University of Cyprus
Department of Economics

Working Paper 10-2017

Endogeneity in Semiparametric Threshold Regression

Andros Kourtellos, Thanasis Stengos, and Yiguo Sun

Department of Economics, University of Cyprus, P.O. Box 20537, 1678 Nicosia, Cyprus
Tel.: +357-22893700, Fax: +357-22895028, Web site: http://www.ucy.ac.cy/econ/en

Endogeneity in Semiparametric Threshold Regression*

Andros Kourtellos[†] **Thanasis Stengos**[‡] **Yiguo Sun**[§]

This Draft: November 29, 2017

Abstract

In this paper, we investigate semiparametric threshold regression models with endogenous threshold variables based on a nonparametric control function approach. Using a series approximation we propose a two-step estimation method for the threshold parameter. For the regression coefficients we consider least-squares estimation in the case of exogenous regressors and two-stage least-squares estimation in the case of endogenous regressors. We show that our estimators are consistent and derive their asymptotic distribution for weakly dependent data. Furthermore, we propose a test for the endogeneity of the threshold variable, which is valid regardless of whether the threshold effect is zero or not. Finally, we assess the performance of our methods using a Monte Carlo simulation.

Keywords: control function, series estimation, threshold regression.

JEL Classification Codes: C14, C24, C51

*The first author acknowledges that this research has received funding from Marie Skodowska-Curie Actions (Work Programme 2016-17) of the Excellence Science Pillar of the Horizon 2020 Research and Innovation Programme of the European Union under REA grant agreement No. 707990. We would also like to thank participants at the 69th European Meeting of the Econometric Society (ESEM 2017) and 26th Midwest Econometric Conference (MEG 2017).

[†]Department of Economics, University of Cyprus, P.O. Box 537, CY 1678 Nicosia, Cyprus, email: andros@ucy.ac.cy.

[‡]Department of Economics and Finance, University of Guelph, Guelph, Ontario N1G 2W1, Canada, email: tstengos@uoguelph.ca.

[§]Department of Economics and Finance, University of Guelph, Guelph, Ontario N1G 2W1, Canada, email: yisun@uoguelph.ca.

1 Introduction

There are several economic theories that suggest threshold-like structures. For example, debt levels that are above a particular threshold value may have different implications for economic growth compared to more moderate levels of debt (e.g., Reinhart and Rogoff (2010)). Another example is motivated by models of intergenerational mobility and poverty traps. Under certain conditions, credit constraints (e.g., Galor and Zeira (1993)) or neighborhood influences (e.g., Durlauf (1996)) may generate a linear transmission of socioeconomic status within a group of individuals, while different levels of credit constraints or neighborhood quality produce different intercepts and slopes. These types of nonlinearities can be modeled by the threshold regression, which sorts the observations on the basis of some observed threshold variable into regimes each of which obeys the same linear model. Threshold-type regressions are also popular in nonlinear time-series. For example, the threshold-autoregressive (TAR) model allows the autoregressive coefficients to change between the regimes while they are constant in each regime based on some time-series which acts as a threshold variable; see Tong (1990) for a review on TAR models.[1]

The first generation of threshold regression models developed inference under the assumption of exogenous or predeteremined threshold variables. Notable examples include the works of Chan (1993), Hansen (2000), Caner and Hansen (2004), Seo and Linton (2007), Gonzalo and Wolf (2005), Yu (2012), and Yu (2013). Recently, there is a growing interest in threshold regression models that accommodate endogenous threshold variables in order to identify the underlying mechanisms of such theories. Kourtellos, Stengos, and Tan (2016)

[1]The threshold regression is also related to a number of smoothed parametric or semiparametric regressions. For example, the threshold regression can be viewed as a special case of the varying coefficient model of Hastie and Tibshirani (1993) when the varying coefficients take the form of indicator functions. Similarly, the Smooth Transition Autoregressive (STAR) model (e.g., Terasvirta, Granger, and Nelson (1986)) replaces the indicator function in TAR with parametric smoothed transition functions such as the logistic and the exponential functions.

propose estimation and inference for a threshold regression model that allows for an endogenous threshold variable as well as for endogenous regressors under certain parametric assumptions and using the diminishing threshold effect asymptotic framework proposed by Hansen (2000) and Caner and Hansen (2004). In particular, in the spirit of Heckman's sample selection method, they account for the endogeneity bias by including regime specific inverse Mills ratio bias correction terms in the threshold regression. Seo and Shin (2016) study a dynamic threshold panel data model, which allows both regressors and threshold effect to be endogenous. In particular, they propose first-difference GMM and two-step least squares estimators as well as a bootstrap-based testing procedure for the presence of threshold effect. An alternative method to deal with endogeneity was proposed by Yu and Phillips (2015) who propose a nonparametric estimator of the threshold parameter, namely the integrated difference kernel estimator. Using the fixed threshold effect framework of Chan (1993) and assuming *i.i.d.* sample, they show that the threshold parameter can be partially identified and estimated without the use of any instruments at the rate n. They also show that while instrumental variables are not necessary for the identification and estimation of the threshold effect parameters at the rate n, regime-specific regression coefficients can only be identified and estimated at the usual $\sqrt{n}$ rate when instrumental variables are available.

In this paper we propose a semiparametric approach to deal with the endogeneity of threshold variable and regressors that avoids the challenges of nonparametric estimators and at the same time relaxes parametric assumptions. Specifically, we propose to estimate the threshold parameter using a concentrated least squares (CLS) method of estimation which includes a regime specific control function estimated by series estimation method based on polynomial and splines.[2] We derive the limiting distribution of our proposed estimators for both the threshold and slope parameters as well as for the smooth control function.

[2] Chen (2007) provides a recent survey of large sample results on nonparametric and semiparametric estimation of econometric models using the method of sieves.

Finally, we propose a test for the endogeneity of the threshold variable and show that limit distribution of the test statistic is the same under the null hypothesis regardless of whether the threshold effect exists or not.

The rest of the paper is organized as follows. In Section 2, we propose a semiparametric threshold model and derive limiting results for the proposed estimator in the case of exogenous regressors. Section 3 extends the result to the case of endogenous regressors. Section 4 considers testing for the endogeneity of the threshold variable. Section 5 reports some Monte Carlo simulation results to assess the finite sample performance of our methods. Section 6 concludes. We delay all the mathematical proofs in the Appendix. Supplementary proofs are given in Kourtellos, Stengos, and Sun (2017)-henceforth, we will refer to this as the Online Appendix.

2 Endogenous threshold variable

We begin by presenting the basic parametric structural threshold regression (or STR) model

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta}_1 + \sigma_1 u_t, \quad q_t \leq \gamma_0 \tag{2.1}$$

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta}_2 + \sigma_2 u_t, \quad q_t > \gamma_0 \tag{2.2}$$

for $t = 1, 2, \ldots, n$, where y_t is dependent variable, x_t is a $d_x \times 1$ vector of regressors, β_1 and β_2 are regime-specific coefficients, and u_t is an error with zero mean and unit variance. q_t is a scalar endogenous threshold variable with γ_0 being the threshold value. A reduced form equation for q_t is given by

$$q_t = \boldsymbol{z}_t'\boldsymbol{\pi}_q + v_{q,t}, \; t = 1, 2, \ldots, n, \tag{2.3}$$

where $\boldsymbol{z}_t$ is a $d_z \times 1$ vector of instrument variables for q_t satisfying $E(v_{q,t}\boldsymbol{z}_t) = 0$ for all t. The endogeneity of the threshold variable q_t comes from the correlation between $\{u_t\}$ and $\{v_{q,t}\}$.

The STR model is analogous to the sample selection model of Heckman (1979) in the way that the endogenous dummy variable is modeled. Specifically, in the sample selection model, the variable q_t that determines the assignment of observations to regimes is latent, but the assignment is known (given by the dummy variable). However, in the STR model, we observe q_t, but the threshold value γ_0 is an unknown parameter to be estimated.

Assuming $E(u_t|\boldsymbol{x}_t, \boldsymbol{z}_t, v_{q,t}) = E(u_t|v_{q,t}) = g(v_{q,t})$ almost surely, we obtain

$$
\begin{aligned}
& E\left(u_t|\boldsymbol{x}_t, \boldsymbol{z}_t, v_{q,t} \leq \gamma_0 - \boldsymbol{z}_t'\boldsymbol{\pi}_q\right) \\
= \; & \frac{E\left[g\left(v_{q,t}\right) I\left(v_{q,t} \leq \gamma_0 - \boldsymbol{z}_t'\boldsymbol{\pi}_q\right)|\boldsymbol{x}_t, \boldsymbol{z}_t\right]}{F_{v|(x,z)}\left(\gamma_0 - \boldsymbol{z}_t'\boldsymbol{\pi}_q\right)} \equiv h_1(\boldsymbol{x}_t, \boldsymbol{z}_t, \gamma_0 - \boldsymbol{z}_t'\boldsymbol{\pi}_q),
\end{aligned}
$$

and similarly $E\left(u_t|\boldsymbol{x}_t, \boldsymbol{z}_t, v_{q,t} > \gamma_0 - \boldsymbol{z}_t'\boldsymbol{\pi}_q\right) \equiv h_2(\boldsymbol{x}_t, \boldsymbol{z}_t, \gamma_0 - \boldsymbol{z}_t'\boldsymbol{\pi}_q)$, where $g(\cdot)$ is an unknown function, $F_{v|(x,z)}(\cdot)$ denotes the conditional cdf of $v_{q,t}$ given $\boldsymbol{x}_t = x$ and $\boldsymbol{z}_t = z$, and $I(A)$ is an indicator function equal to one if event A occurs and zero otherwise. Our estimation method introduced below can be used to estimate the unknown functions, $h_j(\boldsymbol{x}_t, \boldsymbol{z}_t, \gamma_0 - \boldsymbol{z}_t'\boldsymbol{\pi}_q)$, however, it will incur severe curse-of-dimsionality problem. Without loss of essence, we therefore consider the case that $h_j(\boldsymbol{x}_t, \boldsymbol{z}_t, \gamma_0 - \boldsymbol{z}_t'\boldsymbol{\pi}_q) = h_j(\gamma_0 - \boldsymbol{z}_t'\boldsymbol{\pi}_q)$ for $j = 1$, 2 and all t. Thus, we can rewrite model (2.1) and model (2.2), respectively, as

$$
y_t = \boldsymbol{x}_t'\boldsymbol{\beta}_1 + \sigma_1 h_1(\gamma_0 - \boldsymbol{z}_t'\boldsymbol{\pi}_q) + \varepsilon_{1t},\ q_t \leq \gamma_0 \tag{2.4}
$$

$$
y_t = \boldsymbol{x}_t'\boldsymbol{\beta}_2 + \sigma_2 h_2(\gamma_0 - \boldsymbol{z}_t'\boldsymbol{\pi}_q) + \varepsilon_{2t},\ q_t > \gamma_0 \tag{2.5}
$$

where $\varepsilon_{jt} = \sigma_j\left[u_t - h_j(\gamma_0 - \boldsymbol{z}_t'\boldsymbol{\pi}_q)\right]$ for j=1,2.

Assuming that $(u_t, v_{q,t})$ are jointly normally distributed and that $\sigma_1 = \sigma_2$, Kourtellos, Stengos, and Tan (2016) show that the two control functions take the form of inverse Mills ratio bias correction terms

$$h_1(\gamma - \boldsymbol{z}_t'\boldsymbol{\pi}_q) = -\frac{\phi(\gamma - \boldsymbol{z}_t'\boldsymbol{\pi}_q)}{\Phi(\gamma - \boldsymbol{z}_t'\boldsymbol{\pi}_q)} \text{ and } h_2(\gamma - \boldsymbol{z}_t'\boldsymbol{\pi}_q) = \frac{\phi(\gamma - \boldsymbol{z}_t'\boldsymbol{\pi}_q)}{1 - \Phi(\gamma - \boldsymbol{z}_t'\boldsymbol{\pi}_q)} \tag{2.6}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal pdf and cdf, respectively. However, in practice, it is often expected that the joint normality assumption is violated, which leads to misspecification of the two inverse Mills ratio terms. In order to avoid this potential model misspecification problem, this paper is aimed to simultaneously estimate all the unknown parameters appearing in models (2.4) and (2.5) and the unknown inverse Mills ratio bias terms without imposing the joint normality assumption.

As the functional forms of $h_1(\cdot)$ and $h_2(\cdot)$ are both unknown, we cannot identify $(\gamma_0, \sigma_1, \sigma_2)$ from $h_1(\cdot)$ and $h_2(\cdot)$. Therefore, our semiparametric threshold regression model is given by

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta}_1 I(q_t \le \gamma_0) + \boldsymbol{x}_t'\boldsymbol{\beta}_2 I(q_t > \gamma_0) + h_1(\boldsymbol{z}_t'\boldsymbol{\pi}_q) I(q_t \le \gamma_0) + h_2(\boldsymbol{z}_t'\boldsymbol{\pi}_q) I(q_t > \gamma_0) + \varepsilon_t, \tag{2.7}$$

where $\varepsilon_t = \varepsilon_{1t} I(q_t \le \gamma_0) + \varepsilon_{2t} I(q_t > \gamma_0)$ with $\varepsilon_{jt} = \sigma_j u_t - h_j(\boldsymbol{z}_t'\boldsymbol{\pi}_q)$ for $j = 1, 2$. Using the definitions $\boldsymbol{\delta} = \boldsymbol{\beta}_1 - \boldsymbol{\beta}_2$, $\eta(w) = h_1(\omega) - h_2(\omega)$, where $w \in \mathcal{R}$, we can rewrite (2.7) as

$$y_t = \boldsymbol{x}_t'\beta_2 + \boldsymbol{x}_t'\boldsymbol{\delta} I(q_t \le \gamma_0) + h_2(\boldsymbol{z}_t'\boldsymbol{\pi}_q) + \eta(\boldsymbol{z}_t'\boldsymbol{\pi}_q) I(q_t \le \gamma_0) + \varepsilon_t, \tag{2.8}$$

Note that we set $h_1(0) = h_2(0) = 0$ for identification purpose when $\boldsymbol{x}_t$ contains a constant term one. When the threshold variable q_t is exogenous, i.e., $g(v) \equiv 0$, the control functions $h_1(w)$ and $h_2(w)$ are omitted from model (2.7).

As in Hansen (2000), our asymptotic results are derived in the framework of "small threshold" effect; i.e., we assume that the exogenous threshold effect, $\boldsymbol{\delta} = \boldsymbol{\delta}_n$, and the endogenous threshold bias correction term, $\eta(\omega) = \eta_n(w)$, both approach to zero slowly as n diverges. It means that the endogeneity bias vanishes in large samples and that endogenous regime changes temporally exist around a threshold value. Below, we summarize the assumptions that support our model (2.8).

Assumption 1. (i) $\{(\boldsymbol{x}_t', q_t, \boldsymbol{z}_t', u_t)\}$ is a strictly stationary strong mixing sequence with the mixing coefficients of size $-r/(r-2)$ for some $r > 2$;

(ii) $E(\boldsymbol{z}_t v_{q,t}) = 0$, $E\left(v_{q,t}^2\right) = \sigma_v^2$, $E(\|\boldsymbol{z}_t v_{q,t}\|^r) < \infty$ and $E\left(\|\boldsymbol{z}_t\|^{r+\delta}\right) < \infty$ for some $\delta > 0$ and $E(\boldsymbol{z}_t \boldsymbol{z}_t')$ exists and is non-singular;

(iii) (a) $\{(u_t, \mathcal{F}_{n,t})\}_{t=1}^n$ is a martingale difference sequence with $E(u_t^2|\mathcal{F}_{n,t-1}) < \infty$, where $\mathcal{F}_{n,t}$ is the smallest sigma-field generated from $\left\{\left(\boldsymbol{x}_{s+1}', q_s, \boldsymbol{z}_{s+1}', u_s\right) : 1 \leq s < t \leq n\right\}$; (b) $E(u_t|\mathcal{F}_{n,t-1}, v_{q,t} = v) = g(v)$ for any v; (c) $v_{q,t}$ is independent of $\mathcal{F}_{n,t-1}$ for any $t \geq 1$.

(iv) for any $\boldsymbol{\lambda} \neq 0$, there is no measurable function $m(v)$ such that $\boldsymbol{x}'\boldsymbol{\lambda} = m(\boldsymbol{z}'\boldsymbol{\pi}_q)$ when $q_t \leq \gamma_0$ and when $q_t > \gamma_0$;

(v) $\eta_n(\omega) = n^{-\varrho}\eta_0(\omega)$ and $\boldsymbol{\delta}_n = n^{-\varsigma}\boldsymbol{\delta}_0$ for some $0 < \varsigma, \varrho < 1/2$, $\boldsymbol{\delta}_0 \neq 0$, and $\eta_0(\omega) \neq 0$ over at least one non-empty interval.

By Theorem 5.23 in White (2001), Assumptions **1**(i)-(ii) are moment bounds. Assumption **1**(iii) states that $(\boldsymbol{x}_t', \boldsymbol{z}_t')'$ is contemporaneously exogenous in the model (2.1)-(2.2) and (2.3). Because the central limit theorem for martingale difference sequences only requires well-behaved moment conditions up to the second order, and the proofs for the limit distribution of our proposed estimator implicitly incurs unbounded moments beyond the second order, we therefore assume that $\{(\varepsilon_t, \mathcal{F}_{n,t})\}_{t=1}^n$ is a martingale difference sequence which is ensured under Assumption **1**(iii). Assumption **1**(iv) is an identification condition similar to Assumption 2.1 in Newey (2009). It is readily seen that $\boldsymbol{z}'\boldsymbol{\pi}_q$ cannot equal a linear combination of $\boldsymbol{x}$, because $h_j(\omega)$ are unknown for $j = 1$, 2 in model (2.7). Assumption **1**(v) regulates how fast the threshold effects vanish as the sample size increases. The reason for this assumption is that in the case of fixed threshold effect the sampling distribution of the threshold estimator is too complicated for inference. Specifically, Chan (1993) showed that the asymptotic distribution of the threshold estimator depends on a host of nuisance parameters including the marginal distribution of the regressors. To overcome this difficulty, Hansen (2000) employed Assumption **1**(v) that assumes that the difference between the slope coefficients of the two regimes decreases as the sample size grows. That said, the diminishing threshold effects framework is not the only way to deal with this problem. Li and Ling (2012) propose a numerical approach to simulate the limiting distribution of the the estimator of the threshold parameter based on a simulation of a related compound Poisson process. An alternative approach is to introduce smoothness in the objective criterion to achieve asymptotic normality. For example, Seo and Linton (2007) propose a smoothed least squares estimation strategy that leads to asymptotic normality by smoothing the objective function by replacing the indicator function in the objective function with an integrated kernel. In the same spirit, Seo and Shin (2016) exploit the smoothness of the GMM criterion to achieve asymptotic normality of the threshold estimate, regardless of whether the threshold effects are fixed or diminishing.

In the next section, we describe our estimation method and study its asymptotic properties.

2.1 Estimation

Let $\{\phi_1(\omega), \phi_2(\omega), \ldots\}$ be a sequence of orthonormal basis functions in $L_2(-\infty, \infty)$ space if $\boldsymbol{z}_t'\boldsymbol{\pi}_q$ takes value from the real line or $L_2[0,1]$ space if $\boldsymbol{z}_t'\boldsymbol{\pi}_q$ has a finite support. We approximate $h_j(\omega)$ (j=1,2) and $\eta_0(\omega)$ by

$$h_j^*(\omega) = \boldsymbol{\alpha}_{L_n,j}'\boldsymbol{\Phi}_{L_n}(\omega) \text{ and } \eta_0^*(\omega) = \boldsymbol{\alpha}_{L_n,0}'\boldsymbol{\Phi}_{L_n}(\omega)$$

respectively, where $\boldsymbol{\Phi}_{L_n}(\omega) = [\phi_1(\omega), \ldots, \phi_{L_n}(\omega)]'$ denotes an $L_n \times 1$ vector. As $\eta_n(\omega) = h_1(\omega) - h_2(\omega) = n^{-\varrho}\eta_0(\omega)$, we have $n^{-\varrho}\boldsymbol{\alpha}_{L_n,0} = \boldsymbol{\alpha}_{L_n,1} - \boldsymbol{\alpha}_{L_n,2}$ and $\eta_n(\omega)$ is approximated by $\eta_n^*(\omega) = n^{-\varrho}\boldsymbol{\alpha}_{L_n,0}'\boldsymbol{\Phi}_{L_n}(\omega)$. Below, we explain our proposed estimation procedure.

