta_{q0}}{2}\right]f(\nu)\,d\nu\\-\int_\gamma^0\begin{pmatrix}1\\\nu\end{pmatrix}\left[m_1(\nu)-(1,\nu)\overline{\beta}_0\right]f(\nu)\,d\nu-\frac{1}{2}\int_\gamma^0\begin{pmatrix}1\\\nu\end{pmatrix}\nu\delta_{q0}f(\nu)\,d\nu\end{pmatrix}\sim\begin{pmatrix}\gamma^{\alpha-1}+\gamma^2\\\gamma^{\alpha-1}+\gamma^2\\\gamma^{\alpha-1}+\gamma^2\\\gamma^{\alpha-1}+\gamma^2\end{pmatrix},\\S_{\beta}^{+}:=\tfrac{\partial\Psi_+(\beta_0,\gamma)}{\partial\beta}&=\begin{pmatrix}-\int_0^\gamma\begin{pmatrix}1\\\nu\end{pmatrix}\left[m_2(\nu)-(1,\nu)\overline{\beta}_0\right]f(\nu)\,d\nu+\frac{1}{2}\int_0^\gamma\begin{pmatrix}1\\\nu\end{pmatrix}\nu\delta_{q0}f(\nu)\,d\nu\\\int_0^\gamma\begin{pmatrix}1\\\nu\end{pmatrix}\left[m_2(\nu)-(1,\nu)\overline{\beta}_0+\frac{\nu\delta_{q0}}{2}\right]f(\nu)\,d\nu\end{pmatrix}\sim\begin{pmatrix}\gamma^{\alpha-1}+\gamma^2\\\gamma^{\alpha-1}+\gamma^2\\\gamma^{\alpha-1}+\gamma^2\\\gamma^{\alpha-1}+\gamma^2\end{pmatrix}.\end{aligned}$$

[5] Notice that $(1,\gamma_0)\overline{\beta}_0=(1,\gamma_0)\beta_{10}-(1,\gamma_0)\delta_0/2$ and $(1,\gamma_0)\overline{\beta}_0=(1,\gamma_0)\beta_{20}+(1,\gamma_0)\delta_0/2$, so the distance between $m_1(\gamma_0)$ and $(1,\gamma_0)\beta_{10}$ is bounded below by $-(1,\gamma_0)\delta_0/2$ and the distance between $m_2(\gamma_0)$ and $(1,\gamma_0)\beta_{20}$ is bounded above by $(1,\gamma_0)\delta_0/2$ to guarantee $m_1(\gamma_0)$ and $m_2(\gamma_0)$ staying on different sides of $(1,\gamma_0)\overline{\beta}_0$.

[6] In general, we require $P(m_1(x,\gamma_0)\neq m_2(x,\gamma_0))>0$ to guarantee the nondifferentiability of $S(\theta)-S(\theta_0)$ at γ_0 given that $f_{x|q}(x|\gamma)$ is continuous at γ_0.

[7] In general, $P(m_1(x,\gamma_0)=m_2(x,\gamma_0))=1$ is not necessary but sufficient to guarantee the differentiability of $S(\theta)-S(\theta_0)$ at γ_0 in DTR given that $f_{x|q}(x|\gamma)$ is continuous at γ_0

Now, we must distinguish $2 < \alpha < 3$, and $3 \leq \alpha \leq 4$. In the former case, $S^{\pm}_{\beta\gamma^2}$ does not exist while in the later case, it indeed exists and

$$S^{-}_{\beta\gamma^2} = \begin{pmatrix} \frac{f_0\delta_{q0}}{2} \\ \frac{f_0\delta_{q0}}{2} \end{pmatrix} \otimes \begin{pmatrix} 1 \\ \gamma_0 \end{pmatrix} = S^{+}_{\beta\gamma^2} \tag{5}$$

when $3 < \alpha \leq 4$ and

$$\begin{aligned} S^{-}_{\beta\gamma^2} &= \begin{pmatrix} -f_0\left(m_1'(\gamma_0) - \overline{\beta}_{q0}\right) \\ f_0\left(m_1'(\gamma_0) - \overline{\beta}_{q0}\right) \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} \frac{f_0\delta_{q0}}{2} \\ \frac{f_0\delta_{q0}}{2} \end{pmatrix} \otimes \begin{pmatrix} 1 \\ \gamma_0 \end{pmatrix}, \\ S^{+}_{\beta\gamma^2} &= \begin{pmatrix} -f_0\left(m_2'(\gamma_0) - \overline{\beta}_{q0}\right) \\ f_0\left(m_2'(\gamma_0) - \overline{\beta}_{q0}\right) \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} \frac{f_0\delta_{q0}}{2} \\ \frac{f_0\delta_{q0}}{2} \end{pmatrix} \otimes \begin{pmatrix} 1 \\ \gamma_0 \end{pmatrix}, \end{aligned}$$

when $\alpha = 3$, where we assume $f(\gamma)$ is differentiable at γ_0 and $\overline{\beta}_{q0}$ is the second component of $\overline{\beta}_0$.

Now, we can explain why in case I, we do not allow $\alpha > 2$, and in case II, we do not allow $\alpha > 4$. This is mainly because we want to avoid local unidentification. In case I, because $S_{\beta\beta} > 0$, $S^{\pm}_{\beta\gamma} \neq \mathbf{0}$ and $\Lambda_{\pm}(\gamma) \sim |\gamma|^{\alpha}$, we have

$$S(\theta) - S(\theta_0) \sim \left\|\widetilde{\beta}\right\|^2 + \left\|\widetilde{\beta}\right\| |\gamma| + |\gamma|^{\alpha} \tag{6}$$

by Taylor expansion, where $\widetilde{\beta} = \left(\widetilde{\beta}_1', \widetilde{\beta}_2'\right)'$ with $\widetilde{\beta}_{\ell} = \beta_{\ell} - \beta_{\ell 0}$. Since $\left\|\widetilde{\beta}\right\|^2 + |\gamma|^{\alpha} \succeq \left\|\widetilde{\beta}\right\| |\gamma|^{\alpha/2}$, only if $\alpha \leq 2$ we can control the variation of $S(\theta) - S(\theta_0)$ by $\left\|\widetilde{\beta}\right\|^2 + |\gamma|^{\alpha}$ (when $\alpha < 2$, $\left\|\widetilde{\beta}\right\| |\gamma|$ is dominated since $|\gamma|^{\alpha/2} \succ |\gamma|$). Otherwise, the variation in the direction (β, γ) may dominate the variation in the direction β or γ, and γ_0 and/or β_0 may not be identified locally.[8] In case II, we have

$$S(\theta) - S(\theta_0) \sim \left\|\widetilde{\beta}\right\|^2 + \left\|\widetilde{\beta}\right\| |\gamma| + \left\|\widetilde{\beta}\right\|^2 |\gamma| + \left\|\widetilde{\beta}\right\| \left(|\gamma|^2 + |\gamma|^{\alpha-1}\right) + |\gamma|^{\alpha}, \tag{7}$$

where the $\left\|\widetilde{\beta}\right\| |\gamma|$ term appears only if $\alpha = 2$ (and then all other cross terms are dominated by $\left\|\widetilde{\beta}\right\| |\gamma|$ and the analysis is the same as in DTR with $\alpha = 2$), the term $\left\|\widetilde{\beta}\right\|^2 |\gamma|$ is dominated by $\left\|\widetilde{\beta}\right\|^2$, and $\left\|\widetilde{\beta}\right\| |\gamma|^{\alpha-1}$ is dominated by $\left\|\widetilde{\beta}\right\|^2 + |\gamma|^{\alpha}$ since $\left\|\widetilde{\beta}\right\|^2 + |\gamma|^{\alpha} \succeq \left\|\widetilde{\beta}\right\| |\gamma|^{\alpha/2}$ and $\alpha - 1 > \alpha/2$ if $\alpha > 2$. The key term here is $\left\|\widetilde{\beta}\right\| |\gamma|^2$ which is no larger than $O\left(\left\|\widetilde{\beta}\right\|^2 + |\gamma|^4\right)$. By a similar argument as in case I, α cannot be greater than 4. Under these restrictions on the value of α, we have

$$S(\theta) - S(\theta_0) \sim \left\|\widetilde{\beta}\right\|^2 + |\gamma|^{\alpha}$$

in both DTR and CTR.

Second, we check the random part of $S_n(\theta) - S_n(\theta_0)$, which is equal to

$$\begin{aligned} &\tfrac{1}{n}\textstyle\sum_{i=1}^n (\beta_{10} - \beta_1)' \mathbf{x}_i e_{1i} 1(q_i \leq \gamma_0) + \tfrac{1}{n}\sum_{i=1}^n (\beta_{20} - \beta_2)' \mathbf{x}_i e_{2i} 1(q_i > \gamma_0) \\ &+\tfrac{1}{n}\textstyle\sum_{i=1}^n \delta_0' \mathbf{x}_i \varepsilon_{1i} 1(\gamma < q_i \leq \gamma_0) - \tfrac{1}{n}\sum_{i=1}^n \delta_0' \mathbf{x}_i \varepsilon_{2i} 1(\gamma_0 < q_i \leq \gamma) \end{aligned} \tag{8}$$

[8] More rigorously, by Taylor expansion, the main terms of $S(\theta) - S(\theta_0)$ when θ is local to θ_0 are $\frac{1}{4}\widetilde{\beta}'\widetilde{\beta} + \left(S^{-}_{\beta\gamma}\widetilde{\beta}\gamma + \lambda_- |\gamma|^{\alpha}\right) 1(\gamma \leq 0) + \left(S^{+}_{\beta\gamma}\widetilde{\beta}\gamma + \lambda_+\gamma^{\alpha}\right) 1(\gamma > 0)$ whose minimum given $\gamma \neq \gamma_0$ is $-\left[\left(\gamma^2 S^{-}_{\gamma\beta}S^{-}_{\beta\gamma} - \lambda_- |\gamma|^{\alpha}\right) 1(\gamma \leq 0) + \left(\gamma^2 S^{+}_{\gamma\beta}S^{+}_{\beta\gamma} - \lambda_+ |\gamma|^{\alpha}\right) 1(\gamma > 0)\right] < 0$ when $\alpha > 2$ unless $S^{\pm}_{\beta\gamma} = \mathbf{0}$ which is impossible in this simple example; see Section 9 for some concrete calculation. In other words, $\gamma_0 \neq \arg\min_{\gamma} S(\gamma)$ at the beginning. Actually, even when $\alpha = 2$, we need some restrictions on $\lambda_{\pm}$ to guarantee the $\left\|\widetilde{\beta}\right\| |\gamma|$ term not exceeding the other two terms; see Example 3 in Section 5.1.

with $\mathbf{x}_i = (1, q_i)'$. The first two terms contribute to the randomness of $\widehat{\beta}$ and the last two terms contribute to the randomness of $\widehat{\gamma}$. The variance of (8) depends on whether the model is DTR or CTR. In DTR, the variance is $O\left(\frac{\|\beta-\beta_0\|^2}{n} + \frac{|\gamma-\gamma_0|}{n}\right)$, and in CTR, the variance is $O\left(\frac{\|\beta-\beta_0\|^2}{n} + \frac{|\gamma-\gamma_0|^3}{n}\right)$ because $\delta_0'\mathbf{x} \sim q\delta_{q0}$. So in DTR, the random part is $O_p\left(\frac{\|\beta-\beta_0\|+|\gamma-\gamma_0|^{1/2}}{\sqrt{n}}\right)$, and in CTR, the random part is $O_p\left(\frac{\|\beta-\beta_0\|+|\gamma-\gamma_0|^{3/2}}{\sqrt{n}}\right)$. Note further from Yu (2012, 2015) that the randomnesses of $\widehat{\beta}$ and of $\widehat{\gamma}$ are independent asymptotically because the former involves the global information while the latter involves the local information and these two parts of information are independent.

3.2 Determining the Convergence Rates of $\widehat{\beta}$ and $\widehat{\gamma}$

Now, we balance the deterministic part and the random part to determine the convergence rate of $\widehat{\beta}$ and $\widehat{\gamma}$. First of all, we must make $\|\beta - \beta_0\|$ and $|\gamma - \gamma_0|$ have the same scale in $S(\theta) - S(\theta_0)$ to determine the convergence rate, i.e., $\|\beta - \beta_0\| \sim |\gamma - \gamma_0|^{\alpha/2}$. Suppose $\widehat{\beta} - \beta_0 = O_p\left(\kappa_n^{-1}\right)$ and $\widehat{\gamma} - \gamma_0 = O_p\left(\rho_n^{-1}\right)$.

In DTR, the random part is $O_p\left(\frac{\|\beta-\beta_0\|+\left(|\gamma-\gamma_0|^{\alpha/2}\right)^{1/\alpha}}{\sqrt{n}}\right)$. When $\alpha = 1$, the randomness from $\widehat{\beta}$ and $\widehat{\gamma}$ are balanced, while when $1 < \alpha \leq 2$, the randomness from $\widehat{\beta}$ is dominated by that from $\widehat{\gamma}$. So when $\alpha = 1$, solving

$$\rho_n^{-1} = \frac{\rho_n^{-1/2}}{\sqrt{n}} \text{ and } \kappa_n^{-2} = \frac{\kappa_n^{-1}}{\sqrt{n}}$$

to have $\rho_n = n$ and $\kappa_n = \sqrt{n}$. Also, we need to multiply the localized $S_n(\theta) - S_n(\theta_0)$ by $\kappa_n^2 = n$ to have a nondegenerate weak limit. We label this rate as the normalization rate in this subsection. When $1 < \alpha \leq 2$, solving

$$\rho_n^{-\alpha} = \frac{\rho_n^{-1/2}}{\sqrt{n}} \text{ and } \kappa_n^{-2} = \frac{\kappa_n^{-1/\alpha}}{\sqrt{n}}$$

to have $\kappa_n = n^{\frac{\alpha}{2(2\alpha-1)}}$ and $\rho_n = n^{\frac{1}{2\alpha-1}}$. The normalization rate is $\sqrt{n\rho_n}$ (or ρ_n^α or κ_n^2). Especially, when $\alpha = 2$, $\kappa_n = \rho_n = n^{1/3}$ as in BM and the normalization rate is $n^{2/3}$.

In CTR, the random part is $O_p\left(\frac{\|\beta-\beta_0\|+\left(|\gamma-\gamma_0|^{\alpha/2}\right)^{3/\alpha}}{\sqrt{n}}\right)$. When $\alpha = 3$, the randomness from $\widehat{\beta}$ and $\widehat{\gamma}$ are balanced, when $2 < \alpha < 3$, the randomness from $\widehat{\beta}$ dominates that from $\widehat{\gamma}$, while when $3 < \alpha \leq 4$, the randomness from $\widehat{\beta}$ is dominated by that from $\widehat{\gamma}$. So when $\alpha = 3$, solving

$$\rho_n^{-3} = \frac{\rho_n^{-3/2}}{\sqrt{n}} \text{ and } \kappa_n^{-2} = \frac{\kappa_n^{-1}}{\sqrt{n}}$$

to have $\rho_n = n^{1/3}$ and $\kappa_n = \sqrt{n}$ as in HLS. The normalization rate is $\kappa_n^2 = n$. When $2 \leq \alpha < 3$, solving

$$\rho_n^{-\alpha} = \frac{\rho_n^{-\alpha/2}}{\sqrt{n}} \text{ and } \kappa_n^{-2} = \frac{\kappa_n^{-1}}{\sqrt{n}}$$

to have $\rho_n = n^{1/\alpha}$ and $\kappa_n = \sqrt{n}$. Especially, when $\alpha = 2$, $\rho_n = \kappa_n = \sqrt{n}$. The normalization rate is still $\kappa_n^2 = n$. When $3 < \alpha \leq 4$, solving

$$\rho_n^{-\alpha} = \frac{\rho_n^{-3/2}}{\sqrt{n}} \text{ and } \kappa_n^{-2} = \frac{\kappa_n^{-3/\alpha}}{\sqrt{n}}$$

to have $\rho_n = n^{\frac{1}{2\alpha-3}}$ and $\kappa_n = n^{\frac{\alpha}{2(2\alpha-3)}}$. Now, the normalization rate is $\sqrt{n\rho_n^3}$ (or ρ_n^α or κ_n^2). Especially, when $\alpha = 4$, $\kappa_n = n^{2/5}$, $\rho_n = n^{1/5}$ and the normalization rate is $n^{4/5}$. Although both DTR and CTR consider

$\alpha = 2$, the convergence rates of $\widehat{\theta}$ are different because α only indexes the rates of the deterministic part shrinking to zero while the rates of the random part shrinking to zero are different in these two cases.

3.3 Extension and Refinement

When $\alpha = 1$ and 2 in DTR and $\alpha = 2$, 3 and 4 in CTR, the rates in the last subsection are enough. When $1 < \alpha < 2$ in DTR and $2 < \alpha < 3$ and $3 < \alpha < 4$ in CTR, $\Lambda_{\pm}(\gamma)$ need not take the power form of γ, so we need to extend the arguments in the last subsection to this general specification of $\Lambda_{\pm}(\gamma)$. Now, $\|\beta - \beta_0\| \sim \sqrt{\Lambda(\gamma)}$.

When $1 < \alpha < 2$ in DTR, solving

$$\Lambda\left(\frac{1}{\rho_n}\right) \sim \frac{\rho_n^{-1/2}}{\sqrt{n}} \text{ and } \kappa_n^{-2} = \frac{\rho_n^{-1/2}}{\sqrt{n}}$$

to have $\rho_n = n^{\frac{1}{2\alpha-1}} L^*(n)$ and $\kappa_n = n^{\frac{\alpha}{2(2\alpha-1)}} L^*(n)^{1/4}$, where $L^*(n) = L\left(\rho_n^{-1}\right)^{\frac{2}{2\alpha-1}}$. For example, if $L(x) = \log\left(|x|^{-1}\right)$, then $L^*(n) = (\log n)^{\frac{2}{2\alpha-1}}$, $\rho_n = n^{\frac{1}{2\alpha-1}} (\log n)^{\frac{2}{2\alpha-1}}$ and $\kappa_n = n^{\frac{\alpha}{2(2\alpha-1)}} (\log n)^{\frac{1}{2(2\alpha-1)}}$. Also, $\sqrt{n\rho_n}\Lambda_{\pm}\left(\frac{v}{\rho_n}\right) \to \lambda_{\pm}|v|^{\alpha}$ using the definition of slowly varying function at zero. When $2 < \alpha < 3$ in CTR, solving

$$\Lambda\left(\frac{1}{\rho_n}\right) \sim \frac{\sqrt{\Lambda\left(\frac{1}{\rho_n}\right)}}{\sqrt{n}} \text{ and } \kappa_n^{-2} = \frac{\kappa_n^{-1}}{\sqrt{n}}$$

to have $\rho_n = n^{\frac{1}{\alpha}} L^*(n)$ and $\kappa_n = \sqrt{n}$, where $L^*(n) = L(1/\rho_n)^{\frac{1}{\alpha}}$. For example, if $L(x) = \log\left(|x|^{-1}\right)$, then $L^*(n) = (\log n)^{\frac{1}{\alpha}}$ and $\rho_n = (n \log n)^{\frac{1}{\alpha}}$. Also, $n\Lambda_{\pm}\left(\frac{v}{\rho_n}\right) \to \lambda_{\pm}|v|^{\alpha}$. This balancing is the same as in YZ; see their Example 2.1. When $3 < \alpha < 4$ in CTR, solving

$$\Lambda\left(\frac{1}{\rho_n}\right) \sim \frac{\rho_n^{-3/2}}{\sqrt{n}} \text{ and } \kappa_n^{-2} = \frac{\rho_n^{-3/2}}{\sqrt{n}}$$

to have $\rho_n = n^{\frac{1}{2\alpha-3}} L^*(n)$ and $\kappa_n = n^{\frac{\alpha}{2(2\alpha-3)}} L^*(n)^{3/4}$, where $L^*(n) = L\left(\rho_n^{-1}\right)^{\frac{2}{2\alpha-3}}$. For example, if $L(x) = \log\left(|x|^{-1}\right)$, then $\rho_n = n^{\frac{1}{2\alpha-3}} (\log n)^{\frac{2}{2\alpha-3}}$ and $\kappa_n = n^{\frac{\alpha}{2(2\alpha-3)}} (\log n)^{\frac{3}{2(2\alpha-3)}}$. Also, $\sqrt{n\rho_n^3}\Lambda_{\pm}\left(\frac{v}{\rho_n}\right) \to \lambda_{\pm}|v|^{\alpha}$. In the future, ρ_n and κ_n are referred to the rates here and will not specified explicitly.

The convergence rate for $\widehat{\beta}$ when $1 < \alpha < 2$ in DTR and $3 < \alpha < 4$ in CTR and the convergence rate for $\widehat{\gamma}$ when $2 < \alpha < 3$ in CTR derived above are correct but not useful since the asymptotic distributions under these rates will degenerate. We will explain why this can happen below. First check DTR. From (6), only if $\alpha = 2$, the cross term $\widetilde{\beta}' S^{\pm}_{\beta\gamma} \gamma$ will not be dominated. When $\alpha = 1$, because the randomnesses from $\widehat{\beta}$ and $\widehat{\gamma}$ in (8) are balanced and asymptotically independent, and the cross term disappears asymptotically, we expect $\widehat{\beta}$ and $\widehat{\gamma}$ are asymptotically independent. For example, this is indeed the case in Chan (1993) and Hansen (2000) where the model is the CS DTR. When $\alpha = 2$, because the randomness from $\widehat{\beta}$ is dominated but the cross term remains, we expect the asymptotic distribution of $\widehat{\beta}$ is completely determined by $\widehat{\gamma}$ and will not degenerate since it inherits randomness from $\widehat{\gamma}$. When $1 < \alpha < 2$, because the randomness from $\widehat{\beta}$ is still dominated but the cross term disappears, we expect the asymptotic distribution of $\widehat{\beta}$ will degenerate since it cannot inherit randomness from $\widehat{\gamma}$ anymore. In other words, the convergence rate of $\widehat{\beta}$ should be faster. How to obtain this convergence rate? Because the randomness in the γ direction dominates, we cannot search over β and γ jointly; rather, we fix $\widehat{\gamma}$ and concentrate on the randomness in the β direction. Putting in another way, we express $\widehat{\beta}$ as $\widehat{\beta}(\widehat{\gamma})$ and note that $\widehat{\beta}(\widehat{\gamma}) - \beta_0 = \left(\widehat{\beta}(\widehat{\gamma}) - \widehat{\beta}(\gamma_0)\right) + \left(\widehat{\beta}(\gamma_0) - \beta_0\right)$.

It is clear now the convergence rate of $\widehat{\beta}$ is determined by the smaller of the convergence rates $\widehat{\beta}(\widehat{\gamma}) - \widehat{\beta}(\gamma_0)$ and $\widehat{\beta}(\gamma_0) - \beta_0$. Due to the cross term $\widetilde{\beta}' S^{\pm}_{\beta\gamma}\gamma$, it turns out that the effect of estimating γ_0 on $\widehat{\beta}$ is linear in $(\widehat{\gamma} - \gamma_0)$, so the convergence rate of $\widehat{\beta}(\widehat{\gamma}) - \widehat{\beta}(\gamma_0)$ is ρ_n. It is well known that the convergence rate of $\widehat{\beta}(\gamma_0) - \beta_0$ is $\sqrt{n}$, so the convergence rate of $\widehat{\beta}$ is $\min(\rho_n, \sqrt{n})$ with $\rho_n = \sqrt{n}$ when $\alpha = 1.5$.

Second, check CTR. When $\alpha = 2$, because the randomness from $\widehat{\gamma}$ is dominated but the cross term $\left\|\widetilde{\beta}\right\| |\gamma|$ remains, we expect the asymptotic distribution of $\widehat{\gamma}$ is completely determined by $\widehat{\beta}$ and will not degenerate since it can inherit randomness from $\widehat{\beta}$. When $\alpha = 3$, because the random components from $\widehat{\beta}$ and $\widehat{\gamma}$ are balanced and all cross terms in (7) disappear, we expect $\widehat{\beta}$ and $\widehat{\gamma}$ are asymptotically independent. This case is like case I(1). For example, this is indeed the case in HLS where the model is the CS CTR. When $\alpha = 4$, because the randomness from $\widehat{\beta}$ is dominated but the cross term $\left\|\widetilde{\beta}\right\| |\gamma|^2$ remains, we expect the asymptotic distribution of $\widehat{\beta}$ is completely determined by $\widehat{\gamma}$ and will not degenerate since it inherits randomness from $\widehat{\gamma}$. This case is similar to case I(2). When $3 < \alpha < 4$, because the randomness from $\widehat{\beta}$ is dominated and all cross terms disappear, we expect the asymptotic distribution of $\widehat{\beta}$ will degenerate since it cannot inherit randomness from $\widehat{\gamma}$ anymore. This case is similar to case I(α) with $1 < \alpha < 2$, and so can be similarly analyzed. Now, the cross term $\left\|\widetilde{\beta}\right\| |\gamma|$ disappears and the dominating cross term is $\left\|\widetilde{\beta}\right\| |\gamma|^2$, so the effect of estimating γ_0 on $\widehat{\beta}$ is quadratic in $(\widehat{\gamma} - \gamma_0)$ and the convergence rate of $\widehat{\beta}(\widehat{\gamma}) - \widehat{\beta}(\gamma_0)$ is ρ_n^2. All in all, the convergence rate of $\widehat{\beta}$ is $\min\left(\rho_n^2, \sqrt{n}\right)$ with $\rho_n^2 = \sqrt{n}$ when $\alpha = 3.5$. The hardest case is $2 < \alpha < 3$ since there is no explicit-form solution for $\widehat{\gamma}$. Because the randomness from $\widehat{\gamma}$ is dominated and all cross terms disappear, we expect the asymptotic distribution of $\widehat{\gamma}$ will degenerate. Similarly as in case I(α) with $1 < \alpha < 2$, we express $\widehat{\gamma}$ as $\widehat{\gamma}\left(\widehat{\beta}\right)$ and note that $\widehat{\gamma}\left(\widehat{\beta}\right) - \gamma_0 = \left(\widehat{\gamma}\left(\widehat{\beta}\right) - \widehat{\gamma}(\beta_0)\right) + (\widehat{\gamma}(\beta_0) - \gamma_0)$. It is not hard to show that the convergence rate of $\widehat{\gamma}(\beta_0) - \gamma_0$ is ϱ_n, where ϱ_n takes the same formula as in II(α) with $3 \leq \alpha \leq 4$. However, how to characterize the effect of estimating β_0 on $\widehat{\gamma}$ is not an easy task. It turns out that due to the cross term $\left\|\widetilde{\beta}\right\| |\gamma|^{\alpha-1}$ in (7), this effect can be thought of being linear in $\widehat{\beta} - \beta_0$, i.e., the convergence rate of $\widehat{\gamma}\left(\widehat{\beta}\right) - \widehat{\gamma}(\beta_0)$ is $\sqrt{n}$. In summary, the convergence rate of $\widehat{\gamma}$ is $\min(\varrho_n, \sqrt{n})$ with $\varrho_n = \sqrt{n}$ when $\alpha = 2.5$.

We summarize all the discussions on the convergence rates in Figure 2, where we consider only the cases with $\Lambda(|\gamma|)$ taking the form of $|\gamma|$'s power for simplicity. From Figure 2, we have two conclusions. First, the convergence rates of both $\widehat{\gamma}$ and $\widehat{\beta}$ are decreasing in α. This is because a larger α means less identification information for γ so that the convergence rate of $\widehat{\gamma}$ is slower, and a slower convergence rate for $\widehat{\gamma}$ will contaminate the convergence rate of $\widehat{\beta}$. Second, in DTR, the convergence rate of $\widehat{\gamma}$ cannot be slower than that of $\widehat{\beta}$, while in CTR, the converse statement is true. Also, from the discussions above, the asymptotic distributions of $\widehat{\gamma}$ and $\widehat{\beta}$ are asymptotically independent in I(α) with $1 \leq \alpha < 1.5$ and in II(α) with $2.5 < \alpha < 3.5$ and perfectly correlated in I(α) with $1.5 < \alpha \leq 2$ and in II(α) with $2 \leq \alpha < 2.5$ and $3.5 < \alpha \leq 4$; only in some marginal cases I(1.5), II(2.5) and II(3.5), they are partially correlated as in the regular model. This reflects the essential difference in the nature of γ and β. We also expect that when $\rho_n \prec n$, averaging in data is involved and the asymptotic distribution of $\widehat{\gamma}$ will be related to some Gaussian processes rather than some Poisson processes as in I(1) (note that I(1) is the only case where $\rho_n = n$).

For comparison, we also summarize the normalization rates of $LR_n(\gamma)$ (which will be developed in the coming sections) in Figure 3 when $\Lambda(|\gamma|)$ takes the form of $|\gamma|$'s power. Although when the identification power for γ is stronger (i.e., α is smaller) the normalization rate is generally higher, in II(α) with $\alpha \leq 2 < 2.5$, $\widehat{\beta}$ takes in charge and the normalization rate is actually lower for smaller α. Juxtaposing Figures 2 and 3, it is obvious that there is a close relationship between the convergence rate of $\widehat{\gamma}$ and the normalization rate of $LR_n(\gamma)$.

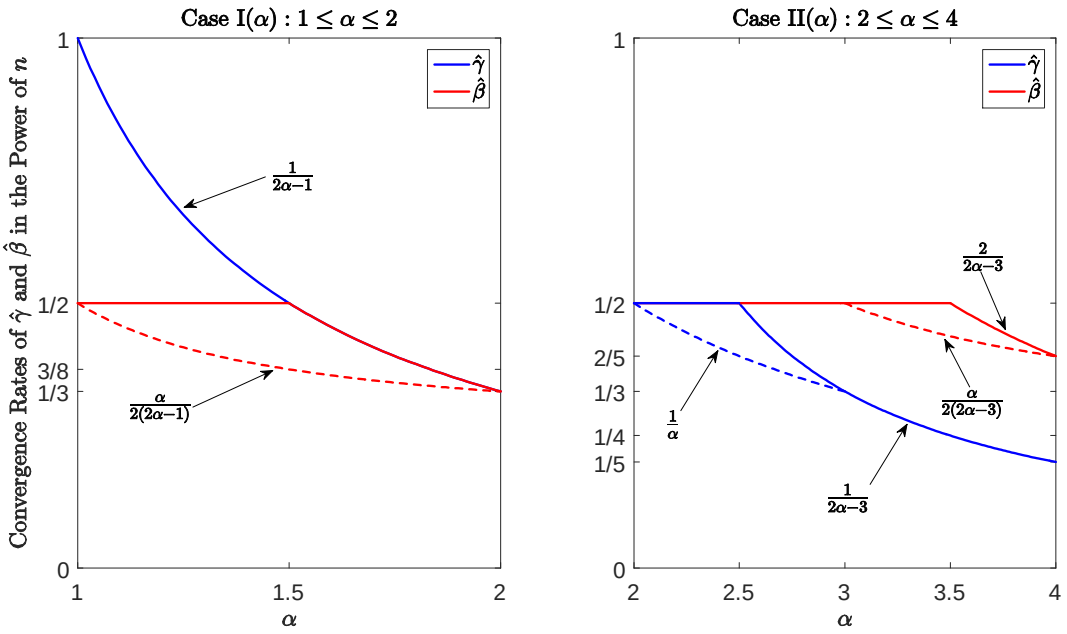

Figure 2: Convergence Rates of $\widehat{\gamma}$ and $\widehat{\beta}$ in the Power of n

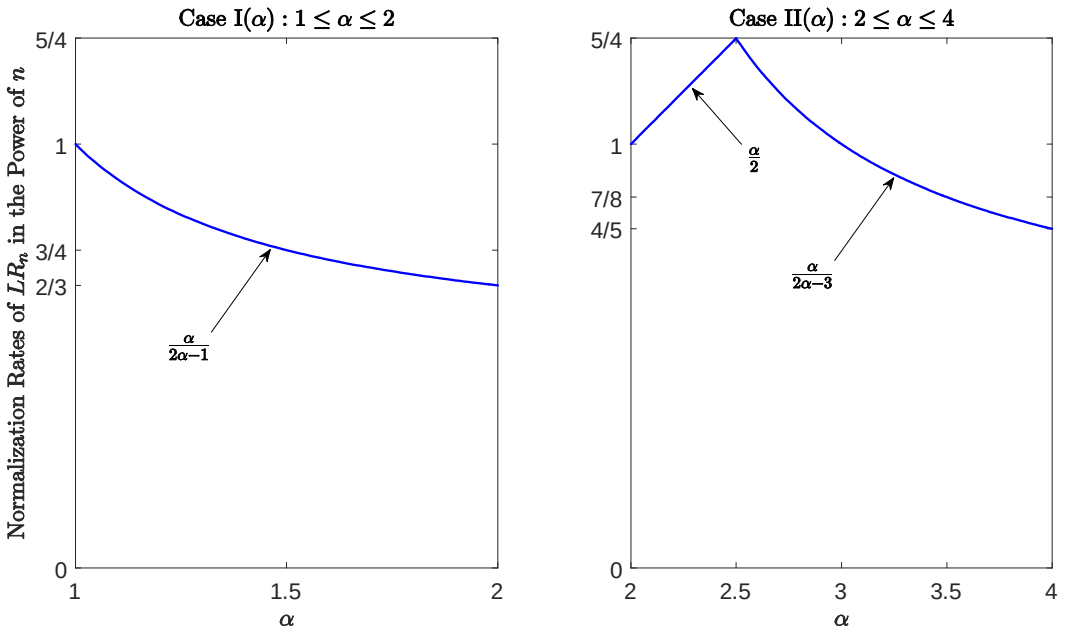

Figure 3: Normalization Rates of LR_n in the Power of n

4 Asymptotics in Discontinuous Threshold Regression: $\alpha = 1$

We will give I(1) a special treatment. This is because (i) the CS DTR is a special case of I(1) and attracts most of attentions of TR research; (ii) we hope the asymptotic results in I(1) provide a benchmark so that we can compare the results in any other case with those in I(1); (iii) we hope the detailed analysis of I(1) can provide a template for other cases so that we can streamline the statements of our asymptotic theory there. Parallel to Chan (1993) and Hansen (2000), we will consider both the fixed-threshold-effect framework and the shrinking-threshold-effect framework.