Step 1. Given instruments $\boldsymbol{z}_t$, the LS estimator from model (2.3) is, $\hat{\boldsymbol{\pi}}_q = \left(\sum_{t=1}^n \boldsymbol{z}_t\boldsymbol{z}_t'\right)^{-1}\sum_{t=1}^n \boldsymbol{z}_t q_t$. Assumptions **1**(i)-(ii) imply that $\hat{\boldsymbol{\pi}}_q$ exists and ensure consistency $\hat{\boldsymbol{\pi}}_q = \boldsymbol{\pi}_q + O_p\left(n^{-1/2}\right)$. We then denote the fitted value of q_t as $\hat{q}_t = \boldsymbol{z}_t'\hat{\boldsymbol{\pi}}_q$ for all t throughout the rest of this paper.

Step 2. For a given $\gamma \in \left[\underline{\gamma}, \overline{\gamma}\right]$, we estimate $\boldsymbol{\theta} = \left(\boldsymbol{\beta}_1', \boldsymbol{\alpha}_{L_n,1}', \boldsymbol{\beta}_2', \boldsymbol{\alpha}_{L_n,2}'\right)'$ from the objective function

$$\hat{\boldsymbol{\theta}}(\gamma) = \arg\min_{\theta}\sum_{t=1}^n \left[y_t - \boldsymbol{x}_{-,t}'\boldsymbol{\beta}_1 - \boldsymbol{\alpha}_{L_n,1}'\boldsymbol{\Phi}_{L_n,\gamma}^-(\hat{q}_t) - \boldsymbol{x}_{+,t}'\boldsymbol{\beta}_2 - \boldsymbol{\alpha}_{L_n,2}'\boldsymbol{\Phi}_{L_n}^+(\hat{q}_t)\right]^2, \tag{2.9}$$

where we denote $\boldsymbol{x}_{-,t} = \boldsymbol{x}_t I(q_t \le \gamma)$, $\boldsymbol{x}_{+,t} = \boldsymbol{x}_t I(q_t > \gamma)$, $\boldsymbol{\Phi}_{L_n,\gamma}^-(\hat{q}_t) = \boldsymbol{\Phi}_{L_n}(\hat{q}_t)I(q_t \le \gamma)$ and $\boldsymbol{\Phi}_{L_n,\gamma}^+(\hat{q}_t) = \boldsymbol{\Phi}_{L_n}(\hat{q}_t)I(q_t > \gamma)$. Denoting an $n \times [2(d_x + L_n)]$ matrix $\mathcal{X}_\gamma = [\mathcal{X}_{-,\gamma}, \mathcal{X}_{+,\gamma}]$, where $\mathcal{X}_{-,\gamma}$ stacks up $\left[\boldsymbol{x}_{-,t}', \boldsymbol{\Phi}_{L_n,\gamma}^{-\prime}(\hat{q}_t)\right]$ and $\mathcal{X}_{+,\gamma}$ stacks up $\left[\boldsymbol{x}_{+,t}', \boldsymbol{\Phi}_{L_n,\gamma}^{+\prime}(\hat{q}_t)\right]$, and solving (2.9) give

$$\hat{\boldsymbol{\theta}}(\gamma) = \left(\mathcal{X}_\gamma'\mathcal{X}_\gamma\right)^{-1}\mathcal{X}_\gamma'\boldsymbol{y}. \tag{2.10}$$

We can estimate the threshold parameter γ by minimizing the concentrated least squares criterion

$$\widehat{\gamma} = \arg \min_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} \sum_{t=1}^{n} \left[y_t - \mathcal{X}_{t,\gamma}' \hat{\boldsymbol{\theta}}(\gamma) \right]^2 \tag{2.11}$$

and then estimate $\boldsymbol{\theta}$ by $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\hat{\gamma})$, where the $\mathcal{X}_{t,\gamma}$ in (3.5) is the t^{th} row of the $\mathcal{X}_\gamma$.

Step 3. Calculating $\tilde{y}_t = y_t - \boldsymbol{x}_{-,t}' \hat{\boldsymbol{\beta}}_1 - \boldsymbol{x}_{+,t}' \hat{\boldsymbol{\beta}}_2$, we can re-estimate $h_2(\omega)$ and $\eta_n(\omega)$ by the local linear regression approach from $\tilde{y}_t = h_2(\hat{q}_t) + \eta_n(\hat{q}_t) I(q_t \leq \hat{\gamma}) + \tilde{\varepsilon}_t$, $t = 1, 2, \ldots, n$. We denote the estimator for $\boldsymbol{\psi}(\omega) = [h_2(\omega), \eta_n(\omega)]'$ by $\tilde{\boldsymbol{\psi}}(\omega) = \left[\tilde{h}_2(\omega), \tilde{\eta}_n(\omega)\right]'$.

Of course, one can obtain the series estimator of $h_1(\cdot)$, $h_2(\cdot)$ and $\eta_n(\cdot)$ once $\hat{\theta}$ is obtained from Step 2. However, estimating these unknown curves by the local linear regression approach enables us to derive the explicit form for the asymptotic bias term and the optimal bandwidth. We therefore add Step 3 in the paper.

2.2 Limiting results

As in Blundell, Chen, and Kristensen (2007), we denote a Hölder space $\Lambda^{\xi}(\mathcal{R})$. For any $h(\cdot)$ in $\Lambda^{\xi}(\mathcal{R})$, $h(\cdot)$ is $[\xi]$-times continuously differentiable over the real line $\mathcal{R}$ and $\left|\nabla^{[\xi]} h(\omega) - \nabla^{[\xi]} h(\omega')\right| \leq M |\omega - \omega'|^{\xi - [\xi]}$ for any $\omega \in \mathcal{R}$ and $\omega' \in \mathcal{R}$, where $[\xi]$ is the largest positive integer less than ξ. Below, we list some regularity conditions used to derive the consistency and limit distribution of our proposed estimators.

Assumption 2. (i) $E(\boldsymbol{x}_t' \boldsymbol{x}_t \boldsymbol{z}_t' \boldsymbol{z}_t) < \infty$, $E\|\boldsymbol{x}_t\|^{2r'} < \infty$ and $E\|\varepsilon_t \boldsymbol{x}_t\|^r < \infty$ for some $r' > r > 2$, where r is defined in Assumption **1** and $||\cdot||$ denotes the Euclidean norm;

(ii) for every L_n and uniformly over $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, there exist constants $\underline{c}$ and $\overline{c}$ such that $0 < \underline{c} \leq \boldsymbol{\lambda}_{\min}(\boldsymbol{\Sigma}_\gamma) \leq \boldsymbol{\lambda}_{\max}(\boldsymbol{\Sigma}_\gamma) \leq \overline{c} < \infty$, where $\boldsymbol{\Sigma}_\gamma = E\left(\mathcal{X}_{t,\gamma}^* \mathcal{X}_{t,\gamma}^{*\prime}\right)$ and $E\left(\varepsilon_t^2 \mathcal{X}_{t,\gamma}^* \mathcal{X}_{t,\gamma}^{*\prime}\right)$, and $\mathcal{X}_{t,\gamma}^*$ equals $\mathcal{X}_{t,\gamma}$ with $\hat{q}_t$ replaced by $\boldsymbol{z}_t' \boldsymbol{\pi}_q$;

(iii) q_t has a probability density function $f_q(q)$ with respect to the Lebesgue measure and $\inf_{q\in\mathcal{R}} f_q(q) > 0$, and $\eta_0(\omega)$, $h_1(\omega)$, $h_2(w)$, and $f_q(q)$ all belong to the Hölder space $\Lambda^{\xi}(\mathcal{R})$ for some $\xi > 2$, and these functions and their first- and second-order derivatives are all uniformly bounded;

(iv) $h_1(\omega)$ and $h_2(\omega)$ are squared integrable, and there exists finite constants $\boldsymbol{\alpha}_{L_n,0}$ and $\boldsymbol{\alpha}_{L_n,j}$ such that $\sup_{\omega\in\mathcal{R}}\left|\eta_0(\omega) - \boldsymbol{\alpha}'_{L_n,0}\boldsymbol{\Phi}_{L_n}(\omega)\right| \leq ML_n^{-\xi}$ and $\sup_{\omega\in\mathcal{R}}\left|h_j(\omega) - \boldsymbol{\alpha}'_{L_n,j}\boldsymbol{\Phi}_{L_n}(\omega)\right| \leq ML_n^{-\xi}$ for j=1,2;

(v) $\{\phi_l(\cdot), l = 1, 2, ...\}$ is a sequence of orthonormal basis functions in $\Lambda^{\xi}(\mathcal{R})$ and uniformly bounded over $\mathcal{R}$. Also, we denote $\sup_{\omega\in\mathcal{R}}\sum_{l=1}^{L_n}\left[\phi_l^{(s)}(\omega)\right]^2 = \|\boldsymbol{\Phi}_{L_n}\|_s^2$ for $s \geq 0$.

Assumption **2** (i)-(ii) ensures the existence of $\hat{\boldsymbol{\theta}}(\gamma)$, which is standard in the literature; see, e.g., Newey (1997) and Ozabaci, Henderson, and Su (2014). As the eigenvalues of a squared matrix are a continuous function of the matrix, the uniform boundness holds over a compact set $[\underline{\gamma}, \overline{\gamma}]$ as long as the eigenvalues are bounded pointwise. Assumption **2**(iii) is a standard smoothness condition in nonparametric estimation. Assumption **2**(iv) restricts the sieve approximation error, and it holds by Theorem 1.1 in Dzyadyk and Shevchuk (2008) if $\boldsymbol{z}_t'\boldsymbol{\pi}_q$ has compact support and $\eta_0(w)$ and $h_j(\omega)$ are all ξ-smooth. If $\boldsymbol{z}_t'\boldsymbol{\pi}_q$ has unbounded support, $(-\infty, \infty)$, Xiang (2012) showed that Assumption **2** (iv) holds for the normalized Hermite orthonormal basis functions if $\eta_0(w)$ and $h_j(w)$ are all p-smooth for some $p > 2(\xi + 1)$. In addition, Assumption **2**(v) describes the properties of the basis functions and implies $\|\boldsymbol{\Phi}_{L_n}\|_0 = O\left(L_n^{1/2}\right)$ and $\|\boldsymbol{\Phi}_{L_n}\|_1 = O\left(L_n^{3/2}\right)$; see, e.g., the normalized Hermite functions and wavelet functions defined in Blundell, Chen, and Kristensen (2007).

Assumption 3. Denote $\vartheta_n = L_n^{-\xi} + \sqrt{L_n/n} + n^{-1/2}\|\boldsymbol{\Phi}_{L_n}\|_1$. (i) $\vartheta_n = o(1)$ and $\|\boldsymbol{\Phi}_{L_n}\|_1^2 L_n/n = o(1)$; (ii) $n^{-1+2[\min(\varsigma,\varrho)-\varrho]}\|\boldsymbol{\Phi}_{L_n}\|_1^2 = o(1)$ and $n^{\min(\varsigma,\varrho)}\vartheta_n = o(1)$; (iii) $\min(\varsigma, \varrho) < 1/4$ and $\sqrt{n}L_n^{-\xi} = o(1)$.

Assumption **3**(i) is used to derive the consistency result of $\widehat{\gamma}$ and $\hat{\boldsymbol{\theta}}$ for Theorem 1, and Assumptions **3**(ii)-(iii) are used to derive below the limit distribution for $\widehat{\gamma}$ and $\widehat{\boldsymbol{\beta}}$ in Theorems 2 and 3, respectively. Below, we give the limit results for $\widehat{\gamma}$ and $\hat{\boldsymbol{\theta}}$.

Theorem 1 *Under Assumptions **1**-**3**(i), we have* $\widehat{\gamma} - \gamma_0 = o_p(1)$ *and* $\left\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right\| = O_p\left(\vartheta_n + n^{-\min(\varsigma,\varrho)}\right)$.

Theorem 1 shows the consistency of $\widehat{\gamma}$ and $\hat{\boldsymbol{\theta}}$. Compared with the conventional convergence rate of series estimator, ϑ_n contains an additional bias term of order $O_p\left(n^{-1/2}\left\|\boldsymbol{\Phi}_{L_n}\right\|_1\right)$, which results from the estimation of $\boldsymbol{\pi}_q$, the parameter appearing in the reduced-form model of the endogenous threshold variable q_t.

Theorem 2 *Under Assumptions **1**-**3**(i)(ii), we have*

$$n^{1-2\min(\varsigma,\varrho)}\left(\widehat{\gamma} - \gamma_0\right) \xrightarrow{d} \varpi T,$$

where we denote $\sigma_j^2 = E\left\{\varepsilon_{jt}^2\left[I(\varsigma \le \varrho)\,\boldsymbol{\delta}_0'\boldsymbol{x}_t + I(\varsigma \ge \varrho)\,\eta_0\left(\boldsymbol{z}_t'\boldsymbol{\pi}_q\right)\right]^2 | q_t = \gamma_0\right\}$ *for* $j = 1,\ 2$, $\omega_0 = E\left\{\left[I(\varsigma \le \varrho)\,\boldsymbol{\delta}_0'\boldsymbol{x}_t + I(\varsigma \ge \varrho)\,\eta_0\left(\boldsymbol{z}_t'\boldsymbol{\pi}_q\right)\right]^2 | q_t = \gamma_0\right\}$,

$$\varpi = \frac{\sigma_1^2}{\omega_0^2 f_q(\gamma_0)} \text{ and } T = \arg\max_{-\infty<r<\infty} T(r)$$

$T(r)$ denotes an asymmetric two-sided Brownian motion on the real line

$$T(r) = \begin{cases} -|r|/2 + W_1(-r), & \text{if } r \le 0 \\ -|r|/2 + \sqrt{\sigma_2^2/\sigma_1^2}\,W_2(r), & \text{if } r > 0 \end{cases}$$

and $W_1(r)$ and $W_2(r)$ are two independent standard Brownian motion processes defined on $[0, \infty)$.

Theorem 2 shows that the asymptotic distribution of the threshold estimate is a Chernoff-type distribution featuring unequal scales for each regime as in Kourtellos, Stengos, and Tan (2016). One difference is that the different scales that reflect the regime-specific heteroskedasticity depend on the nonparametric control functions rather than the inverse Mills ratio functions.

Letting $\hat{\boldsymbol{\beta}} = \left[\hat{\boldsymbol{\beta}}_1', \hat{\boldsymbol{\beta}}_2'\right]'$ and $\boldsymbol{\beta} = [\boldsymbol{\beta}_1', \boldsymbol{\beta}_2']'$, we obtain $\hat{\boldsymbol{\beta}}$'s limit result as follows.

Theorem 3 *Under Assumptions **1**-**3**, we have*

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{J}^{-1}\boldsymbol{\Omega}\boldsymbol{J}^{-1}\right), \tag{2.12}$$

where $\boldsymbol{J}$ and $\boldsymbol{\Omega}$ are defined by (A.20) and (A.23) in Appendix, respectively.

Theorem 3 shows that the parametric part of parameter $\boldsymbol{\beta}$ is root-n consistent and asymptotically normally distributed. Below, we will examine the third-step estimator, $\tilde{\boldsymbol{\psi}}(\omega)$. The following conditions are required for the derivation of the limit distribution of the local linear estimator of $\eta_0(\cdot)$ and $h_2(\cdot)$.

Assumption 4. (i) For some $\delta^* > \delta > 0$, $E\left(|\boldsymbol{z}_t'\boldsymbol{\pi}_q|^{2(2+\delta^*)}\right) < \infty$, $E\left(|y_t|^{2+\delta^*}|\boldsymbol{x}_t = \boldsymbol{x}, \boldsymbol{z}_t = \boldsymbol{z}\right) \leq M < \infty$ for all $x \in \mathcal{S}_x$ in the neighborhood of $\boldsymbol{z}$, and $E\left(y_0^2 + y_t^2|\boldsymbol{z}_0 = \boldsymbol{z}_0, \boldsymbol{x}_0 = \boldsymbol{x}_0, \boldsymbol{z}_t = \boldsymbol{z}, \boldsymbol{x}_t = \boldsymbol{x}_t\right) \leq M < \infty$ for all $(\boldsymbol{z}_0, \boldsymbol{x}_0, \boldsymbol{x}_t) \in \mathcal{S}_z \times \mathcal{S}_x \times \mathcal{S}_x$ in the neighborhood of $\boldsymbol{z}$. (ii) $f_{z\pi}(\omega)$, $E\left[\|\boldsymbol{z}_t\|^j|\boldsymbol{z}_t'\boldsymbol{\pi}_q = w\right]$, $E\left[I(q_t \leq \gamma_0)|\boldsymbol{z}_t'\boldsymbol{\pi}_q = w\right]$, $E\left(\boldsymbol{x}_t|\boldsymbol{z}_t'\boldsymbol{\pi}_q = \omega\right)$, and $E\left[\boldsymbol{z}_t\boldsymbol{x}_t'|\boldsymbol{z}_t'\boldsymbol{\pi}_q = w\right]$ are all twice continuously differentiable up to their second-order derivatives with respect to w, where $f_{z\pi}(\omega)$ is the probability density function of $\boldsymbol{z}_t'\boldsymbol{\pi}_q$, and $j \leq r_k$ with r_k defined in Assumption **5**; (iii) the conditional density function of q_t given $\boldsymbol{z}_t'\boldsymbol{\pi}_q = w$, $f(q|\omega)$, is continuous and uniformly bounded over its domain; (iv) $E\left(\varepsilon_t^{2j}|\boldsymbol{z}_t'\boldsymbol{\pi}_q = \omega\right)$, $E\left(\varepsilon_t^{2j}I(q_t \leq \gamma_0)|\boldsymbol{z}_t'\boldsymbol{\pi}_q = \omega\right)$, and $E\left(|\varepsilon_t|^{2(2+\delta)}|\boldsymbol{z}_t'\boldsymbol{\pi}_q = \omega\right)$are bounded in the neighborhood of w for $j = 1, 2$.

Assumption 5. (i) The kernel function $K(u)$ is a symmetric probability density function with a compact support $[-1, 1]$; (ii) $K(u)$ is continuously differentiable up to order $r_k > 2$; (iii) as $n \to \infty$, $h \to 0$, $nh^{2(r_k+3)/(r_k+2)} \to \infty$, $\|\mathbf{\Phi}_{L_n}\|_1 n^{-1/2} h^{(2-r')/(2r')} \to 0$, and $\lim_{n\to\infty} nh^5 = c_0 > 0$, where $r' > r > 2$. Also, we denote $\kappa_{i,j} = \int K^i(u) u^j du$.

Assumption **4**(i) is Condition A.2 in Cai, Fan, and Yao (2000), and Assumption **4**(ii) is a regularity smoothness condition. As usual, the kernel function with compact support is not essential in Assumption **5**(i), and Assumption **5**(ii) is required to remove the asymptotic impact of the first-step estimation and the estimation of $\boldsymbol{\pi}_q$ on the second-step estimator of $\boldsymbol{\psi}(\omega)$. Assumption **5**(iii) implies that the conventional optimal bandwidth of order $n^{-1/5}$ can be used to calculate $\tilde{\boldsymbol{\psi}}(\omega)$. Below, we give our limit result of $\tilde{\boldsymbol{\psi}}(\omega)$.