First, we define further notations for future use. Let

$$D_\gamma = \mathbb{E}\left[\mathbf{x}\mathbf{x}'|q=\gamma\right], D_0 = D_{\gamma_0}, E_\gamma = \mathbb{E}\left[\mathbf{x}y|q=\gamma\right] \text{ and } E_0 = E_{\gamma_0}.$$

When γ_0 is known, we know

$$\sqrt{n}\begin{pmatrix} \widehat{\beta}_1 - \beta_{10} \\ \widehat{\beta}_2 - \beta_{20} \end{pmatrix} \xrightarrow{d} Z_\beta := S_{\beta\beta}^{-1} W = \begin{pmatrix} M_0^{-1} W_1 \\ \overline{M}_0^{-1} W_2 \end{pmatrix} =: \begin{pmatrix} Z_{\beta_1}, \\ Z_{\beta_2} \end{pmatrix},$$

where

$$W = (W_1', W_2')' \text{ with } W_1 \sim N(0, \Sigma_0), W_2 \sim N\left(0, \overline{\Sigma}_0\right), \text{ and } W_1 \text{ and } W_2 \text{ being independent.} \tag{9}$$

The notations W, Z_{β_1} and Z_{β_2} will be used in the asymptotic distributions of $\widehat{\beta}_1$ and $\widehat{\beta}_2$ in some other cases besides I(1).

4.1 Asymptotics with Fixed Threshold Effects

We now state the asymptotic distribution of $\widehat{\theta}$ when δ_0 is fixed. First, we list the required assumptions.

Assumption I(1): Assumption MA plus

(x) $\mathbb{E}\left[\overline{z}_1|q=\gamma_0\right] > 0$ and $\mathbb{E}\left[\overline{z}_2|q=\gamma_0\right] > 0$, and z_1 and z_2 have absolutely continuous distributions.

Assumption (x) implies $S(\gamma, \beta_0)$ has a kink at γ_0; for all other cases, $\mathbb{E}\left[\overline{z}_\ell|q=\gamma_0\right] = 0$ so $S(\gamma, \beta_0)$ is differentiable at γ_0. Note here that we did not write $\mathbb{E}\left[\overline{z}_1|q=\gamma_0\right]$ as $\mathbb{E}\left[\overline{z}_1|q=\gamma_0-\right]$ and $\mathbb{E}\left[\overline{z}_2|q=\gamma_0\right]$ as $\mathbb{E}\left[\overline{z}_2|q=\gamma_0+\right]$ thanks to Assumption (iii) given that $\overline{z}_\ell$ is a function of x and ε_ℓ.

Define a compound Poisson process $D(\cdot)$ as

$$D(v) = \begin{cases} \sum_{i=1}^{N_1(|v|)} z_{1i}, \text{ if } v \leq 0; \\ \sum_{i=1}^{N_2(v)} z_{2i}, \text{ if } v > 0. \end{cases}$$

In $D(v)$, $\{z_{1i}, z_{2i}\}_{i\geq 1}$, $N_1(\cdot)$ and $N_2(\cdot)$ are independent of each other, $N_\ell(\cdot)$ is a Poisson process with intensity f_0, $z_{1i} = \lim_{\Delta\uparrow 0} \overline{z}_{1i} 1\{\gamma_0 + \Delta < q_i \leq \gamma_0\}$ is the limiting conditional value of $\overline{z}_{1i}$ given $\gamma_0 + \Delta < q_i \leq \gamma_0$, $\Delta < 0$ with $\Delta \uparrow 0$, and $z_{2i} = \lim_{\Delta\downarrow 0} \overline{z}_{2i} 1\{\gamma_0 < q_i \leq \gamma_0 + \Delta\}$ is the limiting conditional value of $\overline{z}_{2i}$ given $\gamma_0 < q_i \leq \gamma_0 + \Delta$, $\Delta > 0$ with $\Delta \downarrow 0$. When $m_\ell(x_i, q_i) = \mathbf{x}_i'\beta_{\ell 0}$, z_{1i} and z_{2i} reduce to the form in Chan (1993) and Yu (2014) (divided by 2) and $\mathbb{E}\left[z_1|q=\gamma_0\right] = \mathbb{E}\left[z_2|q=\gamma_0\right] = \frac{1}{2}\delta_0' D_0 \delta_0 > 0$. Assumption (x) guarantees the uniqueness of $\arg\min_{v\in\mathbb{R}} D(v)$, where following the convention of $\widehat{\gamma}$ in Section 2.1, we takes the mid-point of the minimizing interval of $D(v)$ as the minimizer.

Theorem 1 *Under Assumption I(1),*

$$
\begin{aligned}
&n\left(\widehat{\gamma}-\gamma_0\right) \xrightarrow{d} \arg\min_v D(v) =: Z_\gamma(1), \\
&\sqrt{n}\left(\widehat{\beta}_1-\beta_{10}\right) \xrightarrow{d} Z_{\beta_1}, \\
&\sqrt{n}\left(\widehat{\beta}_2-\beta_{20}\right) \xrightarrow{d} Z_{\beta_2},
\end{aligned}
$$

and $Z_\gamma(1)$, Z_{β_1} and Z_{β_2} are independent.

The asymptotic distribution of the structural break estimator in Proposition 4 of Bai (1997a) can be treated as a special case of Theorem 1 in the structural change context. In Bai's setup, $m(q)$ is piece-wise constant and $\mathbf{x}=1$. GP generalize Bai's setup to the TR context, but they show only that the convergence rate of $\widehat{\gamma}$ is n and do not derive its asymptotic distribution. The asymptotic distribution of $\widehat{\beta}$ is the same as in the case where γ_0 is known, i.e., estimating γ does not affect the asymptotic distribution of $\widehat{\beta}$, just as in the CS model. The following example derives the asymptotic distribution of $\widehat{\gamma}$ in GP's setup with three regimes.

Example 1 *Suppose* $y=(\mathbf{x}'b_{10}+\varepsilon_1)1(q\le\gamma_{10})+(\mathbf{x}'b_{20}+\varepsilon_2)1(\gamma_{10}<q\le\gamma_{20})+(\mathbf{x}'b_{30}+\varepsilon_3)1(q>\gamma_{20})$, *and* $\gamma_0=\arg\min_\gamma S(\gamma)=\gamma_{20}$. *Then* $\beta_{20}=b_{30}$, *and*

$$
\beta_{10}=\mathbb{E}\left[\mathbf{x}\mathbf{x}'_{\le\gamma_{20}}\right]^{-1}\mathbb{E}\left[\mathbf{x}_{\le\gamma_{20}}y\right]=\mathbb{E}\left[\mathbf{x}\mathbf{x}'_{\le\gamma_{20}}\right]^{-1}\left(\mathbb{E}\left[\mathbf{x}\mathbf{x}'_{\le\gamma_{10}}\right]b_{10}+\mathbb{E}\left[\mathbf{x}_{>\gamma_{10}}\mathbf{x}'_{\le\gamma_{20}}\right]b_{20}\right)=:wb_{10}+(I-w)b_{20}
$$

is a weighted average of b_{10} *and* b_{20} *by noticing that* $\mathbb{E}\left[\mathbf{x}\mathbf{x}'_{\le\gamma_{10}}\right]+\mathbb{E}\left[\mathbf{x}_{>\gamma_{10}}\mathbf{x}'_{\le\gamma_{20}}\right]=\mathbb{E}\left[\mathbf{x}\mathbf{x}'_{\le\gamma_{20}}\right]$, *so*

$$
\overline{\beta}_0=\frac{1}{2}\left[wb_{10}+(I-w)b_{20}+b_{30}\right]=w\frac{b_{10}+b_{30}}{2}+(I-w)\frac{b_{20}+b_{30}}{2}=:w\overline{\beta}_{10}+(I-w)\overline{\beta}_{20}
$$

is a weighted average of the the original two $\overline{\beta}$*'s, and*

$$
\delta_0=w(b_{10}-b_{30})+(I-w)(b_{20}-b_{30}):=w\delta_{10}+(I-w)\delta_{20}, \tag{10}
$$

is a weighted average of the original two threshold effects. Now,

$$
\begin{aligned}
\overline{z}_1 &= \left(\mathbf{x}'b_{10}+\varepsilon_1-\mathbf{x}'\overline{\beta}_0\right)\left(\delta_0'\mathbf{x}\right)1(q\le\gamma_{10})+\left(\mathbf{x}'b_{20}+\varepsilon_2-\mathbf{x}'\overline{\beta}_0\right)\left(\delta_0'\mathbf{x}\right)1(\gamma_{10}<q\le\gamma_{20}), \\
\overline{z}_2 &= -\left(\mathbf{x}'b_{30}+\varepsilon_3-\mathbf{x}'\overline{\beta}_0\right)\left(\delta_0'\mathbf{x}\right)1(q>\gamma_{20}),
\end{aligned}
$$

and

$$
\begin{aligned}
z_1 &= \left(\mathbf{x}'b_{20}+\varepsilon_2-\mathbf{x}'\overline{\beta}_0\right)\left(\delta_0'\mathbf{x}\right)|\,(q=\gamma_{20}-)=\left\{\mathbf{x}'\left[w\left(b_{20}-\overline{\beta}_{10}\right)+(I-w)\left(b_{20}-\overline{\beta}_{20}\right)\right]+\varepsilon_2\right\}\left(\delta_0'\mathbf{x}\right)|\,(q=\gamma_{20}-), \\
z_2 &= -\left(\mathbf{x}'b_{30}+\varepsilon_3-\mathbf{x}'\overline{\beta}_0\right)\left(\delta_0'\mathbf{x}\right)|\,(q=\gamma_{20}+)=-\left\{\mathbf{x}'\left[w\left(b_{30}-\overline{\beta}_{10}\right)+(I-w)\left(b_{30}-\overline{\beta}_{20}\right)\right]+\varepsilon_3\right\}\left(\delta_0'\mathbf{x}\right)|\,(q=\gamma_{20}+),
\end{aligned}
$$

where the first term of $\overline{z}_{1i}$ *is neglected because* z_{1i} *is the limit random variable in the left neighborhood of* γ_{20}. *For comparison, if the first and second regimes are combined into* $(\mathbf{x}'b_{20}+\varepsilon_2)1(q\le\gamma_{20})$, *i.e., the model is CS, then*

$$
z_1=\left(\frac{1}{2}\delta_0'\mathbf{x}+\varepsilon_{2i}\right)\left(\delta_0'\mathbf{x}\right)|\,(q=\gamma_{20}-)\ \text{ and }\ z_{2i}=-\left(\varepsilon_3-\frac{1}{2}\delta_0'\mathbf{x}\right)\left(\delta_0'\mathbf{x}\right)|\,(q=\gamma_{20}+);
$$

pro forma, $w\left(b_{30}-\overline{\beta}_{10}\right)+(I-w)\left(b_{30}-\overline{\beta}_{20}\right)=-\frac{1}{2}\delta_0$ *with* δ_0 *defined in (10), but* $w\left(b_{20}-\overline{\beta}_{10}\right)+(I-w)\left(b_{20}-\overline{\beta}_{20}\right)-\frac{1}{2}\delta_0=w\left(b_{20}-b_{10}\right)\neq\mathbf{0}$*, which is because the right regime is CS in the MS model, but the left regime is not.*

4.2 Asymptotics with Shrinking Threshold Effects

Because the distribution of Z_γ in Theorem 1 involves the distribution of $z_{\ell i}$ which is hard to estimate, we employ the shrinking-threshold-effect framework to obtain an accessible asymptotic distribution of $\widehat{\gamma}$. For future reference, we label the shrinking-threshold-effect case as I(1)′. First, we make the following assumptions.

Assumption I(1)′: same as Assumption I(1) except

(iv) (iv) of Assumption MA plus (c) $\sup_{\gamma\in\mathcal{N}}\mathbb{E}\left[\left(\|\mathbf{x}\|\left|\varepsilon_\ell\right|\right)^{2+\epsilon}\left|q=\gamma\right.\right]<\infty$ for some $\epsilon>0$.

(viii) $\varsigma_1(\gamma)=\lim_{n\to\infty}\mathbb{E}\left[\mathbf{x}\left(\frac{m_1(x,q)-\mathbf{x}'\beta_{10}}{\|\delta_n\|}\right)1(q\leq\gamma)\right]$ for $\gamma\in[\underline{\gamma},\gamma_0]$ and $\varsigma_2(\gamma)=\lim_{n\to\infty}\mathbb{E}\left[\mathbf{x}\left(\frac{m_2(x,q)-\mathbf{x}'\beta_{20}}{\|\delta_n\|}\right)1(q>\gamma)\right]$ for $\gamma\in[\gamma_0,\overline{\gamma}]$ exist, and $\arg\min_{\gamma\in\Gamma}S(\gamma)=\gamma_0$ is unique, where $S(\gamma):=\operatorname{plim}_{n\to\infty}\|\delta_n\|^{-2}\left[S_n(\gamma)-S_n(\theta_0)\right]$ is defined in Lemma 3, $\|\delta_n\|:=\|\beta_{10}-\beta_{20}\|\to0$ and $a_n:=n\|\delta_n\|^2\to\infty$.

(x) (a) $\mathbb{E}\left[\left(\frac{m_\ell(x,q)-\mathbf{x}'\beta_{\ell0}}{\|\delta_n\|}\right)^4\middle|q=\gamma\right]$ exists for $\gamma\in\mathcal{N}$; (b) $\dot{\varsigma}_1(\gamma)=\lim_{n\to\infty}\mathbb{E}\left[\mathbf{x}\left(\frac{m_1(x,q)-\mathbf{x}'\beta_{10}}{\|\delta_n\|}\right)\middle|q=\gamma\right]$ and $\dot{\varsigma}_2(\gamma)=-\lim_{n\to\infty}\mathbb{E}\left[\mathbf{x}\left(\frac{m_2(x,q)-\mathbf{x}'\beta_{20}}{\|\delta_n\|}\right)\middle|q=\gamma\right]$ exist for $\gamma\in\mathcal{N}$, $\dot{\varsigma}_\ell(\gamma)$ and D_γ are continuous at γ_0 with $\dot{\varsigma}_{\ell0}=\dot{\varsigma}_\ell(\gamma_0)$; (c) $V_\gamma^\pm$ are continuous at γ_0, where $V_\gamma^-=\mathbb{E}\left[\mathbf{x}\mathbf{x}'\varepsilon_1^2|q=\gamma-\right]$ and $V_\gamma^+=\mathbb{E}\left[\mathbf{x}\mathbf{x}'\varepsilon_2^2|q=\gamma+\right]$; (d) $c'D_0c+2c'\dot{\varsigma}_{10}>0$, $c'D_0c+2c'\dot{\varsigma}_{20}>0$ and $\omega_0^\pm:=c'V_0^\pm c>0$, where $c=\lim_{n\to\infty}\delta_n/\|\delta_n\|$, $V_0^-=V_{\gamma_0}^-$ and $V_0^+=V_{\gamma_0}^+$.

Assumption (iv)(c) is a little bit stronger than (iv)(b) due to a similar reason as Liapounov's condition in Lindeberg-Feller CLT (compared with Lindeberg-Lévy CLT). For example, Hansen (2000) takes $\epsilon=2$ in his Assumption 1.4. Following the discussions in Section 2.5, replacing the finite 4th moments in Assumption (iv)(a) to the finite $(4+\epsilon)$th moments is sufficient for Assumption (iv)(c). The other two assumptions are the counterparts of Assumption I(1) in the framework of shrinking threshold effects. The existence of $\varsigma_\ell(\gamma)$ and $\dot{\varsigma}_\ell(\gamma)$ implies that $m_\ell(x,q)-\mathbf{x}'\beta_{\ell0}\approx O_p(\|\delta_n\|)$. In CS models, $m_\ell(x,q)-\mathbf{x}'\beta_{\ell0}=0$ From Lemma 3, $S(\gamma)$ contains some extra terms beyond those in CS models due to the nonzeroness of $\varsigma_\ell(\gamma)$. Hansen (2000) assumes $\delta_n=cn^{-\alpha}$, $\alpha\in(0,1/2)$, and we extend his setup to the cases where the components of δ_n have different rates and the rates need not take the $n^{-\alpha}$ form (e.g., it can be $n^{-\alpha}\log n$). Our c is the normalized δ_n.

To state the asymptotic distribution of $\widehat{\gamma}$, define

$$\begin{aligned}C(v) &= \begin{cases}\frac{1}{2}f_0\left(c'D_0c+2c'\dot{\varsigma}_{10}\right)|v|+\sqrt{f_0c'V_0^-c}B_1(-v), & \text{if } v\leq0,\\ \frac{1}{2}f_0\left(c'D_0c+2c'\dot{\varsigma}_{20}\right)v+\sqrt{f_0c'V_0^+c}B_2(v), & \text{if } v>0,\end{cases}\\ &= :\begin{cases}\frac{1}{2}\mu_-|v|+\sqrt{\varpi_-}B_1(-v), & \text{if } v\leq0,\\ \frac{1}{2}\mu_+v+\sqrt{\varpi_+}B_2(v), & \text{if } v>0.\end{cases}\end{aligned}$$

where $\varpi_\pm=f_0\omega_0^\pm$. Even if we have assumed the distribution of x given $q=\gamma$ is continuous at γ_0 in Assumption (iii) (which implies D_γ is continuous at γ_0 so that only D_0 rather than $D_0^\pm$ appears in $C(v)$), the slopes of the deterministic part of $C(v)$ on $v\leq0$ and $v>0$ are not the same anymore, which is distinctly different from the CS model. On the other hand, the covariance kernel of $C(v)$ is the same as in the CS model where $\varepsilon_\ell=e_\ell$. As shown in Yu and Phillips (2018), $C(v)$ can be achieved from $D(v)$ by shrinking δ_0 and $m_\ell(x,q)-\mathbf{x}'\beta_{\ell0}$ in $z_{\ell i}$.

Theorem 2 *Under Assumption I(1)′,*

$$a_n\left(\widehat{\gamma}-\gamma_0\right)\xrightarrow{d}\arg\min_v C(v)=\omega\zeta\left(\varphi,\phi;1\right),$$

where $\omega=\frac{\varpi_-}{\mu_-^2}=\frac{c'V_0^-c}{f_0(c'D_0c+2c'\dot{\varsigma}_{10})^2}$ *and* $\varphi=\frac{\mu_+}{\mu_-}=\frac{c'D_0c+2c'\dot{\varsigma}_{20}}{c'D_0c+2c'\dot{\varsigma}_{10}}$ *and* $\phi=\frac{\varpi_+}{\varpi_-}=\frac{c'V_0^+c}{c'V_0^-c}$, $\widehat{\beta}_\ell$ *has the same asymptotic distribution as in Theorem 1, and* $\widehat{\gamma},\widehat{\beta}_1$ *and* $\widehat{\beta}_2$ *are asymptotically independent.*

Proposition 8 of Bai (1997a) can be treated as a special case of Theorem 2 in the structural change context with $m\left(q\right)$ piece-wise constant and $\mathbf{x}=1$. This form of $\zeta\left(\phi,\varphi;1\right)$ also appears in Proposition 3 of Bai (1997b) where the CS structural change model is considered. Note that φ is not equal to 1 in general, which is dramatically different from the case in CS models. On the other hand, the form of ϕ is similar; if the model is homoskedastic in each regime, $\phi=\sigma_2^2/\sigma_1^2$, where $\sigma_\ell^2=\mathbb{E}\left[\varepsilon_\ell^2\right]$, and $\phi=1$ when $\mathbb{E}\left[\mathbf{x}\mathbf{x}'\varepsilon_1^2|q=\gamma_0\right]=\mathbb{E}\left[\mathbf{x}\mathbf{x}'\varepsilon_2^2|q=\gamma_0\right]$ as assumed in Hansen (2000). This theorem can be treated as a misspecification-robust extension of Theorem 1 in Hansen (2000) where the model is CS.

Example 2 (continue of Example 1) *We need only derive the formulae of* $\dot{\varsigma}_{\ell0}$ *and* $V_0^\pm$ *to obtain the asymptotic distribution of* $\widehat{\gamma}$. *Note that*

$$\begin{aligned}\dot{\varsigma}_1\left(\gamma\right) &= \lim_{n\to\infty}\mathbb{E}\left[\mathbf{x}\left(\frac{\mathbf{x}'b_{20}-\mathbf{x}'\beta_{10}}{\|\delta_n\|}\right)\middle|q=\gamma_{20}\right]=D_0\left(\lim_{n\to\infty}\frac{w\left(\delta_{20}-\delta_{10}\right)}{\|\delta_{20}-w\left(\delta_{20}-\delta_{10}\right)\|}\right)=:D_0c_1,\\ \dot{\varsigma}_2\left(\gamma\right) &= \lim_{n\to\infty}\mathbb{E}\left[\mathbf{x}\left(\frac{\mathbf{x}'b_{30}-\mathbf{x}'\beta_{20}}{\|\delta_n\|}\right)\middle|q=\gamma_{20}\right]=\mathbf{0},\end{aligned}$$

so

$$\begin{aligned}c'D_0c+2c'\dot{\varsigma}_{10} &= c'D_0\left(c+2c_1\right)=c'D_0\left(\lim_{n\to\infty}\frac{\delta_{20}}{\|w\delta_{10}+\left(I-w\right)\delta_{20}\|}\right)\\ c'D_0c+2c'\dot{\varsigma}_{20} &= c'D_0c,\end{aligned}$$

where $c'D_0c+2c'\dot{\varsigma}_{20}$ *takes the same form as in the CS model because the right regime is indeed CS. Next,*

$$V_0^+=\mathbb{E}\left[\mathbf{x}\mathbf{x}'\varepsilon_2^2|q=\gamma_{20}-\right]\ \text{and}\ V_0^+=\mathbb{E}\left[\mathbf{x}\mathbf{x}'\varepsilon_3^2|q=\gamma_{20}+\right]$$

take the same form as in the CS model.

To conduct inference on γ, we use the LR-like statistic as mentioned in Section 2.1:

$$LR_n\left(\gamma\right)=\frac{n\left(S_n(\gamma)-S_n\left(\widehat{\gamma}\right)\right)}{\widehat{\eta}^2},$$

where $\widehat{\eta}^2$ is a consistent estimator of $\eta^2=\frac{\varpi_-}{\mu_-}=\frac{c'V_0^-c}{c'D_0c+2c'\dot{\varsigma}_{10}}$, so $\tau_n=n$ and $\widehat{b}=\widehat{\eta}^2$ here.

Corollary 1 *Under Assumption I(1)′,*

$$LR_n\left(\gamma_0\right)\xrightarrow{d}\xi\left(\varphi,\phi;1\right)$$

where the distribution of $\xi\left(\varphi,\phi;1\right)$ *is given in Proposition 2(iii) with* φ *and* ϕ *defined in Theorem 2.*

To make $LR_n(\gamma)$ feasible, we need to estimate η^2, φ and ϕ. From the proof of the Theorem 2,

$$c'V_0^- c = \lim_{n\to\infty} \frac{\mathbb{E}\left[\bar{z}_{1i}^2 | q_i = \gamma_0 -\right]}{\|\delta_n\|^2} = \lim_{n\to\infty} \frac{\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)^2 \left(y_i - \mathbf{x}_i' \overline{\beta}_0\right)^2 | q_i = \gamma_0 -\right]}{\|\delta_n\|^2},^{9}$$
$$c'V_0^+ c = \lim_{n\to\infty} \frac{\mathbb{E}\left[\bar{z}_{2i}^2 | q_i = \gamma_0 +\right]}{\|\delta_n\|^2} = \lim_{n\to\infty} \frac{\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)^2 \left(y_i - \mathbf{x}_i' \overline{\beta}_0\right)^2 | q_i = \gamma_0 +\right]}{\|\delta_n\|^2},$$

and

$$\frac{1}{2}\left(c'D_0 c + 2c'\dot{\varsigma}_{10}\right) = \lim_{n\to\infty} \frac{\mathbb{E}[\bar{z}_{1i} | q_i = \gamma_0 -]}{\|\delta_n\|^2} = \lim_{n\to\infty} \frac{\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)\left(y_i - \mathbf{x}_i' \overline{\beta}_0\right) | q_i = \gamma_0 -\right]}{\|\delta_n\|^2},$$
$$\frac{1}{2}\left(c'D_0 c + 2c'\dot{\varsigma}_{20}\right) = \lim_{n\to\infty} \frac{\mathbb{E}[\bar{z}_{2i} | q_i = \gamma_0 +]}{\|\delta_n\|^2} = \lim_{n\to\infty} \frac{-\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)\left(y_i - \mathbf{x}_i' \overline{\beta}_0\right) | q_i = \gamma_0 +\right]}{\|\delta_n\|^2}.$$

This implies that

$$\eta^2 \approx \frac{\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)^2 \left(y_i - \mathbf{x}_i' \overline{\beta}_0\right)^2 | q_i = \gamma_0 -\right]}{2\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)\left(y_i - \mathbf{x}_i' \overline{\beta}_0\right) | q_i = \gamma_0 -\right]}, \varphi \approx \frac{-\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)\left(y_i - \mathbf{x}_i' \overline{\beta}_0\right) | q_i = \gamma_0 +\right]}{\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)\left(y_i - \mathbf{x}_i' \overline{\beta}_0\right) | q_i = \gamma_0 -\right]} \text{ and } \phi \approx \frac{\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)^2 \left(y_i - \mathbf{x}_i' \overline{\beta}_0\right)^2 | q_i = \gamma_0 +\right]}{\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)^2 \left(y_i - \mathbf{x}_i' \overline{\beta}_0\right)^2 | q_i = \gamma_0 -\right]}. \tag{11}$$

As a result, η^2, φ and ϕ can be consistently estimated by kernel regression or series regression as in Section 3.4 of Hansen (2000) but with δ_n replaced by $\widehat{\delta}$ and $\overline{\beta}_0$ by $\frac{\widehat{\beta}_1 + \widehat{\beta}_2}{2}$ in the formulae of η^2, φ and ϕ. These estimators are robust to misspecification in $\mathbb{E}[y|x,q]$ in TR, just like White's sandwich-form convariance matrix estimator is robust to mispecification in conditional mean $\mathbb{E}[y|\mathbf{x}]$ in linear regression. Given the estimators of η^2, φ and ϕ, say $\widehat{\eta}^2, \widehat{\varphi}$ and $\widehat{\phi}$, the $(1-\alpha)$ LR confidence interval for γ follows by inversion from

$$\left\{\gamma : LR_n(\gamma) \leq \widehat{\text{crit}}\right\},$$

where $\widehat{\text{crit}}$ is the $(1-\alpha)$ quantile of $\xi\left(\widehat{\varphi}, \widehat{\phi}; 1\right)$.

In the CS model,

$$\eta^2 = \frac{\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)^2 e_{1i}^2 | q_i = \gamma_0 -\right]}{\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)^2 | q_i = \gamma_0 -\right]}, \varphi = \frac{\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)^2 | q_i = \gamma_0 +\right]}{\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)^2 | q_i = \gamma_0 -\right]} \text{ and } \phi = \frac{\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)^2 e_{2i}^2 | q_i = \gamma_0 +\right]}{\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)^2 e_{1i}^2 | q_i = \gamma_0 -\right]}, \tag{12}$$

but these formulae are not correct in the MS model. Specifically, the correct formulae should be

$$\eta^2 = \frac{\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)^2 \varepsilon_{1i}^2 | q_i = \gamma_0 -\right]}{\mathbb{E}[(\delta_n' \mathbf{x}_i)(2e_{1i} + \delta_n' \mathbf{x}_i) | q_i = \gamma_0 -]}, \varphi = \frac{\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)\left(\mathbf{x}'\delta_n - 2e_{2i}\right) | q_i = \gamma_0 +\right]}{\mathbb{E}[(\delta_n' \mathbf{x}_i)(2e_{1i} + \delta_n' \mathbf{x}_i) | q_i = \gamma_0 -]} \text{ and } \phi = \frac{\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)^2 \varepsilon_{2i}^2 | q_i = \gamma_0 +\right]}{\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)^2 \varepsilon_{1i}^2 | q_i = \gamma_0 -\right]},$$

while

$$\begin{aligned}
&\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)^2 e_{1i}^2 | q_i = \gamma_0 -\right] - \mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)^2 \varepsilon_{1i}^2 | q = \gamma_0 -\right] = \mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)^2 \left(m_1(x_i, q_i) - \mathbf{x}_i' \beta_{10}\right)^2 | q_i = \gamma_0 -\right] = O\left(\|\delta_n\|^4\right) > 0, \\
&\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)^2 e_{2i}^2 | q_i = \gamma_0 +\right] - \mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)^2 \varepsilon_{2i}^2 | q_i = \gamma_0 +\right] = \mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)^2 \left(m_2(x_i, q_i) - \mathbf{x}_i' \beta_{20}\right)^2 | q_i = \gamma_0 +\right] = O\left(\|\delta_n\|^4\right) > 0, \\
&\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)^2 | q_i = \gamma_0 -\right] - \mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)\left(2e_{1i} + \delta_n' \mathbf{x}_i\right) | q_i = \gamma_0 -\right] = -2\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)\left(m_1(x_i, q_i) - \mathbf{x}_i' \beta_{10}\right) | q_i = \gamma_0 -\right] = O\left(\|\delta_n\|^2\right), \\
&\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)^2 | q_i = \gamma_0 +\right] - \mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)\left(\mathbf{x}'\delta_n - 2e_{2i}\right) | q_i = \gamma_0 +\right] = 2\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)\left(m_2(x_i, q_i) - \mathbf{x}_i' \beta_{20}\right) | q_i = \gamma_0 +\right] = O\left(\|\delta_n\|^2\right).
\end{aligned}$$

Although $\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)^2 e_{1i}^2 | q_i = \gamma_0 -\right]$ and $\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)^2 e_{2i}^2 | q_i = \gamma_0 +\right]$ can replace $\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)^2 \varepsilon_{1i}^2 | q = \gamma_0 -\right]$ and $\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)^2 \varepsilon_{2i}^2 | q_i = \gamma_0 +\right]$ because their differences are $o\left(\|\delta_n\|^2\right)$, $\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)^2 | q_i = \gamma_0 -\right]$ and $\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)^2 | q_i = \gamma_0 +\right]$ cannot replace $\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)\left(2e_{1i} + \delta_n' \mathbf{x}_i\right) | q_i = \gamma_0 -\right]$ and $\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)\left(\mathbf{x}'\delta_n - 2e_{2i}\right) | q_i = \gamma_0 +\right]$. In other words, η^2

[9] Note that $E\left[(\delta_n' \mathbf{x}_i)^2 \left(y_i - \mathbf{x}_i' \overline{\beta}_0\right)^2 | q_i = \gamma_0 -\right] - E\left[(\delta_n' \mathbf{x}_i)^2 \varepsilon_{1i}^2 | q_i = \gamma_0 -\right] = E\left[(\delta_n' \mathbf{x}_i)^2 \left(m_1(x_i, q_i) - \mathbf{x}_i' \overline{\beta}_0\right)^2 | q_i = \gamma_0 -\right] = E\left[(\delta_n' \mathbf{x}_i)^2 \left(m_1(x_i, q_i) - \mathbf{x}_i' \beta_{10} + \mathbf{x}_i' \delta_n / 2\right)^2 | q_i = \gamma_0 -\right] = O\left(\|\delta_n\|^4\right)$.

and φ are biased if the estimation is based on (12). Even for ϕ, the formula in (11) should have a smaller finite-sample bias than that in (12).