Theorem 4 *Under Assumptions **1**-**5**, we have at an interior point* ω

$$\sqrt{nh}\left[\tilde{\boldsymbol{\psi}}(\omega) - \boldsymbol{\psi}(\omega) - \frac{\kappa_{1,2}}{2} h^2 B(\omega)\right] \xrightarrow{d} N\left(0, \frac{\kappa_{2,0}}{f_{z\pi}(\omega)} \mathbf{\Omega}(\omega)^{-1} E\left(\varepsilon_1^2 \mathcal{X}_{1,\gamma_0} \mathcal{X}_{1,\gamma_0}' | \boldsymbol{z}_1' \boldsymbol{\pi}_q = \omega\right) \mathbf{\Omega}(\omega)^{-1}\right).$$

where $\mathcal{X}_{1,\gamma_0} = [1, I(q_1 \le \gamma_0)]'$, $\mathbf{\Omega}(\omega) = E\left(\mathcal{X}_{1,\gamma_0} \mathcal{X}_{1,\gamma_0}' | \boldsymbol{z}_1' \boldsymbol{\pi}_q = \omega\right)$ *and* $B(\omega) = \left[h_2^{(2)}(\omega), n^{-\varrho} \eta_0^{(2)}(\omega)\right] \mathbf{\Omega}(\omega)$.

As expected, Theorem 4 implies that the limit distribution of the local linear estimator, $\tilde{\boldsymbol{\psi}}(\omega)$, is not affected by the estimation of the unknown parameters $(\gamma, \boldsymbol{\beta}_1', \boldsymbol{\beta}_2')$. In addition, let $L_n = cn^{\varphi}$, then Assumptions **3** and **5** imply $\min(\varsigma, \varrho)/\xi < \varphi < \min\left(\xi^{-1}, 1/3 + (2 - r')/(15r')\right)$, where we use $\|\mathbf{\Phi}_{L_n}\|_1^2 = O\left(L_n^{3/2}\right)$ which is true for Hermite basis functions. So, $\varphi < 1/4$.

3 Endogenous threshold variable and regressors

This section considers the case that both q_t and some of variables in $\mathbf{x}_t$ are endogenous.[3] The reduced-form model for $\mathbf{x}_t$ is given by

$$\boldsymbol{x}_t = \boldsymbol{\Pi}_{\boldsymbol{x}} \boldsymbol{z}_t + \boldsymbol{v}_{x,t},\ t = 1, 2, \ldots, n, \tag{3.1}$$

where $\boldsymbol{\Pi}_{\mathbf{x}}$ is a $d_x \times d_z$ parameter matrix, $\boldsymbol{v}_{x,t}$ is a $d_x \times 1$ vector of errors satisfying $E(\boldsymbol{v}_{x,t}|\boldsymbol{z}_t) = \mathbf{0}$ for all t and $d_z \geq d_x + 1$. The endogeneity of the regressor $\mathbf{x}_t$ comes from the correlation between $\{u_t\}$ and $\{\boldsymbol{v}_{x,t}\}$.

Combining (2.1), (2.2) and (3.1) gives

$$\begin{aligned} y_t &= \left[\boldsymbol{\beta}_1' \left(\boldsymbol{\Pi}_{\boldsymbol{x}} \boldsymbol{z}_t + \boldsymbol{v}_{x,t}\right) + \sigma_1 u_t\right] I\left(q_t \leq \gamma_0\right) + \left[\boldsymbol{\beta}_2' \left(\boldsymbol{\Pi}_{\boldsymbol{x}} \boldsymbol{z}_t + \boldsymbol{v}_{x,t}\right) + \sigma_2 u_t\right] I\left(q_t > \gamma_0\right) \\ &= \boldsymbol{\beta}_1' \boldsymbol{\Pi}_{\boldsymbol{x}} \boldsymbol{z}_t I\left(q_t \leq \gamma_0\right) + \boldsymbol{\beta}_2' \boldsymbol{\Pi}_{\boldsymbol{x}} \boldsymbol{z}_t I\left(q_t > \gamma_0\right) + e_t \end{aligned}$$

where $e_t = \left(\boldsymbol{\beta}_1' \boldsymbol{v}_{x,t} + \sigma_1 u_t\right) I\left(v_{q,t} \leq \gamma_0 - \mathbf{z}_t' \boldsymbol{\pi}_q\right) + \left(\boldsymbol{\beta}_2' \boldsymbol{v}_{x,t} + \sigma_2 u_t\right) I\left(v_{q,t} > \gamma_0 - \mathbf{z}_t' \boldsymbol{\pi}_q\right)$. Then, following the discussion given in Section 2.1, we have

$$y_t = \boldsymbol{\beta}_2' \boldsymbol{\Pi}_{\boldsymbol{x}} \boldsymbol{z}_t + \boldsymbol{\delta}_n' \boldsymbol{\Pi}_{\boldsymbol{x}} \boldsymbol{z}_t I\left(q_t \leq \gamma_0\right) + h_2(\boldsymbol{z}_t' \boldsymbol{\pi}_q) + \eta_n(\boldsymbol{z}_t' \boldsymbol{\pi}_q) I\left(q_t \leq \gamma_0\right) + \varepsilon_t, \tag{3.2}$$

where $\varepsilon_t = \varepsilon_{1t} I\left(q_t \leq \gamma_0\right) + \varepsilon_{2t} I\left(q_t \leq \gamma_0\right)$ and $\varepsilon_{jt} = \boldsymbol{\beta}_j' \boldsymbol{v}_{x,t} + \sigma_j u_t - h_j(\mathbf{z}_t' \boldsymbol{\pi}_q)$ for j=1,2. Further descriptions of model (3.2) are given by the following assumption.

Assumption 1'.

(i) $\{(\mathbf{x}_t', q_t, \mathbf{z}_t', u_t)\}$ is a strictly stationary strong mixing sequence with the mixing

[3] Our framework allows for the threshold variable q_t to be included in the set of regressors x_t.

coefficients of size $-r/(r-2)$ for some $r > 2$;

(ii) Assumption **1**(i) holds and $E\left(\mathbf{z}_t \boldsymbol{v}_{x,t}'\right) = \mathbf{0}$, $E\left(\boldsymbol{v}_{x,t}\boldsymbol{v}_{x,t}'\right) = \Omega_x$ is positive definite, and $E\left(\left\|\mathbf{z}_t \boldsymbol{v}_{x,t}'\right\|^r\right) < \infty$;

(iii) (a) $\{(u_t, \mathcal{F}_{n,t})\}_{t=1}^n$ is a martingale difference sequence with $E\left(u_t^2|\mathcal{F}_{n,t-1}\right) < \infty$, where $\mathcal{F}_{n,t}$ is the smallest sigma-field generated from $\left\{\left(\boldsymbol{x}_s', q_s, \boldsymbol{z}_{s+1}', u_s\right) : 1 \leq s < t \leq n\right\}$; (b) $E\left(u_t|\mathcal{F}_{n,t-1}, v_{q,t} = v\right) = g(v)$ for any v; (c) $v_{q,t}$ is independent of $\mathcal{F}_{n,t-1}$ and $v_{q,t} \perp v_{x,t}$ for any $t \geq 1$.

(iv) for any $\boldsymbol{\lambda} \neq 0$, there is no measurable function $m(v)$ such that $\mathbf{z}'\boldsymbol{\lambda} = m\left(\mathbf{z}'\boldsymbol{\pi}_q\right)$ when $q_t \leq \gamma_0$ and when $q_t > \gamma_0$;

(v) Assumption **1**(v) holds.

Assumption **1'**(iii) states that $\mathbf{z}_t$ is contemporaneously exogenous in model (2.1)-(2.2), (2.3), and (3.1). Assumptions **1'**(iv) is an identification condition.

Below, in Section 3.1 we explain our proposed estimation procedure, using the same notation as in Section 2 unless we explicitly define some notation differently.

3.1 Estimation

Step 1: Given instruments $\boldsymbol{z}_t$, we obtain the LS estimates of $\boldsymbol{\pi}_q$ and $\boldsymbol{\Pi}_x$, $\hat{\boldsymbol{\pi}}_q = \left(\sum_{t=1}^n \boldsymbol{z}_t\boldsymbol{z}_t'\right)^{-1}\sum_{t=1}^n \boldsymbol{z}_t q_t$ and $\hat{\boldsymbol{\Pi}}_x = \left(\sum_{t=1}^n \mathbf{z}_t\mathbf{z}_t'\right)^{-1}\sum_{t=1}^n \mathbf{x}_t\mathbf{z}_t'$, from models (2.3) and (3.1), respectively. Assumptions **1**(i)-(ii) and **1'**(i)-(ii) imply existence and consistency of $\hat{\boldsymbol{\pi}}_q = \boldsymbol{\pi}_q + O_p\left(n^{-1/2}\right)$ and $\hat{\boldsymbol{\Pi}}_x = \boldsymbol{\pi}_x + O_p\left(n^{-1/2}\right)$. We then denote the fitted values as $\hat{q}_t = \boldsymbol{z}_t'\hat{\boldsymbol{\pi}}_q$ and $\hat{\mathbf{x}}_t = \hat{\boldsymbol{\Pi}}_x\mathbf{z}_t$ and the estimated residuals as $\hat{v}_{q,t} = q_t - \hat{q}_t$ and $\hat{\boldsymbol{v}}_{x,t} = \mathbf{x}_t - \hat{\mathbf{x}}_t$ for all t throughout the rest of this paper.

Step 2: For a given $\gamma \in \left[\underline{\gamma}, \overline{\gamma}\right]$, we estimate $\boldsymbol{\theta} = \left(\boldsymbol{\beta}_1', \boldsymbol{\alpha}_{L_n,1}', \boldsymbol{\beta}_2', \boldsymbol{\alpha}_{L_n,2}'\right)'$ from the objective function

$$\hat{\boldsymbol{\theta}}(\gamma) = \arg\min_{\theta} \sum_{t=1}^{n} \left[y_t - \hat{\boldsymbol{x}}'_{-,t}\boldsymbol{\beta}_1 - \boldsymbol{\alpha}'_{L_n,1}\boldsymbol{\Phi}^{-}_{L_n,\gamma}(\hat{q}_t) - \hat{\mathbf{x}}'_{+,t}\boldsymbol{\beta}_2 - \boldsymbol{\alpha}'_{L_n,2}\boldsymbol{\Phi}^{+}_{L_n}(\hat{q}_t) \right]^2, \tag{3.3}$$

where we denote $\hat{\mathbf{x}}_{-,t} = \hat{\mathbf{x}}_t I(q_t \leq \gamma)$, $\hat{\mathbf{x}}_{+,t} = \hat{\mathbf{x}}_t I(q_t > \gamma)$. Solving (3.3) yields

$$\hat{\boldsymbol{\theta}}(\gamma) = \left(\mathcal{X}'_{\gamma}\mathcal{X}_{\gamma}\right)^{-1} \mathcal{X}'_{\gamma}\boldsymbol{y} \tag{3.4}$$

where $\mathcal{X}_{\gamma} = [\mathcal{X}_{-,\gamma}, \mathcal{X}_{+,\gamma}]$ and $\mathcal{X}_{-,\gamma}$ and $\mathcal{X}_{+,\gamma}$ are defined the same as in Section 2 with $\mathbf{x}_{+,t}$ and $\mathbf{x}_{-,t}$ replaced with $\hat{\mathbf{x}}_{-,t}$ and $\hat{\mathbf{x}}_{-,t}$, respectively. We then estimate the threshold parameter γ by minimizing the concentrated least squares criterion

$$\widehat{\gamma} = \arg\min_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} \sum_{t=1}^{n} \left[y_t - \mathcal{X}'_{t,\gamma}\hat{\boldsymbol{\theta}}(\gamma) \right]^2 \tag{3.5}$$

and then estimate $\boldsymbol{\theta}$ by $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\hat{\gamma})$.

3.2 Inference

Assumption 2':

(i) $E\left\|\mathbf{z}_t\right\|^{2r'} < \infty$ and $E\left\|\varepsilon_t\mathbf{z}_t\right\|^{r} < \infty$ for some $r' > r > 2$, where r is defined in Assumption **1**;

(ii) for every L_n and uniformly over $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, there exist constants $\underline{c}$ and $\overline{c}$ such that $0 < \underline{c} \leq \boldsymbol{\lambda}_{\min}(\boldsymbol{\Sigma}_{\gamma}) \leq \boldsymbol{\lambda}_{\max}(\boldsymbol{\Sigma}_{\gamma}) \leq \overline{c} < \infty$, where $\boldsymbol{\Sigma}_{\gamma} = E\left(\mathcal{X}^{*}_{t,\gamma}\mathcal{X}^{*\prime}_{t,\gamma}\right)$ and $E\left(\varepsilon_t^2\mathcal{X}^{*}_{t,\gamma}\mathcal{X}^{*\prime}_{t,\gamma}\right)$, and $\mathcal{X}^{*}_{t,\gamma}$ equals $\mathcal{X}_{t,\gamma}$ with $\hat{\mathbf{x}}_t$ and $\hat{q}_t$ replaced with $\boldsymbol{\Pi}_{\mathbf{x}}\mathbf{z}_{\mathbf{t}}$ and $\mathbf{z}'_t\boldsymbol{\pi}_q$, respectively;

(iii) Assumption **2**(iii) holds.

Below, we give the limit results for $\widehat{\gamma}$ and $\hat{\boldsymbol{\theta}}$.

Theorem 5 *Under Assumptions **1', 2'** and **3**(i), we have* $\widehat{\gamma}-\gamma_0=o_p(1)$ *and* $\left\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right\|=O_p\left(\vartheta_n+n^{-\min(\varsigma,\varrho)}\right)$.

Theorem 6 *Under Assumptions **1', 2'** and 3(i)(ii), we have*

$$n^{1-2\min(\varsigma,\varrho)}\left(\widehat{\gamma}-\gamma_0\right)\xrightarrow{d}\varpi T,$$

where we denote $\sigma_j^2=E\left\{\varepsilon_{jt}^2\left[I\left(\varsigma\leq\varrho\right)\boldsymbol{\delta}_0'\mathbf{\Pi_x z_t}+\mathbf{I}\left(\varsigma\geq\varrho\right)\eta_{\mathbf{0}}\left(\mathbf{z_t'}\boldsymbol{\pi}_{\mathbf{q}}\right)\right]^2|q_t=\gamma_0\right\}$ *for* $j=1,\ 2$, $\omega=E\left\{\left[I\left(\varsigma\leq\varrho\right)\boldsymbol{\delta}_0'\boldsymbol{\Pi}_{\boldsymbol{x}}\mathbf{z}_t+I\left(\varsigma\geq\varrho\right)\eta_0\left(\mathbf{z}_t'\boldsymbol{\pi}_q\right)\right]^2|q_t=\gamma_0\right\}$.

Theorem 7 *Under Assumptions **1', 2'** and **3**, we have*

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\xrightarrow{d}N\left(\mathbf{0},\boldsymbol{J}^{-1}\boldsymbol{\Omega}\boldsymbol{J}^{-1}\right),\tag{3.6}$$

where $\boldsymbol{J}$ *and* $\boldsymbol{\Omega}$ *are defined by (A.30) and (A.33) in Appendix, respectively.*

Compared with Theorems 1-3, Theorems 5-7 indicate that the endogeneity of the regressors has impacts on the variation of $\widehat{\gamma}$ and $\hat{\boldsymbol{\beta}}$ not their convergence rates.

4 Testing for the Endogeneity of the Threshold Variable

In this section, we are interested in testing whether the threshold variable, q_t, is endogenous in a linear threshold model (2.1)-(2.2). As the proposed test statistic is applicable regardless of whether $\mathbf{x}_t$ is endogenous or exogenous, we give details for the case that $\mathbf{x}_t$ is exogenous.

Under the null hypothesis, q_t is exogenous, while under the alternative hypothesis, q_t, is endogenous. As it is not necessary to test the endogeneity of q_t if there is no threshold effect at all or $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$, it would be intuitive to define the null and alternative hypotheses as H_0^A : $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ and $h_1(z) = h_2(z) \equiv 0$ vs. H_1^A : not H_0. However, rejecting this null hypothesis will not reveal whether the threshold effect or the threshold variable being endogenous is rejected. Therefore, we define our null and alternative hypotheses as follows

$$\mathrm{H}_0 : h_1(z) = h_2(z) \equiv 0 \text{ vs. } \mathrm{H}_1 : \text{not } \mathrm{H}_0 \tag{4.1}$$

and the working null and alternative hypotheses can be written as, $\mathrm{H}_0' : \boldsymbol{\alpha}_{L_n,1} = \boldsymbol{\alpha}_{L_n,2} = \mathbf{0}_{L_n}$ against H_1' : not H_0'. The null hypothesis defined in (4.1) imposes no extra restriction on $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, other than the restriction given by Assumption **1**(v), so the model under the null hypothesis can be a simple linear regression model or a linear threshold regression model with exogenous threshold variable.