If the model is homoskedastic in each regime,

$$\eta^2 = \frac{c' D_0 c \cdot \sigma_1^2}{c' D_0 c + 2c_{10}'\zeta} \approx \frac{\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)^2 | q_i = \gamma_0 -\right] \mathbb{E}\left[\left(y_i - \mathbf{x}_i' \overline{\beta}_0\right)^2 | q_i = \gamma_0 -\right]}{2\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)\left(y_i - \mathbf{x}_i' \overline{\beta}_0\right) | q_i = \gamma_0 -\right]} \approx \frac{\mathbb{E}\left[\left(y_i - \mathbf{x}_i' \overline{\beta}_0\right)^2 | q_i \le \gamma_0\right]}{2\mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)\left(y_i - \mathbf{x}_i' \overline{\beta}_0\right) | q_i = \gamma_0 -\right] / \mathbb{E}\left[\left(\delta_n' \mathbf{x}_i\right)^2 | q_i = \gamma_0\right]},$$
$$\phi = \frac{\sigma_2^2}{\sigma_1^2} \approx \frac{\mathbb{E}\left[\left(y_i - \mathbf{x}_i' \overline{\beta}_0\right)^2 | q_i = \gamma_0 +\right]}{\mathbb{E}\left[\left(y_i - \mathbf{x}_i' \overline{\beta}_0\right)^2 | q_i = \gamma_0 -\right]} \approx \frac{\mathbb{E}\left[\left(y_i - \mathbf{x}_i' \overline{\beta}_0\right)^2 | q_i > \gamma_0\right]}{\mathbb{E}\left[\left(y_i - \mathbf{x}_i' \overline{\beta}_0\right)^2 | q_i \le \gamma_0\right]},$$

where the second approximation of η^2 and ϕ allows us to use more data to estimate them. Note that η^2 cannot be simplified to σ_1^2 as in the CS model.

5 Asymptotics in Discontinuous Threshold Regression: $1 < \alpha \le 2$

We first state the asymptotic theory of I(2) and then I(α), $1 < \alpha < 2$.

5.1 $\alpha = 2$

First, we specify the required assumptions.

Assumption I(2): same as Assumption I(1) except

(iv) (iv) of Assumption MA plus (c) $\sup_{\gamma \in \mathcal{N}} \mathbb{E}\left[\left|\mathbf{x}_i' \delta_0 \left(y_i - \mathbf{x}_i' \overline{\beta}_0\right)\right|^{2+\epsilon} | q_i = \gamma\right] < \infty$ for some $\epsilon > 0$.

(x) (a) $\Lambda_{\pm}(\gamma) \in RV_2$; (b) $S_{\theta\theta}^{\pm} > 0$, where $S_{\gamma\gamma}^{\pm} := 2\lambda_{\pm}$; (c) $\omega_{\gamma}^{\pm} := \mathbb{E}\left[\left|\mathbf{x}_i' \delta_0 \left(y_i - \mathbf{x}_i' \overline{\beta}_0\right)\right|^2 | q_i = \gamma \pm\right]$ is continuous at γ_0 and $\omega_0^{\pm} := \omega_{\gamma_0}^{\pm} > 0$.[10]

Assumption (iv)(c) is assumed due to a similar reason as in Assumption I(1)′. If $\Lambda_{\pm}(\gamma)$ in Assumption (x)(a) is actually the left and right second-order differentiable, then

$$\begin{aligned} S_{\gamma\gamma}^{-} &= -f_0 \frac{\mathbb{E}[\overline{z}_{1i} | q = \gamma_0 -]}{\partial \gamma} = -f_0 \frac{\partial \mathbb{E}\left[\mathbf{x}' \delta_0 \left(y - \mathbf{x}' \overline{\beta}_0\right) | q = \gamma_0 -\right]}{\partial \gamma} = -f_0 \frac{\partial \mathbb{E}\left[\mathbf{x}' \delta_0 \left(m_1(x,q) - \mathbf{x}' \overline{\beta}_0\right) | q = \gamma_0 -\right]}{\partial \gamma} = 2\lambda_- \\ S_{\gamma\gamma}^{+} &= f_0 \frac{\mathbb{E}[\overline{z}_{2i} | q = \gamma_0 +]}{\partial \gamma} = -f_0 \frac{\partial \mathbb{E}\left[\mathbf{x}' \delta_0 \left(y - \mathbf{x}' \overline{\beta}_0\right) | q = \gamma_0 +\right]}{\partial \gamma} = -f_0 \frac{\partial \mathbb{E}\left[\mathbf{x}' \delta_0 \left(m_2(x,q) - \mathbf{x}' \overline{\beta}_0\right) | q = \gamma_0 +\right]}{\partial \gamma} = 2\lambda_+, \end{aligned} \tag{13}$$

and is even the second-order differentiable, then

$$S_{\gamma\gamma} = -f_0 \frac{\partial \mathbb{E}\left[\mathbf{x}' \delta_0 \left(y - \mathbf{x}' \overline{\beta}_0\right) | q = \gamma_0\right]}{\partial \gamma} = -f_0 \frac{\partial \mathbb{E}\left[\mathbf{x}' \delta_0 \left(m(x,q) - \mathbf{x}' \overline{\beta}_0\right) | q = \gamma_0\right]}{\partial \gamma} = 2\lambda_- = 2\lambda_+, \tag{14}$$

where $f(\gamma)$ is assumed to be differentiable at γ_0. From (14), we need to impose further restrictions on the smoothness of $m(x,q)$ and $f_{x|q}(x|q)$ around $q = \gamma_0$ (beyond the continuity at $q = \gamma_0$) to guarantee the existence of $S_{\gamma\gamma}$. When $\mathbf{x} = (1, q)'$,

$$S_{\gamma} = \mathbb{E}\left[\mathbf{x}' \delta_0 \left(m(q) - \mathbf{x}' \overline{\beta}_0\right) | q = \gamma\right] = (1, \gamma) \delta_0 \left(m(\gamma) - (1, \gamma) \overline{\beta}_0\right), \tag{15}$$

the existence of $S_{\gamma\gamma}$ implies the differentiability of $m(\gamma)$ at γ_0, where $m(\gamma) - (1, \gamma)\overline{\beta}_0$ is termed as the "centered" regresson function in BM. If $m(\gamma)$ has a kink at γ_0, $S_{\gamma\gamma}$ does not exist and only $S_{\gamma\gamma}^{\pm}$ can be used.

[10] Here, "continuous" is understood as "left continuous" for ω_{γ}^{-} and "right continuous" for ω_{γ}^{+}. This convention is applied in the future discussions and will not be repeated again.

In $S^{\pm}_{\theta\theta}$ of Assumption (x)(b),

$$\begin{aligned}
S^{-}_{\beta\gamma} &= f_0 \begin{pmatrix} -\mathbb{E}\left[\mathbf{x}\left(m_1(x,q)-\mathbf{x}'\beta_{10}\right)|q=\gamma_0-\right] \\ \mathbb{E}\left[\mathbf{x}\left(m_1(x,q)-\mathbf{x}'\beta_{20}\right)|q=\gamma_0-\right] \end{pmatrix} = f_0\mathbb{E}\left[\left.\begin{pmatrix} -\mathbf{x}\left(y-\mathbf{x}'\beta_{10}\right) \\ \mathbf{x}\left(y-\mathbf{x}'\beta_{20}\right) \end{pmatrix}\right| q=\gamma_0-\right] \\
&= f_0 \begin{pmatrix} -\mathbb{E}\left[\mathbf{x}\left(m_1(x,q)-\mathbf{x}'\overline{\beta}_0-\mathbf{x}'\frac{\delta_0}{2}\right)|q=\gamma_0-\right] \\ \mathbb{E}\left[\mathbf{x}\left(m_1(x,q)-\mathbf{x}'\overline{\beta}_0+\mathbf{x}'\frac{\delta_0}{2}\right)|q=\gamma_0-\right] \end{pmatrix} =: \begin{pmatrix} S^{-}_{\beta_1\gamma} \\ S^{-}_{\beta_2\gamma} \end{pmatrix}, \\
S^{+}_{\beta\gamma} &= f_0 \begin{pmatrix} -\mathbb{E}\left[\mathbf{x}\left(m_2(x,q)-\mathbf{x}'\beta_{10}\right)|q=\gamma_0+\right] \\ \mathbb{E}\left[\mathbf{x}\left(m_2(x,q)-\mathbf{x}'\beta_{20}\right)|q=\gamma_0+\right] \end{pmatrix} = f_0\mathbb{E}\left[\left.\begin{pmatrix} -\mathbf{x}\left(y-\mathbf{x}'\beta_{10}\right) \\ \mathbf{x}\left(y-\mathbf{x}'\beta_{20}\right) \end{pmatrix}\right| q=\gamma_0+\right] \\
&= f_0 \begin{pmatrix} -\mathbb{E}\left[\mathbf{x}\left(m_2(x,q)-\mathbf{x}'\overline{\beta}_0-\mathbf{x}'\frac{\delta_0}{2}\right)|q=\gamma_0+\right] \\ \mathbb{E}\left[\mathbf{x}\left(m_2(x,q)-\mathbf{x}'\overline{\beta}_0+\mathbf{x}'\frac{\delta_0}{2}\right)|q=\gamma_0+\right] \end{pmatrix} =: \begin{pmatrix} S^{+}_{\beta_1\gamma} \\ S^{+}_{\beta_2\gamma} \end{pmatrix}.
\end{aligned} \tag{16}$$

In the simple example in Section 3, $\mathbf{x}=(1,q)'$, so $\Lambda_{\pm}(\gamma)\in RV_2$ implies

$$\mathbb{E}\left[\left(m_1(x,q)-\mathbf{x}'\overline{\beta}_0\right)\mathbf{x}1\left(\gamma<q\leq\gamma_0\right)\right]\in RV_2 \text{ and } -\mathbb{E}\left[\left(m_2(x,q)-\mathbf{x}'\overline{\beta}_0\right)\mathbf{x}1\left(\gamma_0<q\leq\gamma\right)\right]\in RV_2 \tag{17}$$

or $\mathbb{E}\left[\left(m_1(x,q)-\mathbf{x}'\overline{\beta}_0\right)\mathbf{x}|q=\gamma_0-\right]=\mathbb{E}\left[\left(m_2(x,q)-\mathbf{x}'\overline{\beta}_0\right)\mathbf{x}|q=\gamma_0+\right]=\mathbf{0}$. We do not impose such restrictions in the general case; otherwise, $S^{-}_{\beta\gamma}$ and $S^{+}_{\beta\gamma}$ can be simplified to

$$S^{-}_{\beta\gamma}=S^{+}_{\beta\gamma}=f_0\begin{pmatrix} \frac{1}{2}\mathbb{E}\left[\mathbf{x}\mathbf{x}'|q=\gamma_0\right]\delta_0 \\ \frac{1}{2}\mathbb{E}\left[\mathbf{x}\mathbf{x}'|q=\gamma_0\right]\delta_0 \end{pmatrix}=\frac{f_0}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix}\otimes(D_0\delta_0)=:S_{\beta\gamma}=\begin{pmatrix} S_{\beta_1\gamma} \\ S_{\beta_2\gamma} \end{pmatrix}$$

with $S_{\beta_1\gamma}=S_{\beta_2\gamma}$, and correspondingly, $S^{\pm}_{\theta\theta}$ is simplified to

$$S^{\pm}_{\theta\theta}=\begin{pmatrix} 2\lambda_{\pm} & S_{\gamma\beta} \\ S_{\beta\gamma} & S_{\beta\beta} \end{pmatrix} \text{ with } S_{\gamma\beta}=S'_{\beta\gamma} \text{ and } S_{\beta\beta}=\operatorname{diag}\left\{M_0,\overline{M}_0\right\}.$$

Of course, even if (17) does not hold, $S^{-}_{\beta\gamma}$ can still be equal to $S^{+}_{\beta\gamma}$; in this case,

$$\mathbb{E}\left[\left(m_1(x,q)-\mathbf{x}'\overline{\beta}_0\right)\mathbf{x}|q=\gamma_0-\right]=\mathbb{E}\left[\left(m_2(x,q)-\mathbf{x}'\overline{\beta}_0\right)\mathbf{x}|q=\gamma_0+\right]\neq\mathbf{0},$$

which is implied by the continuity of $m(x,q)$ at γ_0. Now,

$$S_{\beta\gamma}=f_0\mathbb{E}\left[\left.\begin{pmatrix} -\mathbf{x}\left(m(x,q)-\mathbf{x}'\beta_{10}\right) \\ \mathbf{x}\left(m(x,q)-\mathbf{x}'\beta_{20}\right) \end{pmatrix}\right| q=\gamma_0\right]=f_0\mathbb{E}\left[\left.\begin{pmatrix} -\mathbf{x}\left(y-\mathbf{x}'\beta_{10}\right) \\ \mathbf{x}\left(y-\mathbf{x}'\beta_{20}\right) \end{pmatrix}\right| q=\gamma_0\right]=f_0\begin{pmatrix} D_0M_0^{-1}N_0-E_0 \\ E_0-D_0\overline{M}_0^{-1}\overline{N}_0 \end{pmatrix}. \tag{18}$$

In Assumption (x)(c),

$$\begin{aligned}
\omega_0^{-} &= \mathbb{E}\left[\overline{z}_{1i}^2|q_i=\gamma_0-\right]=Var\left(\overline{z}_{1i}|q_i=\gamma_0-\right)=\mathbb{E}\left[\left(m_1(x,q)-\mathbf{x}'\overline{\beta}_0\right)^2\left(\mathbf{x}'\delta_0\right)^2|q=\gamma_0-\right]+\mathbb{E}\left[\left(\mathbf{x}'\delta_0\right)^2\varepsilon_1^2|q=\gamma_0-\right], \\
\omega_0^{+} &= \mathbb{E}\left[\overline{z}_{2i}^2|q_i=\gamma_0+\right]=Var\left(\overline{z}_{2i}|q_i=\gamma_0+\right)=\mathbb{E}\left[\left(m_2(x,q)-\mathbf{x}'\overline{\beta}_0\right)^2\left(\mathbf{x}'\delta_0\right)^2|q=\gamma_0+\right]+\mathbb{E}\left[\left(\mathbf{x}'\delta_0\right)^2\varepsilon_2^2|q=\gamma_0+\right],
\end{aligned} \tag{19}$$

where the second equality is because $\Lambda_{\pm}(\gamma)\in RV_2$ implies $\mathbb{E}\left[\overline{z}_{1i}|q_i=\gamma_0-\right]=0=\mathbb{E}\left[\overline{z}_{2i}|q_i=\gamma_0+\right]$. If $m(x,q)$ is continuous at γ_0 and $\mathbb{E}\left[\varepsilon_1^2|x,q=\gamma_0-\right]=\mathbb{E}\left[\varepsilon_2^2|x,q=\gamma_0+\right]$, then $\omega_0^{-}=\omega_0^{+}$. In the simple example, $\Lambda_{\pm}(\gamma)\in RV_2$ further implies

$$\mathbb{E}\left[\left(m_1(x,q)-\mathbf{x}'\overline{\beta}_0\right)^2\left(\mathbf{x}'\delta_0\right)^2 1(\gamma<q\leq\gamma_0)\right]\in RV_3 \text{ and } \mathbb{E}\left[\left(m_2(x,q)-\mathbf{x}'\overline{\beta}_0\right)^2\left(\mathbf{x}'\delta_0\right)^2 1\left(\gamma_0<q\leq\gamma\right)\right]\in RV_3, \tag{20}$$

so

$$\omega_0^- = \mathbb{E}\left[\left(\mathbf{x}_i'\delta_0\right)^2 \varepsilon_{1i}^2 | q_i = \gamma_0 -\right] \text{ and } \omega_0^+ = \mathbb{E}\left[\left(\mathbf{x}_i'\delta_0\right)^2 \varepsilon_{2i}^2 | q_i = \gamma_0 +\right], \tag{21}$$

but such a simplification cannot happen in general. In other words, the randomness in $\left(m_1(x,q) - \mathbf{x}'\overline{\beta}_0\right)(\mathbf{x}'\delta_0)\,1(\gamma < q \leq \gamma_0)$ and $-\left(m_2(x,q) - \mathbf{x}'\overline{\beta}_0\right)(\mathbf{x}'\delta_0)\,1(\gamma_0 < q \leq \gamma)$ cannot be neglected, which is very different from I(1)′ where the corresponding terms converge to the constants $\frac{1}{2}\mu_\pm |v|$ in mean square (i.e., the L^2-norm). Anyway, since $\varepsilon_{\ell i}$ is not observable, we still need to estimate $\omega_0^\pm$ based on the general formulae even in the simplified case.

The positive-definiteness of $S_{\theta\theta}^\pm$ in Assumption (x)(b) guarantees the local identification of θ_0. The following example shows that $S_{\theta\theta}^\pm > 0$ imposes some restrictions on $\lambda_\pm$.

Example 3 *Suppose* $\mathbf{x} = (1, q)'$, $q \sim U[-0.5, 0.5]$, $\gamma_0 = 0$, *and* $\delta_0 = (\delta_{c0}, \delta_{q0})'$. *Then*

$$S_{\theta\theta}^\pm = \begin{pmatrix} 2\lambda_\pm & S_{\gamma\beta} \\ S_{\beta\gamma} & S_{\beta\beta} \end{pmatrix} = \begin{pmatrix} 2\lambda_\pm & \frac{\delta_{c0}}{2} & 0 & \frac{\delta_{c0}}{2} & 0 \\ \frac{\delta_{c0}}{2} & \frac{1}{2} & -\frac{1}{8} & \multicolumn{2}{c}{\multirow{3}{*}{$\mathbf{0}$}} \\ 0 & -\frac{1}{8} & \frac{1}{24} & & \\ \frac{\delta_{c0}}{2} & \multicolumn{2}{c}{\multirow{2}{*}{$\mathbf{0}$}} & \frac{1}{2} & \frac{1}{8} \\ 0 & & & \frac{1}{8} & \frac{1}{24} \end{pmatrix} > 0$$

implies $\lambda_\pm > 2\delta_{c0}^2$. *Since* $S_{\beta\beta} > 0$ *and* $\lambda_\pm > 0$, *the restriction comes from the appearance of* $S_{\beta\gamma}$. *However,* $S_{\beta\gamma} = \frac{f_0}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix} \otimes (D_0\delta_0)$ *in the current setup, and* $D_0\delta_0 \neq \mathbf{0}$ *in DTR, so* $S_{\beta\gamma}$ *must appear.*

If $\mathbf{x} = 1$ *and* $m(q)$ *is differentiable at* $q = \gamma_0$ *as in BY and BM, then*

$$S_{\theta\theta} = \begin{pmatrix} 2\lambda & S_{\gamma\beta} \\ S_{\beta\gamma} & S_{\beta\beta} \end{pmatrix} = \begin{pmatrix} -f_0\delta_0 m'(\gamma_0) & \frac{f_0}{2}\delta_0 & \frac{f_0}{2}\delta_0 \\ \frac{f_0}{2}\delta_0 & F_0 & 0 \\ \frac{f_0}{2}\delta_0 & 0 & 1 - F_0 \end{pmatrix} = f_0 \begin{pmatrix} -\delta_0 m'(\gamma_0) & \frac{\delta_0}{2} & \frac{\delta_0}{2} \\ \frac{\delta_0}{2} & F_0/f_0 & 0 \\ \frac{\delta_0}{2} & 0 & (1-F_0)/f_0 \end{pmatrix} > 0$$

implies $\delta_0 m'(\gamma_0) < -\frac{f_0\delta_0^2}{4F_0(1-F_0)}\left(\leq -f_0\delta_0^2\right) < 0$ *or* $\lambda > \frac{f_0^2\delta_0^2}{8F_0(1-F_0)}$, *where* $F_0 = P(q \leq \gamma_0)$, *and* $(1-F_0)/f_0$ *is the reciprocal of the hazard function of* q *at* γ_0. *This is actually the assumption* $b > 0$ *in Theorem 2.1 of BM; the form of* $S_{\theta\theta}$ *(times 2) also appears in BM (p. 571). Note that* $m'(\gamma_0) \neq 0$ *as assumed in BY's (A2)(i) and BM's (A2) implies* $\Lambda(\gamma) \in RV_2$.

Before stating the asymptotic distribution of $\widehat{\theta}$, define $\mu_\pm = 2\lambda_\pm - S_{\gamma\beta_1}^\pm M_0^{-1} S_{\beta_1\gamma}^\pm - S_{\gamma\beta_2}^\pm \overline{M}_0^{-1} S_{\beta_2\gamma}^\pm$, and $\varpi_\pm = f_0\omega_0^\pm$.

Theorem 3 *Under Assumption I(2),*

$$n^{1/3}\left(\widehat{\gamma} - \gamma_0\right) \xrightarrow{d} \omega^{1/3}\zeta(\varphi, \phi; 2) =: Z_\gamma(2),$$

and

$$n^{1/3}\left(\widehat{\beta}_1 - \beta_{10}\right) \xrightarrow{d} -M_0^{-1}\left[S_{\beta_1\gamma}^- Z_\gamma(2)_\ominus + S_{\beta_1\gamma}^+ Z_\gamma(2)_\oplus\right],$$
$$n^{1/3}\left(\widehat{\beta}_2 - \beta_{20}\right) \xrightarrow{d} -\overline{M}_0^{-1}\left[S_{\beta_2\gamma}^- Z_\gamma(2)_\ominus + S_{\beta_2\gamma}^+ Z_\gamma(2)_\oplus\right],$$

where $\omega = \frac{\varpi_-}{\mu_-^2}$, $\varphi = \frac{\mu_+}{\mu_-}$, *and* $\phi = \frac{\varpi_+}{\varpi_-} = \frac{\omega_0^+}{\omega_0^-}$.

If $S^-_{\beta\gamma} = S^+_{\beta\gamma} = S_{\beta\gamma}$, $\omega^-_0 = \omega^+_0 = \omega_0$ and $\lambda_- = \lambda_+ = \lambda = \frac{S_{\gamma\gamma}}{2}$, then $\zeta(\varphi, \phi; 2)$ in $Z_\gamma(2)$ reduces to $\zeta_{1/2}$, where ζ_c is defined in Section 2.4. At the same time,

$$n^{1/3}\left(\widehat{\beta} - \beta_0\right) \xrightarrow{d} \begin{pmatrix} -M_0^{-1} S_{\beta_1\gamma} Z_\gamma(2) \\ -\overline{M}_0^{-1} S_{\beta_2\gamma} Z_\gamma(2) \end{pmatrix}. \tag{22}$$

This is actually Theorem 2 of Seo (2015) after some manipulations. The cube-root convergence rate of $\widehat{\gamma}$ implies that γ cannot be estimated precisely even as a sample mean, which is very different from what happens in I(1). When $\mathbf{x} = (1, q)'$, $m(q)$ must be continuous but may have a kink at γ_0. If we fit a CTR as in Hansen (2017), then $\widehat{\gamma}$ is $\sqrt{n}$-consistent, i.e., restrictions imposed in CTR help to improve the preciseness of $\widehat{\gamma}$. As Kim and Siegmund (1989) point out, the estimation of an abrupt change when only a gradual change exists can exaggerate the magnitude of the possible change or introduce unwanted bias into an estimate of its location.

Section 2 of BM considers the case where $\mathbf{x} = 1$, $\varepsilon_1 = \varepsilon_2$ and $m(q)$ is differentiable at $q = \gamma_0$. In this case, $\omega^+_0 = \omega^-_0 = \delta_0^2 \sigma^2(\gamma_0) =: \omega_0$ with $\sigma^2(\gamma) = \mathbb{E}\left[\varepsilon^2 | q = \gamma\right]$, $S_{\gamma\gamma} - S_{\gamma\beta} S^{-1}_{\beta\beta} S_{\beta\gamma} = f_0\left(-\delta_0 m'(\gamma_0) - \frac{f_0 \delta_0^2}{4F_0(1-F_0)}\right)$ as shown in Example 3, $-M_0^{-1} S_{\beta_1\gamma} = -\frac{f_0\delta_0}{2F_0}$ and $-\overline{M}_0^{-1} S_{\beta_2\gamma} = -\frac{f_0\delta_0}{2(1-F_0)}$. In summary,

$$\begin{aligned} n^{1/3}\left(\widehat{\theta} - \theta_0\right) &\xrightarrow{d} \begin{pmatrix} 1 \\ -\frac{f_0\delta_0}{2F_0} \\ -\frac{f_0\delta_0}{2(1-F_0)} \end{pmatrix} \arg\min_v \left\{ \sqrt{f_0 \sigma^2(\gamma_0) \delta_0^2} B(v) - \tfrac{v^2}{2} f_0 \left(-\delta_0 m'(\gamma_0) - \tfrac{f_0\delta_0^2}{4F_0(1-F_0)}\right) \right\} \\ &= \begin{pmatrix} 1 \\ -\frac{f_0\delta_0}{2F_0} \\ -\frac{f_0\delta_0}{2(1-F_0)} \end{pmatrix} \arg\min_v \left\{ \sqrt{f_0 \sigma^2(\gamma_0)} B(v) - \tfrac{v^2}{2} f_0 \left(-\left|m'(\gamma_0)\right| - \tfrac{f_0|\delta_0|}{4F_0(1-F_0)}\right) \right\}, \end{aligned}$$

which is exactly the same as that in Theorem 2.1 of BM, where $-\delta_0 m'(\gamma_0) / |\delta_0| = |m'(\gamma_0)|$. Section 3 of BM also considers the case where q is the only regressor and a possibly nonlinear function of q, say $g_\ell(q)$, is used in the conditional mean of y in each regime. In this case, $\overline{z}_1 = (m(q) - \overline{g}(q)) \Delta g(q)$ and $\overline{z}_2 = -(m(q) - \overline{g}(q)) \Delta g(q)$, where $\Delta g(q) = g_1(q) - g_2(q)$. They assume $\Delta g(\gamma_0) \neq 0$, $m(\gamma_0) = \overline{g}(\gamma_0)$ and $m'(\gamma_0) \neq \overline{g}'(\gamma_0)$, which guarantees the model is a DTR, and $\Lambda(\gamma) \in RV_2$ but $\Lambda(\gamma) \notin RV_\alpha$ for $\alpha > 2$. Their Theorem 3.1 can be treated as a special case of our Theorem 3 where $S(\theta)$ is the second-order differentiable, and the regressors in each regime are completely different but all functions of q.

From (22), the randomness in $\widehat{\beta} - \beta_0$ is completely determined by that in $\widehat{\gamma} - \gamma_0$ asymptotically; in other words, there is perfect correlation between $\widehat{\beta} - \beta_0$ and $\widehat{\gamma} - \gamma_0$ asymptotically, or the asymptotic distribution of $n^{1/3}\left(\widehat{\theta} - \theta_0\right)$ concentrates on a line through the origin. This is very different from case I(1) and I(1)′, where the randomness in the former is independent of that in the latter. If γ_0 were known, $\widehat{\beta}(\gamma_0) - \beta_0$ is $O_p\left(n^{-1/2}\right)$; however, when γ_0 is unknown, it contaminates the estimation of β such that $\widehat{\beta} - \beta_0$ is $O_p\left(n^{-1/3}\right)$. As a result, the randomness in $\widehat{\beta} - \beta_0 = \widehat{\beta}(\widehat{\gamma}) - \beta_0 = \left(\widehat{\beta}(\widehat{\gamma}) - \widehat{\beta}(\gamma_0)\right) + \left(\widehat{\beta}(\gamma_0) - \beta_0\right)$ is dominated by that in $\widehat{\beta}(\widehat{\gamma}) - \widehat{\beta}(\gamma_0)$. To get more intuitions, note that

$$\begin{aligned} & n^{1/3}\left(\widehat{\beta}_1(\widehat{\gamma}) - \widehat{\beta}_1(\gamma_0)\right) \\ &= n^{1/3}\left[\left(\tfrac{1}{n} X'_{\le\widehat{\gamma}} X_{\le\widehat{\gamma}}\right)^{-1}\left(\tfrac{1}{n} X'_{\le\widehat{\gamma}} Y\right) - \left(\tfrac{1}{n} X'_{\le\gamma_0} X_{\le\gamma_0}\right)^{-1}\left(\tfrac{1}{n} X'_{\le\gamma_0} Y\right)\right] \\ &\approx n^{1/3}\left(\mathbb{E}\left[\mathbf{x}\mathbf{x}' 1(q \le \widehat{\gamma})\right]^{-1} \mathbb{E}\left[\mathbf{x} y 1(q \le \widehat{\gamma})\right] - \mathbb{E}\left[\mathbf{x}\mathbf{x}' 1(q \le \gamma_0)\right]^{-1} \mathbb{E}\left[\mathbf{x} y 1(q \le \gamma_0)\right]\right) \\ &\approx n^{1/3}\left(-f_0 M_0^{-1} D_0 M_0^{-1} N_0 + f_0 M_0^{-1} E_0\right)(\widehat{\gamma} - \gamma_0) \\ &= f_0 M_0^{-1}\left(E_0 - D_0 M_0^{-1} N_0\right) n^{1/3}(\widehat{\gamma} - \gamma_0) = -M_0^{-1} S_{\beta_1\gamma} n^{1/3}(\widehat{\gamma} - \gamma_0), \end{aligned} \tag{23}$$

where the first approximation is because $\frac{1}{n}X'_{\leq\widehat{\gamma}}X_{\leq\widehat{\gamma}} = n^{-1/2}\mathbb{G}_n\left(\mathbf{x}\mathbf{x}'1(q\leq\widehat{\gamma})\right)+\mathbb{E}\left[\mathbf{x}\mathbf{x}'1(q\leq\widehat{\gamma})\right] = O_p\left(n^{-1/2}\right)+\mathbb{E}\left[\mathbf{x}\mathbf{x}'1(q\leq\widehat{\gamma})\right]$[11] and similarly $\frac{1}{n}X'_{\leq\widehat{\gamma}}Y = O_p\left(n^{-1/2}\right)+\mathbb{E}\left[\mathbf{x}y1(q\leq\widehat{\gamma})\right]$ so that the $O_p\left(n^{-1/2}\right)$ part can be neglected given that the pre-multiplying term is $n^{1/3}$, the second approximation is from the calculus of matrix derivative, and the last equality is from the alternative formula of $S_{\beta_1\gamma}$ in (18). Similar analyses apply to $\widehat{\beta}_2$.