In Section 2, applying series approximation to model (2.7) gives

$$y_t = \boldsymbol{\beta}_1' \mathbf{x}_{-,t} + \boldsymbol{\beta}_1' \mathbf{x}_{+,t} + \boldsymbol{\alpha}_{L_n,1}' \boldsymbol{\Phi}_{L_n,\gamma}^{-}(\hat{q}_t) + \boldsymbol{\alpha}_{L_n,2}' \boldsymbol{\Phi}_{L_n,\gamma}^{+}(\hat{q}_t) + v_{\gamma,t} \tag{4.2}$$

where we denote $v_{\gamma,t} = [h_1(\mathbf{z}_t'\boldsymbol{\pi}_q) - h_1^*(\mathbf{z}_t'\hat{\boldsymbol{\pi}}_q)] I(\boldsymbol{q}_t \leq \gamma) + [h_2(\mathbf{z}_t'\boldsymbol{\pi}_q) - h_2^*(\mathbf{z}_t'\hat{\boldsymbol{\pi}}_q)] I(q_t > \gamma) + \varepsilon_t$. Denoting $\boldsymbol{Q}_\gamma$ be an $n \times (2d_x)$ matrix with its t^{th} row equal to $\left[\mathbf{x}_t' I_{\gamma,t}^{-}, \mathbf{x}_t' I_{\gamma,t}^{+}\right]$ and $\boldsymbol{M}_{\gamma,\boldsymbol{Q}} = \mathbf{I}_n - \boldsymbol{Q}_\gamma \left(\boldsymbol{Q}_\gamma' \boldsymbol{Q}_\gamma\right)^{-1} \boldsymbol{Q}_\gamma$ and multiplying $\boldsymbol{M}_{\gamma,\boldsymbol{Q}}$ to the both sides of eq. (4.2) gives

$$\boldsymbol{y}^* = \hat{\boldsymbol{\Phi}}_{L_n,\gamma}^{-,*} \boldsymbol{\alpha}_{L_n,1} + \hat{\boldsymbol{\Phi}}_{L_n,\gamma}^{+,*} \boldsymbol{\alpha}_{L_n,2} + \boldsymbol{v}^*$$

where $\boldsymbol{y}^* = \boldsymbol{M}_{\gamma,\boldsymbol{Q}} \boldsymbol{y}$, $\hat{\boldsymbol{\Phi}}_{L_n,\gamma}^{-,*} = \boldsymbol{M}_{\gamma,\boldsymbol{Q}} \hat{\boldsymbol{\Phi}}_{L_n,\gamma}^{-}$, $\hat{\boldsymbol{\Phi}}_{L_n,\gamma}^{+,*} = \boldsymbol{M}_{\gamma,\boldsymbol{Q}} \hat{\boldsymbol{\Phi}}_{L_n,\gamma}^{+}$, and $\boldsymbol{v}^* = \boldsymbol{M}_{\boldsymbol{Q}} \boldsymbol{v}$; $\boldsymbol{y}$ and $\boldsymbol{v}$ are $n \times 1$ vector with typical element y_t and v_t, respectively; $\hat{\boldsymbol{\Phi}}_{L_n,\gamma}^{-}$ and $\hat{\boldsymbol{\Phi}}_{L_n,\gamma}^{+}$ are $n \times L_n$ matrix with t^{th} row equal to $\mathbf{\Phi}_{L_n,\gamma}^{-\prime}(\hat{q}_t)$ and $\mathbf{\Phi}_{L_n}^{+\prime}(\hat{q}_t)$, respectively. Then given $\gamma \in [\underline{\gamma}, \bar{\gamma}]$, we have

$$\hat{\boldsymbol{\alpha}}_{L_n}(\gamma) = \left(\hat{\mathbf{\Phi}}_{L_n,\gamma}^{*\prime} \hat{\mathbf{\Phi}}_{L_n,\gamma}^{*}\right)^{-1} \hat{\mathbf{\Phi}}_{L_n,\gamma}^{*\prime} \boldsymbol{y}^{*} \tag{4.3}$$

where we denote $\hat{\mathbf{\Phi}}_{L_n,\gamma}^{*} = \left[\hat{\mathbf{\Phi}}_{L_n,\gamma}^{-,*}, \hat{\mathbf{\Phi}}_{L_n,\gamma}^{+,*}\right]$ and $\hat{\boldsymbol{\alpha}}_{L_n}(\gamma) = \left[\hat{\boldsymbol{\alpha}}_{L_n,1}'(\gamma), \hat{\boldsymbol{\alpha}}_{L_n,2}'(\gamma)\right]'$. We then construct a Wald statistic

$$W_n(\gamma) = \hat{\boldsymbol{\alpha}}_{L_n}(\gamma)' \hat{\mathbf{\Phi}}_{L_n,\gamma}^{*\prime} \hat{\mathbf{\Phi}}_{L_n,\gamma}^{*} \left(\hat{\mathbf{\Phi}}_{L_n,\gamma}^{*\prime} \hat{\boldsymbol{\varepsilon}}_{\gamma} \hat{\boldsymbol{\varepsilon}}_{\gamma}' \hat{\mathbf{\Phi}}_{L_n,\gamma}^{*}\right)^{-1} \hat{\mathbf{\Phi}}_{L_n,\gamma}^{*\prime} \hat{\mathbf{\Phi}}_{L_n,\gamma}^{*} \hat{\boldsymbol{\alpha}}_{L_n}(\gamma) \tag{4.4}$$

where $\hat{\boldsymbol{\varepsilon}}_{\gamma}$ is an $n \times 1$ vector of residuals calculated from the alternative hypothesis and its t^{th} element equals $\hat{\boldsymbol{\varepsilon}}_{\gamma,t} = y_t - \mathbf{x}_{-,t}'\hat{\boldsymbol{\beta}}_1 - \mathbf{x}_{+,t}'\hat{\boldsymbol{\beta}}_2 - \hat{\boldsymbol{\alpha}}_{L_n,1}'\mathbf{\Phi}_{L_n,\gamma}^{-}(\hat{q}_t) - \hat{\boldsymbol{\alpha}}_{L_n,2}'\mathbf{\Phi}_{L_n}^{+}(\hat{q}_t)$.

Next, let $\hat{\gamma}$ be the estimate of γ under the null hypothesis. That is,

$$\hat{\gamma} = \arg \min_{\gamma \in [\underline{\gamma}, \bar{\gamma}]} \boldsymbol{y}' \boldsymbol{M}_{\gamma,\boldsymbol{Q}} \boldsymbol{y}. \tag{4.5}$$

Our final test statistic is defined as $W_n(\hat{\gamma})$. Motivated by Gonzalo and Pitarakis (2016), we will show that $W_n(\hat{\gamma})$ has the same limit distribution under the null hypothesis regardless of whether $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ or $\boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$.

Theorem 8 *Under Assumptions **1-3**(i)(ii) and H_0 and $E(u_t^4) < M < \infty$, we have (i) $n^{1-2\min(\varsigma,\varrho)}(\widehat{\gamma} - \gamma_0) \xrightarrow{d} \varpi T$ when $\boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$, (ii) $\widehat{\gamma} \xrightarrow{d} \gamma^*$, where γ^* is defined in (A.34) when $\sigma_1^2 = \sigma_2^2 = \sigma^2$ and $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$, and (iii) $\widehat{\gamma} \xrightarrow{a.s.} \gamma^*$, where $\gamma^* = \underline{\gamma} I(\sigma_1^2 > \sigma_2^2) + \bar{\gamma} I(\sigma_1^2 < \sigma_2^2)$ when $\sigma_1^2 \neq \sigma_2^2$ and $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$.*

Theorem 9 *Under Assumptions **1-3** and H_0, $W_n(\hat{\gamma}) \xrightarrow{d} \chi_{2L_n}^2$ holds (i) if $\boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$ or (ii) if $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$, $\sigma_1^2 = \sigma_2^2$, and $\{u_t\}$ is independent of $\{(\mathbf{x}_t, \mathbf{z}_t)\}$.*

As shown in the Appendix, out test statistic does not converge to a chi-squared distribution when $\beta_1 = \beta_2$ and $\sigma_1^2 \neq \sigma_2^2$.[4]

5 Monte Carlo Simulations

5.1 Threshold and slope parameters

Athreya and Pantula (1986) provide theoretical argument on the strong mixing properties of stationary ARMA processes. Following Chen (2007), we can use the Hermite functions as the basis functions for the series approximation of $h(\cdot)$.

We first consider a model with endogeneity only in the threshold variable:

$$y_i = \beta_1 + \beta_2 x_i + (\delta_1 + \delta_2 x_i) I\{q_i \leq \gamma\} + u_i, \tag{5.1}$$

where

$$q_i = 2 + z_{qi} + v_{qi}. \tag{5.2}$$

The threshold parameter is set at the center of the distribution of q_i, hence $\gamma = 2$. The instrumental variable z_{qi} is given by

$$z_{qi} = (x_i + \varsigma_{zi}) / 2\sqrt{2} \tag{5.3}$$

and

$$u_i = 0.1\varsigma_{ui} + \kappa v_{qi}, \tag{5.4}$$

where x_i, v_{qi}, ς_{zi}, and ς_{ui} are independent $i.i.d.$ $N(0,1)$ random variables. The degree of endogeneity of the threshold is controlled by κ. We fix $\beta_1 = \beta_2 = 1$, and $\delta_1 = 0$ and vary δ_2

[4]For details, please read the proof given in the Appendix.

over the values of $1, 2, 3, 4, 5$, which correspond to a range of small to large threshold effects. We also vary κ over the values of 0.05, 0.50, 0.95 that correspond to low, medium, and large degrees of endogeneity of the threshold variable.

Our second DGP adds an endogenous regressor to model (5.1)

$$y_i = \beta_1 + \beta_2 x_{1i} + \beta_3 x_{2i} + (\delta_1 + \delta_2 x_{1i} + \delta_3 x_{2i}) I\{q_i \leq \gamma\} + u_i, \tag{5.5}$$

where

$$x_{1i} = z_{xi} + v_{xi},$$

with

$$z_{xi} = (w x_{2i} + (1-w)\varsigma_{zi}) \,/\, \sqrt{w^2 + (1-w)^2}, \tag{5.6}$$

and

$$u_i = (c_{xu} v_{xi} + c_{qu} v_{qi} + (1 - c_{xu} - c_{qu})\varsigma_{ui}) \,/\, \sqrt{c_{xu}^2 + c_{qu}^2 + (1 - c_{xu} - c_{qu})^2}, \tag{5.7}$$

where x_{2i}, ς_{zi} and ς_{ui} are independent *i.i.d.* $N(0,1)$ random variables. The degree of endogeneity of the threshold variable is controlled by the correlation coefficient between u_i and v_{qi} given by $c_{qu}/\sqrt{c_{xu}^2 + c_{qu}^2 + (1 - c_{xu} - c_{qu})^2}$. Similarly, the degree of endogeneity of x_{1i} is determined by the correlation between u_i and v_{xi} given by $c_{xu}/\sqrt{c_{xu}^2 + c_{qu}^2 + (1 - c_{xu} - c_{qu})^2}$. We vary δ_3 and fix c_{xu}, $w = 0.5$, $\beta_1 = \beta_2 = 1$, and $\delta_1 = \delta_2 = 0$. We set c_{qu} at 0.45, which corresponds to correlation of 0.7.

We begin by assessing the performance of our estimators for the threshold parameter and the threshold effect by considering sample sizes of 100, 250, 500, and 1000 using 1000 Monte Carlo replications simulations. Tables 1 and 2 present the quantiles of the distribution of γ and δ_2 by varying the threshold effect δ_2 over the values 1, 2, 3, and 4 using a 6th order

Hermite basis function for models (5.1) and (5.5), respectively.[5] We see that the performance of the estimators for both the threshold parameter and the threshold effect improve as the threshold effect and the sample size increase. Specifically, the 50th quantile approaches the true threshold parameter, $\gamma = 2$, as the sample size n increases and the width of the distribution becomes smaller as δ_2 increases.

5.2 Size and power of the Wald statistic

We assess the size and power of the Wald statistic in equation (4.4) and γ estimated by the objective function defined by (4.5), which tests for the endogeneity of the threshold variable. Table 3 provides the results for the case of the DGP in equation (5.1).[6] We present the size ($\rho = 0$) and the power ($\rho > 0$) for various orders L_n of Hermite basis functions and sample sizes. Panel A presents results of the test statistic defined in equation (4.4), which is based on a White covariance estimator. Panel B present results of a homoskedastic version of the test statistic and Panel C shows results based on Andrews (1991) covariance estimator based on the principle of leave-one-out cross-validation.

Our simulations reveal several things. In general, we find that our test exhibits good size and power properties, especially when the number of basis functions is small. However, we see that the basis functions that correspond to higher order polynomials are likely to lead an oversized test. This size problem appears to go away when we employ a homoskedastic version of the test statistic and is mitigated when we use the Andrews covariance estimator at the cost of lower power.[7]

[5]In Figures 1 and 2 of the Online Appendix we also show the corresponding Monte Carlo kernel densities of the threshold estimator for a small threshold effect ($\delta_2 = 1$) and a large threshold effect ($\delta_2 = 4$).

[6]We also investigated models that impose the restriction that $h_1 = h_2$. As expected both the size and power of the test improve using this extra information. All results including those for the DGP in equation (5.5) are available on request.

[7]We also investigated a finite-sample correction of the White estimator as well as Horn, Horn, and Duncan (1975) estimator but the results were not better than the Andrews estimator. An important factor in

6 Conclusion

In this paper we propose different types of semiparametric threshold regression models with endogenous threshold variables based on a nonparametric control function approach. Using a series approximation we propose to estimate the threshold parameter using a concentrated least squares which includes a regime specific control function. We develop estimation and inference for weakly dependent data for the estimators of both the threshold and slope parameters. Furthermore, we propose a test for the endogeneity of the threshold variable, which is valid regardless of whether the threshold effect is zero or not. Finally, we assess the performance of the proposed estimation method using a Monte Carlo simulation.

achieving substantial improvements in both size and power is the restriction $h_1 = h_2$.

Table 1: Threshold Parameter and Threshold Effect - Exogenous Regressor

This table presents Monte Carlo quantiles of the estimates of the true threshold parameter $\gamma = 2$ and true threshold effect $\delta_2 = 1, 2, 3, 4$ in the case of exogenous regressor and endogenous threshold variable using a 6th order Hermite basis function.

	Threshold Parameter			Threshold Effect		
Quantile	5th	50th	95th	5th	50th	95th
Sample size						
			$\delta_2 = 1$			
100	1.472	1.957	2.307	0.531	0.931	1.326
250	1.758	1.991	2.182	0.760	0.974	1.179
500	1.873	1.996	2.087	0.822	0.980	1.126
1000	1.931	1.998	2.039	0.883	0.994	1.096
			$\delta_2 = 2$			
100	1.752	1.974	2.150	1.616	1.973	2.355
250	1.911	1.992	2.058	1.794	1.998	2.204
500	1.960	1.996	2.028	1.841	1.993	2.131
1000	1.979	1.998	2.014	1.891	2.000	2.100
			$\delta_2 = 3$			
100	1.831	1.976	2.092	2.641	2.987	3.361
250	1.937	1.991	2.032	2.794	3.005	3.208
500	1.970	1.996	2.014	2.842	2.994	3.131
1000	1.985	1.998	2.009	2.893	3.000	3.101
			$\delta_2 = 4$			
100	1.851	1.975	2.063	3.658	3.991	4.372
250	1.947	1.991	2.024	3.794	4.005	4.211
500	1.974	1.996	2.010	3.843	3.994	4.132
1000	1.987	1.998	2.006	3.894	4.000	4.101

Table 2: Threshold Parameter and Threshold Effect - Endogenous Regressor

This table presents Monte Carlo quantiles of the estimates of the true threshold parameter $\gamma = 2$ and true threshold effect $\delta_2 = 1, 2, 3, 4$ in the case of both endogenous regressor and endogenous threshold variable using a 6th order Hermite basis function.

	Threshold Parameter			Threshold Effect		
Quantile	5th	50th	95th	5th	50th	95th
Sample size						
			$\delta_2 = 1$			
100	0.648	1.836	2.980	0.149	0.883	1.554
250	0.908	1.922	2.913	0.499	0.910	1.231
500	1.266	1.962	2.509	0.613	0.932	1.165
1000	1.467	1.973	2.277	0.729	0.948	1.105
			$\delta_2 = 2$			
100	1.011	1.934	2.529	1.120	1.892	2.517
250	1.592	1.983	2.209	1.584	1.974	2.251
500	1.809	1.992	2.079	1.765	1.979	2.185
1000	1.913	1.997	2.040	1.842	1.985	2.125
			$\delta_2 = 3$			
100	1.538	1.969	2.278	2.352	2.979	3.506
250	1.867	1.989	2.088	2.700	2.999	3.252
500	1.929	1.994	2.037	2.800	2.990	3.189
1000	1.971	1.998	2.019	2.857	2.991	3.132
			$\delta_2 = 4$			
100	1.717	1.975	2.151	3.444	3.992	4.507
250	1.904	1.990	2.050	3.720	4.005	4.263
500	1.956	1.995	2.026	3.806	3.992	4.193
1000	1.980	1.998	2.014	3.863	3.992	4.135

Table 3: Size and Power

This table presents the size ($\rho = 0$) and the power ($\rho > 0$) for various orders of Hermite basis functions L_n and sample sizes. Panel A presents results of the test statistic defined in equation (4.4), which is based on a White covariance estimator. Panel B present results of a homoskedastic version of the test statistic and Panel C use Andrews (1991) covariance estimator based on the principle of leave-one-out cross-validation.

Panel A: White covariance matrix

ρ Sample size	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
					$L_n = 2$					
100	0.21	0.62	0.89	0.94	0.95	0.96	0.96	0.97	0.97	0.97
250	0.10	0.60	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00
500	0.08	0.55	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1000	0.06	0.56	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					$L_n = 3$					
100	0.38	0.75	0.93	0.96	0.97	0.98	0.98	0.98	0.98	0.98
250	0.16	0.65	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00
500	0.12	0.61	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1000	0.08	0.56	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					$L_n = 4$					
100	0.61	0.86	0.96	0.98	0.99	0.99	0.99	0.99	0.99	0.99
250	0.31	0.74	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00
500	0.19	0.66	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1000	0.11	0.58	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					$L_n = 5$					
100	0.78	0.93	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.99
250	0.52	0.84	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
500	0.31	0.74	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1000	0.17	0.63	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					$L_n = 6$					
100	0.90	0.97	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
250	0.70	0.90	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
500	0.51	0.84	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1000	0.29	0.71	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 3 continued

Panel B: Homoskedastic covariance matrix

ρ Sample size	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
					$L_n = 2$					
100	0.05	0.40	0.72	0.81	0.84	0.86	0.88	0.88	0.89	0.89
250	0.05	0.50	0.94	0.99	1.00	1.00	1.00	1.00	1.00	1.00
500	0.06	0.50	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1000	0.06	0.51	0.98	0.99	1.00	1.00	1.00	1.00	1.00	1.00
					$L_n = 3$					
100	0.06	0.40	0.71	0.81	0.83	0.85	0.86	0.86	0.87	0.87
250	0.05	0.48	0.94	0.99	1.00	1.00	1.00	1.00	1.00	1.00
500	0.05	0.51	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1000	0.06	0.52	0.98	0.99	1.00	1.00	1.00	1.00	1.00	1.00
					$L_n = 4$					
100	0.06	0.37	0.68	0.77	0.81	0.82	0.84	0.84	0.84	0.84
250	0.04	0.46	0.91	0.99	1.00	1.00	1.00	1.00	1.00	1.00
500	0.06	0.49	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1000	0.05	0.48	0.98	0.99	0.99	1.00	1.00	1.00	1.00	1.00
					$L_n = 5$					
100	0.06	0.35	0.65	0.75	0.78	0.79	0.80	0.81	0.81	0.81
250	0.05	0.42	0.90	0.98	0.99	1.00	1.00	1.00	1.00	1.00
500	0.06	0.44	0.96	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1000	0.05	0.45	0.98	0.99	0.99	1.00	1.00	1.00	1.00	1.00
					$L_n = 6$					
100	0.07	0.33	0.62	0.72	0.76	0.77	0.77	0.78	0.78	0.78
250	0.05	0.40	0.89	0.98	0.99	0.99	1.00	1.00	1.00	1.00
500	0.07	0.43	0.96	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1000	0.05	0.42	0.98	0.99	0.99	0.99	1.00	1.00	1.00	1.00

Table 3 continued

Panel C: Andrews leave-one-out cross-validation

ρ Sample size	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
					$L_n = 2$					
100	0.11	0.48	0.80	0.88	0.91	0.92	0.93	0.93	0.93	0.93
250	0.06	0.53	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00
500	0.06	0.51	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1000	0.06	0.54	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					$L_n = 3$					
100	0.17	0.53	0.81	0.88	0.92	0.92	0.93	0.93	0.93	0.94
250	0.09	0.54	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00
500	0.08	0.53	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1000	0.06	0.54	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					$L_n = 4$					
100	0.23	0.54	0.79	0.87	0.90	0.91	0.92	0.92	0.93	0.93
250	0.13	0.55	0.94	0.99	1.00	1.00	1.00	1.00	1.00	1.00
500	0.10	0.54	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1000	0.07	0.52	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					$L_n = 5$					
100	0.31	0.60	0.80	0.86	0.89	0.90	0.91	0.91	0.91	0.92
250	0.16	0.54	0.93	0.99	1.00	1.00	1.00	1.00	1.00	1.00
500	0.14	0.51	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1000	0.10	0.51	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					$L_n = 6$					
100	0.39	0.66	0.84	0.89	0.90	0.92	0.93	0.93	0.93	0.93
250	0.23	0.54	0.92	0.98	1.00	1.00	1.00	1.00	1.00	1.00
500	0.17	0.51	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1000	0.11	0.50	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00

References

Andrews, D. W. K., 1991, Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models, *Econometrica* 59, 307–345.

Athreya, K.B., and S.G. Pantula, 1986, A Note on Strong Mixing of ARMA Processes, *Statistics and Probability Letters* 4, 187–190.

Blundell, R., X. Chen, and D. Kristensen, 2007, Semiparametric IV Estimation of Shape Invariant Engel Curves, *Econometrica* 75, 1613–1669.

Cai, Z., J. Fan, and Q. Yao, 2000, Functional-Coefficient Regression Models for Nonlinear Time Series., *Journal of American Statistical Association* 95, 941–956.

Caner, M., and B. Hansen, 2004, Instrumental Variable Estimation of a Threshold Model, *Econometric Theory* 20, 813–843.

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

Chen, X., 2007, Large Sample Sieve Estimation of Semi-nonparametric Models, in J.J. Heckman, and E.E. Leamer, ed.: *Handbook of Econometrics* vol. 6.

Durlauf, S. N., 1996, A Theory of Persistent Income Inequality, *Journal of Economic Growth* 1, 75–93.

Dzyadyk, V. K., and I. A. Shevchuk, 2008, *Theory of Uniform Approximation of Functions by Polynomials* (Walter de Gruyter: London).

Galor, O., and J. Zeira, 1993, Income Distribution and Macroeconomics, *Review of Economic Studies* 60, 35–52.

Gonzalo, J., and J.-Y. Pitarakis, 2016, Inferring the predictability induced by a persistent regressor in a predictive threshold model, *Journal of Business and Economic Statistics*.

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

Hansen, B. E., 2000, Sample Splitting and Threshold Estimation, *Econometrica* 68, 575–603.

Hastie, T., and R. Tibshirani, 1993, Varying Parameter Models, *Journal of the Royal Statistical Society, Series B* 55, 757–796.

Heckman, J., 1979, Sample Selection Bias as a Specification Error, *Econometrica* 47, 153–161.

Horn, S.D., R. A. Horn, and D.B. Duncan, 1975, Estimating Heteroscedastic Variances in Linear Model, *Journal of the American Statistical Association* 70, 380–385.

Kourtellos, A., T. Stengos, and Y. Sun, 2017, Endogeneity in Semiparametric Threshold Regression: Supplementary Online Appendix, The University of Cyprus, mimeo.

Kourtellos, A., T. Stengos, and C. M. Tan, 2016, Structural Threshold Regression, *Econometric Theory* 32, 827–860.

Li, D., and S. Ling, 2012, On the Least Squares Estimation of Multiple Regime Threshold Autoregressive Models, *Journal of Econometrics* 167, 240–253.

Newey, W. K., 1997, Convergence Rates and Asymptotic Normality for Series Estimators, *Journal of Econometrics* 79, 147168.

——— , 2009, Two-step Series Estimation of Sample Selection Models, *The Econometrics Journal* 12, 217–229.

Ozabaci, D., D. J. Henderson, and L. Su, 2014, Additive Nonparametric Regression in the Presence of Endogenous Regressors, *IZA Working Paper 8144*.

Reinhart, C.M., and K.S. Rogoff, 2010, Growth in Time of Debt, *American Economic Review* 100, 573–578.

Seo, M. H., and O. Linton, 2007, A Smoothed Least Squares Estimator for Threshold Regression Models, *Journal of Econometrics* 141, 704–735.

Seo, M. H., and Y. Shin, 2016, Dynamic Panels with Threshold Effect and Endogeneity, *Journal of Econometrics* 195, 169–186.

Terasvirta, T., W.J. Granger, and R. R. Nelson, 1986, Aspects of modelling nonlinear time series, in T. Terasvirta, W.J. Granger, and R. R. Nelson, ed.: *Handbook of Econometrics* . chap. 15, pp. 2917–2957 (Elsevier).

Tong, H., 1990, *Nonlinear Time Series: A Dynamical System Approach* (Oxford University Press).

White, H., 2001, *Asymptotic Theory for Econometricians* (Academic Press: London).

Xiang, S., 2012, Asymptotics on Laguerre or Hermite Polynomial Expansions and Their Applications in Gauss Quadrature, *Journal of Mathematical Analysis and Applications* J393, 434–444.

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

——— , 2013, Adaptive Estimation of the Threshold Point in Threshold Regression, *Journal of Econometrics* (forthcoming).

———, and C. B. Phillips, 2015, Threshold Regression with Endogeneity, Journal of Econometrics (forthcoming).

Appendix

Proof of Theorem 1: Denote $\boldsymbol{P}_{\gamma} = \mathcal{X}_{\gamma}\left(\mathcal{X}_{\gamma}'\mathcal{X}_{\gamma}\right)^{-1}\mathcal{X}_{\gamma}'$, $\boldsymbol{P}_{\gamma}^{-} = \mathcal{X}_{-,\gamma}\left(\mathcal{X}_{-,\gamma}'\mathcal{X}_{-,\gamma}\right)^{-1}\mathcal{X}_{-,\gamma}'$, and $\boldsymbol{P}_{\gamma}^{+} = \mathcal{X}_{+,\gamma}\left(\mathcal{X}_{+,\gamma}'\mathcal{X}_{+,\gamma}\right)^{-1}\mathcal{X}_{+,\gamma}'$, where $\mathcal{X}_{-,\gamma}$ and $\mathcal{X}_{+,\gamma}$ are defined in Section 2.1. Applying simple calculation gives $\boldsymbol{P}_{\gamma} = \boldsymbol{P}_{\gamma}^{-} + \boldsymbol{P}_{\gamma}^{+}$. As $\left(\boldsymbol{I}_n - \boldsymbol{P}_{\gamma}\right)\mathcal{X}_{\gamma} = \boldsymbol{0}$, by (2.10), we have

$$S_n(\gamma) = \sum_{t=1}^{n}\left[y_t - \mathcal{X}_{t,\gamma}'\hat{\boldsymbol{\theta}}(\gamma)\right]^2 = \boldsymbol{y}'\left(\boldsymbol{I}_n - \boldsymbol{P}_{\gamma}\right)\boldsymbol{y}, \tag{A.1}$$

where $\boldsymbol{I}_n$ denotes the $n \times n$ identity matrix. Below, we will show that

$$n^{-1}S_n(\gamma) = S(\gamma) + o_p(1) \tag{A.2}$$

holds uniformly over $\gamma \in \left[\underline{\gamma}, \overline{\gamma}\right]$, where $S(\gamma) = S_1(\gamma)\, I(\gamma > \gamma_0) + S_2(\gamma)\, I(\gamma \leq \gamma_0)$, $S_1(\gamma)$ is a strictly increasing function of γ over the interval of $[\gamma_0, \overline{\gamma}]$, $S_2(\gamma)$ is a strictly decreasing function of γ over the interval of $\left[\underline{\gamma}, \gamma_0\right]$ and both $S_1(\gamma)$ and $S_2(\gamma)$ are continuous over $\gamma \in \left[\underline{\gamma}, \overline{\gamma}\right]$. Therefore, $S(\gamma)$ is a continuous function of γ and is uniquely minimized at γ_0, and we then obtain $\hat{\gamma} \xrightarrow{p} \gamma_0$ by Theorem 2.1 in Newey and McFadden (1994) if we can show that $S(\gamma)$ is uniquely minimized at point γ_0.

Specifically, applying simple algebras give

$$\begin{aligned}
S_n(\gamma) &= \left(\boldsymbol{X}_{-}\boldsymbol{\delta}_n + \boldsymbol{\eta}_{-}(\mathbf{z}'\boldsymbol{\pi}_q) + \hat{\varepsilon}\right)'\left(\boldsymbol{I}_n - \boldsymbol{P}_{\gamma}\right)\left(\boldsymbol{X}_{-}\boldsymbol{\delta}_n + \boldsymbol{\eta}_{-}(\mathbf{z}'\boldsymbol{\pi}_q) + \hat{\varepsilon}\right) \\
&= \boldsymbol{\delta}_n'\boldsymbol{X}_{-}'\left(\boldsymbol{I}_n - \boldsymbol{P}_{\gamma}\right)\boldsymbol{X}_{-}\boldsymbol{\delta}_n + \boldsymbol{\eta}_{-}(\boldsymbol{z}\boldsymbol{\pi}_q)'\left(\boldsymbol{I}_n - \boldsymbol{P}_{\gamma}\right)\boldsymbol{\eta}_{-}(\mathbf{z}'\boldsymbol{\pi}_q) + \hat{\varepsilon}'\left(\boldsymbol{I}_n - \boldsymbol{P}_{\gamma}\right)\hat{\varepsilon} \\
&\quad + 2\boldsymbol{\delta}_n'\boldsymbol{X}_{-}'\left(\boldsymbol{I}_n - \boldsymbol{P}_{\gamma}\right)\boldsymbol{\eta}_{-}(\boldsymbol{z}\boldsymbol{\pi}_q) + 2\boldsymbol{\delta}_n'X_{-}'\left(\boldsymbol{I}_n - \boldsymbol{P}_{\gamma}\right)\hat{\varepsilon} + 2\boldsymbol{\eta}_{-}(\mathbf{z}\boldsymbol{\pi}_q)'\left(\boldsymbol{I}_n - \boldsymbol{P}_{\gamma}\right)\hat{\varepsilon} \\
&= S_{n1} + S_{n2} + S_{n3} + 2\left(S_{n4} + S_{n5} + S_{n6}\right),
\end{aligned}$$

where $\boldsymbol{X}_{-}$ is an $n \times d_x$ matrix with its t^{th} row equal to $\mathbf{x}_t' I(q_t \leq \gamma_0)$, and $\boldsymbol{\eta}_{-}(\boldsymbol{z}\boldsymbol{\pi}_q)$ is an

$n\times 1$ vector with its t^{th} element equal to $\eta_n(\mathbf{z}_t'\boldsymbol{\pi}_q)\,I(q_t\le\gamma_0)$, $\hat{\boldsymbol{\varepsilon}}$ is an $n\times 1$ vector stacking up $\hat{\varepsilon}_t = h_2(\mathbf{z}_t'\boldsymbol{\pi}_q) - h_2^*(\mathbf{z}_t'\hat{\boldsymbol{\pi}}_q) + \varepsilon_t$, and the subscript j in S_{nj} (j=1,…,6) is labelled according to the ordering of appearance.

(i) Under Assumption **1**(v), we have

$$
\begin{aligned}
S_{n1} &= \boldsymbol{\delta}_n'\boldsymbol{X}_-'\left(\boldsymbol{I}_n-\boldsymbol{P}_\gamma\right)\boldsymbol{X}_-\boldsymbol{\delta}_n = n^{-2\varsigma}\boldsymbol{\delta}_0'\boldsymbol{X}_-'\left(\boldsymbol{I}_n-\boldsymbol{P}_\gamma\right)\boldsymbol{X}_-\boldsymbol{\delta}_0,\\
S_{n2} &= \boldsymbol{\eta}_-\left(\boldsymbol{z}\boldsymbol{\pi}_q\right)'\left(\boldsymbol{I}_n-\boldsymbol{P}_\gamma\right)\boldsymbol{\eta}_-\left(\boldsymbol{z}\boldsymbol{\pi}_q\right) = n^{-2\varrho}\boldsymbol{\eta}_{0,-}\left(\boldsymbol{z}\boldsymbol{\pi}_q\right)'\left(\boldsymbol{I}_n-\boldsymbol{P}_\gamma\right)\boldsymbol{\eta}_{0,-}\left(\boldsymbol{z}\boldsymbol{\pi}_q\right),\\
S_{n4} &= \boldsymbol{\delta}_n'\boldsymbol{X}_-'\times\left(\boldsymbol{I}_n-\boldsymbol{P}_\gamma\right)\boldsymbol{\eta}_-\left(\boldsymbol{z}\boldsymbol{\pi}_q\right) = n^{-\varsigma-\varrho}\boldsymbol{\delta}_0'\boldsymbol{X}_-'\left(\boldsymbol{I}_n-\boldsymbol{P}_\gamma\right)\boldsymbol{\eta}_{0,-}\left(\mathbf{z}\boldsymbol{\pi}_q\right),
\end{aligned}
$$

where $\boldsymbol{\eta}_{0,-}(\boldsymbol{z}\boldsymbol{\pi}_q)$ is an $n\times 1$ vector with its t^{th} element equal to $\eta_0(\mathbf{z}_t'\boldsymbol{\pi}_q)\,I(q_t\le\gamma_0)$. By Lemmas 1 and 2 of the Online Appendix, we have

$$
\begin{aligned}
\max_{\gamma\in[\underline{\gamma},\overline{\gamma}]}\left|n^{-1}\boldsymbol{\delta}_0'\boldsymbol{X}_-'\boldsymbol{P}_\gamma\boldsymbol{X}_-\boldsymbol{\delta}_0-\boldsymbol{\delta}_0'\boldsymbol{g}_1'(\gamma)\,\boldsymbol{\Sigma}^{-1}_{\mathcal{X}^*\mathcal{X}^{*\prime},\gamma}\boldsymbol{g}_1(\gamma)\,\boldsymbol{\delta}_0\right| &= o_p(1), &\text{(A.3)}\\
\max_{\gamma\in[\underline{\gamma},\overline{\gamma}]}\left|n^{-1}\boldsymbol{\eta}_{0,-}(\boldsymbol{z}\boldsymbol{\pi}_q)'\boldsymbol{P}_\gamma\boldsymbol{\eta}_{0,-}(\boldsymbol{z}\boldsymbol{\pi}_q)-\boldsymbol{g}_2'(\gamma)\,\boldsymbol{\Sigma}^{-1}_{\mathcal{X}^*\mathcal{X}^{*\prime},\gamma}\boldsymbol{g}_2(\gamma)\right| &= o_p(1), &\text{(A.4)}\\
\max_{\gamma\in[\underline{\gamma},\overline{\gamma}]}\left|n^{-1}\boldsymbol{\delta}_0'\boldsymbol{X}_-'\boldsymbol{P}_\gamma\boldsymbol{\eta}_{0,-}(\boldsymbol{z}\boldsymbol{\pi}_q)-\boldsymbol{\delta}_0'\boldsymbol{g}_1'(\gamma)\,\boldsymbol{\Sigma}^{-1}_{\mathcal{X}^*\mathcal{X}^{*\prime},\gamma}\boldsymbol{g}_2(\gamma)\right| &= o_p(1). &\text{(A.5)}
\end{aligned}
$$

Under Assumption 1(i), $\{(\mathbf{x}_t', q_t, \mathbf{z}_t', u_t)\}$ is ergodic by Proposition 3.44 in White (2001). As $E(\|\mathbf{x}_t\mathbf{x}_t'\|)<\infty$ under Assumption 2(i) and a uniformly bounded $\eta_0(\cdot)$ under Assumption 2(iii), we apply the law of large numbers for stationary ergodic time series data and obtain

$$
\begin{aligned}
n^{-1}\boldsymbol{\delta}_0'\boldsymbol{X}_-'\boldsymbol{X}_-\boldsymbol{\delta}_0 \overset{a.s.}{\to} \boldsymbol{\delta}_0'E\left[\mathbf{x}_t\mathbf{x}_t'I(q_t\le\gamma_0)\right]\boldsymbol{\delta}_0 &\equiv \boldsymbol{\delta}_0'\boldsymbol{m}_1(\gamma_0)\,\boldsymbol{\delta}_0, &\text{(A.6)}\\
n^{-1}\boldsymbol{\eta}_{0,-}(\boldsymbol{z}\boldsymbol{\pi}_q)'\boldsymbol{\eta}_{0,-}(\boldsymbol{z}\boldsymbol{\pi}_q) \overset{a.s.}{\to} E\left[\eta_0^2(\mathbf{z}_t'\boldsymbol{\pi}_q)\,I(q_t\le\gamma_0)\right] &\equiv \boldsymbol{m}_2(\gamma_0), &\text{(A.7)}\\
n^{-1}\boldsymbol{\delta}_0'\boldsymbol{X}_-'\boldsymbol{\eta}_{0,-}(\boldsymbol{z}\boldsymbol{\pi}_q) \overset{a.s.}{\to} \boldsymbol{\delta}_0'E\left[\mathbf{x}_t\eta_0(\mathbf{z}_t'\boldsymbol{\pi}_q)\,I(q_t\le\gamma_0)\right] &\equiv \boldsymbol{\delta}_0'\boldsymbol{m}_3(\gamma_0). &\text{(A.8)}
\end{aligned}
$$

(ii) We consider $S_{n,3}(\gamma) = \hat{\boldsymbol{\varepsilon}}'\left(\boldsymbol{I}_n-\boldsymbol{P}_\gamma\right)\hat{\boldsymbol{\varepsilon}} = \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}-\hat{\boldsymbol{\varepsilon}}'\boldsymbol{P}_\gamma\hat{\boldsymbol{\varepsilon}}$. Note that $h_2(\mathbf{z}_t'\boldsymbol{\pi}_q) - h_2^*(\mathbf{z}_t'\hat{\boldsymbol{\pi}}_q)$

$= [h_2(\mathbf{z}_t'\boldsymbol{\pi}_q) - h_2^*(\mathbf{z}_t'\boldsymbol{\pi}_q)] + [h_2^*(\mathbf{z}_t'\boldsymbol{\pi}_q) - h_2^*(\mathbf{z}_t'\hat{\boldsymbol{\pi}}_q)]$, where the first term is uniformly bounded by $O\left(L_n^{-\xi}\right)$ for all t by Assumption **2**(iv), and the second term $h_2^*(\mathbf{z}_t'\boldsymbol{\pi}_q) - h_2^*(\mathbf{z}_t'\hat{\boldsymbol{\pi}}_q)$ $= \boldsymbol{\alpha}_{L_n,2}'[\boldsymbol{\Phi}_{L_n}(\mathbf{z}_t'\boldsymbol{\pi}_q) - \boldsymbol{\Phi}_{L_n}(\mathbf{z}_t'\hat{\boldsymbol{\pi}}_q)] = \boldsymbol{\alpha}_{L_n,2}'\boldsymbol{\Phi}_{L_n}^{(1)}(\mathbf{z}_t'\tilde{\pi}_q)\mathbf{z}_t'(\boldsymbol{\pi}_q - \hat{\boldsymbol{\pi}}_q)$ with $\mathbf{z}_t'\tilde{\boldsymbol{\pi}}_q$ lying between $\mathbf{z}_t'\boldsymbol{\pi}_q$ and $\mathbf{z}_t'\hat{\boldsymbol{\pi}}_q$. It therefore follows

$$\frac{1}{n}\sum_{t=1}^{n}[h_2(\mathbf{z}_t'\boldsymbol{\pi}_q) - h_2^*(\mathbf{z}_t'\hat{\boldsymbol{\pi}}_q)]^2 = O_p\left(L_n^{-2\xi} + n^{-1}\|\boldsymbol{\Phi}_{L_n}\|_1^2\right) \tag{A.9}$$

under Assumptions 1(i)-(ii) and 2(iv)-(v) as $\sum_{l=1}^{\infty}\alpha_{l,2}^2 = \int h_2^2(\omega)\,dw < \infty$ by Parseval's equality if $h_2(\cdot)$ is squared integrable over its domain. Therefore, we have

$$n^{-1}\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} = n^{-1}\sum_{t=1}^{n}\varepsilon_t^2 + O\left(L_n^{-2\xi}\right) + O_p\left(n^{-1/2}\|\boldsymbol{\Phi}_{L_n}\|_1\right) \tag{A.10}$$

under Assumption **1**(iii). In addition, we have $\hat{\boldsymbol{\varepsilon}}'\boldsymbol{P}_\gamma\hat{\boldsymbol{\varepsilon}} = \|\boldsymbol{P}_\gamma\hat{\boldsymbol{\varepsilon}}\|^2 \le \|h_2(\boldsymbol{z}\boldsymbol{\pi}_q) - h_2^*(\mathbf{z}\hat{\boldsymbol{\pi}}_q)\|^2$ $+ \|\boldsymbol{P}_\gamma\boldsymbol{\varepsilon}\|^2$ by the triangular inequality, $|\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}| \le \boldsymbol{\lambda}_{\max}(\boldsymbol{A})\boldsymbol{x}'\boldsymbol{x}$, and an idempotent matrix's eigenvalues equal to either zeros or ones. Applying Lemma 3 of the Online Appendix gives $\max_{\gamma\in[\underline{\gamma},\overline{\gamma}]} n^{-1}\boldsymbol{\varepsilon}'\boldsymbol{P}_\gamma\boldsymbol{\varepsilon} = O_p(L_n/n)$. Therefore, we obtain

$$\max_{\gamma\in[\underline{\gamma},\overline{\gamma}]}\left|n^{-1}S_{n,3}(\gamma) - n^{-1}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\right| = O_p\left(\vartheta_n^2\right), \tag{A.11}$$

where we denote $\vartheta_n = L_n^{-\xi} + \sqrt{L_n/n} + n^{-1/2}\|\boldsymbol{\Phi}_{L_n}\|_1$.