Finally, consider the LR-inference on γ. Define

$$LR_n(\gamma) = \frac{n^{2/3}\left(S_n(\gamma) - S_n\left(\widehat{\gamma}\right)\right)}{\widehat{\eta}^{2/3}},$$

where $\widehat{\eta}^2$ is a consistent estimator of $\eta^2 = \frac{\varpi_-^2}{\mu_-}$.

Corollary 2 *Under Assumption I(2),*

$$LR_n(\gamma_0) \xrightarrow{d} \xi(\varphi,\phi;2),$$

where $\xi(\varphi,\phi;2)$ is defined in Proposition 2(ii) with φ and ϕ defined in Theorem 3.

When $\varphi=\phi=1$, $\xi(\varphi,\phi;2) = \xi_{1/2}$ with ξ_c defined in Section 2.4, which is essentially the asymptotic distribution appearing in Corollary 3 of Seo (2015) where $\ln S_n(\cdot)$ rather than $S_n(\cdot)$ is used in $LR_n(\gamma)$. The $\xi_{1/2}$ distribution also appears in Theorem 2.2 of BM after some manipulations. BM further suggest in their Theorem 2.3 an alternative LR statistic which has a larger $\mu_-(=2\lambda_-)$ to stabilize the inversion of $LR_n(\gamma)$.

To make the LR inference feasible, we need to estimate φ, ϕ and η^2 which are functions of $\lambda_\pm$, $S^\pm_{\gamma\beta_\ell}$, M_0, $\overline{M}_0$ and $\varpi_\pm$. The latter four objects can be estimated by their sample analogs, i.e., $S^\pm_{\gamma\beta_\ell}$ is estimated based on (16), and $\varpi_\pm$ is based on (19). Here, we provide more details on the estimation of $\lambda_\pm$ because the method here will be used in other cases. Although we can estimate $\lambda_\pm$ based on (13), the estimator is specific to II(2) and is not easy to extend to other scenarios. Alternatively, in I(α) with $1<\alpha\leq 2$, suppose $\Lambda(|\gamma|)$ takes the form of $|\gamma|$'s power for simplicity; then by observing

$$\mathbb{E}\left[\overline{z}_i 1(\gamma_0 - h < q_i \leq \gamma_0)\right] \approx \lambda_- h^\alpha \text{ and } \mathbb{E}\left[-\overline{z}_i 1(\gamma_0 < q_i \leq \gamma_0 + h)\right] \approx \lambda_+ h^\alpha$$

for some bandwidth h, we can estimate $\lambda_\pm$ by

$$\widehat{\lambda}_- = \frac{n^{-1}\sum_{i=1}^n \widehat{z}_i 1(\widehat{\gamma}-h<q_i\leq\widehat{\gamma})}{h^\alpha} \text{ and } \widehat{\lambda}_+ = \frac{-n^{-1}\sum_{i=1}^n \widehat{z}_i 1(\widehat{\gamma}<q_i\leq\widehat{\gamma}+h)}{h^\alpha.}, \tag{24}$$

where $\widehat{z}_i = \mathbf{x}_i'\widehat{\delta}\left(y_i - \mathbf{x}_i'\widehat{\overline{\beta}}\right)$ with $\widehat{\delta} = \widehat{\beta}_1 - \widehat{\beta}_2$ and $\widehat{\overline{\beta}} = \frac{\widehat{\beta}_1+\widehat{\beta}_2}{2}$. When $\alpha=1$, $\widehat{\lambda}_\pm$ reduce to the kernel estimator with the uniform boundary kernel as Section 3.4 of Yu et al. (2019). This estimator of $\lambda_\pm$ can be used for all other cases and will not be repeated in the future.

5.2 $1<\alpha<2$

As indicated in the intuition of Section 3.3, we expect the convergence rate of $\widehat{\beta}$ to be $\min\left(n^{1/2},\rho_n\right)$; when $\Lambda(|\gamma|)$ takes the form of $|\gamma|$'s power,

$$\min\left(n^{1/2},\rho_n\right) = \begin{cases} n^{1/2}, & \text{if } 1<\alpha\leq\frac{3}{2}, \\ n^{\frac{1}{2\alpha-1}}, & \text{if } \frac{3}{2}<\alpha<2, \end{cases}$$

[11] Here, $\mathbb{G}_n\left(\mathbf{x}\mathbf{x}'1(q\leq\gamma)\right)$ is the empirical process indexed by γ, and $E\left[\mathbf{x}\mathbf{x}'1(q\leq\widehat{\gamma})\right] := E\left[\mathbf{x}\mathbf{x}'1(q\leq\gamma)\right]|_{\gamma=\widehat{\gamma}}$.

which is faster than the usual balancing rate $n^{\frac{\alpha}{2(2\alpha-1)}}$. This rate can also be seen through a similar analysis as in (23). When $\alpha < \frac{3}{2}$, the convergence rate of $\widehat{\beta}(\widehat{\gamma}) - \widehat{\beta}(\gamma_0)$ is determined by $\widehat{\gamma} - \gamma_0$ whose convergence rate is faster than $n^{1/2}$, so the asymptotic distribution of $\widehat{\beta}$ is completely determined by $\widehat{\beta}(\gamma_0) - \beta_0$. When $\alpha > \frac{3}{2}$, the converse happens, and the asymptotic distribution of $\widehat{\beta}$ is just a linear transformation of that of $\widehat{\gamma}$ as indicated in (23). Only when $\alpha = \frac{3}{2}$, both $\widehat{\beta}(\widehat{\gamma}) - \widehat{\beta}(\gamma_0)$ and $\widehat{\beta}(\gamma_0) - \beta_0$ will contribute to the asymptotic distribution of $\widehat{\beta}$.

We next specify the required assumptions.

Assumption I(α) $[1 < \alpha < 2]$: same as Assumption I(2) except

(x) (a) $\Lambda_{\pm}(\gamma) \in RV_\alpha$; (b) $\omega_\gamma^{\pm} := \mathbb{E}\left[\left|\mathbf{x}_i'\delta_0\left(y_i - \mathbf{x}_i'\overline{\beta}_0\right)\right|^2 | q_i = \gamma\pm\right]$ is continuous at γ_0 and $\omega_0^{\pm} := \omega_{\gamma_0}^{\pm} > 0$.

If we assume (20) but replace RV_3 by $RV_{2\alpha-1}$, then we can simplify $\omega_0^{\pm}$ to (21), but we will keep this general form of $\omega_0^{\pm}$ here. Similarly, if RV_2 is replaced by RV_α in (17), $S_{\beta\gamma}^{\pm}$ can be simplified as there.

Before stating the asymptotic distribution of $\widehat{\theta}$, define $\mu_{\pm} = 2\lambda_{\pm}$ and $\varpi_{\pm} = f_0\omega_0^{\pm}$.

Theorem 4 *Under Assumption I(α), $1 < \alpha < 2$,*

$$\rho_n(\widehat{\gamma} - \gamma_0) \xrightarrow{d} \omega^{\frac{1}{2\alpha-1}}\zeta(\varphi, \phi; \alpha) =: Z_\gamma(\alpha),$$

where $\omega = \frac{\varpi_-}{\mu_-^2} = \frac{f_0\omega_0^-}{4\lambda_-^2}$, $\varphi = \frac{\mu_+}{\mu_-} = \frac{\lambda_+}{\lambda_-}$, and $\phi = \frac{\varpi_+}{\varpi_-} = \frac{\omega_0^+}{\omega_0^-}$, when $1 < \alpha < 1.5$,

$$\sqrt{n}\left(\widehat{\beta}_1 - \beta_{10}\right) \xrightarrow{d} Z_{\beta_1},$$
$$\sqrt{n}\left(\widehat{\beta}_2 - \beta_{20}\right) \xrightarrow{d} Z_{\beta_2},$$

when $1.5 < \alpha < 2$,

$$\rho_n\left(\widehat{\beta}_1 - \beta_{10}\right) \xrightarrow{d} -M_0^{-1}\left[S_{\beta_1\gamma}^- Z_\gamma(\alpha)_\ominus + S_{\beta_1\gamma}^+ Z_\gamma(\alpha)_\oplus\right],$$
$$\rho_n\left(\widehat{\beta}_2 - \beta_{20}\right) \xrightarrow{d} -\overline{M}_0^{-1}\left[S_{\beta_2\gamma}^- Z_\gamma(\alpha)_\ominus + S_{\beta_2\gamma}^+ Z_\gamma(\alpha)_\oplus\right],$$

and when $\alpha = 1.5$,

$$\sqrt{n}\left(\widehat{\beta}_1 - \beta_{10}\right) \xrightarrow{d} Z_{\beta_1} - M_0^{-1}\left[S_{\beta_1\gamma}^- Z_\gamma(1.5)_\ominus + S_{\beta_1\gamma}^+ Z_\gamma(1.5)_\oplus\right],$$
$$\sqrt{n}\left(\widehat{\beta}_2 - \beta_{20}\right) \xrightarrow{d} Z_{\beta_2} - \overline{M}_0^{-1}\left[S_{\beta_2\gamma}^- Z_\gamma(1.5)_\ominus + S_{\beta 2\gamma}^+ Z_\gamma(1.5)_\oplus\right],$$

where $Z_\gamma(\alpha), Z_{\beta_1}$ and Z_{β_2} are independent.

Comparing Theorems 2, 3 and 4, we can see that the asymptotic distributions of $\widehat{\gamma}$ take a unified form with the key difference lying in the definitions of ω, φ and ϕ. In I(1)′, using the notations in this subsection, we have $\mu_{\pm} = \lim_{n\to\infty} 2\lambda_{\pm}/\|\delta_n\|^2$, $\omega_0^- = \lim_{n\to\infty}\mathbb{E}\left[\overline{z}_{1i}^2 | q_i = \gamma_0-\right]/\|\delta_n\|^2$ and $\omega_0^+ = \lim_{n\to\infty}\mathbb{E}\left[\overline{z}_{2i}^2 | q_i = \gamma_0+\right]/\|\delta_n\|^2$, where note that $\lambda_- = f_0\mathbb{E}\left[\overline{z}_{1i} | q_i = \gamma_0-\right]$ and $\lambda_+ = f_0\mathbb{E}\left[\overline{z}_{2i} | q_i = \gamma_0+\right]$. Dividing $\|\delta_n\|^2$ is to ensure $\mu_{\pm}$ and $\omega_0^{\pm}$ nondegenerate given that we assume $\|\delta_n\| \to 0$ as $n \to 0$; apart from this, the formulae in Theorems 2 and 4 can be unified. Compared with Theorem 3, $Z_\gamma(\alpha)$ in Theorem 4 replaces $\mu_{\pm}$ by $2\lambda_{\pm}$ because the cross terms are dominated as shown in Section 3.1. As expected, when $1 < \alpha < 1.5$, the asymptotic distribution of $\widehat{\beta}$ is not affected by $\widehat{\gamma}$ and is exactly the same as in I(1) and I(1)′. When $1.5 < \alpha < 2$, it is completely determined by $\widehat{\gamma}$ and takes the same form as in I(2). When $\alpha = 1.5$, it is the sum of both components. BM notice that BY made a mistake in claiming $\widehat{\beta}$ is $\sqrt{n}$-consistent (in

their proof of Theorem 3.1) which is used to show that $\widehat{\beta}$ will not affect the asymptotics of $\widehat{\gamma}$. Actually, $\widehat{\beta}$ is $\sqrt{n}$-consistent when $1<\alpha\leq 1.5$ and it will not affect the asymptotics of $\widehat{\gamma}$ as long as $\alpha<2$.

Finally, consider the LR-inference on γ. Define

$$LR_n(\gamma)=\frac{\sqrt{n\rho_n}\left(S_n(\gamma)-S_n\left(\widehat{\gamma}\right)\right)}{\widehat{\eta}^{2/(2\alpha-1)}},$$

where $\widehat{\eta}^2$ is a consistent estimator of $\eta^2=\frac{\varpi_-^\alpha}{\mu_-}=\frac{\left(f_0\omega_0^-\right)^\alpha}{2\lambda_-}$. Note that $n^{2/3}\prec\sqrt{n\rho_n}\prec n$.

Corollary 3 *Under Assumption I(α), $1<\alpha<2$,*

$$LR_n(\gamma_0)\xrightarrow{d}\xi(\varphi,\phi;\alpha),$$

where $\xi(\varphi,\phi;\alpha)$ is defined in Proposition 2(ii) with φ and ϕ defined in Theorem 4.

The form and asymptotic distribution of the LR statistic can also be unified for I(1)$'$ and I(α) with $1<\alpha\leq 2$ with properly defined ω, φ and η^2.

6 Asymptotics in Continuous Threshold Regression

We will discuss the asymptotic theory in the following order: II(2), II(3), II(4), II(α) with $3<\alpha<4$, and II(α) with $2<\alpha<3$. We first discuss II(2), II(3), II(4) to make them anchors of others (as I(1) and I(2) in DTR), and discuss II(α) with $3<\alpha<4$ before II(α) with $2<\alpha<3$ because we will use some results of the former in the latter.

In CTR, besides $\Lambda_\pm(\gamma)\in RV_\alpha$, we must assume a key assumption throughout. This key assumption is satisfied in the simple example of Section 3. Specifically, we need to extend (17) to

$$\mathbb{E}\left[\left(m_1(x,q)-\mathbf{x}'\overline{\beta}_0\right)\mathbf{x}1(\gamma<q\leq\gamma_0)\right]\in RV_{\alpha-1}\text{ and }-\mathbb{E}\left[\left(m_2(x,q)-\mathbf{x}'\overline{\beta}_0\right)\mathbf{x}1(\gamma_0<q\leq\gamma)\right]\in RV_{\alpha-1}. \quad (25)$$

Here, we implicitly assume these two objects are positive; otherwise, a negative sign is added given that the range of regularly varying functions must be $(0,\infty)$; the point is that only the rates of these two objects shrinking to zero matter. Although it seems innocent to assume (25) given that $\Lambda_\pm(\gamma)\in RV_\alpha$, i.e., $\mathbb{E}\left[\left(m_1(x,q)-\mathbf{x}'\overline{\beta}_0\right)q\delta_{q0}1(\gamma<q\leq\gamma_0)\right]\in RV_\alpha$ and $-\mathbb{E}\left[\left(m_2(x,q)-\mathbf{x}'\overline{\beta}_0\right)q\delta_{q0}1(\gamma_0<q\leq\gamma)\right]\in RV_\alpha$, it indeed excludes many cases when $\mathbf{x}$ includes nonconstant regressors besides q. When $\alpha=2$, this assumption is trivial, but when $\alpha>2$, it indeed has some contents.

Example 4 *Suppose $-(m_2(x,q)-\mathbf{x}'\overline{\beta}_0)=g(x)$, $x\in\mathbb{R}$, in the neighborhood of $q=\gamma_0$ i.e., q is offsetted; then $\mathbb{E}[g(x)q1(\gamma_0<q\leq\gamma)]\sim\gamma^\alpha$ need not imply $\mathbb{E}[g(x)x1(\gamma_0<q\leq\gamma)]\sim\gamma^{\alpha-1}$. To be concrete, let $g(x)=x$ and $\alpha=3$; $\mathbb{E}[g(x)q1(\gamma_0<q\leq\gamma)]\sim\gamma^3$ implies $\mathbb{E}[x|q=\gamma]\sim\gamma$, but this need not imply $\mathbb{E}[g(x)x|q=\gamma]=\mathbb{E}\left[x^2|q=\gamma\right]\sim\gamma$ (which would imply $\mathbb{E}[g(x)x1(\gamma_0<q\leq\gamma)]\sim\gamma^2$); e.g., if $(x,q)'\sim N\left(0,\left(\sigma_x^2,\sigma_{xq};\sigma_{qx},\sigma_q^2\right)\right)$ with $\sigma_{qx}\neq 0$, then $\mathbb{E}[x|q=\gamma]=\frac{\sigma_{qx}}{\sigma_x^2}\gamma\sim\gamma$ but $\mathbb{E}\left[x^2|q=\gamma\right]=\left(\frac{\sigma_{qx}}{\sigma_x^2}\right)^2\gamma^2+\left(\sigma_x^2-\frac{\sigma_{qx}^2}{\sigma_q^2}\right)\sim 1$.*

Nevertheless, without this assumption, we can only analyze II(2) because as shown in Section 3, the model in II(α) with $2<\alpha\leq 4$ may be locally unidentified given that $\left\|\widetilde{\beta}\right\|^2+|\gamma|^\alpha$ need not dominate the cross term $\left\|\widetilde{\beta}\right\||\gamma|$. Here, note that $\left\|\widetilde{\beta}\right\||\gamma|$ would appear if (25) does not hold such that $S_{\widetilde{\beta}\gamma}^\pm\neq\mathbf{0}$; to be concrete, from

the formulae of $\Psi_{\pm}(\beta,\gamma)$ in Section 2.1, we have

$$\begin{aligned}
S_{\beta}^{-}(\gamma) = \tfrac{d\Psi_{-}(\beta_0,\gamma)}{d\beta} &= \begin{pmatrix} \mathbb{E}\left[\mathbf{x}\left(m_1(x,q)-\mathbf{x}'\overline{\beta}_0\right)1(\gamma<q\leq\gamma_0)\right]-\frac{1}{2}\mathbb{E}\left[\mathbf{x}q\delta_{q0}1(\gamma<q\leq\gamma_0)\right] \\ -\mathbb{E}\left[\mathbf{x}\left(m_1(x,q)-\mathbf{x}'\overline{\beta}_0\right)1(\gamma<q\leq\gamma_0)\right]-\frac{1}{2}\mathbb{E}\left[\mathbf{x}q\delta_{q0}1(\gamma<q\leq\gamma_0)\right] \end{pmatrix} \\
&= \begin{pmatrix} \mathbb{E}\left[\mathbf{x}\left(y-\mathbf{x}'\beta_{10}\right)1(\gamma<q\leq\gamma_0)\right] \\ \mathbb{E}\left[-\mathbf{x}\left(y-\mathbf{x}'\beta_{20}\right)1(\gamma<q\leq\gamma_0)\right] \end{pmatrix} =: \begin{pmatrix} S_{\beta_1}^{-}(\gamma) \\ S_{\beta_2}^{-}(\gamma) \end{pmatrix}, \\
S_{\beta}^{+}(\gamma) = \tfrac{d\Psi_{+}(\beta_0,\gamma)}{d\beta} &= \begin{pmatrix} -\mathbb{E}\left[\mathbf{x}\left(m_2(x,q)-\mathbf{x}'\overline{\beta}_0\right)1(\gamma_0<q\leq\gamma)\right]+\frac{1}{2}\mathbb{E}\left[\mathbf{x}q\delta_{q0}1(\gamma_0<q\leq\gamma)\right] \\ \mathbb{E}\left[\mathbf{x}\left(m_2(x,q)-\mathbf{x}'\overline{\beta}_0\right)1(\gamma_0<q\leq\gamma)\right]+\frac{1}{2}\mathbb{E}\left[\mathbf{x}q\delta_{q0}1(\gamma_0<q\leq\gamma)\right] \end{pmatrix} \\
&= -\begin{pmatrix} \mathbb{E}\left[\mathbf{x}\left(y-\mathbf{x}'\beta_{10}\right)1(\gamma_0<q\leq\gamma)\right] \\ \mathbb{E}\left[-\mathbf{x}\left(y-\mathbf{x}'\beta_{20}\right)1(\gamma_0<q\leq\gamma)\right] \end{pmatrix} =: \begin{pmatrix} S_{\beta_1}^{+}(\gamma) \\ S_{\beta_2}^{+}(\gamma) \end{pmatrix},
\end{aligned} \tag{26}$$

where the second terms of $S_{\beta}^{\pm}(\gamma)$ are definitely RV_2, but the first terms may be only RV_1 such that $S_{\beta\gamma}^{\pm} \neq \mathbf{0}$ if (25) is not imposed.

6.1 $\alpha = 2$

This case is not discussed in the literature. First, we specify the required assumptions.

Assumption II(2): Assumption MA plus

(x) (a) $\Lambda_{\pm}(\gamma) \in RV_2$; (b) $S_{\theta\theta}^{\pm} > 0$, where $S_{\gamma\gamma}^{\pm} := 2\lambda_{\pm}$.

Assumption (x)(a) implies $P\left(m_1(x,\gamma_0) \neq (1,x',\gamma_0)\,\overline{\beta}_0 \neq m_2(x,\gamma_0)\right) > 0$. As to $S_{\theta\theta}^{\pm} > 0$, the discussions in Section 5.1 can still be applied here, but since $\mathbf{x}'\delta_0 = q\delta_{q0}$, $S_{\gamma\gamma}^{\pm}$ and $S_{\beta\gamma}^{\pm}$ can be simplified. For example, if $\Lambda_{\pm}(\gamma)$ is the left and right second-order differentiable, then from (13),

$$\begin{aligned}
S_{\gamma\gamma}^{-} &= -f_0\tfrac{\partial\mathbb{E}\left[\mathbf{x}'\delta_0\left(y-\mathbf{x}'\overline{\beta}_0\right)|q=\gamma_0-\right]}{\partial\gamma} = -f_0\tfrac{\partial\mathbb{E}\left[q\delta_{q0}\left(e_{1i}+q\frac{\delta_{q0}}{2}\right)|q_i=\gamma_0-\right]}{\partial\gamma} = -f_0\tfrac{\partial\mathbb{E}[q\delta_{q0}e_{1i}|q_i=\gamma_0-]}{\partial\gamma} \\
&= -f_0\delta_{q0}\mathbb{E}\left[e_{1i}|q_i=\gamma_0-\right] = -f_0\delta_{q0}\mathbb{E}\left[m_1(x,q)-\mathbf{x}'\beta_{10}|q_i=\gamma_0-\right] = 2\lambda_{-},
\end{aligned}$$

and similarly,

$$S_{\gamma\gamma}^{+} = -f_0\delta_{q0}\mathbb{E}\left[e_{2i}|q_i=\gamma_0+\right] = -f_0\delta_{q0}\mathbb{E}\left[m_2(x,q)-\mathbf{x}'\beta_{20}|q_i=\gamma_0+\right] = 2\lambda_{+},$$

where $f(\gamma)$ is assumed to be differentiable at γ_0; when $\Lambda_{\pm}(\gamma)$ is even the second-order differentiable, then

$$S_{\gamma\gamma} = -f_0\delta_{q0}\mathbb{E}\left[e_i|q_i=\gamma_0\right] = -f_0\delta_{q0}\mathbb{E}\left[m(x,q)-\mathbf{x}'\beta_{\ell 0}|q_i=\gamma_0\right] = 2\lambda_{-} = 2\lambda_{+} =: 2\lambda.$$

Similarly, we can simplify (16) as

$$\begin{aligned}
S_{\beta\gamma}^{-} &= f_0\begin{pmatrix} -\mathbb{E}\left[\left(m_1(x,q)-\mathbf{x}'\beta_{10}\right)\mathbf{x}|q=\gamma_0-\right] \\ \mathbb{E}\left[\left(m_1(x,q)-\mathbf{x}'\beta_{10}\right)\mathbf{x}|q=\gamma_0-\right] \end{pmatrix} = f_0\begin{pmatrix} -\mathbb{E}\left[\mathbf{x}e_{1i}|q=\gamma_0-\right] \\ \mathbb{E}\left[\mathbf{x}e_{1i}|q=\gamma_0-\right] \end{pmatrix} \\
&= f_0\begin{pmatrix} -\mathbb{E}\left[\left(m_1(x,q)-\mathbf{x}'\overline{\beta}_0\right)\mathbf{x}|q=\gamma_0-\right] \\ \mathbb{E}\left[\left(m_1(x,q)-\mathbf{x}'\overline{\beta}_0\right)\mathbf{x}|q=\gamma_0-\right] \end{pmatrix} =: \begin{pmatrix} S_{\beta_1\gamma}^{-} \\ S_{\beta_2\gamma}^{-} \end{pmatrix}, \\
S_{\beta\gamma}^{+} &= f_0\begin{pmatrix} -\mathbb{E}\left[\left(m_2(x,q)-\mathbf{x}'\beta_{20}\right)\mathbf{x}|q=\gamma_0+\right] \\ \mathbb{E}\left[\left(m_2(x,q)-\mathbf{x}'\beta_{20}\right)\mathbf{x}|q=\gamma_0+\right] \end{pmatrix} = f_0\begin{pmatrix} -\mathbb{E}\left[\mathbf{x}e_{2i}|q=\gamma_0+\right] \\ \mathbb{E}\left[\mathbf{x}e_{2i}|q=\gamma_0+\right] \end{pmatrix} \\
&= f_0\begin{pmatrix} -\mathbb{E}\left[\left(m_2(x,q)-\mathbf{x}'\overline{\beta}_0\right)\mathbf{x}|q=\gamma_0+\right] \\ \mathbb{E}\left[\left(m_2(x,q)-\mathbf{x}'\overline{\beta}_0\right)\mathbf{x}|q=\gamma_0+\right] \end{pmatrix} =: \begin{pmatrix} S_{\beta_1\gamma}^{+} \\ S_{\beta_2\gamma}^{+} \end{pmatrix}.
\end{aligned}$$

where note that $S^{\pm}_{\beta_2\gamma} = -S^{\pm}_{\beta_1\gamma}$. In CS models, $S^{\pm}_{\gamma\gamma} = 0$ so that $S^{\pm}_{\theta\theta} > 0$ cannot hold, i.e., II(2) includes only MS models. It is not hard to see $S^{\pm}_{\beta\gamma}$ is also equal to $\mathbf{0}$ in CS models. Different from Assumption I(2)(x), we do not need to define $\omega_0^{\pm}$ since it is zero in CTR; $\omega_0^{\pm}$ represents the randomness in $\widehat{\gamma}$, but we know from Section 3 that this part of randomness is dominated by the randomness in $\widehat{\beta}$.

Theorem 5 *Under Assumption II(2),*

$$\sqrt{n}\left(\widehat{\beta}-\beta_0\right) \xrightarrow{d} \begin{cases} \left(S_{\beta\beta}-S^{-}_{\beta\gamma}\left(S^{-}_{\gamma\gamma}\right)^{-1}S^{-}_{\gamma\beta}\right)^{-1}W, \text{ if } W\in R_1, \\ \left(S_{\beta\beta}-S^{+}_{\beta\gamma}\left(S^{+}_{\gamma\gamma}\right)^{-1}S^{+}_{\gamma\beta}\right)^{-1}W, \text{ if } W\in \overline{R}_1, \end{cases}$$

and

$$\sqrt{n}\left(\widehat{\gamma}-\gamma_0\right) \xrightarrow{d} \begin{cases} -\left(S^{-}_{\gamma\gamma}\right)^{-1}S^{-}_{\gamma\beta}\left(S_{\beta\beta}-S^{-}_{\beta\gamma}\left(S^{-}_{\gamma\gamma}\right)^{-1}S^{-}_{\gamma\beta}\right)^{-1}W, \text{ if } W\in R_1\cap R_2, \\ -\left(S^{+}_{\gamma\gamma}\right)^{-1}S^{+}_{\gamma\beta}\left(S_{\beta\beta}-S^{-}_{\beta\gamma}\left(S^{-}_{\gamma\gamma}\right)^{-1}S^{-}_{\gamma\beta}\right)^{-1}W, \text{ if } W\in R_1\cap \overline{R}_2, \\ -\left(S^{-}_{\gamma\gamma}\right)^{-1}S^{-}_{\gamma\beta}\left(S_{\beta\beta}-S^{+}_{\beta\gamma}\left(S^{+}_{\gamma\gamma}\right)^{-1}S^{+}_{\gamma\beta}\right)^{-1}W, \text{ if } W\in \overline{R}_1\cap R_3, \\ -\left(S^{+}_{\gamma\gamma}\right)^{-1}S^{+}_{\gamma\beta}\left(S_{\beta\beta}-S^{+}_{\beta\gamma}\left(S^{+}_{\gamma\gamma}\right)^{-1}S^{+}_{\gamma\beta}\right)^{-1}W, \text{ if } W\in \overline{R}_1\cap \overline{R}_3, \end{cases}$$

where W is defined in (9),

$$\begin{aligned} R_1 &= \left\{W| -\tfrac{1}{2}W'\left(S_{\beta\beta}-S^{-}_{\beta\gamma}\left(S^{-}_{\gamma\gamma}\right)^{-1}S^{-}_{\gamma\beta}\right)^{-1}W \le -\tfrac{1}{2}W'\left(S_{\beta\beta}-S^{+}_{\beta\gamma}\left(S^{+}_{\gamma\gamma}\right)^{-1}S^{+}_{\gamma\beta}\right)^{-1}W\right\}, \\ R_2 &= \left\{W|W'\left(S_{\beta\beta}-S^{-}_{\beta\gamma}\left(S^{-}_{\gamma\gamma}\right)^{-1}S^{-}_{\gamma\beta}\right)^{-1}\left[S^{-}_{\beta\gamma}\left(S^{-}_{\gamma\gamma}\right)^{-1}S^{-}_{\gamma\beta}-S^{+}_{\beta\gamma}\left(S^{+}_{\gamma\gamma}\right)^{-1}S^{+}_{\gamma\beta}\right]\left(S_{\beta\beta}-S^{-}_{\beta\gamma}\left(S^{-}_{\gamma\gamma}\right)^{-1}S^{-}_{\gamma\beta}\right)^{-1}W\ge 0\right\}, \\ R_3 &= \left\{W|W'\left(S_{\beta\beta}-S^{+}_{\beta\gamma}\left(S^{+}_{\gamma\gamma}\right)^{-1}S^{+}_{\gamma\beta}\right)^{-1}\left[S^{-}_{\beta\gamma}\left(S^{-}_{\gamma\gamma}\right)^{-1}S^{-}_{\gamma\beta}-S^{+}_{\beta\gamma}\left(S^{+}_{\gamma\gamma}\right)^{-1}S^{+}_{\gamma\beta}\right]\left(S_{\beta\beta}-S^{+}_{\beta\gamma}\left(S^{+}_{\gamma\gamma}\right)^{-1}S^{+}_{\gamma\beta}\right)^{-1}W\ge 0\right\}, \end{aligned}$$

and $\overline{R}_1$, $\overline{R}_2$ and $\overline{R}_3$ are their negations.