(iii) We consider $S_{n,5}(\gamma) = \delta_n'\boldsymbol{X}_-'(\boldsymbol{I}_n - \boldsymbol{P}_\gamma)\hat{\boldsymbol{\varepsilon}} = n^{-\varsigma}\boldsymbol{\delta}_0'\boldsymbol{X}_-'(\boldsymbol{I}_n - P_\gamma)\hat{\boldsymbol{\varepsilon}}$, where

$$\begin{aligned} n^{-1}\boldsymbol{\delta}_0'\boldsymbol{X}_-'\hat{\boldsymbol{\varepsilon}} &= n^{-1}\sum_{t=1}^{n}\boldsymbol{\delta}_0'\mathbf{x}_t'I(q_t\le\gamma_0)[h_2(\mathbf{z}_t'\boldsymbol{\pi}_q) - h_2^*(\mathbf{z}_t'\hat{\boldsymbol{\pi}}_q) + \varepsilon_t] \\ &= O_p\left(L_n^{-\xi} + \|\boldsymbol{\Phi}_{L_n}\|_1 n^{-1/2} + n^{-1/2}\right). \end{aligned} \tag{A.12}$$

As $|\boldsymbol{x}'\boldsymbol{A}\boldsymbol{y}| \leq \|\boldsymbol{A}\boldsymbol{x}\| \|\boldsymbol{A}\boldsymbol{y}\|$ for any conformable vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, and matrix $\boldsymbol{A}$, we have

$$\max_{\gamma\in[\underline{\gamma},\overline{\gamma}]} \frac{\boldsymbol{\delta}_0'\boldsymbol{X}_-'\boldsymbol{P}_\gamma\hat{\boldsymbol{\varepsilon}}}{n} \leq \frac{1}{n} \max_{\gamma\in[\underline{\gamma},\overline{\gamma}]} \|\boldsymbol{P}_\gamma\boldsymbol{X}_-\boldsymbol{\delta}_0\| \|\boldsymbol{P}_\gamma\hat{\boldsymbol{\varepsilon}}\| = O_p(\vartheta_n). \tag{A.13}$$

It follows that

$$\max_{\gamma\in[\underline{\gamma},\overline{\gamma}]} \left|n^{-1}S_{n,5}(\gamma)\right| = O_p\left(n^{-\varsigma}\vartheta_n\right). \tag{A.14}$$

Similarly, for $S_{n,6}(\gamma) = \boldsymbol{\eta}_-(\mathbf{z}'\boldsymbol{\pi}_q)'(\boldsymbol{I}_n - \boldsymbol{P}_\gamma)\hat{\boldsymbol{\varepsilon}} = n^{-\varrho}\boldsymbol{\eta}_{0,-}(\mathbf{z}'\boldsymbol{\pi}_q)'(\boldsymbol{I}_n - \boldsymbol{P}_\gamma)\hat{\boldsymbol{\varepsilon}}$,we obtain

$$\max_{\gamma\in[\underline{\gamma},\overline{\gamma}]} \left|n^{-1}S_{n,6}(\gamma)\right| = O_p\left(n^{-\varrho}\vartheta_n\right). \tag{A.15}$$

(iv) Taking together (A.3)-(A.15) gives

$$\max_{\gamma\in[\underline{\gamma},\overline{\gamma}]} \left|n^{-1}S_n(\gamma) - S_1(\gamma)\right| = o_p(1) \tag{A.16}$$

where

$S_1(\gamma) = \bar{\sigma}_\varepsilon^2 + n^{-2\varsigma}\boldsymbol{\delta}_0'\boldsymbol{m}_1(\gamma_0)\boldsymbol{\delta}_0 + n^{-2\varrho}\boldsymbol{m}_2(\gamma_0) + 2n^{-\varsigma-\varrho}\boldsymbol{\delta}_0'\boldsymbol{m}_3(\gamma_0) - \mu(\gamma)$,

$\bar{\sigma}_\varepsilon^2 = \lim_{n\to\infty} n^{-1}\sum_{t=1}^n E(\varepsilon_t^2)$, and

$\mu(\gamma) = [n^{-\varsigma}\boldsymbol{g}_1(\gamma)\boldsymbol{\delta}_0 + n^{-\varrho}\boldsymbol{g}_2(\gamma)]'\boldsymbol{\Sigma}_{\mathcal{X}^*\mathcal{X}^{*\prime},\gamma}^{-1}[n^{-\varsigma}\boldsymbol{g}_1(\gamma)\boldsymbol{\delta}_0 + n^{-\varrho}\boldsymbol{g}_2(\gamma)]$. Evidently, $S_1(\gamma)$ is continuous in γ.

(v) Denote $\boldsymbol{D}(\gamma) = E[\boldsymbol{\chi}_t^*\boldsymbol{\chi}_t^{*\prime}I(q_t \leq \gamma)]$ and $\boldsymbol{M}(\gamma) = E[\boldsymbol{\chi}_t^*\eta_0(\mathbf{z}_t'\boldsymbol{\pi}_q)I(q_t \leq \gamma)]$, where $\boldsymbol{\chi}_t^* = \left[\mathbf{x}_t', \boldsymbol{\Phi}_{L_n}'(\mathbf{z}_t'\boldsymbol{\pi}_q)\right]'$. Then, we have

$$\boldsymbol{\Sigma}_{\mathcal{X}^*\mathcal{X}^{*\prime},\gamma} = E\left(\mathcal{X}_{t,\gamma}^*\mathcal{X}_{t,\gamma}^{*\prime}\right) = \begin{bmatrix} \boldsymbol{D}(\gamma) & \mathbf{0} \\ \mathbf{0}' & \boldsymbol{\Sigma}_{\chi^*\chi^{*\prime}} - \boldsymbol{D}(\gamma) \end{bmatrix},$$

and for $\gamma \in [\gamma_0, \overline{\gamma}]$, $\boldsymbol{g}_1'(\gamma) = [\boldsymbol{D}_1'(\gamma), \mathbf{0}']$ and $\boldsymbol{g}_2'(\gamma) = [\boldsymbol{M}'(\gamma), \mathbf{0}']$, where $\boldsymbol{D}_1(\gamma) =$

$E\left[\boldsymbol{\chi}_t^* \mathbf{x}_t' I\left(q_t \leq \gamma\right)\right]$. As for any random variable z, $d\left\{E\left[zI\left(q \leq \gamma\right)\right]\right\}/d\gamma = E\left(z|q = \gamma\right) f_q\left(\gamma\right)$, we have $\boldsymbol{D}^{(1)}\left(\gamma\right) = E\left(\boldsymbol{\chi}_t^* \boldsymbol{\chi}_t^{*\prime}|q_t = \gamma\right) f_q\left(\gamma\right)$, $\boldsymbol{M}^{(1)}\left(\gamma\right) = E\left(\boldsymbol{\chi}_t^* \eta_0\left(\mathbf{z}_t'\boldsymbol{\pi}_q\right)|q_t = \gamma\right) f_q\left(\gamma\right)$, and $\partial\left(\boldsymbol{\Sigma}_{\mathcal{X}^*\mathcal{X}^{*\prime},\gamma}\right)/\partial\gamma = diag\left\{1, -1\right\} \otimes \boldsymbol{D}^{(1)}\left(\gamma\right)$. Moreover, applying Propositions 17.3(a) and 17.25 in Seber (2008), for any differentiable function $\boldsymbol{a}\left(\gamma\right)$, we have

$$\frac{d\left(\boldsymbol{a}\left(\gamma\right)\boldsymbol{\Sigma}_{\mathcal{X}^*\mathcal{X}^{*\prime},\gamma}^{-1}\boldsymbol{a}\left(\gamma\right)\right)}{d\gamma} = 2\boldsymbol{a}'\left(\gamma\right)\boldsymbol{\Sigma}_{\mathcal{X}^*\mathcal{X}^{*\prime},\gamma}^{-1}\frac{d\left(\boldsymbol{a}\left(\gamma\right)\right)}{d\gamma} - \boldsymbol{a}'\left(\gamma\right)\boldsymbol{\Sigma}_{\mathcal{X}^*\mathcal{X}^{*\prime},\gamma}^{-1}\frac{d\left(\boldsymbol{\Sigma}_{\mathcal{X}^*\mathcal{X}^{*\prime},\gamma}\right)}{d\gamma}\boldsymbol{\Sigma}_{\mathcal{X}^*\mathcal{X}^{*\prime},\gamma}^{-1}\boldsymbol{a}\left(\gamma\right).$$

Hence, we have

$$-\frac{d\mu\left(\gamma\right)}{d\gamma} = \left[n^{-\varsigma}\boldsymbol{D}_1\left(\gamma\right)\boldsymbol{\delta}_0 + n^{-\varrho}\boldsymbol{M}\left(\gamma\right)\right]'\boldsymbol{D}^{-1}\left(\gamma\right)\boldsymbol{D}^{(1)}\left(\gamma\right)\boldsymbol{D}^{-1}\left(\gamma\right)\left[n^{-\varsigma}\boldsymbol{D}_1\left(\gamma\right)\boldsymbol{\delta}_0 + n^{-\varrho}\boldsymbol{M}\left(\gamma\right)\ \right] > 0$$

as $\boldsymbol{D}^{-1}\left(\gamma\right)\boldsymbol{D}^{(1)}\left(\gamma\right)\boldsymbol{D}^{-1}\left(\gamma\right)$ is a p.d.f. matrix uniformly over γ under Assumption **2**(ii). Therefore, $S_1\left(\gamma\right)$ is a strictly increasing function over $\gamma \in \left[\gamma_0, \overline{\gamma}\right]$.

By symmetry, we can rewrite $S_n\left(\gamma\right)$ as

$$S_n\left(\gamma\right) = \left[-\boldsymbol{X}_+\boldsymbol{\delta} - \boldsymbol{\eta}_+\left(\mathbf{z}'\boldsymbol{\pi}_q\right) + \hat{\varepsilon}\right]'\left(\boldsymbol{I}_n - \boldsymbol{P}_\gamma\right)\left[-\boldsymbol{X}_+\boldsymbol{\delta} - \boldsymbol{\eta}_+\left(\mathbf{z}'\boldsymbol{\pi}_q\right) + \hat{\varepsilon}\right]$$

where $\boldsymbol{X}_+$ is an $n \times d_x$ matrix with its t^{th} row equal to $\mathbf{x}_t' I\left(q_t > \gamma_0\right)$, and $\boldsymbol{\eta}_+\left(\mathbf{z}'\boldsymbol{\pi}_q\right)$ is an $n \times 1$ vector with its t^{th} element equal to $\eta\left(\mathbf{z}_t'\boldsymbol{\pi}_q\right) I\left(q_t > \gamma_0\right)$, $\hat{\boldsymbol{\varepsilon}}$ is an $n \times 1$ vector stacking up $\hat{\varepsilon}_t = h_1\left(\mathbf{z}_t'\boldsymbol{\pi}_q\right) - h_1^*\left(\mathbf{z}_t'\hat{\boldsymbol{\pi}}_q\right) + \varepsilon_t$. Applying the same proof method used above, we can show that

$$\max_{\gamma \in \left[\underline{\gamma}, \overline{\gamma}\right]}\left|n^{-1}S_n\left(\gamma\right) - S_2\left(\gamma\right)\right| = o_p\left(1\right) \tag{A.17}$$

where $S_2\left(\gamma\right)$ equals $S_1\left(\gamma\right)$ with $I\left(q_t \leq \gamma_0\right)$ replaced by $I\left(q_t > \gamma_0\right)$. For $\gamma \in \left[\underline{\gamma}, \gamma_0\right]$, we have

$\boldsymbol{g}_1'(\gamma) = \left[\mathbf{0}', \boldsymbol{\Sigma}'_{\boldsymbol{\chi}^*\boldsymbol{\chi}^{*\prime}} - \boldsymbol{D}'(\gamma)\right]$ and $\boldsymbol{g}_2'(\gamma) = \left[\mathbf{0}', E\left[\boldsymbol{\chi}_t^*\eta_0\left(\mathbf{z}_t'\boldsymbol{\pi}_q\right)I\left(q_t > \gamma\right)\right]\right]$, so that

$$
\begin{aligned}
& -\frac{d\mu(\gamma)}{d\gamma} = -\left\{\left[\boldsymbol{\Sigma}_{\boldsymbol{\chi}^*\boldsymbol{\chi}^{*\prime}} - \boldsymbol{D}(\gamma)\right]\boldsymbol{\delta}_0 + E\left[\boldsymbol{\chi}_t^{*\prime}\eta_0\left(\mathbf{z}_t'\boldsymbol{\pi}_q\right)I\left(q_t > \gamma\right)\right]\right\}'\left[\boldsymbol{\Sigma}_{\boldsymbol{\chi}^*\boldsymbol{\chi}^{*\prime}} - \boldsymbol{D}(\gamma)\right]^{-1}\boldsymbol{D}^{(1)}(\gamma) \\
& \times\left[\boldsymbol{\Sigma}_{\boldsymbol{\chi}^*\boldsymbol{\chi}^{*\prime}} - \boldsymbol{D}(\gamma)\right]^{-1}\left\{\left[\boldsymbol{\Sigma}_{\boldsymbol{\chi}^*\boldsymbol{\chi}^{*\prime}} - \boldsymbol{D}(\gamma)\right]\boldsymbol{\delta}_0 + E\left[\boldsymbol{\chi}_t^{*\prime}\eta_0\left(\mathbf{z}_t'\boldsymbol{\pi}_q\right)I\left(q_t > \gamma\right)\right]\right\} \\
< \;& 0.
\end{aligned}
$$

Therefore, $S_2(\gamma)$ is a strictly decreasing function of $\gamma \in \left[\underline{\gamma}, \gamma_0\right]$.

To sum up, we have

$$
\max_{\gamma\in\left[\underline{\gamma},\overline{\gamma}\right]}\left|n^{-1}S_n(\gamma) - S(\gamma)\right| = o_p(1) \tag{A.18}
$$

where $S(\gamma) = S_1(\gamma)I(\gamma > \gamma_0) + S_2(\gamma)I(\gamma \leq \gamma_0)$ is continuous function of γ and is uniquely minimized at γ_0. It then follows $\hat{\gamma} \xrightarrow{p} \gamma_0$.

Finally, we verify $\left\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right\| = O_p\left(\vartheta_n + n^{-\varsigma}\right)$ in Lemma 4 of the Online Appendix. This completes the proof of this theorem.

Proof of Theorem 2: In matrix form, we have $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}_2 + \boldsymbol{X}_{\gamma_0}\boldsymbol{\delta}_n + \boldsymbol{h}_2(\boldsymbol{z}\boldsymbol{\pi}_q) + \boldsymbol{\eta}_{n,\gamma_0}(\boldsymbol{z}\boldsymbol{\pi}_q) + \boldsymbol{\varepsilon}$, and

$$
\begin{aligned}
& \boldsymbol{y} - \mathcal{X}_\gamma\hat{\boldsymbol{\theta}} \\
= \;& \boldsymbol{X}\boldsymbol{\beta}_2 + \boldsymbol{X}_{\gamma_0}\boldsymbol{\delta}_n + \boldsymbol{h}_2(\mathbf{z}\boldsymbol{\pi}_q) + \boldsymbol{\eta}_{n,\gamma_0}(\boldsymbol{z}\boldsymbol{\pi}_q) + \boldsymbol{\varepsilon} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_2 - \boldsymbol{X}_\gamma\hat{\boldsymbol{\delta}}_n - \hat{\boldsymbol{h}}_2^*(\boldsymbol{z}\hat{\boldsymbol{\pi}}_q) - \hat{\boldsymbol{\eta}}_{n,\gamma}^*(\boldsymbol{z}\hat{\boldsymbol{\pi}}_q) \\
= \;& \boldsymbol{\varepsilon} + \boldsymbol{\Delta}_n - \Delta\boldsymbol{X}_\gamma\hat{\boldsymbol{\delta}}_n - \Delta\hat{\boldsymbol{\eta}}_{n,\gamma}^*(\boldsymbol{z}\hat{\boldsymbol{\pi}}_q)
\end{aligned}
$$

where $\Delta\boldsymbol{X}_\gamma = \boldsymbol{X}_\gamma - \boldsymbol{X}_{\gamma_0}$, $\Delta\hat{\boldsymbol{\eta}}_{n,\gamma}^* = \hat{\boldsymbol{\eta}}_{n,\gamma}^*(\boldsymbol{z}\hat{\boldsymbol{\pi}}_q) - \hat{\boldsymbol{\eta}}_{n,\gamma_0}^*(\boldsymbol{z}\hat{\boldsymbol{\pi}}_q)$, $\boldsymbol{\Delta}_n = \boldsymbol{X}\left(\boldsymbol{\beta}_2 - \hat{\boldsymbol{\beta}}_2\right) + \boldsymbol{X}_{\gamma_0}\left(\boldsymbol{\delta}_n - \hat{\boldsymbol{\delta}}_n\right) + \boldsymbol{h}_2(\boldsymbol{z}\boldsymbol{\pi}_q) - \hat{\boldsymbol{h}}_2^*(\boldsymbol{z}\hat{\boldsymbol{\pi}}_q) + \boldsymbol{\eta}_{n,\gamma_0}(\boldsymbol{z}\boldsymbol{\pi}_q) - \hat{\boldsymbol{\eta}}_{n,\gamma_0}^*(\boldsymbol{z}\hat{\boldsymbol{\pi}}_q)$, and the typical element of $\boldsymbol{X}$, $\boldsymbol{X}_\gamma$, $\boldsymbol{h}_2(\mathbf{z}\boldsymbol{\pi}_q)$, $\hat{\boldsymbol{h}}_2^*(\boldsymbol{z}\hat{\boldsymbol{\pi}}_q)$, $\boldsymbol{\eta}_{n,\gamma}(\boldsymbol{z}\boldsymbol{\pi}_q)$, and $\hat{\boldsymbol{\eta}}_{n,\gamma}^*(\boldsymbol{z}\hat{\boldsymbol{\pi}}_q)$ are $\mathbf{x}_t$, $\mathbf{x}_tI(q_t \leq \gamma)$, $\boldsymbol{h}_2(\mathbf{z}_t'\boldsymbol{\pi}_q)$,