In some special cases, the asymptotic distributions of $\widehat{\beta}$ and $\widehat{\gamma}$ can be simplified. For example, if $S^{-}_{\beta\gamma} = S^{+}_{\beta\gamma} = S_{\beta\gamma}$, then $R_1 = R_2 = R_3 = \mathbb{R}^{2d+2}$ when $\lambda_- \le \lambda_+$ and $\emptyset$ otherwise. As a result,

$$\sqrt{n}\left(\widehat{\beta}-\beta_0\right) \xrightarrow{d} \begin{cases} \left(S_{\beta\beta}-S_{\beta\gamma}\left(S^{-}_{\gamma\gamma}\right)^{-1}S_{\gamma\beta}\right)^{-1}W, \text{ if } \lambda_-\le\lambda_+, \\ \left(S_{\beta\beta}-S_{\beta\gamma}\left(S^{+}_{\gamma\gamma}\right)^{-1}S_{\gamma\beta}\right)^{-1}W, \text{ if } \lambda_->\lambda_+, \end{cases}$$

and

$$\sqrt{n}\left(\widehat{\gamma}-\gamma_0\right) \xrightarrow{d} \begin{cases} -\left(S^{-}_{\gamma\gamma}\right)^{-1}S_{\gamma\beta}\left(S_{\beta\beta}-S_{\beta\gamma}\left(S^{-}_{\gamma\gamma}\right)^{-1}S_{\gamma\beta}\right)^{-1}W, \text{ if } \lambda_-\le\lambda_+, \\ -\left(S^{+}_{\gamma\gamma}\right)^{-1}S_{\gamma\beta}\left(S_{\beta\beta}-S_{\beta\gamma}\left(S^{+}_{\gamma\gamma}\right)^{-1}S_{\gamma\beta}\right)^{-1}W, \text{ if } \lambda_->\lambda_+. \end{cases}$$

When $\lambda_- = \lambda_+ = \lambda$ which implies $S^{-}_{\gamma\gamma} = S^{+}_{\gamma\gamma} = 2\lambda =: S_{\gamma\gamma}$, the formulae can be further simplified and are the same as the case where $S(\theta)$ is the second-order differentiable:

$$\begin{aligned} \sqrt{n}\left(\widehat{\beta}-\beta_0\right) &\xrightarrow{d} \left(S_{\beta\beta}-\tfrac{S_{\beta\gamma}S_{\gamma\beta}}{S_{\gamma\gamma}}\right)^{-1}W =: Z_\beta, \\ \sqrt{n}\left(\widehat{\gamma}-\gamma_0\right) &\xrightarrow{d} -\tfrac{S_{\gamma\beta}}{S_{\gamma\gamma}}\left(S_{\beta\beta}-\tfrac{S_{\beta\gamma}S_{\gamma\beta}}{S_{\gamma\gamma}}\right)^{-1}W = -\tfrac{S_{\beta\gamma}}{S_{\gamma\gamma}}Z_\beta, \end{aligned} \tag{27}$$

i.e., $\widehat{\beta}$ is asymptotically normal and $\widehat{\gamma}-\gamma_0$ is a linear transformation of $\widehat{\beta}-\beta_0$ asymptotically.

Compared with I(2), $\omega_0^{\pm}$ (and $\varpi_\pm$) in Theorem 3 equals zero, so $n^{1/3}\left(\widehat{\theta}-\theta_0\right) = o_p(1)$; in other words,

$\widehat{\theta}$ has a faster convergence rate than $n^{1/3}$. From Theorem 5, the convergence rate of $\widehat{\theta}$ is actually $n^{1/2}$. Also, different from the asymptotic distribution in Theorem 1, $\widehat{\gamma}$ and $\widehat{\beta}$ are not asymptotically independent; actually, even $\widehat{\beta}_1$ and $\widehat{\beta}_2$ are not asymptotically independent. The asymptotic distribution of $\widehat{\gamma}$ is completely determined by that of $\widehat{\beta}$, so the asymptotic distribution of $\sqrt{n}\left(\widehat{\theta}-\theta_0\right)$ concentrates on a hyperplane with dimension $2d+2$. This asymptotic distribution is also different from that in CT and Hansen (2017) where although $\widehat{\gamma}$ and $\widehat{\beta}$ are asymptotically jointly normal and not asymptotically independent, the asymptotic distribution of $\widehat{\gamma}$ is not fully determined by $\widehat{\beta}$.[12] Another way to see why the randomness related to $\widehat{\gamma}$ disappears is to check the localized objective function in the direction of γ, $\sum_{i=1}^{n}\overline{z}_{1i}1\left(\gamma_0+\frac{v}{\sqrt{n}}<q_i\leq\gamma_0\right)+\sum_{i=1}^{n}\overline{z}_{2i}1\left(\gamma_0<q_i\leq\gamma_0+\frac{v}{\sqrt{n}}\right)$, whose variance goes to zero in CTR, i.e., the $O\left(n^{-1/2}\right)$ neighborhood is too small to accumulate randomness for γ.

Finally, consider the LR-inference on γ. Define

$$LR_n(\gamma)=2n\left(S_n(\gamma)-S_n\left(\widehat{\gamma}\right)\right),$$

which takes the same form as the LR statistic in the regular model.

Corollary 4 *Under Assumption II(2),*

$$LR_n(\gamma_0)\xrightarrow{d}\max\left\{\frac{W'S_{\beta\beta}^{-1}S_{\beta\gamma}^{-}S_{\gamma\beta}^{-}S_{\beta\beta}^{-1}W}{S_{\gamma\gamma}^{-}-S_{\gamma\beta}^{-}S_{\beta\beta}^{-1}S_{\beta\gamma}^{-}},\frac{W'S_{\beta\beta}^{-1}S_{\beta\gamma}^{+}S_{\gamma\beta}^{+}S_{\beta\beta}^{-1}W}{S_{\gamma\gamma}^{+}-S_{\gamma\beta}^{+}S_{\beta\beta}^{-1}S_{\beta\gamma}^{+}}\right\}.$$

Note that

$$\begin{aligned}&\frac{W'S_{\beta\beta}^{-1}S_{\beta\gamma}^{-}S_{\gamma\beta}^{-}S_{\beta\beta}^{-1}W}{S_{\gamma\gamma}^{-}-S_{\gamma\beta}^{-}S_{\beta\beta}^{-1}S_{\beta\gamma}^{-}}=W'S_{\beta\beta}^{-1}S_{\beta\gamma}^{-}\left(S_{\gamma\gamma}^{-}-S_{\gamma\beta}^{-}S_{\beta\beta}^{-1}S_{\beta\gamma}^{-}\right)^{-1}S_{\gamma\beta}^{-}S_{\beta\beta}^{-1}W\\&=Z'\mathrm{diag}\left\{\left(\Sigma_0\right)^{1/2},\left(\overline{\Sigma}_0\right)^{1/2}\right\}S_{\beta\beta}^{-1}S_{\beta\gamma}^{-}\left(S_{\gamma\gamma}^{-}-S_{\gamma\beta}^{-}S_{\beta\beta}^{-1}S_{\beta\gamma}^{-}\right)^{-1}S_{\gamma\beta}^{-}S_{\beta\beta}^{-1}\mathrm{diag}\left\{\left(\Sigma_0\right)^{1/2},\left(\overline{\Sigma}_0\right)^{1/2}\right\}Z=:Z'\Omega^{-}Z,\end{aligned}$$

where $Z=\mathrm{diag}\left\{\left(\Sigma_0\right)^{-1/2},\left(\overline{\Sigma}_0\right)^{-1/2}\right\}W\sim N\left(\mathbf{0},I_{2d+2}\right)$. Since $\Omega^{-}\geq 0$ and has rank 1, we can decompose it as $H^{-\prime}\Pi^{-}H^{-}$ for an orthogonal matrix H^{-} and a diagonal matrix $\Pi^{-}=\mathrm{diag}\{\pi^{-},0,\cdots,0\}$. As a result, $\frac{W'S_{\beta\beta}^{-1}S_{\beta\gamma}^{-}S_{\gamma\beta}^{-}S_{\beta\beta}^{-1}W}{S_{\gamma\gamma}^{-}-S_{\gamma\beta}^{-}S_{\beta\beta}^{-1}S_{\beta\gamma}^{-}}=\pi^{-}z_1^{-2}$ follows a scaled χ_1^2 distribution, where $\left(z_1^{-},\cdots,z_{2d+2}^{-}\right)'=H^{-}Z\sim N\left(\mathbf{0},I_{2d+2}\right)$. In summary,

$$LR_n(\gamma_0)\xrightarrow{d}\max\left\{\pi^{-}z_1^{-2},\pi^{+}z_1^{+2}\right\}=:\max\left\{\pi^{-}\chi_1^{-2},\pi^{+}\chi_1^{+2}\right\},$$

where π^{+} and z_j^{+} are parallelly defined as π^{-} and z_j^{-}, and $\mathbb{E}\left[\left(z_1^{-},\cdots,z_{2d+2}^{-}\right)'\left(z_1^{+},\cdots,z_{2d+2}^{+}\right)\right]=\mathbb{E}\left[H^{-}ZZ'H^{+\prime}\right]=H^{-}H^{+\prime}$ whose $(1,1)$ element need not be 1, i.e., the two χ_1^2 distributions need not be the same. If $S_{\beta\gamma}^{+}=S_{\beta\gamma}^{-}=S_{\beta\gamma}$, then

$$LR_n(\gamma_0)\xrightarrow{d}\frac{W'S_{\beta\beta}^{-1}S_{\beta\gamma}S_{\gamma\beta}S_{\beta\beta}^{-1}W}{2\min\{\lambda_{-},\lambda_{+}\}-S_{\gamma\beta}S_{\beta\beta}^{-1}S_{\beta\gamma}}=\left(\pi^{-}\vee\pi^{+}\right)\chi_1^2,$$

where note that $\chi_1^{-2}=\chi_1^{+2}=:\chi_1^2$, and $\pi^{\pm}$ is the nonzero eigenvalue of Ω divided by $\left(2\lambda^{\pm}-S_{\gamma\beta}S_{\beta\beta}^{-1}S_{\beta\gamma}\right)$, where $\Omega=\mathrm{diag}\left\{\left(\Sigma_0\right)^{1/2},\left(\overline{\Sigma}_0\right)^{1/2}\right\}S_{\beta\beta}^{-1}S_{\beta\gamma}S_{\gamma\beta}S_{\beta\beta}^{-1}\mathrm{diag}\left\{\left(\Sigma_0\right)^{1/2},\left(\overline{\Sigma}_0\right)^{1/2}\right\}$. When $\lambda_{-}=\lambda_{+}=\lambda$, $\pi^{-}=\pi^{+}=:\pi$, and $LR_n(\gamma_0)\xrightarrow{d}\pi\chi_1^2$ which is close to the asymptotic distribution of the standard LR test.

[12] Note that in CT and Hansen (2017), the regressors are $\left(1,x,(q-\gamma)_{\ominus},(q-\gamma)_{\oplus}\right)$, and $\beta=\left(\beta_{1c}+\beta_{1q}\gamma,\beta_x',\beta_{1q},\beta_{2q}\right)'\in\mathbb{R}^{d+2}$ since d restrictions $\delta_{x0}=0$ and $\delta_{c0}+\delta_{q0}\gamma_0=0$ are imposed on the model.

6.2 $\alpha = 3$

First, we specify the required assumptions.

Assumption II(3): same as Assumption II(2) except

(iv) (iv) of Assumption MA plus (c) $\sup_{\gamma\in\mathcal{N}} \mathbb{E}\left[\left|\left(y_i - \mathbf{x}_i'\overline{\beta}_0\right)\right|^{2+\epsilon} |q_i = \gamma\right] < \infty$ for some $\epsilon > 0$.

(x) (a) $\Lambda_{\pm}(\gamma) \in RV_3$; (b) (25) holds with $\alpha = 3$; (c) $\omega_\gamma^{\pm} := \mathbb{E}\left[\left(y - \mathbf{x}'\overline{\beta}_0\right)^2 |q = \gamma\pm\right]$ is continuous at γ_0 and $\omega_0^{\pm} := \omega_{\gamma_0}^{\pm} > 0$.

Assumption (iv)(c) is assumed due to a similar reason as in Assumption I(2) but here $\mathbf{x}_i'\delta_0\left(y_i - \mathbf{x}_i'\overline{\beta}_0\right)$ is replaced by $\left(y_i - \mathbf{x}_i'\overline{\beta}_0\right)$ as $\mathbf{x}_i'\delta_0 = q_i\delta_{10}$. Actually, Assumption (iv)(c) is implied by Assumption (iv)(b), but we state it here for comparison with Assumption I(2). Correspondingly, $\omega_\gamma^{\pm}$ in Assumption I(2) takes the new form. Parallel to (19),

$$\begin{aligned}\omega_0^- &= \mathbb{E}\left[\left(y - \mathbf{x}'\overline{\beta}_0\right)^2 |q = \gamma_0-\right] = \mathbb{E}\left[\left(m_1(x,q) - \mathbf{x}'\overline{\beta}_0\right)^2 |q = \gamma_0-\right] + \mathbb{E}\left[\varepsilon_1^2 | q = \gamma_0-\right],\\ \omega_0^+ &= \mathbb{E}\left[\left(y - \mathbf{x}'\overline{\beta}_0\right)^2 |q = \gamma_0+\right] = \mathbb{E}\left[\left(m_2(x,q) - \mathbf{x}'\overline{\beta}_0\right)^2 |q = \gamma_0+\right] + \mathbb{E}\left[\varepsilon_2^2 | q = \gamma_0+\right].\end{aligned}$$

Actually, $\omega_0^{\pm}$ are still from the variances of $\overline{z}_{1i}$ and $\overline{z}_{2i}$ in the neighborhood of $q = \gamma_0$ but caution is taken since $\left(y - \mathbf{x}'\overline{\beta}_0\right)(\mathbf{x}'\delta_0) = \left(y - \mathbf{x}'\overline{\beta}_0\right)(q\delta_{q0})$ now. In the simple example of Section 3, $\Lambda_{\pm}(\gamma) \in RV_3$ implies

$$\mathbb{E}\left[\left(m_1(x_i,q_i) - \mathbf{x}_i'\overline{\beta}_0\right)^2 1(\gamma < q_i \le \gamma_0)\right] \in RV_3 \text{ and } \mathbb{E}\left[\left(m_2(x_i,q_i) - \mathbf{x}_i'\overline{\beta}_0\right)^2 1(\gamma_0 < q_i \le \gamma)\right] \in RV_3, \quad (28)$$

so

$$\omega_0^- = \mathbb{E}\left[\varepsilon_1^2 | q = \gamma_0-\right] \text{ and } \omega_0^+ = \mathbb{E}\left[\varepsilon_2^2 | q = \gamma_0+\right].$$

In general, (28) does not hold and the simplification will not happen. HLS is a special case of II(3) with $\Lambda_{\pm}(\gamma)$ taking the special form (4), and $\omega_0^{\pm} = \mathbb{E}\left[\varepsilon^2 | q = \gamma_0\right]$, where the simplification of $\omega_0^{\pm}$ can be seen from the facts that

$$\mathbb{E}\left[\left(m_1(x,q) - \mathbf{x}'\overline{\beta}_0\right)^2 |q = \gamma_0-\right] = \mathbb{E}\left[\left(\tfrac{q\delta_{q0}}{2}\right)^2 |q = 0-\right] = 0$$

and similarly $\mathbb{E}\left[\left(m_2(x,q) - \mathbf{x}'\overline{\beta}_0\right)^2 |q = \gamma_0+\right] = 0$, and $\varepsilon_1 = \varepsilon_2 = \varepsilon$ is assumed. Another form of $\omega_0^{\pm}$ is

$$\begin{aligned}\omega_0^+ &= \mathbb{E}\left[\left(y - \mathbf{x}'\beta_{10} + \mathbf{x}'\delta_0/2\right)^2 |q = \gamma_0-\right] = \mathbb{E}\left[e_1^2 | q = \gamma_0-\right],\\ \omega_0^+ &= \mathbb{E}\left[\left(y - \mathbf{x}'\beta_{20} - \mathbf{x}'\delta_0/2\right)^2 |q = \gamma_0-\right] = \mathbb{E}\left[e_2^2 | q = \gamma_0+\right].\end{aligned}$$

Before stating the asymptotic distribution of $\widehat{\theta}$, define $\mu_{\pm} = 2\lambda_{\pm}$ and $\varpi_{\pm} = \frac{f_0\delta_{q0}^2\omega_0^{\pm}}{3}$.

Theorem 6 *Under Assumption II(3),*

$$n^{1/3}\left(\widehat{\gamma} - \gamma_0\right) \xrightarrow{d} \left[\omega\zeta(\varphi,\phi;1)\right]^{1/3} =: Z_\gamma(3),$$

and

$$\begin{aligned}n^{1/2}\left(\widehat{\beta}_1 - \beta_{10}\right) &\xrightarrow{d} Z_{\beta_1},\\ n^{1/2}\left(\widehat{\beta}_2 - \beta_{20}\right) &\xrightarrow{d} Z_{\beta_2},\end{aligned}$$

where $\omega = \frac{\varpi_-}{\mu_-^2} = \frac{f_0\delta_{q0}^2\omega_0^{\pm}}{12\lambda_-^2}$, $\varphi = \frac{\mu_+}{\mu_-} = \frac{\lambda_+}{\lambda_-}$ *and* $\phi = \frac{\varpi_+}{\varpi_-} = \frac{\omega_0^+}{\omega_0^-}$, *and* $Z_\gamma(3)$, Z_{β_1} *and* Z_{β_2} *are independent.*

As noticed in footnote 2 of HLS, the asymptotic distribution in Gonzalo and Wolf (2005)'s Theorem A.1 and Remark A.1 are not properly developed because the Hessian matrix of $S(\theta)$ is degenerate. Interestingly, the constraints $\delta_{x0} = \mathbf{0}$ and $\delta_{c0} + \delta_{q0}\gamma_0 = 0$ imposed in CTR help to improve the convergence rates of $\widehat{\gamma}$; a similar phenomenon appears in I(2). Although $\widehat{\gamma}$ is $n^{1/3}$-consistent in both BM and HLS, the reasons for the cube-root rate are different because BM is a special I(2) and HLS is a special II(3) and the balancings for the convergence rate of $\widehat{\gamma}$ are different. Also, $\widehat{\gamma}$ and $\widehat{\beta}$ are perfectly correlated asymptotically in BM, while they are asymptotically independent in HLS; $\widehat{\beta}$ even has different convergence rates and asymptotic distributions in BM and HLS. Note that we do not need the model to be CS to achieve the cube-root rate. Note also that the intuition in (23) can still be applied here. But now $n^{1/3}\left(\widehat{\beta}_1(\widehat{\gamma}) - \widehat{\beta}(\gamma_0)\right) \approx -M_0^{-1}S_{\beta_1\gamma}n^{1/3}(\widehat{\gamma} - \gamma_0) = \mathbf{0}$ under Assumption (x)(b), so $\widehat{\beta}_1(\widehat{\gamma}) - \widehat{\beta}(\gamma_0) = o_p\left(n^{-1/3}\right)$. Actually, Theorem 6 shows that $\widehat{\beta}_1(\widehat{\gamma}) - \widehat{\beta}(\gamma_0) = o_p\left(n^{-1/2}\right)$ because $\widehat{\beta}_1 - \beta_{10}$ and $\widehat{\beta}_1(\gamma_0) - \beta_{10}$ have the same asymptotic distribution. The same arguments apply to $\widehat{\beta}_2$.

Finally, consider the LR-inference on γ. Define

$$LR_n(\gamma) = \frac{n\left(S_n(\gamma) - S_n(\widehat{\gamma})\right)}{\widehat{\eta}^2},$$

where $\widehat{\eta}^2$ is a consistent estimator of $\eta^2 = \frac{\varpi_-}{\mu_-} = \frac{f_0\delta_{q0}^2\omega_0^-}{6\lambda_-}$ which reduces to ω_0^- in HLS.

Corollary 5 *Under Assumption II(3),*

$$LR_n(\gamma_0) \xrightarrow{d} \xi(\varphi, \phi; 1),$$

where the distribution of $\xi(\varphi, \phi; 1)$ *is given in Proposition 2(iii) with* φ *and* ϕ *defined in Theorem 6.*

6.3 $\alpha = 4$

This case is not discussed in the literature. First, we specify the required assumptions.

Assumption II(4): same as Assumption II(3) except

(vii) (a) $f(\gamma)$ is differentiable at γ_0, and $0 < \underline{f} \le f_0 \le \overline{f} < \infty$; (b) $\mathbb{E}[\mathbf{x}|q = \gamma]$ is differentiable at γ_0.

(x) (a) $\Lambda_\pm(\gamma) \in RV_4$; (b) (25) holds with $\alpha = 4$; (c) $\omega_\gamma^\pm := \mathbb{E}\left[\left(y - \mathbf{x}'\overline{\beta}_0\right)^2 | q = \gamma\pm\right]$ is continuous at γ_0 and $\omega_0^\pm := \omega_{\gamma_0}^\pm > 0$; (d)

$$\mathbb{S}^- := \begin{pmatrix} 2\lambda_- & \frac{1}{2}S'_{\beta_1\gamma^2} & \frac{1}{2}S'_{\beta_2\gamma^2} \\ \frac{1}{2}S^-_{\beta_1\gamma^2} & S_{\beta_1\beta_1} & \mathbf{0} \\ \frac{1}{2}S^-_{\beta_2\gamma^2} & \mathbf{0} & S_{\beta_2\beta_2} \end{pmatrix} =: \begin{pmatrix} 2\lambda_- & \frac{1}{2}S'_{\beta\gamma^2} \\ \frac{1}{2}S_{\beta\gamma^2} & S_{\beta\beta} \end{pmatrix} > 0,$$

$$\mathbb{S}^+ := \begin{pmatrix} 2\lambda_+ & \frac{1}{2}S^{+\prime}_{\beta_1\gamma^2} & \frac{1}{2}S^{+\prime}_{\beta_2\gamma^2} \\ \frac{1}{2}S^+_{\beta_1\gamma^2} & S_{\beta_1\beta_1} & \mathbf{0} \\ \frac{1}{2}S^+_{\beta_2\gamma^2} & \mathbf{0} & S_{\beta_2\beta_2} \end{pmatrix} =: \begin{pmatrix} 2\lambda_+ & \frac{1}{2}S'_{\beta\gamma^2} \\ \frac{1}{2}S_{\beta\gamma^2} & S_{\beta\beta} \end{pmatrix} > 0,$$

where $S^\pm_{\beta_1\gamma^2} = S^\pm_{\beta_2\gamma^2} = \frac{\delta_{q0}f_0}{2}\mathbb{E}[\mathbf{x}|q = \gamma_0] =: S_{\beta_\ell\gamma^2}$, and $S_{\beta\gamma^2} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \otimes S_{\beta_\ell\gamma^2}$.

Only Assumption (x)(d) needs some explanations. $\mathbb{S}^{\pm}$ is parallel to $S^{\pm}_{\theta\theta}$ only with γ^2 replacing γ. From Assumption (x)(b), only the second terms of (26) are involved in calculating $S^{\pm}_{\beta_\ell\gamma^2}$. Take S^{+}_{β} as an example. We need to calculate

$$\frac{\partial^2\mathbb{E}\left[\mathbf{x}q\delta_{q0}1(\gamma_0<q\leq\gamma)\right]}{\partial\gamma^2}=\delta_{q0}\frac{\partial^2\int_0^\gamma\mathbb{E}\left[\mathbf{x}|q\right]qf(q)dq}{\partial\gamma^2}=\delta_{q0}\frac{\partial\mathbb{E}\left[\mathbf{x}|q=\gamma\right]\gamma f(\gamma)}{\partial\gamma}=\delta_{q0}f_0\mathbb{E}\left[\mathbf{x}|q=\gamma_0\right],$$

where the last equality uses Assumption (vii). Parallel to Example 3, the following example shows that $\mathbb{S}^{\pm}>0$ imposes some restrictions on $\lambda_{\pm}$.

Example 5 *When* $\mathbf{x}=(1,q)'$, $q\sim U\left[-0.5,0.5\right]$, $\gamma_0=0$, *and* $\delta_0=(\delta_{c0},\delta_{q0})'$. *Then*

$$\mathbb{S}^{\pm}=\begin{pmatrix} 2\lambda_{\pm} & \frac{\delta_{q0}}{2} & 0 & \frac{\delta_{q0}}{2} & 0 \\ \frac{\delta_{q0}}{2} & \frac{1}{2} & -\frac{1}{8} & & \\ 0 & -\frac{1}{8} & \frac{1}{24} & \mathbf{0} & \\ \frac{\delta_{q0}}{2} & & & \frac{1}{2} & \frac{1}{8} \\ 0 & \mathbf{0} & & \frac{1}{8} & \frac{1}{24} \end{pmatrix}>0$$

implies $\lambda_{\pm}>2\delta^2_{q0}$. *Note that since* $\delta_{q0}\neq0$, $S_{\beta_\ell\gamma^2}=\frac{\delta_{q0}f_0}{2}\mathbb{E}\left[\mathbf{x}|q=\gamma_0\right]\neq0$, *which implies the restriction on* $\lambda_{\pm}$.

Before stating the asymptotic distribution of $\widehat{\theta}$, define $\mu_{\pm}=2\lambda_{\pm}-S_{\gamma^2\beta_\ell}\left(M_0^{-1}+\overline{M}_0^{-1}\right)S_{\beta_\ell\gamma^2}/4$ and $\varpi_{\pm}=\frac{f_0\delta^2_{q0}\omega^{\pm}_0}{3}$, which play similar roles and take similar forms as $\mu_{\pm}$ and $\varpi_{\pm}$ in I(2).

Theorem 7 *Under Assumption II*(4),

$$n^{1/5}\left(\widehat{\gamma}-\gamma_0\right)\xrightarrow{d}\omega^{1/5}\zeta\left(\varphi,\phi;4/3\right)^{1/3}=:Z_\gamma\left(4\right)$$

and

$$\begin{pmatrix} n^{2/5}\left(\widehat{\beta}_1-\beta_{10}\right) \\ n^{2/5}\left(\widehat{\beta}_2-\beta_{20}\right)\end{pmatrix}\xrightarrow{d}\begin{pmatrix} -\frac{1}{2}M_0^{-1}S_{\beta_\ell\gamma^2}Z_\gamma\left(4\right)^2 \\ -\frac{1}{2}\overline{M}_0^{-1}S_{\beta_\ell\gamma^2}Z_\gamma\left(4\right)^2\end{pmatrix},$$

where $\omega=\frac{\varpi_-}{\mu_-^2}$, $\varphi=\frac{\mu_+}{\mu_-}$ *and* $\phi=\frac{\varpi_+}{\varpi_-}=\frac{\omega^+_0}{\omega^-_0}$.

When $\lambda_-=\lambda_+$, $\mu_-=\mu_+$ so $\varphi=1$; if $\omega^-_0=\omega^+_0$, then $\phi=1$; and $\zeta\left(\varphi,\phi;4/3\right)$ will reduces to $\zeta\left(1,1;4/3\right)$. As in I(2), the asymptotic distribution of $\widehat{\beta}$ is completely determined by $\widehat{\gamma}$; actually, the asymptotic distribution of $\widehat{\theta}$ concentrates on a quadratic line through the origin. The intuition in (23) can still be applied here, but now $n^{1/5}\left(\widehat{\beta}_1\left(\widehat{\gamma}\right)-\widehat{\beta}\left(\gamma_0\right)\right)\approx-M_0^{-1}S_{\beta_1\gamma}n^{1/5}\left(\widehat{\gamma}-\gamma_0\right)=\mathbf{0}$. To get a nondegenerate distribution for $\widehat{\beta}$, we need to expand $\widehat{\beta}\left(\widehat{\gamma}\right)$ around γ_0 to the second order. It turns out that

$$n^{2/5}\left(\widehat{\beta}_1\left(\widehat{\gamma}\right)-\widehat{\beta}\left(\gamma_0\right)\right)\approx-\frac{1}{2}M_0^{-1}S_{\beta_\ell\gamma^2}n^{2/5}\left(\widehat{\gamma}-\gamma_0\right)^2, \tag{29}$$

which results in the asymptotic distribution in Theorem 7. The same arguments apply to $\widehat{\beta}_2$.

Finally, consider the LR-inference on γ. Define

$$LR_n\left(\gamma\right)=\frac{n^{4/5}\left(S_n(\gamma)-S_n\left(\widehat{\gamma}\right)\right)}{\widehat{\eta}^{6/5}},$$

where $\widehat{\eta}^2$ is a consistent estimator of $\eta^2=\frac{\varpi_-^{4/3}}{\mu_-}$.

Corollary 6 *Under Assumption II(4),*

$$LR_n(\gamma_0) \xrightarrow{d} \xi(\varphi, \phi; 4/3),$$

where $\xi(\varphi, \phi; 4/3)$ is defined in Proposition 2(ii) with φ and ϕ defined in Theorem 7.

6.4 $3 < \alpha < 4$

From the intuition of Section 3.3, we expect the convergence rate of $\widehat{\beta}$ to be $\min\left(n^{1/2}, \rho_n^2\right)$; when $\Lambda(|\gamma|)$ takes the form of $|\gamma|$'s power,

$$\min\left(n^{1/2}, \rho_n\right) = \begin{cases} n^{1/2}, & \text{if } 3 < \alpha \leq \frac{7}{2}, \\ n^{\frac{2}{2\alpha-3}}, & \text{if } \frac{7}{2} < \alpha < 4, \end{cases}$$

which is faster than the usual balancing rate $n^{\frac{\alpha}{2(2\alpha-3)}}$. This rate can also be seen through a similar analysis as in (29). When $\alpha < \frac{7}{2}$, the convergence rate of $\widehat{\beta}(\widehat{\gamma}) - \widehat{\beta}(\gamma_0)$ is determined by $(\widehat{\gamma} - \gamma_0)^2$ whose convergence rate is faster than $n^{1/2}$, so the asymptotic distribution of $\widehat{\beta}$ is completely determined by $\widehat{\beta}(\gamma_0) - \beta_0$. When $\alpha > \frac{7}{2}$, the converse happens, and the asymptotic distribution of $\widehat{\beta}$ is just a quadratic transformation of that of $\widehat{\gamma}$ as indicated in (29). Only when $\alpha = \frac{7}{2}$, both $\widehat{\beta}(\widehat{\gamma}) - \widehat{\beta}(\gamma_0)$ and $\widehat{\beta}(\gamma_0) - \beta_0$ will contribute to the asymptotic distribution of $\widehat{\beta}$.

We next specify the required assumptions.

Assumption II(α) $[3 < \alpha < 4]$: same as Assumption II(4) except

(x) (a) $\Lambda_\pm(\gamma) \in RV_\alpha$; (b) (25) holds; (c) $\omega_\gamma^\pm := \mathbb{E}\left[\left(y - \mathbf{x}'\overline{\beta}_0\right)^2 | q = \gamma\pm\right]$ is continuous at γ_0 and $\omega_0^\pm := \omega_{\gamma_0}^\pm > 0$.

As in (28), we can assume

$$\mathbb{E}\left[\left(m_1(x,q) - \mathbf{x}'\overline{\beta}_0\right)^2 1(\gamma < q \leq \gamma_0)\right] \in RV_{2\alpha-3} \text{ and } \mathbb{E}\left[\left(m_2(x,q) - \mathbf{x}'\overline{\beta}_0\right)^2 1(\gamma_0 < q \leq \gamma)\right] \in RV_{2\alpha-3} \tag{30}$$

to simplify $\omega_0^\pm$, but we will keep this general form of $\omega_0^\pm$ here.

Before stating the asymptotic distribution of $\widehat{\theta}$, define $\mu_\pm = 2\lambda_\pm$ and $\varpi_\pm = \frac{f_0 \delta_{q0}^2 \omega_0^\pm}{3}$, which are the same as in II(3).

Theorem 8 *Under Assumption II(α), $3 < \alpha < 4$,*

$$\rho_n(\widehat{\gamma} - \gamma_0) \xrightarrow{d} \omega^{\frac{1}{2\alpha-3}} \zeta(\varphi, \phi; \alpha/3)^{1/3} =: Z_\gamma(\alpha),$$

where $\omega = \frac{\varpi_-}{4\lambda_-^2} = \frac{f_0 \delta_{q0}^2 \omega_0^-}{12\lambda_-^2}$, $\varphi = \frac{\mu_+}{\mu_-} = \frac{\lambda_+}{\lambda_-}$ and $\phi = \frac{\varpi_+}{\varpi_-} = \frac{\omega_0^+}{\omega_0^-}$, when $3 < \alpha < 3.5$,

$$\sqrt{n}\left(\widehat{\beta}_1 - \beta_{10}\right) \xrightarrow{d} Z_{\beta_1},$$
$$\sqrt{n}\left(\widehat{\beta}_2 - \beta_{20}\right) \xrightarrow{d} Z_{\beta_2},$$

when $3.5 < \alpha < 4$,

$$\rho_n^2\left(\widehat{\beta}_1 - \beta_{10}\right) \xrightarrow{d} -\frac{1}{2}M_0^{-1}S_{\beta_\ell\gamma^2}Z_\gamma(\alpha)^2,$$
$$\rho_n^2\left(\widehat{\beta}_2 - \beta_{20}\right) \xrightarrow{d} -\frac{1}{2}\overline{M}_0^{-1}S_{\beta_\ell\gamma^2}Z_\gamma(\alpha)^2,$$

and when $\alpha = 3.5$,

$$\sqrt{n}\left(\widehat{\beta}_1 - \beta_{10}\right) \xrightarrow{d} Z_{\beta_1} - \frac{1}{2}M_0^{-1}S_{\beta_\ell\gamma^2}Z_\gamma(3.5)^2,$$
$$\sqrt{n}\left(\widehat{\beta}_2 - \beta_{20}\right) \xrightarrow{d} Z_{\beta_2} - \frac{1}{2}\overline{M}_0^{-1}S_{\beta_\ell\gamma^2}Z_\gamma(3.5)^2,$$

where Z_{β_1}, Z_{β_2} *and* $Z_\gamma(\alpha)$ *are independent.*

Comparing Theorems 6, 7 and 8, we can see that the asymptotic distributions of $\widehat{\gamma}$ take a unified form except in II(4) where $\mu_\pm$ includes some extra term. These extra cross terms are dominated in II(α) with $3 \leq \alpha < 4$ as shown in Section 3.1. As expected, when $3 < \alpha < 3.5$, the asymptotic distribution of $\widehat{\beta}$ is not affected by $\widehat{\gamma}$ and is exactly the same as in case II(3), which is similar to $\widehat{\beta}$ in case I(α) with $1 < \alpha < 1.5$. When $3.5 < \alpha < 4$, it is completely determined by $\widehat{\gamma}$ and takes the same form as in II(4), which is similar to $\widehat{\beta}$ in case I(α) with $1.5 < \alpha < 2$, but takes a quadratic instead of linear form of $Z_\gamma(\alpha)$. When $\alpha = 3.5$, it is the sum of both components, which is similar to $\widehat{\beta}$ in I(1.5).

Finally, consider the LR-inference on γ. Define

$$LR_n(\gamma) = \frac{\sqrt{n\rho_n^3}\left(S_n(\gamma) - S_n(\widehat{\gamma})\right)}{\widehat{\eta}^{\frac{6}{2\alpha-3}}},$$

where $\widehat{\eta}^2$ is a consistent estimator of $\eta^2 = \frac{\varpi_-^{\alpha/3}}{2\lambda_-}$. Note that $\sqrt{n\rho_n^3} \prec n$.

Corollary 7 *Under Assumption II(α),* $3 < \alpha < 4$,

$$LR_n(\gamma_0) \xrightarrow{d} \xi(\varphi, \phi; \alpha/3),$$

where $\xi(\varphi, \phi; \alpha/3)$ *is defined in Proposition 2(ii) with* φ *and* ϕ *defined in Theorem 8.*

The form of $LR_n(\gamma)$ and the asymptotic distribution $LR_n(\gamma_0)$ take unified forms when $3 \leq \alpha \leq 4$.

6.5 $2 < \alpha < 3$

From the analyses in previous sections, we can see that II(α) with $3 \leq \alpha \leq 4$ are parallel to I(α) with $1 \leq \alpha \leq 2$ in some sense, but II(α) with $2 \leq \alpha < 3$ are new. From the intuitions in Section 3.3 and in I(α) with $1 < \alpha < 2$, we expect the convergence rate of $\widehat{\gamma}$ to be $\min\left(n^{1/2}, \varrho_n\right)$, where ϱ_n is determined from $\sqrt{n\rho_n^3}\Lambda\left(\rho_n^{-1}\right) = 1$ and is the convergence rate of $\widehat{\gamma}(\beta_0) - \gamma_0$; when $\Lambda(|\gamma|)$ takes the form of $|\gamma|$'s power,

$$\min\left(n^{1/2}, \varrho_n\right) = \begin{cases} n^{1/2}, & \text{if } 2 < \alpha \leq \frac{5}{2}, \\ n^{\frac{1}{2\alpha-3}}, & \text{if } \frac{5}{2} < \alpha < 3, \end{cases}$$

which is faster than the usual balancing rate $n^{\frac{1}{\alpha}}$.

We now specify the required assumptions.

Assumption II(α) [$2 < \alpha < 3$]: same as Assumption II(3) except

(iv) (iv) of Assumption MA plus (c) $\sup_{\gamma\in\mathcal{N}} \mathbb{E}\left[\left|\left(y - \mathbf{x}'\overline{\beta}_0\right)\right|^{2+\epsilon} |q = \gamma\right] < \infty$ for some $\epsilon > 0$ when $2.5 < \alpha < 3$ and $\sup_{\gamma\in\mathcal{N}} \mathbb{E}\left[\left\|\left(y - \mathbf{x}'\overline{\beta}_0\right)\mathbf{x}\right\|^{2+\epsilon} |q = \gamma\right] < \infty$ for some $\epsilon > 0$ when $\alpha = 2.5$.

(x) (a) $\Lambda_{\pm}(\gamma) \in RV_\alpha$; (b) (25) holds, and $\lim_{|v|\downarrow 0} \frac{S^{\pm}_{\beta}(v)}{|v|^{\alpha-1}L(v)} = \psi_{\pm}$; (c) when $2.5 \leq \alpha < 3$, $\omega^{\pm}_{\gamma} := \mathbb{E}\left[\left(y - \mathbf{x}'\overline{\beta}_0\right)^2 |q = \gamma\pm\right]$ is continuous at γ_0 and $\omega^{\pm}_0 := \omega^{\pm}_{\gamma_0} > 0$, and when $\alpha = 2.5$, $\Omega^{\pm}_{\gamma} := \mathbb{E}\left[\left(y - \mathbf{x}'\overline{\beta}_0\right)^2 \mathbf{x}\mathbf{x}' |q = \gamma\pm\right]$ and $\Upsilon^{\pm}_{\gamma} := \mathbb{E}\left[\left(y - \mathbf{x}'\overline{\beta}_0\right)^2 \mathbf{x} |q = \gamma\pm\right]$ are continuous at γ_0 and $\Omega^{\pm}_0 := \Omega^{\pm}_{\gamma_0} > 0$.

In Assumption (iv)(c), $\sup_{\gamma\in\mathcal{N}} \mathbb{E}\left[\left\|\left(y - \mathbf{x}'\overline{\beta}_0\right)\mathbf{x}\right\|^{2+\epsilon} |q = \gamma\right] < \infty$ implies $\sup_{\gamma\in\mathcal{N}} \mathbb{E}\left[\left|\left(y - \mathbf{x}'\overline{\beta}_0\right)\right|^{2+\epsilon} |q = \gamma\right] < \infty$ because 1 is the first element of $\mathbf{x}$, i.e., we need a stronger assumption when $\alpha = 2.5$; $\sup_{\gamma\in\mathcal{N}} \mathbb{E}\left[\left\|\left(y - \mathbf{x}'\overline{\beta}_0\right)\mathbf{x}\right\|^{2+\epsilon} |q = \gamma\right] < \infty$ is also stronger than $\sup_{\gamma\in\mathcal{N}} \mathbb{E}\left[\left|\mathbf{x}_i'\delta_0\left(y_i - \mathbf{x}_i'\overline{\beta}_0\right)\right|^{2+\epsilon} |q_i = \gamma\right] < \infty$ in Assumption I(2). Similarly in Assumption (x)(c), we need a stronger assumption for $\alpha = 2.5$ by noticing that $\omega^{\pm}_0$ is the $(1,1)$ element of $\Omega^{\pm}_0$. Parallel to $\varpi_{\pm} = \frac{f_0\delta^2_{q0}\omega^{\pm}_0}{3}$, define

$$\Omega_{\pm} = f_0\Omega^{\pm}_0 \text{ and } \Upsilon_{\pm} = -\frac{f_0\delta_{q0}}{2}\Upsilon^{\pm}_{\gamma_0}.$$

As mentioned in Section 3.3, we need to characterize the effect of $\widehat{\beta}$ on $\widehat{\gamma}$ to derive the asymptotic distribution of $\widehat{\gamma}$; this is why we impose Assumption (x)(b) on $S^{\pm}_{\beta}(\gamma)$. First of all, (25) implies $S^{\pm}_{\beta}(\gamma) \in RV_{\alpha-1}$, so the limit in Assumption (x)(b) is meaningful; the only thing that deserves caution is the same $L(\cdot)$ as in $\Lambda(|\gamma|)$ appearing in the normalization rate, but when $\Lambda(|\gamma|)$ takes the form of $|\gamma|$'s power, $L(\cdot) \sim 1$ and such an assumption is innocent. From the formulae of $S^{\pm}_{\beta}(\gamma)$ in (26), only the first terms contribute to the limit $\psi_{\pm}|v|^{\alpha-1}$ but these terms in $S^{\pm}_{\beta_1}(\gamma)$ and $S^{\pm}_{\beta_2}(\gamma)$ are exactly the same with opposite signs; in other words, the last $(d+1)$ components of $\psi_{\pm}$ are the negative of its first $(d+1)$ components.

Example 6 *Take $\psi_+ v^{\alpha-1}$ as an example, and let $L(\cdot) = 1$, $\mathbf{x} = (1, q)'$, $\gamma_0 = 0$, $\delta_{q0} > 0$, and $m_2(q) - \mathbf{x}'\overline{\beta}_0 = Aq^{\alpha-2}$. Because $\Lambda_{\pm}(\gamma) \in RV_\alpha$, we have*

$$-\mathbb{E}\left[\left(m_2(q) - \mathbf{x}'\overline{\beta}_0\right) q\delta_{q0}1(0 < q \leq v)\right] \approx -\frac{A\delta_{q0}f_0}{\alpha}v^\alpha = \lambda_+ v^\alpha,$$

for v around 0, which implies

$$\begin{aligned} -\mathbb{E}\left[\left(m_2(q) - \mathbf{x}'\overline{\beta}_0\right)1(0 < q \leq v)\right] &\approx -\frac{Af_0}{\alpha-1}v^{\alpha-1} = \frac{\lambda_+}{\delta_{q0}}\frac{\alpha}{\alpha-1}v^{\alpha-1}, \\ -\mathbb{E}\left[\left(m_2(q) - \mathbf{x}'\overline{\beta}_0\right)q1(0 < q \leq v)\right] &\approx \frac{\lambda_+}{\delta_{q0}}v^\alpha, \end{aligned}$$

so we have

$$\psi_+ = \begin{pmatrix} \frac{\lambda_+}{\delta_{q0}}\frac{\alpha}{\alpha-1} \\ 0 \\ -\frac{\lambda_+}{\delta_{q0}}\frac{\alpha}{\alpha-1} \\ 0 \end{pmatrix},$$

where the first element of ψ_+ is positive and the third is negative.

To connect $\psi_\pm$ *with* $S^\pm_{\beta\gamma}$ *in Theorem 5, note that when* $\alpha = 2$, $S^\pm_\beta(\gamma) \in RV_1$ *and* $\lim_{|v|\downarrow 0} L(v) = 1$, *so*

$$\frac{S^\pm_\beta(v)}{L(v)} \approx S^\pm_\beta(v) \approx S^\pm_{\beta\gamma} v,$$

i.e., $\psi_- = -S^-_{\beta\gamma}$ *and* $\psi_+ = S^+_{\beta\gamma}$.

We next state the asymptotic distributions of $\widehat{\beta}_\ell$ and $\widehat{\gamma}$.

Theorem 9 *Under Assumption II(*α*),* $2 < \alpha < 3$,

$$\sqrt{n}\left(\widehat{\beta}_1 - \beta_{10}\right) \xrightarrow{d} Z_{\beta_1},$$
$$\sqrt{n}\left(\widehat{\beta}_2 - \beta_{20}\right) \xrightarrow{d} Z_{\beta_2},$$

when $2 < \alpha < 2.5$,

$$\sqrt{n}(\widehat{\gamma} - \gamma_0) \xrightarrow{d} \arg\min_v \left\{ \left[Z'_\beta \psi_- |v|^{\alpha-1} + \lambda_- |v|^\alpha \right] 1(v \le 0) + \left[Z'_\beta \psi_+ v^{\alpha-1} + \lambda_+ v^\alpha \right] 1(v \ge 0) \right\}$$
$$= \begin{cases} \frac{\alpha-1}{\alpha} \frac{\psi'_-}{\lambda_-} Z_\beta, & \text{if } Z_\beta \in R_1, \\ -\frac{\alpha-1}{\alpha} \frac{\psi'_+}{\lambda_+} Z_\beta, & \text{if } Z_\beta \in R_2, \\ 0, & \text{if } Z_\beta \in R_3, \end{cases}$$

where

$$R_1 = \left\{ Z_\beta | \psi'_- Z_\beta < 0,\ \psi'_+ Z_\beta < 0 \text{ and } \left|\psi'_- Z_\beta\right| \ge \left(\tfrac{\lambda_-}{\lambda_+}\right)^{1-\frac{1}{\alpha}} \left|\psi'_+ Z_\beta\right| \text{ OR } \psi'_- Z_\beta < 0 \text{ and } \psi'_+ Z_\beta \ge 0 \right\},$$
$$R_2 = \left\{ Z_\beta | \psi'_- Z_\beta < 0,\ \psi'_+ Z_\beta < 0 \text{ and } \left|\psi'_- Z_\beta\right| < \left(\tfrac{\lambda_-}{\lambda_+}\right)^{1-\frac{1}{\alpha}} \left|\psi'_+ Z_\beta\right| \text{ OR } \psi'_- Z_\beta \ge 0 \text{ and } \psi'_+ Z_\beta < 0 \right\},$$
$$R_3 = \mathbb{R}^{2d+2} \backslash (R_1 \cup R_2) = \left\{ Z_\beta | \psi'_- Z_\beta \ge 0 \text{ and } \psi'_+ Z_\beta \ge 0 \right\},$$

when $2.5 < \alpha < 3$,

$$\varrho_n(\widehat{\gamma} - \gamma_0) \xrightarrow{d} \arg\min_v \begin{cases} \lambda_- |v|^\alpha + \Xi^-_2(|v|), & \text{if } v \le 0, \\ \lambda_+ v^\alpha + \Xi^+_2(v), & \text{if } v > 0, \end{cases} = \omega^{\frac{1}{2\alpha-3}} \zeta(\varphi, \phi; \alpha/3)^{1/3},$$

where $\omega = \frac{\varpi_-}{4\lambda^2_-} = \frac{f_0 \delta^2_{q0} \omega^-_0}{12\lambda^2_-}$, $\varphi = \frac{\lambda_+}{\lambda_-}$ *and* $\phi = \frac{\omega^+_0}{\omega^-_0}$, *and when* $\alpha = 2.5$,

$$\sqrt{n}(\widehat{\gamma} - \gamma_0) \xrightarrow{d} \arg\min_v \left\{ \left[Z'_\beta \psi_- |v|^{3/2} + \left(Z_{\beta_1} - Z_{\beta_2}\right)' \Xi^-_1(|v|) + \lambda_- |v|^{5/2} + \Xi^-_2(|v|) \right] 1(v \le 0) \right.$$
$$\left. + \left[Z'_\beta \psi_+ v^{3/2} - \left(Z_{\beta_1} - Z_{\beta_2}\right)' \Xi^+_1(v) + \lambda_+ v^{5/2} + \Xi^+_2(v) \right] 1(v \ge 0) \right\},$$

where $\left(\Xi^\pm_1(v)', \Xi^\pm_2(v)\right)'$ *is a* $(d+2)$*-dimensional zero-mean Gaussian process on* $[0, \infty)$ *with the covariance kernel*

$$\begin{pmatrix} \Omega_\pm(v_1 \wedge v_2) & \Upsilon_\pm(v_1 \wedge v_2)^2 \\ \Upsilon'_\pm(v_1 \wedge v_2)^2 & \varpi_\pm(v_1 \wedge v_2)^3 \end{pmatrix},$$

and $Z_{\beta_1}, Z_{\beta_2}, \left(\Xi^-_1(\cdot)', \Xi^-_2(\cdot)\right)'$ *and* $\left(\Xi^+_1(\cdot)', \Xi^+_2(\cdot)\right)'$ *are independent.*

When $2<\alpha<2.5$, if $\psi_+=\psi=-\psi_-$ as in II(2) with $S^+_{\beta\gamma}=S^-_{\beta\gamma}$, then

$$\sqrt{n}\left(\widehat{\gamma}-\gamma_0\right)\xrightarrow{d}\begin{cases}-\frac{\alpha-1}{\alpha}\frac{\psi'}{\lambda_-}Z_\beta, & \text{if } \psi'Z_\beta\geq 0,\\ -\frac{\alpha-1}{\alpha}\frac{\psi'}{\lambda_+}Z_\beta, & \text{if } \psi'Z_\beta<0,\end{cases}=-\frac{\alpha-1}{\alpha}\left[\frac{\left(\psi'Z_\beta\right)_\oplus}{\lambda_-}+\frac{\left(\psi'Z_\beta\right)_\ominus}{\lambda_+}\right],$$

which is a mixture of two half normals; if $\lambda_+=\lambda_-=:\lambda$, this asymptotic distribution further reduces to $-\frac{\alpha-1}{\alpha}\frac{\psi'Z_\beta}{\lambda}$, which is close to the asymptotic distribution of $\widehat{\gamma}$ in (27). Inspecting Theorems 6, 7, 8 and 9, we can see that the asymptotic distributions of $\widehat{\gamma}$ in II(α) with $2.5<\alpha\leq 4$ can be unified with only difference lying on the values of α, φ, ϕ and ω; except II(4), even the formulae of φ,ϕ and ω are the same.

Comparing II(α), $2<\alpha<2.5$, with II(2) where $\widehat{\gamma}$ is also fully determined by $\widehat{\beta}$ asymptotically, we can see that the asymptotic distribution of $\widehat{\gamma}$ in the former may have a point mass at zero because $P\left(Z_\beta\in R_3\right)\geq 0$ while is continuous in the former. This difference is not because the relationship between $\widehat{\gamma}$ and $\widehat{\beta}$ has dramatically changed, but because the interaction between $\widehat{\beta}$ and $\widehat{\gamma}$ in II(2) makes $\widehat{\beta}$'s asymptotic distribution not be Z_β anymore. In other words, $\widehat{\gamma}$ indeed has some effect on $\widehat{\beta}$ in II(2) but has no effect in II(α) with $2<\alpha<2.5$. Similar phenomena happen in I(2) and II(4) where $\widehat{\beta}$ is fully determined by $\widehat{\gamma}$ asymptotically but $\widehat{\beta}$ indeed has some effects on $\widehat{\gamma}$ by observing that $\widehat{\gamma}\left(\widehat{\beta}\right)$ and $\widehat{\gamma}\left(\beta_0\right)$ have different asymptotic distributions.

In II(α) with $2<\alpha<2.5$, the randomness of $\widehat{\gamma}$ comes completely from $\widehat{\beta}$ asymptotically, which is similar to $\widehat{\beta}$ in II(α) with $3.5<\alpha<4$; in II(α) with $2.5<\alpha<3$, the asymptotic distribution of $\widehat{\gamma}$ takes the same form as in II(3), which is similar to $\widehat{\beta}$ in II(α) with $3<\alpha<3.5$. However, different from II(3.5) where the asymptotic randomness of $\widehat{\beta}$ is a linear combination of those in II(α) with $3<\alpha<3.5$ and II(α) with $3.5<\alpha<4$, the randomness of $\widehat{\gamma}$ in II(2.5) has some extra elements beyond those in II(α) with $2<\alpha<2.5$ and II(α) with $2.5<\alpha<3$. These extra elements come from $\left(Z_{\beta_1}-Z_{\beta_2}\right)'\Xi_1^-\left(|v|\right)1(v\leq 0)-\left(Z_{\beta_1}-Z_{\beta_2}\right)'\Xi_1^+\left(v\right)1(v\geq 0)$, especially, the $\Xi_1^\pm\left(|v|\right)$ components.

Recall that $\widehat{\gamma}-\gamma_0=\left(\widehat{\gamma}\left(\widehat{\beta}\right)-\widehat{\gamma}\left(\beta_0\right)\right)+\left(\widehat{\gamma}\left(\beta_0\right)-\gamma_0\right)$. When $2<\alpha<2.5$, the first term dominates, and the asymptotic distribution of $\sqrt{n}\left(\widehat{\gamma}-\gamma_0\right)$ indicates the effect of estimating β_0 on $\widehat{\gamma}$; that effect depends on α. When $2.5<\alpha<3$, the second term dominates and the asymptotic distribution of $\sqrt{n}\left(\widehat{\gamma}-\gamma_0\right)$ is as if β_0 were known. When $\alpha=2.5$, both terms contribute. Since

$$\sqrt{n}\left(\widehat{\gamma}\left(\beta_0\right)-\gamma_0\right)\xrightarrow{d}\arg\min_v\left\{\left[\lambda_-|v|^{5/2}+\Xi_2^-\left(|v|\right)\right]1(v\leq 0)+\left[\lambda_+v^{5/2}+\Xi_2^+\left(v\right)\right]1(v>0)\right\},$$

we have

$$\begin{aligned}\sqrt{n}\left(\widehat{\gamma}\left(\widehat{\beta}\right)-\widehat{\gamma}\left(\beta_0\right)\right)\xrightarrow{d}\arg\min_v\Big\{&\left[Z'_\beta\psi_-|v|^{3/2}+\left(Z_{\beta_1}-Z_{\beta_2}\right)'\Xi_1^-\left(|v|\right)+\lambda_-|v|^{5/2}+\Xi_2^-\left(|v|\right)\right]1(v\leq 0)\\&+\left[Z'_\beta\psi_+v^{3/2}-\left(Z_{\beta_1}-Z_{\beta_2}\right)'\Xi_1^+\left(v\right)+\lambda_+v^{5/2}+\Xi_2^+\left(v\right)\right]1(v\geq 0)\Big\}\\&-\arg\min_v\left\{\left[\lambda_-|v|^{5/2}+\Xi_2^-\left(|v|\right)\right]1(v\leq 0)+\left[\lambda_+v^{5/2}+\Xi_2^+\left(v\right)\right]1(v>0)\right\};\end{aligned}$$

that is, the effect of estimating β_0 on $\widehat{\gamma}$ indeed depends on α and the form of the effect when $\alpha=2.5$ is different from that when $2<\alpha<2.5$.

Finally, consider the LR-inference on γ. Define

$$LR_n\left(\gamma\right)=\begin{cases}\frac{n^{\alpha/2}\left(S_n(\gamma)-S_n(\widehat{\gamma})\right)}{L\left(\frac{1}{n^{1/2}}\right)}, \text{ if } 2<\alpha<2.5,\\ n^{5/4}\left(S_n(\gamma)-S_n\left(\widehat{\gamma}\right)\right), \text{ if } \alpha=2.5,\\ \frac{\sqrt{n\varrho_n^3}\left(S_n(\gamma)-S_n(\widehat{\gamma})\right)}{\widehat{\eta}^{\frac{6}{2\alpha-3}}}, \text{ if } 2.5<\alpha<3,\end{cases}$$

where $\widehat{\eta}^2$ is a consistent estimator of $\eta^2 = \frac{\varpi_-^{\alpha/3}}{2\lambda_-}$, and $\sqrt{n\varrho_n^3} \succ n$. Note that the normalization rate of $S_n(\gamma) - S_n(\widehat{\gamma})$ is faster than n for any $2 < \alpha < 3$, which is different from all other cases in both DTR and CTR.

Corollary 8 *Under Assumption II(α), when $2 < \alpha < 2.5$,*

$$\begin{aligned} LR_n(\gamma_0) \xrightarrow{d} & -\left[Z_\beta'\psi_- |Z_\gamma|^{\alpha-1} + \lambda_- |Z_\gamma|^\alpha\right] 1(Z_\gamma \le 0) - \left[Z_\beta'\psi_+ Z_\gamma^{\alpha-1} + \lambda_+ Z_\gamma^\alpha\right] 1(Z_\gamma \ge 0) \\ = & \frac{(\alpha-1)^{\alpha-1}}{\alpha^\alpha} \frac{\left(-Z_\beta'\psi_-\right)^\alpha}{\lambda_-^{\alpha-1}} 1(Z_\beta \in R_1) + \frac{(\alpha-1)^{\alpha-1}}{\alpha^\alpha} \frac{\left(-Z_\beta'\psi_+\right)^\alpha}{\lambda_+^{\alpha-1}} 1(Z_\beta \in R_2), \end{aligned}$$

where R_1 and R_2 are defined in Theorem 9, when $2.5 < \alpha < 3$,

$$LR_n(\gamma_0) \xrightarrow{d} \xi(\varphi, \phi; \alpha/3),$$

where $\xi(\varphi, \phi; \alpha/3)$ is defined in Proposition 2(ii) with φ and ϕ defined in Theorem 9, and when $\alpha = 2.5$,

$$\begin{aligned} LR_n(\gamma_0) \xrightarrow{d} & -\left[Z_\beta'\psi_- |Z_\gamma|^{3/2} + \left(Z_{\beta_1} - Z_{\beta_2}\right)' \Xi_1^-(|Z_\gamma|) + \lambda_- |Z_\gamma|^{5/2} + \Xi_2^-(|Z_\gamma|)\right] 1(Z_\gamma \le 0) \\ & - \left[Z_\beta'\psi_+ Z_\gamma^{3/2} - \left(Z_{\beta_1} - Z_{\beta_2}\right)' \Xi_1^+(Z_\gamma) + \lambda_+ Z_\gamma^{5/2} + \Xi_2^+(Z_\gamma)\right] 1(Z_\gamma \ge 0), \end{aligned}$$

where Z_γ follows the asymptotic distribution of $\sqrt{n}(\widehat{\gamma} - \gamma_0)$ in Theorem 9 when $2 < \alpha \le 2.5$.

When $2 < \alpha < 2.5$, the asymptotic distribution of $LR_n(\gamma_0)$ has a point mass at 0. When $2.5 < \alpha < 3$, the form of $LR_n(\gamma)$ and the asymptotic distribution of $LR_n(\gamma_0)$ take the unified form in II(α) with $3 \le \alpha \le 4$. To make the LR inference feasible, we need to estimate the nuisance parameters. Given the discussions at the end of Section 5.1, only $\psi_\pm$ deserve further attention. Suppose $L(v) \sim 1$ for simplicity; then $S_\beta^\pm(v) \approx \psi_\pm |v|^{\alpha-1}$ for v in a neighborhood of zero. As a result, $\psi_\pm$ can be similarly estimated as $\widehat{\lambda}_\pm$ in (24), only replacing $\widehat{z}_i$ by $\begin{pmatrix} \mathbf{x}_i\left(y_i - \mathbf{x}_i'\widehat{\beta}_1\right) \\ -\mathbf{x}_i\left(y_i - \mathbf{x}_i'\widehat{\beta}_2\right) \end{pmatrix}$ and h^α by $h^{\alpha-1}$.

7 Asymptotics Without Point Identification

When γ cannot be point identified, it can be either partial identified or unidentified. An example of the former is the multiple-regime TR considered in GP (see also Bai (1997a) in the structural change context with $\mathbf{x} = 1$). As noted in the Introduction, the minimizer can only be achieved among the original threshold points. If at two or more threshold points, the limit objective function $S(\gamma)$ has the same value, then γ is only partially identified. An example of the later is $\delta_0 = \mathbf{0}$ which implies $S(\beta_{10}, \beta_{20}, \gamma) = S(\beta_{10}, \beta_{20}, \gamma_0)$ for any $\gamma \in \Gamma$ regardless of the model is CS or MS; in this case, the identified set is Γ and the model is unidentified.

Assumption III: Assumptions MA(i) and (iv)(a) plus

(v-vi) $\Sigma_{\gamma\gamma} > 0$, where $\Sigma_{\gamma_1\gamma_2} := Cov\left(\begin{pmatrix} \mathbf{x}_{\le\gamma_1} y \\ vec\left(\mathbf{x}\mathbf{x}_{\le\gamma_1}\right) \end{pmatrix}\left(\mathbf{x}_{\le\gamma_2}' y, vec\left(\mathbf{x}\mathbf{x}_{\le\gamma_2}\right)'\right)\right)$ for $\gamma_1, \gamma_2 \in \Gamma_o^\epsilon$, Γ_o^ϵ is the ϵ-enlargement of Γ_o with Γ_o defined in (viii) below, and $vec(\cdot)$ is the vec operator.

(vii) $0 < \underline{f} \le f(\gamma) \le \overline{f} < \infty$ for $\gamma \in \Gamma$.

(viii) $\arg\min_\gamma S(\gamma) = \Gamma_o$, a set.

Note that because $S(\gamma)$ is continuous under Assumption (vii), Γ_o is a compact set.

Theorem 10 *Under Assumption III, $\widehat{\gamma}$ is consistent to Γ_o in the sense that*

$$\lim_{n\to\infty} P\left(\widehat{\gamma} \in \Gamma_o^{\epsilon}\right) = 1$$

for any $\epsilon > 0$. (i) If $\beta_{\ell\gamma} \neq \beta_{\ell 0}$ on Γ_o or $\beta_{\ell\gamma} = \beta_{\ell 0}$ but $\beta_{10} \neq \beta_{20}$,

$$\widehat{\gamma} \xrightarrow{d} \arg\max_{\gamma\in\Gamma_o} \Xi(\gamma) =: Z_\gamma,$$

where

$$\Xi(\gamma) = \left(2\beta_{1\gamma}'\mathbb{B}_{1\gamma}^{\mathbf{x}y} - \beta_{1\gamma}'\mathbb{B}_{1\gamma}^{\mathbf{xx}}\beta_{1\gamma}\right) + \left(2\beta_{2\gamma}'\mathbb{B}_{2\gamma}^{\mathbf{x}y} - \beta_{2\gamma}'\mathbb{B}_{2\gamma}^{\mathbf{xx}}\beta_{2\gamma}\right),$$

$\left(\mathbb{B}_{1\gamma}^{\mathbf{x}y\prime}, vec\left(\mathbb{B}_{1\gamma}^{\mathbf{xx}}\right)'\right)'$ *is a $(d+1)(d+2)$-dimensional zero-mean Gaussian process with the covariance kernel equal to $\Sigma_{\gamma_1\gamma_2}$, $\mathbb{B}_{2\gamma}^{\mathbf{x}y} = \mathbb{B}_{1\infty}^{\mathbf{x}y} - \mathbb{B}_{1\gamma}^{\mathbf{x}y}$, and $vec\left(\mathbb{B}_{2\gamma}^{\mathbf{xx}}\right) = vec\left(\mathbb{B}_{1\infty}^{\mathbf{xx}}\right) - vec\left(\mathbb{B}_{1\gamma}^{\mathbf{xx}}\right)$. (ii) If $\beta_{\ell\gamma} = \beta_{\ell 0}$ and $\beta_{10} = \beta_{20}$ on Γ_o,*

$$\widehat{\gamma} \xrightarrow{d} \arg\max_{\gamma\in\Gamma_o} \widetilde{\Xi}(\gamma) =: \widetilde{Z}_\gamma,$$

where

$$\widetilde{\Xi}(\gamma) = \mathbb{B}_{1\gamma}^{\mathbf{x}e\prime} M_\gamma^{-1}\mathbb{B}_{1\gamma}^{\mathbf{x}e} + \mathbb{B}_{2\gamma}^{\mathbf{x}e\prime}\overline{M}_\gamma^{-1}\mathbb{B}_{2\gamma}^{\mathbf{x}e},$$

$\mathbb{B}_{1\gamma}^{\mathbf{x}e}$ *is a $(d+1)$ zero-mean Gaussian process with the covariance kernel equal to $\mathbb{E}\left[\mathbf{xx}'_{\leq\gamma_1\wedge\gamma_2}e^2\right]$, and $\mathbb{B}_{2\gamma}^{\mathbf{x}e} = \mathbb{B}_{1\infty}^{\mathbf{x}e} - \mathbb{B}_{1\gamma}^{\mathbf{x}e}$. (iii) If $\beta_{\ell\gamma} \neq \beta_{\ell 0}$ on Γ_o,*

$$\begin{aligned}
\lim_{n\to\infty} P\left(\widehat{\beta}_1 \leq b_1\right) &= P\left(Z_\gamma \in \Gamma_o^{b_1}\right), \\
\lim_{n\to\infty} P\left(\widehat{\beta}_2 \leq b_2\right) &= P\left(Z_\gamma \in \Gamma_o^{b_2}\right),
\end{aligned}$$

if $\beta_{\ell\gamma} = \beta_{\ell 0}$ but $\beta_{10} \neq \beta_{20}$ on Γ_o,

$$\begin{aligned}
\sqrt{n}\left(\widehat{\beta}_1 - \beta_{\ell 0}\right) &\xrightarrow{d} M_{Z_\gamma}^{-1}\mathbb{B}_{1Z_\gamma}^{\mathbf{x}e}, \\
\sqrt{n}\left(\widehat{\beta}_2 - \beta_{\ell 0}\right) &\xrightarrow{d} \overline{M}_{Z_\gamma}^{-1}\mathbb{B}_{2Z_\gamma}^{\mathbf{x}e},
\end{aligned}$$

and if $\beta_{\ell\gamma} = \beta_{\ell 0}$ and $\beta_{10} = \beta_{20}$ on Γ_o,

$$\begin{aligned}
\sqrt{n}\left(\widehat{\beta}_1 - \beta_{\ell 0}\right) &\xrightarrow{d} M_{\widetilde{Z}_\gamma}^{-1}\mathbb{B}_{1\widetilde{Z}_\gamma}^{\mathbf{x}e}, \\
\sqrt{n}\left(\widehat{\beta}_2 - \beta_{\ell 0}\right) &\xrightarrow{d} \overline{M}_{\widetilde{Z}_\gamma}^{-1}\mathbb{B}_{2\widetilde{Z}_\gamma}^{\mathbf{x}e},
\end{aligned}$$

where $\Gamma_o^{b_1} = \left\{\gamma \in \Gamma_o | \beta_{1\gamma} \leq b_1\right\}$, and $\Gamma_o^{b_2} = \left\{\gamma \in \Gamma_o | \beta_{2\gamma} \leq b_2\right\}$.