$\hat{\boldsymbol{h}}_2^*(\mathbf{z}_t'\hat{\boldsymbol{\pi}}_q)$, $\boldsymbol{\eta}_n(\mathbf{z}_t'\boldsymbol{\pi}_q) I(q_t \leq \gamma)$, and $\hat{\boldsymbol{\eta}}_{n,\gamma}^*(\mathbf{z}_t'\hat{\boldsymbol{\pi}}_q) I(q_t \leq \gamma)$, respectively. Denote $S_n(\gamma) = \left(\boldsymbol{y} - \mathcal{X}_\gamma \hat{\boldsymbol{\theta}}\right)' \left(\boldsymbol{y} - \mathcal{X}_\gamma \hat{\boldsymbol{\theta}}\right)$. Then $\hat{\gamma}$ minimizes

$$
\begin{aligned}
& S_n(\gamma) - S_n(\gamma_0) \\
= \; & \left(\Delta \hat{\boldsymbol{\eta}}_{n,\gamma}^*(\boldsymbol{z}\hat{\boldsymbol{\pi}}_q) + \Delta \boldsymbol{X}_\gamma \hat{\boldsymbol{\delta}}_n\right)' \left(\Delta \hat{\boldsymbol{\eta}}_{n,\gamma}^*(\boldsymbol{z}\hat{\boldsymbol{\pi}}_q) + \Delta \boldsymbol{X}_\gamma \hat{\boldsymbol{\delta}}_n\right) \\
& -2(\varepsilon + \boldsymbol{\Delta}_n)' \left(\Delta \hat{\boldsymbol{\eta}}_{n,\gamma}^*(\boldsymbol{z}\hat{\boldsymbol{\pi}}_q) + \Delta \boldsymbol{X}_\gamma \hat{\boldsymbol{\delta}}_n\right) \\
= \; & \kappa_n' \sum_{t=1}^n \boldsymbol{\chi}_t \boldsymbol{\chi}_t' d_t^2(\gamma, \gamma_0) \kappa_n - 2\hat{\kappa}_n' \sum_{t=1}^n \varepsilon_t \boldsymbol{\chi}_t d_t(\gamma, \gamma_0) \\
& -2\hat{\kappa}_n' \sum_{t=1}^n \left[\left(\boldsymbol{\beta}_2 - \hat{\boldsymbol{\beta}}_2\right)' \mathbf{x}_t + \left(\boldsymbol{\delta}_n - \hat{\boldsymbol{\delta}}_n\right)' \mathbf{x}_t I(q_t \leq \gamma_0)\right. \\
& \left. + h_2(\mathbf{z}_t'\boldsymbol{\pi}_q) - \hat{h}_2^*(\mathbf{z}_t'\hat{\boldsymbol{\pi}}_q) + n^{-\varrho}\left(\eta_0(\mathbf{z}_t'\boldsymbol{\pi}_q) - \hat{\eta}_0^*(\mathbf{z}_t'\hat{\boldsymbol{\pi}}_q)\right) I(q_t \leq \gamma_0)\right] \boldsymbol{\chi}_t d_t(\gamma, \gamma_0) \\
& + (\hat{\boldsymbol{\kappa}}_n + \boldsymbol{\kappa}_n)' \sum_{t=1}^n \boldsymbol{\chi}_t \boldsymbol{\chi}_t' d_t^2(\gamma, \gamma_0) (\hat{\boldsymbol{\kappa}}_n - \boldsymbol{\kappa}_n) \\
= \; & S_{n,1}^*(\gamma) - 2S_{n,2}^*(\gamma) - 2S_{n,3}^*(\gamma) + S_{n,4}^*(\gamma)
\end{aligned} \tag{A.19}
$$

where we denote $\boldsymbol{\chi}_t = \left[\mathbf{x}_t', \boldsymbol{\Phi}_{L_n}'(\mathbf{z}_t'\hat{\boldsymbol{\pi}}_q)\right]'$, $\hat{\boldsymbol{\kappa}}_n = \left[\hat{\boldsymbol{\delta}}_n', n^{-\varrho}\hat{\boldsymbol{\alpha}}_{L_n,0}'\right]'$ and $\boldsymbol{\kappa}_n = \left[n^{-\varsigma}\boldsymbol{\delta}_0', n^{-\varrho}\boldsymbol{\alpha}_{L_n,0}'\right]'$ with $\hat{\boldsymbol{\delta}}_n = \hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2$, $n^{-\varrho}\hat{\boldsymbol{\alpha}}_{L_n,0} = \hat{\boldsymbol{\alpha}}_{L_n,1} - \hat{\boldsymbol{\alpha}}_{L_n,2}$, $n^{-\varsigma}\boldsymbol{\delta}_0 = \boldsymbol{\beta}_1 - \boldsymbol{\beta}_2$ and $n^{-\varrho}\boldsymbol{\alpha}_{L_n,0} = \boldsymbol{\alpha}_{L_n,1} - \boldsymbol{\alpha}_{L_n,2}$. Closely following the proof of A.9 in Hansen (2000), we can show that

$$
a_n(\hat{\gamma} - \gamma_0) = \arg\max_v \boldsymbol{Q}_n(v) = O_p(1),
$$

where $a_n = n^{1-2\min(\varsigma,\varrho)}$ and $\boldsymbol{Q}_n(v) = S_n(\gamma_0) - S_n(\gamma_0 + v/a_n)$.

Now, we consider $S_{n,1}^*(\gamma) = \boldsymbol{\kappa}_n' \sum_{t=1}^n \boldsymbol{\chi}_t \boldsymbol{\chi}_t' d_t^2(\gamma, \gamma_0) \boldsymbol{\kappa}_n$. For any given $v \in [\underline{v}, \overline{v}]$, a finite interval, we have

$$\begin{aligned}
S_{n,1}^{*}(v) &= \boldsymbol{\kappa}_n' \sum_{t=1}^{n} \boldsymbol{\chi}_t \boldsymbol{\chi}_t' d_t^2 \left(\gamma_0 + \frac{v}{a_n}, \gamma_0 \right) \boldsymbol{\kappa}_n \\
&= G_n(v) + 2 \sum_{t=1}^{n} \boldsymbol{\kappa}_n' \left(\boldsymbol{\chi}_t - \boldsymbol{\chi}_t^{*} \right) \boldsymbol{\chi}_t^{*\prime} \boldsymbol{\kappa}_n d_t^2 \left(\gamma_0 + \frac{v}{a_n}, \gamma_0 \right) \\
&\quad + \sum_{t=1}^{n} \left[\boldsymbol{\kappa}_n' \left(\boldsymbol{\chi}_t - \boldsymbol{\chi}_t^{*} \right) \right]^2 d_t^2 \left(\gamma_0 + \frac{v}{a_n}, \gamma_0 \right) \\
&= G_n(v) + 2n^{-2\varrho} \sum_{t=1}^{n} \boldsymbol{\alpha}_{L_n,0}' \left[\boldsymbol{\Phi}_{L_n}\left(\mathbf{z}_t' \hat{\boldsymbol{\pi}}_q \right) - \boldsymbol{\Phi}_{L_n}\left(\mathbf{z}_t' \boldsymbol{\pi}_q \right) \right] \eta_0^{*}\left(\mathbf{z}_t' \boldsymbol{\pi}_q \right) d_t^2 \left(\gamma_0 + \frac{v}{a_n}, \gamma_0 \right) \\
&\quad + n^{-2\varrho} \sum_{t=1}^{n} \left(\boldsymbol{\alpha}_{L_n,0}' \left[\boldsymbol{\Phi}_{L_n}\left(\mathbf{z}_t' \hat{\boldsymbol{\pi}}_q \right) - \boldsymbol{\Phi}_{L_n}\left(\mathbf{z}_t' \boldsymbol{\pi}_q \right) \right] \right)^2 d_t^2 \left(\gamma_0 + \frac{v}{a_n}, \gamma_0 \right) \\
&= G_n(v) + A_n(v),
\end{aligned}$$

where we denote $\boldsymbol{\chi}_t^{*} = \left[\mathbf{x}_t', \boldsymbol{\Phi}_{L_n}'\left(\mathbf{z}_t' \boldsymbol{\pi}_q \right) \right]'$, and $G_n(v) = \sum_{t=1}^{n} \left(\boldsymbol{\kappa}_n' \boldsymbol{\chi}_t^{*} \right)^2 d_t^2 \left(\gamma_0 + v/a_n, \gamma_0 \right)$ is uniformly bounded in probability over $v \in [\underline{v}, \overline{v}]$ by Lemma 5 of the Online Appendix. As $\hat{\boldsymbol{\pi}}_q - \boldsymbol{\pi}_q = O_p\left(n^{-1/2} \right)$ and $\max_{1 \le t \le n} \left| \eta_0^{*}\left(\mathbf{z}_t' \boldsymbol{\pi}_q \right) \right| < \infty$ under Assumptions 2(iii)-(iv), we have

$$\begin{aligned}
&\left| A_n(v) \right| \\
\le\ & M n^{-1/2-2\varrho} \left\| \boldsymbol{\Phi}_{L_n} \right\|_1 \sum_{t=1}^{n} \left\| \mathbf{z}_t \right\| d_t^2 \left(\gamma_0 + \frac{v}{a_n}, \gamma_0 \right) + M n^{-1-2\varrho} \left\| \boldsymbol{\Phi}_{L_n} \right\|_1^2 \sum_{t=1}^{n} d_t^2 \left(\gamma_0 + \frac{v}{a_n}, \gamma_0 \right) \\
=\ & O_p\left(n^{1/2-2\varrho} \left\| \boldsymbol{\Phi}_{L_n} \right\|_1 a_n^{-1} \right) + O_p\left(n^{-2\varrho} \left\| \boldsymbol{\Phi}_{L_n} \right\|_1^2 a_n^{-1} \right) \\
=\ & O_p\left(n^{-1+2[\min(\varsigma,\varrho)-\varrho]} \left\| \boldsymbol{\Phi}_{L_n} \right\|_1^2 \right) = o_p(1)
\end{aligned}$$

under Assumption **3**. Also, closely following the interval split method used in the proof of Lemma 1 of the Online Appendix, we can show that $A_n(v) = o_p(1)$ holds uniformly over $v \in [\underline{v}, \overline{v}]$. Hence, $G_n(v)$ is the leading term of $S_{n,1}^{*}(v)$ for any $v \in [\underline{v}, \overline{v}]$.

Secondly, we consider $S_{n,2}^{*}(\gamma)=\hat{\boldsymbol{\kappa}}_{n}^{\prime}\sum_{t=1}^{n}\varepsilon_{t}\boldsymbol{\chi}_{t}d_{t}(\gamma,\gamma_{0})$. For any given $v\in[\underline{v},\overline{v}]$, we have

$$
\begin{aligned}
S_{n,2}^{*}(v) &= \hat{\boldsymbol{\kappa}}_{n}^{\prime}\sum_{t=1}^{n}\varepsilon_{t}\boldsymbol{\chi}_{t}d_{t}\left(\gamma_{0}+\frac{v}{a_{n}},\gamma_{0}\right)\\
&= R_{n}(v)+\hat{\boldsymbol{\kappa}}_{n}^{\prime}\sum_{t=1}^{n}\varepsilon_{t}\left(\boldsymbol{\chi}_{t}-\boldsymbol{\chi}_{t}^{*}\right)d_{t}\left(\gamma_{0}+\frac{v}{a_{n}},\gamma_{0}\right)\\
&\quad+\left(\hat{\boldsymbol{\kappa}}_{n}-\boldsymbol{\kappa}_{n}\right)^{\prime}\sum_{t=1}^{n}\varepsilon_{t}\boldsymbol{\chi}_{t}^{*}d_{t}\left(\gamma_{0}+\frac{v}{a_{n}},\gamma_{0}\right)\\
&= R_{n}(v)\left[1+o_{p}(1)\right]+O_{p}\left(n^{-1/2-\varrho+\min(\varrho,\varsigma)}\left\|\boldsymbol{\Phi}_{L_{n}}\right\|_{1}\right)
\end{aligned}
$$

if $\|\hat{\boldsymbol{\kappa}}_{n}-\boldsymbol{\kappa}_{n}\|=o_{p}(1)$, where $R_{n}(v)=\sum_{t=1}^{n}\varepsilon_{t}\boldsymbol{\kappa}_{n}^{\prime}\boldsymbol{\chi}_{t}^{*}d_{t}(\gamma_{0}+v/a_{n},\gamma_{0})=O_{e}(1)$ holds uniformly over $v\in[\underline{v},\overline{v}]$ by Lemma 5 of the Online Appendix. Therefore, we show that the leading term of $S_{n,2}^{*}(v)$ is $R_{n}(v)$ under Assumption **3.** Note that we can improve the result in Lemma 4 of the Online Appendix to $\left\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right\|=O_{p}(\vartheta_{n})$ as $\hat{\gamma}=\gamma_{0}+O_{p}(a_{n})$.

Thirdly, we can show that $S_{n,4}^{*}(v)=(\hat{\kappa}_{n}+\kappa_{n})^{\prime}\sum_{t=1}^{n}\boldsymbol{\chi}_{t}\boldsymbol{\chi}_{t}^{\prime}d_{t}^{2}(\gamma_{0}+v/a_{n},\gamma_{0})(\hat{\kappa}_{n}-\kappa_{n})$ $=o_{p}\left(S_{n,1}^{*}(v)\right)$ and $S_{n,3}^{*}(v)=o_{p}\left(S_{n,1}^{*}(v)\right)$. And, taking above results together with Lemma 5 of the Online Appendix, we have $Q_{n}(v)=-G_{n}(v)+2R_{n}(v)+o_{p}(1)$ and

$$
Q(v)=-\mu|v|+2\sqrt{\sigma_{1}^{2}}W_{1}(v)I(\underline{v}\leq v\leq 0)+\sqrt{\sigma_{2}^{2}}W_{2}(v)I(0<v\leq\overline{v}).
$$

Following the proof of Theorem 1 in Kourtellos, Stengos, and Tan (2016), we complete the proof of this theorem.

Proof of Theorem 3: Denote $\boldsymbol{\Delta}_{v}=[\boldsymbol{\Delta}_{-,v},\boldsymbol{\Delta}_{+,v}]$ and $\boldsymbol{X}_{v}=[\boldsymbol{X}_{-,v},\boldsymbol{X}_{+,v}]$, where the t^{th} row vector of $\boldsymbol{\Delta}_{-,v}$, $\boldsymbol{\Delta}_{+,v}$, $\boldsymbol{X}_{-,v}$ and $\boldsymbol{X}_{+,v}$ are $\boldsymbol{\Delta}_{v,t}^{-}=\boldsymbol{\Phi}_{L_{n}}(\mathbf{z}_{t}^{\prime}\hat{\boldsymbol{\pi}}_{q})I(q_{t}\leq\gamma_{0}+v/a_{n})$, $\boldsymbol{\Delta}_{v,t}^{+}=\boldsymbol{\Phi}_{L_{n}}(\mathbf{z}_{t}^{\prime}\hat{\boldsymbol{\pi}}_{q})I(q_{t}>\gamma_{0}+v/a_{n})$, $\boldsymbol{X}_{v,t}^{-}=\mathbf{x}_{t}I(q_{t}\leq\gamma_{0}+v/a_{n})$, and $\boldsymbol{X}_{v,t}^{+}=\mathbf{x}_{t}I(q_{t}>\gamma_{0}+v/a_{n})$, respectively. Also, denote $\boldsymbol{P}_{v}=\boldsymbol{\Delta}_{v}(\boldsymbol{\Delta}_{v}^{\prime}\boldsymbol{\Delta}_{v})^{-1}\boldsymbol{\Delta}_{v}^{\prime}$ and $\hat{v}=a_{n}(\hat{\gamma}-\gamma_{0})$. Applying the partitioned least squares gives $\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}=\left[\boldsymbol{X}_{\hat{v}}^{\prime}(\boldsymbol{I}_{n}-\boldsymbol{P}_{\hat{v}})\boldsymbol{X}_{\hat{v}}\right]^{-1}\boldsymbol{X}_{\hat{v}}^{\prime}(\boldsymbol{I}_{n}-\boldsymbol{P}_{\hat{v}})(\boldsymbol{y}-\boldsymbol{X}_{\hat{v}}\boldsymbol{\beta})$.

Firstly, we consider $\boldsymbol{A}_n(v) = \boldsymbol{X}_v'(\boldsymbol{I}_n - \boldsymbol{P}_v)\boldsymbol{X}_v = \boldsymbol{X}_v'\boldsymbol{X}_v - \boldsymbol{X}_v'\boldsymbol{P}_v\boldsymbol{X}_v$. Denote $\boldsymbol{A}_{n1}(v) = n^{-1}\sum_{t=1}^{n}\mathbf{x}_t\mathbf{x}_t'I(q_t \leq \gamma_0 + v/a_n)$, $\boldsymbol{A}_{n2}(v) = n^{-1}\sum_{t=1}^{n}\boldsymbol{\Phi}_{L_n}(\mathbf{z}_t'\hat{\boldsymbol{\pi}}_q)\mathbf{x}_t'I(q_t \leq \gamma_0 + v/a_n)$, and $\boldsymbol{A}_{n3}(v) = n^{-1}\sum_{t=1}^{n}\boldsymbol{\Phi}_{L_n}(\mathbf{z}_t'\hat{\boldsymbol{\pi}}_q)\boldsymbol{\Phi}_{L_n}(\mathbf{z}_t'\hat{\boldsymbol{\pi}}_q)\ I(q_t \leq \gamma_0 + v/a_n)$. By Lemma 1 of the Online Appendix, we have

$$\begin{aligned}
\boldsymbol{A}_{n1}(v) &= E[\boldsymbol{A}_{n1}(v)] + o_p(1) = E[\mathbf{x}_t\mathbf{x}_t'I(q_t \leq \gamma_0)]\left[1 + O\left(a_n^{-1}\right)\right] + o_p(1) \\
\boldsymbol{A}_{n2}^{*}(v) &= E[\boldsymbol{A}_{n2}^{*}(v)] + o_p(1) = E[\boldsymbol{\Phi}_{L_n}(\mathbf{z}_t'\boldsymbol{\pi}_q)\mathbf{x}_t'I(q_t \leq \gamma_0)]\left[1 + O\left(a_n^{-1}\right)\right] + o_p(1)
\end{aligned}$$

uniformly over $v \in [\underline{v}, \overline{v}]$, where $\boldsymbol{A}_{n2}^{*}(v)$ equals $\boldsymbol{A}_{n2}(v)$ with $\hat{\boldsymbol{\pi}}_q$ replaced with $\boldsymbol{\pi}_q$. In addition, by equation (B.3) of the Online Appendix we can show that

$$\max_{v\in[\underline{v},\overline{v}]}\|\boldsymbol{A}_{n2}(v) - \boldsymbol{A}_{n2}^{*}(v)\| = O_p\left(\|\boldsymbol{\Phi}_{L_n}\|_1 n^{-1/2}\right) = o_p(1)$$

under Assumption **3**. And, from the proof of Lemma 2 of the Online Appendix, we have

$$\boldsymbol{A}_{n3}(v) = E[\boldsymbol{\Phi}_{L_n}(\mathbf{z}_t'\boldsymbol{\pi}_q)\boldsymbol{\Phi}_{L_n}(\mathbf{z}_t'\boldsymbol{\pi}_q)I(q_t \leq \gamma_0)]\left[1 + O\left(a_n^{-1}\right)\right] + o_p(1)$$

uniformly over $v \in [\underline{v}, \overline{v}]$. Hence, we obtain $n^{-1}\boldsymbol{X}_{\hat{v}}'\boldsymbol{X}_{\hat{v}} \xrightarrow{p} \boldsymbol{\Sigma}_{\mathbf{x}\mathbf{x}',\gamma_0}$, $n^{-1}\boldsymbol{X}_{\hat{v}}'\boldsymbol{\Delta}_{\hat{v}} \xrightarrow{p} \boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{\Phi}_{L_n}',\gamma_0}$, and $n^{-1}\boldsymbol{\Delta}_{\hat{v}}'\boldsymbol{\Delta}_{\hat{v}} \xrightarrow{p} \boldsymbol{\Sigma}_{\boldsymbol{\Phi}_{L_n}\boldsymbol{\Phi}_{L_n}',\gamma_0}$, where

$$\begin{aligned}
\boldsymbol{\Sigma}_{\mathbf{x}\mathbf{x}',\gamma_0} &\equiv \begin{bmatrix} E[\mathbf{x}_t\mathbf{x}_t'I(q_t \leq \gamma_0)] & \mathbf{0} \\ \mathbf{0}' & E[\mathbf{x}_t\mathbf{x}_t'I(q_t > \gamma_0)] \end{bmatrix}, \\
\boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{\Phi}_{L_n}',\gamma_0} &\equiv \begin{bmatrix} E\left[\mathbf{x}_t\boldsymbol{\Phi}_{L_n}(\mathbf{z}_t'\boldsymbol{\pi}_q)'I(q_t \leq \gamma_0)\right] & \mathbf{0} \\ \mathbf{0}' & E[\boldsymbol{\Phi}_{L_n}(\mathbf{z}_t'\boldsymbol{\pi}_q)\mathbf{x}_t'I(q_t > \gamma_0)] \end{bmatrix}, \\
\boldsymbol{\Sigma}_{\boldsymbol{\Phi}_{L_n}\boldsymbol{\Phi}_{L_n}',\gamma_0} &\equiv \begin{bmatrix} E\left[\boldsymbol{\Phi}_{L_n}(\mathbf{z}_t'\boldsymbol{\pi}_q)\boldsymbol{\Phi}_{L_n}(\mathbf{z}_t'\boldsymbol{\pi}_q)'I(q_t \leq \gamma_0)\right] & \mathbf{0} \\ \mathbf{0}' & E\left[\boldsymbol{\Phi}_{L_n}(\mathbf{z}_t'\boldsymbol{\pi}_q)\boldsymbol{\Phi}_{L_n}'(\mathbf{z}_t'\boldsymbol{\pi}_q)I(q_t > \gamma_0)\right] \end{bmatrix}.
\end{aligned}$$

It then follows

$$n^{-1}\boldsymbol{A}_n(\hat{v}) \xrightarrow{p} \boldsymbol{\Sigma}_{\mathbf{x}\mathbf{x}',\gamma_0} - \boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{\Phi}'_{Ln},\gamma_0}\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\Phi}_{Ln}\boldsymbol{\Phi}'_{Ln},\gamma_0}\boldsymbol{\Sigma}_{\boldsymbol{\Phi}_{Ln}\mathbf{x}',\gamma_0} \equiv \boldsymbol{J}. \tag{A.20}$$

Secondly, we consider $\boldsymbol{B}_n(\hat{v})$, where $\boldsymbol{B}_n(v) = n^{-1}\boldsymbol{X}'_v(\boldsymbol{I}_n - \boldsymbol{P}_v)(\boldsymbol{y} - \boldsymbol{X}_v\boldsymbol{\beta})$, $y_t - \boldsymbol{X}'_{v,t}\boldsymbol{\beta}$ $= \eta_{v,0,t} + h_1(\mathbf{z}'_t\boldsymbol{\pi}_q)I(q_t \leq \gamma_0) + h_2(\mathbf{z}'_t\boldsymbol{\pi}_q)I(q_t > \gamma_0) + \varepsilon_t$ and $\eta_{v,0,t} = -\boldsymbol{\delta}'_n\mathbf{x}_t d_t(\gamma_0 + v/a_n, \gamma_0)$.