From Theorem 10, $\widehat{\gamma}$ is not consistent to a point but converges to a random variable on Γ_o. This partial identifiability of γ is different from that in the usual partial identification literature, e.g., the moment inequalities, where the estimator is not random on a set in finite samples and that random set converges to a fixed set in limit. On the contrary, $\widehat{\gamma}$ is random on any set in finite samples and only the randomness on a specific set Γ_o will not disappear even letting n go to infinity. When $\beta_{\ell\gamma} = \beta_{\ell 0}$ and $\beta_{10} = \beta_{20}$ on Γ_o, $\Xi(\gamma)$ will degenerate to a random variable which does not depend on γ, i.e., $\Xi(\gamma)$ is not useful in deriving the asymptotic distribution of $\widehat{\gamma}$. Specifically,

$$\Xi(\gamma) = \left(2\beta_{\ell 0}'\mathbb{B}_{1\gamma}^{\mathbf{x}y} - \beta_{\ell 0}'\mathbb{B}_{1\gamma}^{\mathbf{xx}}\beta_{\ell 0}\right) + \left(2\beta_{\ell 0}'\mathbb{B}_{2\gamma}^{\mathbf{x}y} - \beta_{\ell 0}'\mathbb{B}_{2\gamma}^{\mathbf{xx}}\beta_{\ell 0}\right) = 2\beta_{\ell 0}'\mathbb{B}_{1\infty}^{\mathbf{x}y} - \beta_{\ell 0}'\mathbb{B}_{2\infty}^{\mathbf{xx}}\beta_{\ell 0}$$

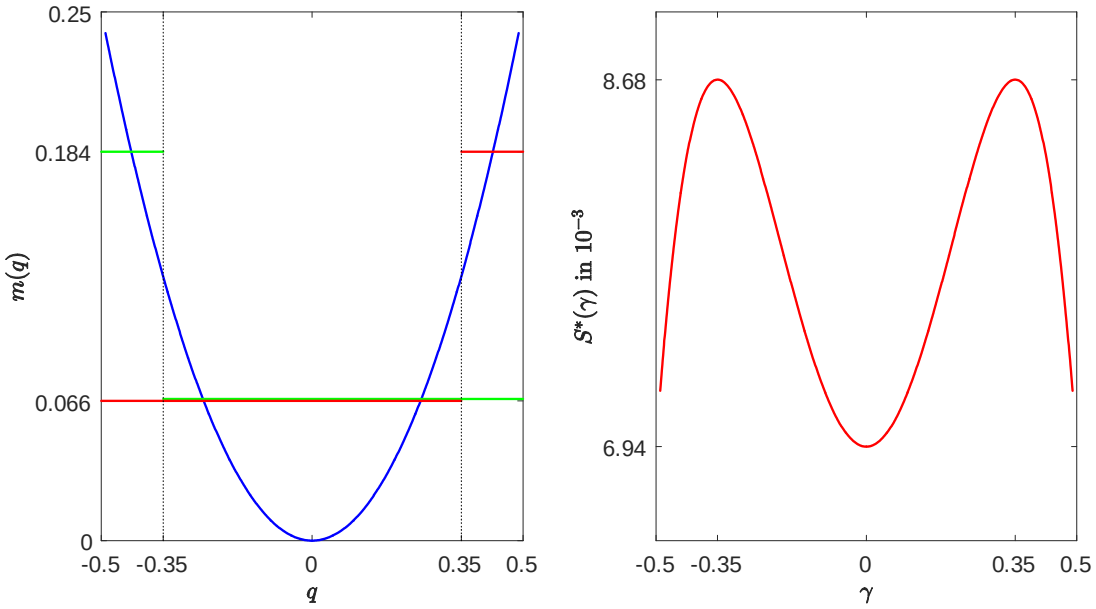

Figure 4: Approximation of $m(q) = q^2$ by $\beta_1 1(q \leq \gamma) + \beta_2 1(q > \gamma)$ and the Limit Objective Function $S^*(\gamma)$

does not involve γ, where we use $\beta_{\ell 0}$ to denote the common β_{10} and β_{20}. In Theorem 10(ii), we refine the asymptotic distribution of $\widehat{\gamma}$, where note that Γ_o must be Γ as mentioned at the beginning of this section, and $e = y - \mathbf{x}'\beta_{\ell 0}$.

When $\beta_{\ell\gamma} \neq \beta_{\ell 0}$ on Γ_o, the distribution of $\widehat{\beta}$ is completely determined by the distribution of $\widehat{\gamma}$. There may be a point mass in the asymptotic distribution of $\widehat{\beta}$. For example, if $\beta_{\ell\gamma} = a$ for $\gamma \in \Gamma_o^s$ with Γ_o^s being a subset of Γ_o and $P(Z_\gamma \in \Gamma_o^s) > 0$, then $P\left(\widehat{\beta}_\ell = a\right) \to P(Z_\gamma \in \Gamma_o^s) > 0$. If $\beta_{\ell\gamma} = \beta_{\ell 0}$ for all $\gamma \in \Gamma_o$, then $\widehat{\beta}_\ell$ converges to a point mass at $\beta_{\ell 0}$ and we need to refine the distribution of $\widehat{\beta}$. It turns out that the asymptotic distribution of $\widehat{\beta}_\ell$ is a mixture normal with the mixing probability depending on $\beta_{10} = \beta_{20}$ or not.

In Bai (1997a), Γ_o is a set of two points, and β_γ is different at these two γ values, so Theorem 10(i) can be applied. Specially, $\widehat{\gamma}$ converges to each of the two points with probability 1/2 as shown in his Proposition 3, and $\widehat{\beta}$ should converge to each of its two different possible values with probability 1/2. Actually, Bai (1997a) refines this result in this special scenario, e.g., $\widehat{\gamma}$ converges to these two points at rate of n, and it is easy to see that $\widehat{\beta}$ converges to its two possible values at rate of $\sqrt{n}$. Yu and Phillips (2019) consider a case where the model is CS but $\delta_0 = \mathbf{0}$; then by Theorem 10(iii),

$$\widehat{\gamma} \xrightarrow{d} \arg\max_{\gamma \in \Gamma} \left\{ \mathbb{B}_{1\gamma}^{\mathbf{x}\varepsilon\prime} M_\gamma^{-1} \mathbb{B}_{1\gamma}^{\mathbf{x}\varepsilon} + \mathbb{B}_{2\gamma}^{\mathbf{x}\varepsilon\prime} \overline{M}_\gamma^{-1} \mathbb{B}_{2\gamma}^{\mathbf{x}\varepsilon} \right\}, \tag{31}$$

where $\mathbb{B}_{1\gamma}^{\mathbf{x}\boldsymbol{\epsilon}} = \mathbb{B}_{1\gamma}^{\mathbf{x}\varepsilon}$ in CS models, $\mathbb{B}_{1\gamma}^{\mathbf{x}\varepsilon}$ is a $(d+1)$-dimensional Gaussian process with the covariance kernel equal to $\mathbb{E}\left[\mathbf{x}\mathbf{x}'_{\leq \gamma_1 \wedge \gamma_2} \varepsilon^2\right]$, and $\mathbb{B}_{2\gamma}^{\mathbf{x}\varepsilon} = \mathbb{B}_{1\infty}^{\mathbf{x}\varepsilon} - \mathbb{B}_{1\gamma}^{\mathbf{x}\varepsilon}$. This asymptotic distribution is exactly the same as that in Yu and Phillips (2019). Our conclusion is that we can refine our results in Theorem 10 when the model is known to have some structures. In practice, we should explore these structures on a case-by-case basis.

The following example shows that partial identification can happen even if $m(q)$ does not take the piecewise constant form as in Bai (1997a). This example is inspired by the example in Remark 1 of BM.

Example 7 *Suppose* $y = |q|^2 + \varepsilon$, *where* $q \sim U[-0.5, 0.5]$, $\varepsilon \sim N(0,1)$, *and* $\mathbf{x} = 1$.[13] *From the proof of Theorem 10, we have*

$$\begin{aligned}\arg\min_{\gamma} S(\gamma) &= \arg\max_{\gamma} S^*(\gamma) := \arg\max_{\gamma} \left\{\beta_{1\gamma}^2 F(\gamma) + \beta_{2\gamma}^2 \overline{F}(\gamma)\right\} \\ &= \arg\max_{\gamma} \left\{ \left(\frac{1+8\gamma^3}{24\gamma+12}\right)^2 (\gamma + 0.5) + \left(\frac{1-8\gamma^3}{12-24\gamma}\right)^2 (0.5-\gamma) \right\} = \{\gamma_{1o}, \gamma_{2o}\},\end{aligned}$$

where $F(\gamma) = \gamma + 0.5$ *is the cdf of* q, *and* $\overline{F}(\gamma) = 1 - F(\gamma) = 0.5 - \gamma$ *is the survival function. From the right panel of Figure 4,* $\gamma_{1o} = -\sqrt{2}/4 \approx -0.35$ *and* $\gamma_{2o} = -\gamma_{1o}$. *As a result,* $\beta_{1\gamma_{1o}} = \beta_{2\gamma_{2o}} = \frac{3+\sqrt{2}}{24} \approx 0.184$ *and* $\beta_{2\gamma_{1o}} = \beta_{1\gamma_{2o}} = \frac{3+\sqrt{2}}{24} \approx 0.066$, *which are shown in the left panel of Figure 4. To derive the asymptotic distribution of* $\widehat{\gamma} = \gamma_{1o}$ *and* $\widehat{\gamma} = \gamma_{2o}$, *we apply Theorem 10(i). Note that* $\Xi(\gamma_{1o}) - \Xi(\gamma_{2o})$ *follows a mean-zero normal distribution, so* $P(\Xi(\gamma_{1o}) - \Xi(\gamma_{2o}) > 0) = \frac{1}{2}$; *as a result,* $\widehat{\gamma}$ *converges in distribution to a random variable with equal mass at* γ_{1o} *and* γ_{2o}.[14] *Interestingly, these probabilities are the same as in Proposition 3 of Bai (1997a). Given the asymptotic distribution of* $\widehat{\gamma}$, *it is not hard to see that* $\widehat{\beta}_\ell$ *converges in distribution to a random variable with equal mass at* $\beta_{\ell\gamma_{1o}}$ *and* $\beta_{\ell\gamma_{2o}}$.

The next example shows that unidentification can happen even if the model is MS. This example is inspired by Example 1 of Hidalgo (1995).

Example 8 *Suppose* $y = x + x^2 + \varepsilon$, *where* $\varepsilon \sim N(0,1)$, $\mathbf{x} = x \sim N(0,1)$, $q \sim U[0,1]$, *and* x, q *and* ε *are independent of each other. Then it is not hard to see* $\beta_{1\gamma} = \beta_{2\gamma} = 1$ *for any* γ *because* x *is symmetrically distributed. As a result,*

$$\gamma_0 = \arg\max_{\gamma} \left\{\beta_{1\gamma}^2 \mathbb{E}\left[x^2_{\leq\gamma}\right] + \beta_{2\gamma}^2 \mathbb{E}\left[x^2_{>\gamma}\right]\right\} = \arg\max_{\gamma} \left\{\mathbb{E}\left[x^2_{\leq\gamma}\right] + \mathbb{E}\left[x^2_{>\gamma}\right]\right\} = \arg\max_{\gamma} \left\{\mathbb{E}\left[x^2\right]\right\} = \Gamma.$$

We need to apply Theorem 10(ii) to derive the asymptotic distribution of $\widehat{\gamma}$:

$$\begin{aligned}\widehat{\gamma} \xrightarrow{d} \arg\max_{\gamma\in\Gamma_o} \left\{ \frac{\left(\mathbb{B}^{\mathbf{x}e}_{1\gamma}\right)^2}{M_\gamma} + \frac{\left(\mathbb{B}^{\mathbf{x}e}_{2\gamma}\right)^2}{\overline{M}_\gamma} \right\} &= \arg\max_{\gamma\in\Gamma_o} \left\{ 16\left(\frac{\mathbb{B}^{\mathbf{x}e}_{1\gamma}}{4\sqrt{\gamma}}\right)^2 + 16\left(\frac{\mathbb{B}^{\mathbf{x}e}_{2\gamma}}{4\sqrt{1-\gamma}}\right)^2 \right\} \\ &= \arg\max_{\gamma\in\Gamma_o} \left\{\chi^2_{1\gamma} + \chi^2_{2\gamma}\right\},\end{aligned}$$

where the covariance kernel of $\mathbb{B}^{\mathbf{x}e}_{1\gamma}$ *is* $\mathbb{E}\left[\mathbf{x}\mathbf{x}'_{\leq\gamma_1\wedge\gamma_2} e^2\right] = \mathbb{E}\left[x^2_{\leq\gamma_1\wedge\gamma_2}(y - x\beta_{\ell 0})^2\right] = 16(\gamma_1 \wedge \gamma_2)$, $M_\gamma = \gamma$, $\overline{M}_\gamma = 1 - \gamma$, *and* $\chi^2_{1\gamma} = \left(\frac{\mathbb{B}^{\mathbf{x}e}_{1\gamma}}{4\sqrt{\gamma}}\right)^2$ *and* $\chi^2_{2\gamma} = \left(\frac{\mathbb{B}^{\mathbf{x}e}_{2\gamma}}{4\sqrt{1-\gamma}}\right)^2$ *are two chi-square processes. If the model is CS, i.e.,* $y = x + \varepsilon$, *then* $\mathbb{B}^{\mathbf{x}e}_{\ell\gamma} = \mathbb{B}^{\mathbf{x}\varepsilon}_{\ell\gamma}$ *as in (31) and* $\mathbb{E}\left[\mathbf{x}\mathbf{x}'_{\leq\gamma_1\wedge\gamma_2}\varepsilon^2\right] = \gamma_1 \wedge \gamma_2$.

Finally, consider the LR-inference on γ. Define

$$LR_n(\gamma) = \begin{cases} 2\sqrt{n}\left(S_n(\gamma) - S_n(\widehat{\gamma})\right), & \text{if } \beta_{\ell\gamma} \neq \beta_{\ell 0} \text{ on } \Gamma_o \text{ or } \beta_{\ell\gamma} = \beta_{\ell 0} \text{ but } \beta_{10} \neq \beta_{20}, \\ 2n\left(S_n(\gamma) - S_n(\widehat{\gamma})\right), & \text{if } \beta_{\ell\gamma} = \beta_{\ell 0} \text{ and } \beta_{10} = \beta_{20} \text{ on } \Gamma_o. \end{cases}$$

Note that the normalization rate depends on whether $\beta_{\ell\gamma} = \beta_{\ell 0}$ and $\beta_{10} = \beta_{20}$ on Γ_o.

[13] Actually, we can show that as long as the power of $|q|$ in $m(q)$ is positive, $S(\gamma)$ has two minimizers and the arguments below apply.

[14] Note that this does not mean $\lim_{n\to\infty} P(\widehat{\gamma} = \gamma_{1o}) = 1/2 = \lim_{n\to\infty} P(\widehat{\gamma} = \gamma_{2o})$. Also, when there are more than two maximizers of $S^*(\gamma)$, the asymptotic distribution of $\widehat{\gamma}$ need not put equal mass on each maximizer.

Corollary 9 *Under Assumption III, if $\beta_{\ell\gamma} \neq \beta_{\ell 0}$ on Γ_o or $\beta_{\ell\gamma} = \beta_{\ell 0}$ but $\beta_{10} \neq \beta_{20}$,*

$$LR_n(\gamma_0) \xrightarrow{d} \max_{\gamma\in\Gamma_o} \Xi(\gamma) - \Xi(\gamma_0),$$

and if $\beta_{\ell\gamma} = \beta_{\ell 0}$ and $\beta_{10} = \beta_{20}$ on Γ_o,

$$LR_n(\gamma_0) \xrightarrow{d} \max_{\gamma\in\Gamma_o} \widetilde{\Xi}(\gamma) - \widetilde{\Xi}(\gamma_0)$$

where $\Xi(\gamma)$ and $\widetilde{\Xi}(\gamma)$ are defined in Theorem 10.

In Example 7, $\max_{\gamma\in\Gamma_o} \Xi(\gamma) - \Xi(\gamma_0)$ is the max of two correlated zero-mean normal random variables minus one of them, and in Example 8, $\max_{\gamma\in\Gamma_o} \widetilde{\Xi}(\gamma) - \widetilde{\Xi}(\gamma_0)$ is the maximum of $\widetilde{\Xi}(\gamma)$, which is the sum of two chi-square processes, minus the value of $\widetilde{\Xi}(\gamma)$ at the hypothetical γ_0.

The CI by inverting $LR_n(\gamma)$ will cover each γ_0 in Γ_o with the prespecified level, but may not cover Γ_o with that level. Anyway, the critical value depends on Γ_o; however, if we know Γ_o, then we do not need a CI for the point in Γ_o anymore. In practice, we can choose a set Γ_n that includes Γ_o almost surely in determining the critical value based on Corollary 9. Such a critical value would be conservative but avoids the precise knowledge on Γ_o. For example, in Bai (1997a), if there are three threshold points (i.e., four regimes) in the original model, and we suspect Γ_o includes two of them but are not sure which two, then we can replace Γ_o by a three-points set (which needs to be estimated) to obtain a conservative critical value. Of course, the most conservative critical value is achieved by replacing Γ for Γ_o, which is also the appropriate critical value when $\beta_{\ell\gamma} = \beta_{\ell 0}$ and $\beta_{10} = \beta_{20}$ on Γ_o.

8 Discussions

In this section, we discuss some extensions of our asymptotic theory and also some unsolved problems in this paper.

First, there seems a gap between the asymptotic theory of DTR and CTR. For example, the range of α in DTR is $[1, 2]$ while in CTR is $[2, 4]$. From the intuitions in Section 3, we can see this is because $\mathbf{x}'\delta_0 = q\delta_{q0}$ in CTR where the power of q is 1. If we replace q by q^τ, $0 < \tau < 1$, when we can transfer smoothly from DTR to CTR. Now, in Proposition 1, $\delta_0'\mathbb{E}[\mathbf{x}\mathbf{x}'|q = \gamma_0]\delta_0 = 0$ if and only if $\delta_{x0} = 0$ and $\delta_{c0} + \delta_{q0}^\tau\gamma_0 = 0$ (or $\delta_{c0} = 0$ when $\gamma_0 = 0$). Because such a regressor seems rare in practice, we will not study this setup in this paper.

Second, in Section 2.3, we assume the rates of $\Lambda_\pm(\gamma)$ shrinking to zero are the same; what will happen if these two rates are different? Actually, by a similar argument as in Theorem 3.2 of YZ, the convergence rate of $\widehat{\gamma}$ is determined by the neighborhood with less identification information for γ. Specifically, suppose in Sections 3.2 and 3.3 the convergence rate of $\widehat{\gamma}$ determined by $\Lambda_-(\cdot)$ (instead of $\Lambda(\cdot)$) is ρ_n^- and by $\Lambda_+(\cdot)$ os ρ_n^+, then the ultimate convergence rate of $\widehat{\gamma}$ is $\rho_n := \rho_n^- \wedge \rho_n^+$. If $\rho_n = \rho_n^+$, then in all theorems the information in the left neighborhood of γ_0 can be neglected because $\widehat{\gamma}$ cannot fall in the left neighborhood of γ_0 asymptotically, and vice versa. To be concrete, suppose $\alpha_- = 1$, $\alpha_+ = 1.5$ and $L_\pm(\cdot) = 1$ in DTR; then $\rho_n = n \wedge \sqrt{n} = \sqrt{n}$. We now need to revise Theorem 4 as

$$\begin{aligned}
&\sqrt{n}(\widehat{\gamma} - \gamma_0) \xrightarrow{d} \arg\max_{v\geq 0}\left\{-\lambda_+ v^\alpha + \sqrt{\varpi_+}B_2(v)\right\} =: Z_\gamma(1.5), \\
&\sqrt{n}\begin{pmatrix}\widehat{\beta}_1 - \beta_{10}\end{pmatrix} \xrightarrow{d} Z_{\beta_1} - M_0^{-1}S_{\beta_1\gamma}^+ Z_\gamma(1.5), \\
&\sqrt{n}\begin{pmatrix}\widehat{\beta}_2 - \beta_{20}\end{pmatrix} \xrightarrow{d} Z_{\beta_2} - \overline{M}_0^{-1}S_{\beta_2\gamma}^+ Z_\gamma(1.5).
\end{aligned}$$

Third, when the joint asymptotic distribution of $\widehat{\beta}$ and $\widehat{\gamma}$ is degenerate, we can combine them in an appropriate way to develop a nondegenerate asymptotic distribution. For example, in I(α) with $1.5 < \alpha \leq 2$, combine $\widehat{\beta}_\ell$ and $\widehat{\gamma}$ as

$$\rho_n \left[\left(\widehat{\beta}_1 - \beta_{10} \right) + M_0^{-1} S^-_{\beta_1 \gamma} \left(\widehat{\gamma} - \gamma_0 \right)_- + M_0^{-1} S^+_{\beta_1 \gamma} \left(\widehat{\gamma} - \gamma_0 \right)_+ \right],$$
$$\rho_n \left[\left(\widehat{\beta}_2 - \beta_{20} \right) + \overline{M}_0^{-1} S^-_{\beta_2 \gamma} \left(\widehat{\gamma} - \gamma_0 \right)_- + \overline{M}_0^{-1} S^+_{\beta_2 \gamma} \left(\widehat{\gamma} - \gamma_0 \right)_+ \right];$$

in II(α) with $3.5 < \alpha \leq 4$, combine $\widehat{\beta}_\ell$ and $\widehat{\gamma}$ as

$$\rho_n^2 \left[\left(\widehat{\beta}_1 - \beta_{10} \right) + \frac{1}{2} M_0^{-1} S_{\beta_\ell \gamma^2} \left(\widehat{\gamma} - \gamma_0 \right)^2 \right],$$
$$\rho_n^2 \left[\left(\widehat{\beta}_2 - \beta_{20} \right) + \frac{1}{2} \overline{M}_0^{-1} S_{\beta_\ell \gamma^2} \left(\widehat{\gamma} - \gamma_0 \right)^2 \right];$$

in II(2), combine $\widehat{\beta}$ and $\widehat{\gamma}$ as

$$\sqrt{n} \left[\left(\widehat{\gamma} - \gamma_0 \right) + \left(\left(S^-_{\gamma\gamma} \right)^{-1} S^-_{\beta\gamma} 1_{R_1 \cap R_2} + \left(S^+_{\gamma\gamma} \right)^{-1} S^+_{\beta\gamma} 1_{R_1 \cap \overline{R}_2} + \left(S^-_{\gamma\gamma} \right)^{-1} S^-_{\beta\gamma} 1_{\overline{R}_1 \cap R_3} + \left(S^+_{\gamma\gamma} \right)^{-1} S^+_{\beta\gamma} 1_{\overline{R}_1 \cap \overline{R}_3} \right) \left(\widehat{\beta} - \beta_0 \right) \right],$$

where the subscript of the indicator function signifies the area of $\mathrm{diag}\{M_0, \overline{M}_0\} \sqrt{n} \left(\widehat{\beta} - \beta_0 \right)$ staying; in II(α) with $2 < \alpha < 2.5$, combine $\widehat{\beta}$ and $\widehat{\gamma}$ as

$$\sqrt{n} \left[\left(\widehat{\gamma} - \gamma_0 \right) + \left(-\frac{\alpha - 1}{\alpha} \frac{\psi'_-}{\lambda_-} 1_{R_1} + \frac{\alpha - 1}{\alpha} \frac{\psi'_+}{\lambda_+} 1_{R_2} \right) \left(\widehat{\beta} - \beta_0 \right) \right],$$

where the subscript of the indicator function signifies the area of $\sqrt{n} \left(\widehat{\beta} - \beta_0 \right)$ staying. Because we conduct inference on γ based on the LR statistic which is nondegenerate, the developments of such refinements seem unnecessary for our purpose.

Fourth, as mentioned in the Introduction, YZ is closely related to this paper but $f(\gamma)$ there can converge to zero or diverge to infinity as γ converges to γ_0. Combing YZ and this paper would be an interesting exercise, but it seems reasonable to assume $f(\gamma)$ to be finite in a neighborhood of γ_0 in practice.

Fifth, we assume δ_0 shrinks to zero in I(1)$'$ to obtain accessible asymptotic distributions for $\widehat{\gamma}$ and the LR statistic. In all other cases of both DTR and CTR, we can also assume shrinking threshold effects, but it seems unnecessary because the asymptotic distributions in all these cases involve only Gaussian processes and can be simulated at least in principle.

Sixth, the techniques used in the paper can be extended to study misspecification in quantile threshold regression. If the model is CS, $\widehat{\gamma}$ based on any quantile index should converge to the same value, so it is a sign of misspecification if quantile threshold regression based on different quantile indices generates different threshold estimates; see Galvao et al (2011) for some evidences in threshold quantile autoregressive models.

Seventh, as mentioned in the Introduction, the TAR model is proposed initially to approximate more general time series, so it is desirable to extend the results in this paper to time series. By extending the techniques of Hansen (2000), we expect the results in this paper still hold for stationary ergodic time series. But we will not investigate this extension in this paper because our proofs are already quite complicated; adding time dependency to the DGP will dramatically lengthen the proofs without essentially changing the main results.

Eighth, when $\Lambda(\cdot)$ is known, we can conduct LR inference on γ as detailed in the main text, but if $\Lambda(\cdot)$ is unknown, then ρ_n is unknown and it is hard to formulate the LR statistic because the normalization rate

τ_n and normalization constant $\widehat{b}$ are hard to determine. In other words, we must have some a priori (at least quantitative) knowledge on the behavior of $m(x,q)$ around $q=\gamma_0$ to apply the LR inference in this paper. For example, if it is believed that we are using TR to approximate some phenomena involving discontinuity, then the LR inference in Section 4.2 is appropriate. Generally speaking, a challenging problem is how to conduct uniform inference on γ without any knowledge on $\Lambda(\cdot)$.

BY (p. 940) point out that the bootstrap is not valid, while BM show that the subsampling still works in I(2). HLS unify the LR inference on γ in CS I(1)$'$ and CS II(3) by the grid bootstrap. However, the grid bootstrap implicitly assumes the model is CS. For example, conditional on $\{\mathbf{x}_i\}_{i=1}^n$, the grid bootstrap generates the bootstrap samples $\{y_i^*\}_{i=1}^n$ by

$$y_i^* = \begin{cases} \mathbf{x}_i'\widehat{\beta}_1 + \widehat{e}_{1i}\epsilon_i, & \text{if } q_i \le \widehat{\gamma}, \\ \mathbf{x}_i'\widehat{\beta}_2 + \widehat{e}_{2i}\epsilon_i, & \text{if } q_i > \widehat{\gamma}, \end{cases}$$

where $\widehat{e}_{\ell i} = y_i - \mathbf{x}_i'\widehat{\beta}_\ell$, and $\{\epsilon_i\}_{i=1}^n$ are i.i.d. zero mean random variables with unit variance and finite fourth moments. Obviously, in the bootstrap world, the conditional mean of y_i^* is linear in $\mathbf{x}_i$ in each regime; in other words, the model is CS and only I(1)$'$ and II(3) can happen. As shown in Section 4.2, φ is generally not equal to 1 in MS I(1)$'$, but it is equal to 1 in CS I(1)$'$; furthermore, the normalization constant η^2 should be $\mathbb{E}\left[\left(\delta_n'\mathbf{x}_i\right)^2 e_{1i}^2 | q_i = \gamma_0-\right] / \mathbb{E}\left[\left(\delta_n'\mathbf{x}_i\right)^2 | q_i = \gamma_0\right]$ in CS I(1)$'$, so we need change to this formula of η^2 in the grid bootstrap, where note that the error variances in the bootstrap world are $\widehat{e}_{\ell i}^2 \approx e_{\ell i}^2$ in each regime, which is the reason of $\varepsilon_{\ell i}^2$ being replaced by $e_{\ell i}^2$. Using this η^2, the asymptotic bootstrap distribution is $\xi(1,\phi;1)$ with a correct ϕ but a wrong φ. In other words, the asymptotic bootstrap distribution does not match the original asymptotic distribution of the LR statistic, so the grid bootstrap is not consistent in MS I(1)$'$. In MS II(3), $\varphi = \lambda_+/\lambda_- \neq 1$ in general, and $\phi = \mathbb{E}\left[e_2^2 | q=\gamma_0+\right] / \mathbb{E}\left[e_1^2 | q=\gamma_0-\right] = \omega_0^+/\omega_0^-$ can be consistently estimated by $\frac{n^{-1}\sum_{i=1}^n \left(\mathbf{x}_i'\widehat{\delta}\right)^2 \widehat{e}_2^2 K_h^+(q_i-\widehat{\gamma})}{n^{-1}\sum_{i=1}^n \left(\mathbf{x}_i'\widehat{\delta}\right)^2 \widehat{e}_1^2 K_h^-(q_i-\widehat{\gamma})}$ by extending Proposition 3 of HLS, where $K_h^\pm(\cdot) = h^{-1}K^\pm(\cdot/h)$ for some bandwidth h and boundary kernel functions $K^\pm(\cdot)$. By setting $\eta^2 = \omega_0^- = \mathbb{E}\left[e_1^2 | q=\gamma_0-\right]$, which can be consistently estimated by $\frac{n^{-1}\sum_{i=1}^n \left(\mathbf{x}_i'\widehat{\delta}\right)^2 \widehat{e}_2^2 K_h^-(q_i-\widehat{\gamma})}{n^{-1}\sum_{i=1}^n \left(\mathbf{x}_i'\widehat{\delta}\right)^2 K_h(q_i-\widehat{\gamma})}$ as shown in Proposition 3 of HLS, as in CS II(3), we have the asymptotic bootstrap distribution as $\xi(1,\phi;1)$, so it is still that φ is wrong, where $K_h(\cdot)$ is similarly defined as $K_h^\pm(\cdot)$. In summary, HLS's grid bootstrap procedure has the asymptotic bootstrap distribution $\xi(1,\phi;1)$, so is not valid in both MS I(1)$'$ and MS II(3).

9 Numerical Examples

In this section, we consider some concrete DGPs to illustrate the asymptotic distributions of $\widehat{\gamma}$. For simplicity, we normalize $\mathbf{x}'\overline{\beta}_0 = 0$ and set $q \sim U[-0.5, 0.5]$, $\varepsilon \sim N(0,1)$ independent of q, and $\gamma_0 = 0$. Also, due to the symmetricity in the specification of β_{10} vs. β_{20} and $m_1(q)$ vs. $m_2(q)$, $\lambda_- = \lambda_+ =: \lambda$ and all asymptotic distributions are symmetric about zero. In DTR, we consider I(1), I(1.5), I(2) and I(3), and in CTR, we consider II(2), II(2.5), II(3) and II(4).

9.1 DTR

Suppose $\mathbf{x} = 1$ to further simplify the discussion. Let $\beta_{10} = 0.5$ and $\beta_{20} = -0.5$, which implies $\delta_0 = 1$,

$$m_1(q) = a + b\,|q|^{\alpha-1} \text{ and } m_2(q) = -a - bq^{\alpha-1}$$

with $a \geq 0$, $b > 0$, and $\alpha > 1$. Now, $\beta_{10} = 0.5$ and $\beta_{20} = -0.5$ implies $b = (0.5 - a)\,\alpha 2^{\alpha-1}$, and

$$\Lambda_{\pm}(\gamma) = a|\gamma| + \frac{b}{\alpha}|\gamma|^{\alpha}.$$

$m(q)$ and $\Lambda_{\pm}(\gamma)$ are shown in the first and second rows of Figure 5. In I(1), set $\alpha = 2$, $a = 0.25$ and $b = 1$; in I(1.5), set $\alpha = 1.5$, $a = 0$ and $b = 0.75\sqrt{2}$; in I(2), set $\alpha = 2$, $a = 0$ and $b = 2$; in I(3), set $\alpha = 3$, $a = 0$ and $b = 6$. In the asymptotic distribution of I(1), $z_1 = 0.25 + \varepsilon_1$ and $z_2 = 0.25 - \varepsilon_2$ with ε_1 and ε_2 being i.i.d. copies of ε. From Appendix D of Yu (2012), we can derive the distribution of $Z_\gamma(1)$, which is shown in the $(3,1)$ panel of Figure 5. In I(1.5), $Z_\gamma(1.5) = \sqrt{\omega}\zeta(1,1;1.5)$ with $\omega = \frac{\varpi_-}{(2\lambda_-)^2} = \frac{1}{2}$. Because there is no closed-form density for $\zeta(1,1;1.5)$, we simulate it; the resulting density of $Z_\gamma(1.5)$ is shown in the $(3,2)$ panel of Figure 5. In I(2), $\varphi = \phi = 1$, and $\omega = \frac{\varpi_-}{\mu_-^2} = 1$ from Example 3, so $Z_\gamma(2) = \zeta_{1/2} = (1/2)^{-2/3}\zeta_1$. Dykstra and Carolan (1999) suggest the approximation $N(0,(0.52)^2)$ for ζ_1. Such an approximation turns out to be fairly accurate, as evidenced by the results of Groeneboom and Wellner (2001, Table 2). The density of $(1/2)^{-2/3}\,N(0,(0.52)^2)$ is shown in the $(3,3)$ panel of Figure 5. Comparing the densities of $\widehat{\gamma}$ in I(1), I(1.5) and I(2), we can see that the asymptotic density at zero gets smoother as α gets larger.

In Section 3, we show that γ_0 may be locally unidentified in I(3) because the cross term $\left\|\widetilde{\beta}\right\||\gamma|$ may not be dominated by $\left\|\widetilde{\beta}\right\|^2 + |\gamma|^3$. To show this is indeed the case in our example, we scrutinize $S(\theta) - S(\theta_0)$. First,

$$\begin{aligned}
\Psi_-(\beta,\gamma) &= \begin{pmatrix} -\mathbb{E}\left[\left(6q^2 - \frac{\beta_{10}+\beta_1}{2}\right)(\beta_{10}-\beta_1)\,1(\gamma < q \leq 0)\right] \\ \mathbb{E}\left[\left(6q^2 - \frac{\beta_{10}+\beta_2}{2}\right)(\beta_{10}-\beta_2)\,1(\gamma < q \leq 0)\right] - 2|\gamma|^3 \end{pmatrix} = \begin{pmatrix} -2\gamma^3\widetilde{\beta}_1 + \frac{\beta_{10}+\beta_1}{2}\gamma\widetilde{\beta}_1 \\ 2\gamma^3\widetilde{\beta}_2 + \frac{\widetilde{\beta}_2}{2}\gamma\left(1 - \widetilde{\beta}_2\right) \end{pmatrix}, \\
\Psi_+(\beta,\gamma) &= \begin{pmatrix} \mathbb{E}\left[\left(-6q^2 - \frac{\beta_{20}+\beta_1}{2}\right)(\beta_{20}-\beta_1)\,1(0 < q \leq \gamma)\right] - 2\gamma^3 \\ -\mathbb{E}\left[\left(-6q^2 - \frac{\beta_{20}+\beta_2}{2}\right)(\beta_{20}-\beta_2)\,1(0 < q \leq \gamma)\right] \end{pmatrix} = \begin{pmatrix} 2\gamma^3\widetilde{\beta}_1 + \frac{\widetilde{\beta}_1}{2}\gamma\left(1 + \widetilde{\beta}_1\right) \\ -2\gamma^3\widetilde{\beta}_2 - \frac{\beta_{20}+\beta_2}{2}\gamma\widetilde{\beta}_2 \end{pmatrix},
\end{aligned}$$

which implies

$$S^-_{\beta\gamma} = S^+_{\beta\gamma} = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} =: S_{\beta\gamma}.$$

Second,

$$\Phi(\beta) = \begin{pmatrix} \mathbb{E}\left[\left(6q^2 - \frac{\beta_{10}+\beta_1}{2}\right)(\beta_{10}-\beta_1)\,1(q \leq 0)\right] \\ \mathbb{E}\left[\left(-6q^2 - \frac{\beta_{20}+\beta_2}{2}\right)(\beta_{20}-\beta_2)\,1(q > 0)\right] \end{pmatrix} = \begin{pmatrix} \frac{1}{4}\widetilde{\beta}_1^2 \\ \frac{1}{4}\widetilde{\beta}_2^2 \end{pmatrix}.$$

So locally,

$$S(\theta) - S(\theta_0) \approx \Phi(\beta_1) + \overline{\Phi}(\beta_2) + \widetilde{\beta}' S_{\beta\gamma}\gamma + \Lambda_{\pm}(\gamma) = \tfrac{1}{4}\left(\widetilde{\beta}_1^2 + \widetilde{\beta}_2^2 + 2\gamma\widetilde{\beta}_1 + 2\gamma\widetilde{\beta}_2 + 8|\gamma|^3\right) =: \overline{S}(\theta) - S(\theta_0),$$

whose minimum with γ fixed is achieved at $\widetilde{\beta}_{1\gamma} = \widetilde{\beta}_{2\gamma} = -\gamma$ with the minimum equal to

$$\overline{S}(\gamma) - S(\gamma_0) := -\frac{1}{2}\gamma^2 + 2|\gamma|^3 < 0,$$

i.e., 0 is not the minimizer of $S(\gamma)$. The cross terms $\gamma\widetilde{\beta}_1$ and $\gamma\widetilde{\beta}_2$ play a key role to make this happen. Including higher order terms in, we have

$$S(\theta) - S(\theta_0) = \overline{S}(\theta) - S(\theta_0) + 2|\gamma|^3\widetilde{\beta}_1 + \frac{1}{2}\gamma\widetilde{\beta}_1^2 - 2|\gamma|^3\widetilde{\beta}_2 - \frac{1}{2}\gamma\widetilde{\beta}_2^2,$$

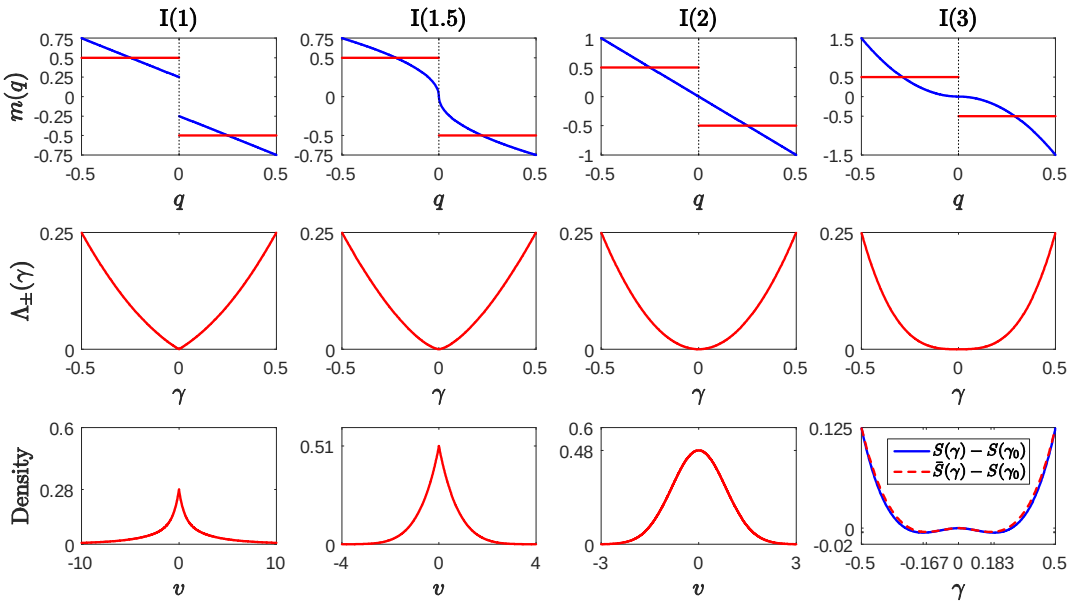

Figure 5: $m(q)$, $\Lambda_{\pm}(\gamma)$ and LSE Asymptotic Distributions in DTR

whose minimum with γ fixed is achieved at $\widetilde{\beta}_{1\gamma} = -\frac{\gamma+4|\gamma|^3}{1+2\gamma}$ and $\widetilde{\beta}_{2\gamma} = -\frac{\gamma-4|\gamma|^3}{1-2\gamma}$ with the minimum equal to

$$S(\gamma) - S(\gamma_0) = -\frac{1}{4}\frac{\left(\gamma + 4|\gamma|^3\right)^2}{1+2\gamma} - \frac{1}{4}\frac{\left(\gamma - 4|\gamma|^3\right)^2}{1-2\gamma} + 2|\gamma|^3.$$

Both $S(\gamma) - S(\gamma_0)$ and $\overline{S}(\gamma) - S(\gamma_0)$ are shown in the $(3,4)$ panel of Figure 5. From Figure 5, we can see $\overline{S}(\gamma) - S(\gamma_0)$ approximates $S(\gamma) - S(\gamma_0)$ very well, and γ_0 is not the local minimizer but the local maximizer of $S(\gamma) - S(\gamma_0)$; both $\overline{S}(\gamma) - S(\gamma_0)$ and $S(\gamma) - S(\gamma_0)$ have two global minimizers at $\pm 1/6 \approx \pm 0.167$ and $\pm\left(\sqrt{3}-1\right)/4 \approx \pm 0.183$, respectively. From Example 7, $\widehat{\gamma}$ converges in distribution to a random variable with equal mass at $\pm\left(\sqrt{3}-1\right)/4$. This setup also shows that monotonicity of $m(q)$ does not guarantee γ to be point identified as hinted in Remark 1 of BM, and combined with the setup in I(2), shows that a strictly increasing transformation of $m(q)$ need not imply the same γ_0 as claimed on page 551 of BM.

9.2 CTR

Suppose $\mathbf{x} = (1, q)'$ and let $\beta_{10} = \left(0, \frac{1}{2}\right)'$ and $\beta_{20} = \left(0, -\frac{1}{2}\right)$, which implies $\delta_0 = (0, 1)$ and $\mathbf{x}'\delta_0 = q$. In II(3), set

$$m_1(q) = (q+a)^2 - a^2 - bq^4 \text{ and } m_2(q) = (q-a)^2 - a^2 - bq^4$$

with $a = 1/3$ and $b = 10/3$, where $m_1'(0) = 2/3 \neq 1/2$ and $m_2'(0) = -2/3 \neq -1/2$; this setup indicates that II(3) does not require the model to be CS. Now,

$$\Lambda_{\pm}(\gamma) = \frac{|\gamma|^3\left(8 - 9|\gamma| + 20|\gamma|^3\right)}{36},$$

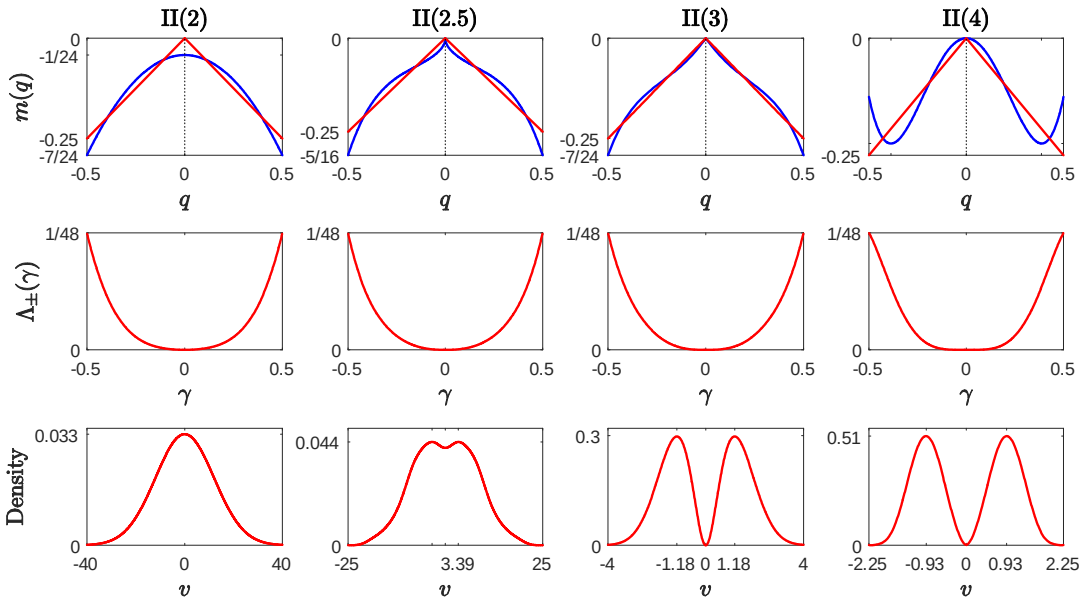

Figure 6: $m(q)$, $\Lambda_{\pm}(\gamma)$ and LSE Asymptotic Distributions in CTR

$\varpi = \frac{f_0\delta_{q0}^2\mathbb{E}\left[\varepsilon^2|q=\gamma_0\pm\right]}{3} = \frac{1}{3}$, and $\lambda = \frac{2}{9}$ (different from $\frac{1}{6}f_0\delta_{q0}^2 = \frac{1}{6}$ – the λ in HLS), which implies $\varphi = \phi = 1$ and $\omega = \frac{\varpi}{4\lambda^2} = \frac{27}{16}$. For other α values, set

$$m_1(q) = a - b_1|q|^{\alpha-2} - b_2q^4 \text{ and } m_2(q) = a - b_1q^{\alpha-2} - b_2q^4$$

with $a \leq 0$ and $b_1 > 0$. Now,

$$\Lambda_{\pm}(\gamma) = -\frac{a}{2}|\gamma|^2 + \frac{b_1}{\alpha}|\gamma|^{\alpha} + \frac{b_2}{6}|\gamma|^6.$$

In II(2), set $\alpha = 4$, $a = -\frac{1}{24}$, and $b_1 = 1$, $b_2 = 0$. This setup indicates that II(2) does not require $m_1(\gamma_0) \neq m_2(\gamma_0)$ as in I(1). The asymptotic distribution of $\widehat{\gamma}$ is

$$-\frac{S_{\gamma\beta}}{S_{\gamma\gamma}}\left(S_{\beta\beta} - \frac{S_{\beta\gamma}S_{\gamma\beta}}{S_{\gamma\gamma}}\right)^{-1}W,$$

where $S_{\gamma\beta} = \left(\frac{1}{24}, 0, -\frac{1}{24}, 0\right)$, $S_{\gamma\gamma} = \frac{1}{24}$, $M_0 = \begin{pmatrix} 1/2 & -1/8 \\ -1/8 & 1/24 \end{pmatrix}$, $\overline{M}_0 = \begin{pmatrix} 1/2 & 1/8 \\ 1/8 & 1/24 \end{pmatrix}$, and

$$\begin{aligned}
\Sigma_0 &= \mathbb{E}\left[\begin{pmatrix} 1 \\ q \end{pmatrix}(1,q)\left(\varepsilon - \frac{1}{24} - q^2 - \frac{1}{2}q\right)^2 1(q \leq 0)\right] = M_0 + \begin{pmatrix} 1/5760 & -1/23040 \\ -1/23040 & 1/60480 \end{pmatrix}, \\
\overline{\Sigma}_0 &= \mathbb{E}\left[\begin{pmatrix} 1 \\ q \end{pmatrix}(1,q)\left(\varepsilon - \frac{1}{24} - q^2 + \frac{1}{2}q\right)^2 1(q > 0)\right] = \overline{M}_0 + \begin{pmatrix} 1/5760 & 1/23040 \\ 1/23040 & 1/60480 \end{pmatrix}.
\end{aligned}$$

Note that there is an extra terms besides M_0 and $\overline{M}_0$ in Σ_0 and $\overline{\Sigma}_0$ due to misspecification. In II(2.5), set $\alpha = 2.5$, $a = 0$, $b_1 = \frac{15\sqrt{2}}{112}$, and $b_2 = \frac{20}{7}$. In the asymptotic distribution of $\widehat{\gamma}$,

$$\Omega_{\pm} = f_0 \mathbb{E}\left[\varepsilon^2 \mathbf{x}\mathbf{x}' | q = \gamma_0 \pm\right] = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \ \Upsilon_{\pm} = -\frac{f_0 \delta_{q0}}{2} \mathbb{E}\left[\varepsilon^2 \mathbf{x} | q = \gamma_0 \pm\right] = \begin{pmatrix} -\frac{1}{2} \\ 0 \end{pmatrix} \text{ and } \varpi = \frac{1}{3}.$$

From the form of $\Omega_{\pm}$ and $\Upsilon_{\pm}$, $\left(\Xi_1^{\pm}(v)', \Xi_2^{\pm}(v)\right)'$ degenerates to a two-dimensional Brownian motion on $[0, \infty)$. From Example 6, $\lambda = \frac{b_1 f_0 \delta_{q0}}{\alpha} = \frac{b_1}{\alpha}$, and $\psi_+ = \psi = \begin{pmatrix} \frac{\lambda}{\delta_{q0}} \frac{\alpha}{\alpha - 1} \\ 0 \\ -\frac{\lambda}{\delta_{q0}} \frac{\alpha}{\alpha - 1} \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{b_1}{\alpha - 1} \\ 0 \\ -\frac{b_1}{\alpha - 1} \\ 0 \end{pmatrix} = -\psi_-$. Combining the form of ψ and the degeneration of $\left(\Xi_1^{\pm}(v)', \Xi_2^{\pm}(v)\right)'$, we can see that only the first component of Z_{β_1} and Z_{β_2} contributes to the asymptotic distribution of $\widehat{\gamma}$. Similarly as in II(2), there is an extra term besides M_0 and $\overline{M}_0$ in Σ_0 and $\overline{\Sigma}_0$. Because there is no closed-form density for this asymptotic distribution, we simulate it. In II(4), set $\alpha = 4$, $a = 0$, $b_1 = 3$, and $b_2 = -10$; then $S_{\beta_\ell \gamma^2} = \frac{1}{2}(1, 0)'$, $\lambda = \frac{b_1}{\alpha} = \frac{3}{4}$, and $\varpi = \frac{1}{3}$, so $\mu = 2\lambda - S_{\gamma^2 \beta_\ell}\left(M_0^{-1} + \overline{M}_0^{-1}\right) S_{\beta_\ell \gamma^2}/4 = \frac{1}{2}$, which implies $\varphi = \phi = 1$ and $\omega = \frac{\varpi}{\mu^2} = \frac{4}{3}$. As in Figure 5, we show $m(q)$ in the first row, $\Lambda_{\pm}(\gamma)$ in the second row, and the asymptotic density of $\widehat{\gamma}$ in the third row of Figure 6.

Before discussing the asymptotic density of $\widehat{\gamma}$, first check some of its quantitative properties. When $2 \leq \alpha < 2.5$, the asymptotic distribution is normal and has a closed-form density.[15] When $2.5 < \alpha \leq 4$,

$$P\left(Z_\gamma(\alpha) \leq x\right) = P\left(\omega^{\frac{1}{2\alpha - 3}} \zeta(\varphi, \phi; \alpha/3)^{1/3} \leq x\right) = P\left(\zeta(\varphi, \phi; \alpha/3) \leq x^3 / \omega^{\frac{3}{2\alpha - 3}}\right) := F_{\alpha/3}\left(x^3 / \omega^{\frac{3}{2\alpha - 3}}; \varphi, \phi\right),$$

so the density of $Z_\gamma(\alpha)$ at x is $f_{\alpha/3}\left(\frac{x^3}{\omega^{\frac{3}{2\alpha - 3}}}; \varphi, \phi\right) \frac{3x^2}{\omega^{\frac{3}{2\alpha - 3}}}$, where $f_{\alpha/3}(\cdot; \varphi, \phi)$ is the density of $\zeta(\varphi, \phi; \alpha/3)$. In other words, $Z_\gamma(\alpha)$ should be bimodal and has a density zero at 0, which contrasts the usual asymptotic density which is unimoal and the mode is at zero. Note that $5/6 < \alpha/3 \leq 4/3$, so the density of $\zeta(\varphi, \phi; \alpha/3)$ should have a cusp at 0 from Figure 5, but the density of $Z_\gamma(\alpha)$ is zero at 0. The graphs in the third row of Figure 6 satisfy these properties, where $Z_\gamma(3)$ has a closed-form density but $Z_\gamma(4)$ does not so we simulate it. From Figure 6, we can also see that II(2.5) is a turning point from a unimodal asymptotic distribution to a bimodal asymptotic distribution; when $\alpha > 2.5$, the asymptotic distribution is not only bimodal, but the density at 0 is zero.

10 Conclusion

In this paper, we develop the asymptotic theory for the least squares estimator in threshold regression under misspecification. It turns out that this asymptotic distribution depends on the fitted model being DTR or CTR and also on the rate of the limit objective function shrinking to zero in the direction of threshold parameter. Our asymptotic theory includes many theories developed in the literature as special cases; actually, only three special cases are discussed until now. Besides the point identified model, we also discuss the partial identified and fully unidentified models. For inference on the threshold point, we focus on the LR statistic whose asymptotic null distribution is derived regardless of the model is point identified or not and

[15] When $2 < \alpha < 2.5$, the asymptotic distribution of $\widehat{\gamma}$ is $-\frac{\alpha - 1}{\alpha} \frac{\psi' Z_\beta}{\lambda}$ which is a zero-mean normal as in II(2) so is not shown in Figure 6.

is DTR or CTR. Although our asymptotic theory is thorough in the sense that all cases are discussed when we know which case we are in, there is an important unsolved problem – how to conduct uniform inference on the threshold point without knowing the form of misspecification. Recently, Yu (2020) tries to do this work.

References

Angrist, J., V. Chernozhukov and I. Fernández-Val., 2006, Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure, *Econometrica*, 74, 539-563.

Bai, J., 1997a, Estimating Multiple Breaks One At a Time, *Econometric Theory*, 13, 315-352.

Bai, J., 1997b, Estimation of a Change Point in Multiple Regression Models, *Review of Economics and Statistics*, 79, 551-563.

Bai, J., H. Chen, T.T.-L. Chong and S.X. Wang, 2008, Generic Consistency of the Break-Point Estimators under Specification Errors in a Multiple-Break Model, *Econometrics Journal*, 11, 287-307.

Banerjee, M. and I.W. McKeague, 2007, Confidence Sets for Split Points in Decision Trees, *Annals of Statistics*, 35, 543-574.

Breiman, L., J.H. Friedman, R.A. Olshen and C.J. Stone, 1984, *Classification and Regression Trees*, Belmont, CA: Wadsworth.

Bühlmann, P. and B. Yu, 2002, Analyzing Bagging, *Annals of Statistics*, 30, 927-961.

Card, D., D. Lee, P. Pei and W. Weber, 2015, Inference on Causal Effects in a Generalized Regression Kind Design, *Econometrica*, 83, 2453-2483.

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

Chan, K.S. and R.S. Tsay, 1998, Limiting Properties of the Least Squares Estimator of a Continuous Threshold Autoregressive Model, *Biometrika*, 85, 413-426.

Chong, T.T.-L., 1995, Partial Parameter Consistency in a Misspecified Structural Change Model, *Economics Letters*, 49, 351–357.

Chong, T.T.-L., 2003, Generic Consistency of the Break-Point Estimator under Specification Errors, *Econometrics Journal*, 6, 167-192.

Dykstra, R. and C. Carolan, 1999, The Distribution of the Argmax of Two-Sided Brownian Motion with Quadratic Drift, *Journal of Statistical Computation and Simulation*, 63, 47-58.

Efron, B. and T.J. Hastie, 2016, *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science*, New York: Cambridge University Press.

Fan, J., and W.-Y. Zhang, 2008, Statistical Methods with Varying Coefficient Models, *Statistics and Its Interface*, 1, 179-195.

Feder, P.I., 1975a, On Asymptotic Distribution Theory in Segmented Regression Problems - Identified Case, *Annals of Statistics*, 3, 49-83.

Feder, P.I., 1975b, The Log Likelihood Ratio in Segmented Regression, *Annals of Statistics*, 3, 84-97.

Friedman, J.H., 1991, Multivariate Adaptive Regression Splines (with discussion), *Annals of Statistics*, 19, 1-141.

Galvao, A.F., G. Montes-Rojas and J. Olmo, 2011, Threshold Quantile Autoregressive Models, *Journal of Time Series Analysis*, 32, 253-267.

Gonzalo, J. and J.-Y. Pitarakis, 2002, Estimation and Model Selection Based Inference in Single and Multiple Threshold Models, *Journal of Econometrics*, 110, 319-352.

Gonzalo, J. and M. Wolf, 2005, Subsampling Inference in Threshold Autoregressive Models, *Journal of Econometrics*, 127, 201-224.

Groeneboom, P., 1989, Brownian Motion with a Parabolic Drift and Airy Functions, *Probability Theory and Related Fields*, 81, 79-109.

Groeneboom, P. and J.A. Wellner, 2001, Computing Chernoff's Distribution, *Journal of Computational and Graphical Statistics*, 10, 388-400.

Hall, A.R. and A. Inoue, 2003, The Large Sample Behavior of the Generalized Method of Moments Estimator in Misspecified Models, *Journal of Econometrics*, 114, 361-394.

Hansen, B.E., 2000, Sample Splitting and Threshold Estimation, *Econometrica*, 575-603.

Hansen, B.E., 2017, Regression Kink With an Unknown Threshold, *Journal of Business & Economic Statistics*, 35, 228-240.

Hastie, T.J., R.J. Tibshirani and J.H. Friedman, 2009, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd edition, New York: Springer Verlag.

Heckman, J.J. and E.J. Vytlacil, 2005, Structural Equations, Treatment Effects, and Econometric Policy Evaluation, *Econometrica*, 73, 669-738.

Hidalgo, J., 1995, A Nonparametric Conditional Moment Test for Structural Stability, *Econometric Theory*, 11, 671-698.

Hidalgo, J., J. Lee and M.H. Seo, 2019, Robust Inference for Threshold Regression Models, *Journal of Econometrics*, 210, 291-309.

Kim, H.J. and D. Siegmund, 1989, The Likelihood Ratio Test for a Change-point in Single Linear Regression, *Biometrika*, 76, 409-423.

Koo, B. and M.H. Seo, 2015, Structural-break Models Under Mis-specification: Implications for Forecasting, *Journal of Econometrics*, 166-181.

Perron, P. and Y. Yamamoto, 2015, Using OLS to Estimate and Test for Structural Changes in Models with Endogenous Regressors, *Journal of Applied Econometrics*, 30, 119-144.

Petruccelli, J.D., 1992, On the Approximation of Time Series by Threshold Autoregressive Models, *Sankhya Series B*, 54, 54-61.

Porter, J. and P. Yu, 2015, Regression Discontinuity Designs with Unknown Discontinuity Points: Testing and Estimation, *Journal of Econometrics*, 189, 132-147.

Resnick, S.I., 1987, *Extreme Values, Regular Variation, and Point Processes*, New York: Springer-Verlag.

Seneta, E., 1976, Regularly Varying Functions. *Lecture Notes in Mathematics*, Vol. 508, Berlin: Springer Verlag.

Seo, M.H., 2015, Threshold Regression Under Misspecification, mimeo, LSE.

Tong, H., 1982, Discontinuous Decision Processes and Threshold Autoregressive Time Series Modelling, *Biometrika*, 69, 274-276.

Tong, H. and K.S. Lim, 1980, Threshold Autoregression, Limit Cycles and Cyclical Data, *Journal of the Royal Statistical Society, Series B*, 42, 245-292.

White, H., 1980, Using Least Squares to Approximate Unknown Regression Functions, *International Economic Review*, 21, 149-170.

White, H., 1981, Consequences and Detection of Misspecified Nonlinear Regression Models, *Journal of the American Statistical Association*, 76, 419–433.

White, H., 1982, Maximum Likelihood Estimation of Misspecified Models, *Econometrica*, 50, 1-25.

Yu, P., 2012, Likelihood Estimation and Inference in Threshold Regression, *Journal of Econometrics*, 2012, 167, 274-294.

Yu, P., 2013, Inconsistency of 2SLS Estimators in Threshold Regression with Endogeneity, *Economics Letters*, 120, 532-536.

Yu, P., 2014, The Bootstrap in Threshold Regression, *Econometric Theory*, 30, 676-714.

Yu, P., 2015a, Adaptive Estimation of the Threshold Point in Threshold Regression, *Journal of Econometrics*, 189, 83-100.

Yu, P., 2015b, Consistency of the Least Squares Estimator in Threshold Regression with Endogeneity, *Economics Letters*, 131, 41-46.

Yu, P., 2016, Treatment Effects Estimators Under Misspecification, mimeo.

Yu, P., 2019, On Inferences Based on OLS in Structural Change Models with Endogenous Regressors, mimeo.

Yu, P., 2020, Misspecification-Robust Inference in Threshold Regression, work in progress.

Yu, P., Q. Liao and P.C.B. Phillips, 2018, Inferences and Specification Testing in Threshold Regression with Endogeneity, mimeo.

Yu, P., Q. Liao and P.C.B. Phillips, 2019, New Control Function Approaches in Threshold Regression with Endogeneity, mimeo.

Yu, P. and P.C.B. Phillips, 2018a, Threshold Regression with Endogeneity, *Journal of Econometrics*, 203, 50-68.

Yu, P. and P.C.B. Phillips, 2018b, Threshold Regression Asymptotics: From the Compound Poisson Process to Two-Sided Brownian Motion, *Economics Letters*, 172, 123-126.

Yu, P. and P.C.B. Phillips, 2019, Calibrating the Confidence Intervals in Threshold Regression, mimeo.

Yu, P. and Y. Zhao, 2013, Asymptotics for Threshold Regression Under General Conditions, *Econometrics Journal*, 16, 430-462.