(i) We will show $n^{-1}\boldsymbol{X}'_{\hat{v}}(\boldsymbol{I}_n - \boldsymbol{P}_{\hat{v}})\boldsymbol{\eta}_{\hat{v},0} = o_p\left(n^{-1/2}\right)$, where $\boldsymbol{\eta}_{v,0}$ is an $n \times 1$ vector with its t^{th} element equal to $\eta_{v,0,t}$. By Lemma 1 of the Online Appendix, $n^{-1}\sum_{t=1}^n \boldsymbol{\delta}'_n\mathbf{x}_t\mathbf{x}'_t d_t(\gamma_0 + v/a_n, \gamma_0) = n^{-\varsigma}E[\boldsymbol{\delta}'_0\mathbf{x}_t\mathbf{x}'_t d_t(\gamma_0 + v/a_n, \gamma_0)][1 + o_p(1)] = O_p(a_n^{-1}n^{-\varsigma})$ holds uniformly over $v \in [\underline{v}, \overline{v}]$. Hence, we obtain $n^{-1}\boldsymbol{X}'_{\hat{v}}\boldsymbol{\eta}_{\hat{v},0} = O_p(a_n^{-1}n^{-\varsigma})$. Moreover, we have

$$\begin{aligned}
\left\|n^{-1}\boldsymbol{X}'_{\hat{v}}\boldsymbol{P}_{\hat{v}}\boldsymbol{\eta}_{\hat{v},0}\right\| &\leq \left\|n^{-1}\boldsymbol{X}'_{\hat{v}}\boldsymbol{\Delta}_{\hat{v}}(\boldsymbol{\Delta}'_{\hat{v}}\boldsymbol{\Delta}_{\hat{v}}/n)^{-1/2}\right\|_{sp}\left\|n^{-1}(\boldsymbol{\Delta}'_{\hat{v}}\boldsymbol{\Delta}_{\hat{v}}/n)^{-1/2}\boldsymbol{\Delta}'_{\hat{v}}\boldsymbol{\eta}_{\hat{v},0}\right\|_{sp} \\
&\leq \lambda_{\max}^{1/2}\left(\boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{\Phi}'_{Ln},\gamma_0}\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\Phi}_{Ln}\boldsymbol{\Phi}'_{Ln},\gamma_0}\boldsymbol{\Sigma}_{\boldsymbol{\Phi}_{Ln}\mathbf{x}',\gamma_0}\right)[1 + o_p(1)]\sqrt{n^{-1}\boldsymbol{\eta}'_{\hat{v},0}\boldsymbol{\eta}_{\hat{v},0}} = O_p\left(a_n^{-1/2}n^{-\varsigma}\right).
\end{aligned}$$

Therefore, we obtain $n^{-1}\boldsymbol{X}'_{\hat{v}}(\boldsymbol{I}_n - \boldsymbol{P}_{\hat{v}})\boldsymbol{\eta}_{\hat{v},0} = O_p\left(a_n^{-1/2}n^{-\varsigma}\right) = o_p\left(n^{-1/2}\right)$.

(ii) We will consider $n^{-1}\boldsymbol{X}'_{\hat{v}}(\boldsymbol{I}_n - \boldsymbol{P}_{\hat{v}})\boldsymbol{h}_{1,\gamma_0}(\boldsymbol{z}\boldsymbol{\pi}_q)$, where $\boldsymbol{h}_{1,\gamma_0}(\boldsymbol{z}\boldsymbol{\pi}_q)$ denotes an $n \times 1$ vector and its t^{th} element equals $\boldsymbol{h}_1(\mathbf{z}'_t\boldsymbol{\pi}_q)I(q_t \leq \gamma_0)$. As $\boldsymbol{I}_n - \boldsymbol{P}_{\hat{v}}$ removes any linear combination of $\boldsymbol{\Phi}_{Ln}(\mathbf{z}'_t\hat{\boldsymbol{\pi}}_q)I(q_t \leq \gamma_0 + \hat{v}/a_n)$, we have $(\boldsymbol{I}_n - \boldsymbol{P}_{\hat{v}})\boldsymbol{h}_{1,\gamma_0}(\boldsymbol{z}\boldsymbol{\pi}_q)$ $= (\boldsymbol{I}_n - \boldsymbol{P}_{\hat{v}})\left[\boldsymbol{h}_{1,\gamma_0}(\mathbf{z}\pi_q) - \boldsymbol{h}^*_{1,\hat{v}}(\boldsymbol{z}\hat{\boldsymbol{\pi}}_q)\right]$, where the t^{th} element of $\boldsymbol{h}^*_{1,v}(\mathbf{z}\hat{\pi}_q)$ equals

$h_1^*\left(\mathbf{z}_t'\hat{\boldsymbol{\pi}}_q\right)I\left(q_t\le\gamma_0+v/a_n\right)$ and

$$\begin{aligned}&h_1\left(\mathbf{z}_t'\boldsymbol{\pi}_q\right)I\left(q_t\le\gamma_0\right)-h_1^*\left(\mathbf{z}_t'\hat{\boldsymbol{\pi}}_q\right)I\left(q_t\le\gamma_0+v/a_n\right)\\=&\left[h_1\left(\mathbf{z}_t'\boldsymbol{\pi}_q\right)-h_1^*\left(\mathbf{z}_t'\boldsymbol{\pi}_q\right)\right]I\left(q_t\le\gamma_0\right)-h_1^*\left(\mathbf{z}_t'\hat{\boldsymbol{\pi}}_q\right)d_t\left(\gamma_0+v/a_n,\gamma_0\right)\\&+\left[h_1^*\left(\mathbf{z}_t'\boldsymbol{\pi}_q\right)-h_1^*\left(\mathbf{z}_t'\hat{\boldsymbol{\pi}}_q\right)\right]I\left(q_t\le\gamma_0\right)\\=&\boldsymbol{\eta}_{v,1,t}+\boldsymbol{\eta}_{v,2,t},\end{aligned}$$

where we denote $\eta_{v,2,t}=\left[h_1^*\left(\mathbf{z}_t'\boldsymbol{\pi}_q\right)-h_1^*\left(\mathbf{z}_t'\hat{\boldsymbol{\pi}}_q\right)\right]I\left(q_t\le\gamma_0\right)$. Let $\boldsymbol{\eta}_{v,j}=\left[\eta_{v,j,1},\ldots,\eta_{v,1,n}\right]'$ for j=1,2. Applying again Lemma 1 of the Online Appendix, we show that $n^{-1}\boldsymbol{X}_v'\boldsymbol{\eta}_{v,1}$ and $n^{-1}\boldsymbol{\Delta}_v'\boldsymbol{\eta}_{v,1}$ are both of order $O_p\left(L_n^{-\xi}+a_n^{-1}\right)$ uniformly over $v\in[\underline{v},\overline{v}]$. Hence, we obtain $n^{-1}\boldsymbol{X}_{\hat{v}}'\left(\boldsymbol{I}_n-\boldsymbol{P}_{\hat{v}}\right)\boldsymbol{\eta}_{\hat{v},1}=O_p\left(L_n^{-\xi}+a_n^{-1}\right)=o_p\left(n^{-1/2}\right)$ if $\min\left(\varsigma,\varrho\right)<1/4$ and $\sqrt{n}L_n^{-\xi}=o\left(1\right)$.

Next, as $h_1^*\left(\mathbf{z}_t'\boldsymbol{\pi}_q\right)-h_1^*\left(\mathbf{z}_t'\hat{\boldsymbol{\pi}}_q\right)=h_1\left(\mathbf{z}_t'\boldsymbol{\pi}_q\right)-h_1\left(\mathbf{z}_t'\hat{\boldsymbol{\pi}}_q\right)+O\left(L_n^{-\xi}\right)=h_1^{(1)}\left(\mathbf{z}_t'\tilde{\boldsymbol{\pi}}_q\right)\mathbf{z}_t'\left(\boldsymbol{\pi}_q-\hat{\boldsymbol{\pi}}_q\right)+O\left(L_n^{-\xi}\right)$, closely following the proof of Lemma 1 of the Online Appendix, we have

$$\max_{v\in[\underline{v},\overline{v}]}\left\|n^{-1}\sum_{t=1}^n\mathbf{w}_te_th_1^{(1)}\left(\mathbf{z}_t'\tilde{\boldsymbol{\pi}}_q\right)\mathbf{z}_t'I\left(q_t\le\gamma_0\right)-E\left[\mathbf{w}_te_th_1^{(1)}\left(\mathbf{z}_t'\boldsymbol{\pi}_q\right)\mathbf{z}_t'I\left(q_t\le\gamma_0\right)\right]\right\|=o_p\left(1\right)$$

where $\mathbf{w}_t=\mathbf{x}_t$ or $\boldsymbol{\Phi}_{L_n}\left(\mathbf{z}_t'\boldsymbol{\pi}_q\right)$ and $e_t=I\left(q_t\le\gamma_0+v/a_n\right)$ or $e_t=I\left(q_t>\gamma_0+v/a_n\right)$. It then follows $n^{-1}\boldsymbol{X}_{-,\hat{v}}'\boldsymbol{\eta}_{\hat{v},2}=\boldsymbol{\Gamma}_{\mathbf{x},1}\left(\boldsymbol{\pi}_q-\hat{\boldsymbol{\pi}}_q\right)+O_p\left(L_n^{-\xi}\right)$, $n^{-1}\boldsymbol{X}_{+,\hat{v}}'\boldsymbol{\eta}_{\hat{v},2}=O_p\left(a_n^{-1}\left\|\boldsymbol{\Phi}_{L_n}\right\|_1/\sqrt{n}+L_n^{-\xi}\right)$, $n^{-1}\boldsymbol{\Delta}_{-,\hat{v}}'\boldsymbol{\eta}_{\hat{v},2}=\boldsymbol{\Gamma}_{\boldsymbol{\Phi}_{L_n},1}\left(\boldsymbol{\pi}_q-\hat{\boldsymbol{\pi}}_q\right)+O_p\left(L_n^{-\xi}\right)$, $n^{-1}\boldsymbol{\Delta}_{+,\hat{v}}'\boldsymbol{\eta}_{\hat{v},2}=O_p\left(a_n^{-1}\left\|\boldsymbol{\Phi}_{L_n}\right\|_1/\sqrt{n}+L_n^{-\xi}\right)$, where we denote $\boldsymbol{\Gamma}_{\mathbf{x},1}=E\left[\mathbf{x}_th_1^{(1)}\left(\mathbf{z}_t'\boldsymbol{\pi}_q\right)\mathbf{z}_t'I\left(q_t\le\gamma_0\right)\right]$ and $\boldsymbol{\Gamma}_{\boldsymbol{\Phi}_{L_n},1}=E\left[\boldsymbol{\Phi}_{L_n}\left(\mathbf{z}_t'\boldsymbol{\pi}_q\right)h_1^{(1)}\left(\mathbf{z}_t'\boldsymbol{\pi}_q\right)\mathbf{z}_t'I\left(q_t\le\gamma_0\right)\right]$. Hence, we have

$$n^{-1}\boldsymbol{X}_{\hat{v}}'\left(\boldsymbol{I}_n-\boldsymbol{P}_{\hat{v}}\right)\boldsymbol{h}_{1,\gamma_0}\left(\boldsymbol{z}\boldsymbol{\pi}_q\right)=\boldsymbol{B}_1\left(\boldsymbol{\pi}_q-\hat{\boldsymbol{\pi}}_q\right)+o_p\left(n^{-1/2}\right)$$

where

$$\boldsymbol{B}_1=\left[\boldsymbol{\Gamma}_{\mathbf{x},1}',\mathbf{0}'\right]'-\boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{\Phi}_{L_n}',\gamma_0}\boldsymbol{\Sigma}_{\boldsymbol{\Phi}_{L_n}\boldsymbol{\Phi}_{L_n}',\gamma_0}^{-1}\left[\Gamma_{\boldsymbol{\Phi}_{L_n},1}',\mathbf{0}'\right]'. \tag{A.21}$$

(iii) Let $\boldsymbol{h}_{2,\gamma_0}(\boldsymbol{z}\boldsymbol{\pi}_q)$ denote an $n \times 1$ vector and its t^{th} element equals $h_2(\mathbf{z}_t'\boldsymbol{\pi}_q)\, I(q_t > \gamma_0)$. Closely following the proof for (ii) above, we obtain that the leading term of $n^{-1}\boldsymbol{X}_{\hat{v}}'(\boldsymbol{I}_n - \boldsymbol{P}_{\hat{v}})\,\boldsymbol{h}_{2,\gamma_0}(\boldsymbol{z}\boldsymbol{\pi}_q)$ is

$$\boldsymbol{B}_2 = \left[\mathbf{0}', \boldsymbol{\Gamma}_{\mathbf{x},2}'\right]' - \boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{\Phi}_{L_n}',\gamma_0}\boldsymbol{\Sigma}_{\boldsymbol{\Phi}_{L_n}\boldsymbol{\Phi}_{L_n}',\gamma_0}^{-1}\left[\mathbf{0}', \boldsymbol{\Gamma}_{\boldsymbol{\Phi}_{L_n},2}'\right]' \tag{A.22}$$

where

$$\boldsymbol{\Gamma}_{\mathbf{x},2} = E\left[\mathbf{x}_t h_2^{(1)}(\mathbf{z}_t'\boldsymbol{\pi}_q)\,\mathbf{z}_t' I(q_t > \gamma_0)\right]$$

and

$$\boldsymbol{\Gamma}_{\boldsymbol{\Phi}_{L_n},2} = E\left[\boldsymbol{\Phi}_{L_n}(\mathbf{z}_t'\boldsymbol{\pi}_q)\, h_2^{(1)}(\mathbf{z}_t'\boldsymbol{\pi}_q)\,\mathbf{z}_t' I(q_t > \gamma_0)\right],$$

Taking together all the results above, we have

$$\begin{aligned}
\boldsymbol{B}_n(\hat{v}) &= n^{-1}\boldsymbol{X}_{\hat{v}}'(\boldsymbol{I}_n - \boldsymbol{P}_{\hat{v}})(\boldsymbol{y} - \boldsymbol{X}_{\hat{v}}\boldsymbol{\beta}) \\
&= -\boldsymbol{B}(\hat{\boldsymbol{\pi}}_q - \boldsymbol{\pi}_q) + n^{-1}\boldsymbol{X}_{\hat{v}}'(\boldsymbol{I}_n - \boldsymbol{P}_{\hat{v}})\,\boldsymbol{\varepsilon} + o_p\left(n^{-1/2}\right) \\
&= -\boldsymbol{B}(\hat{\boldsymbol{\pi}}_q - \boldsymbol{\pi}_q) + n^{-1}\left(\boldsymbol{X}_0 - \boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{\Phi}_{L_n}',\gamma_0}\boldsymbol{\Sigma}_{\boldsymbol{\Phi}_{L_n}\boldsymbol{\Phi}_{L_n}',\gamma_0}^{-1}\boldsymbol{\Delta}_0\right)\boldsymbol{\varepsilon} + o_p\left(n^{-1/2}\right)
\end{aligned}$$

where $\boldsymbol{B} = \boldsymbol{B}_1 + \boldsymbol{B}_2 = \left[\boldsymbol{\Gamma}_{\mathbf{x},1}', \boldsymbol{\Gamma}_{\mathbf{x},2}'\right]' - \boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{\Phi}_{L_n}',\gamma_0}\boldsymbol{\Sigma}_{\boldsymbol{\Phi}_{L_n}\boldsymbol{\Phi}_{L_n}',\gamma_0}^{-1}\left[\boldsymbol{\Gamma}_{\boldsymbol{\Phi}_{L_n},1}', \boldsymbol{\Gamma}_{\boldsymbol{\Phi}_{L_n},2}'\right]'$. Applying Wooldridge and White's central limit theorem for strong mixing process (White (2001), Th. 5.2, p.130), we obtain

$$\begin{pmatrix} n^{-1/2}\sum_{t=1}^{n}\mathbf{z}_t v_{q,t} \\ n^{-1/2}\left(\boldsymbol{X}_0 - \boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{\Phi}_{L_n}',\gamma_0}\boldsymbol{\Sigma}_{\boldsymbol{\Phi}_{L_n}\boldsymbol{\Phi}_{L_n}',\gamma_0}^{-1}\boldsymbol{\Delta}_0\right)\boldsymbol{\varepsilon} \end{pmatrix} \xrightarrow{d} N\left(\mathbf{0}, \begin{pmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{12}' & \boldsymbol{\Omega}_{22} \end{pmatrix}\right)$$

where $\boldsymbol{\Omega}_{11} = \lim_{n\to\infty} Var\left(n^{-1/2}\sum_{t=1}^{n}\mathbf{z}_t v_{q,t}\right)$, $\boldsymbol{\Omega}_{12} = \lim_{n\to\infty} n^{-1}\sum_{t=2}^{n}\sum_{s=1}^{t-1} E\left(\mathbf{z}_t\boldsymbol{\varphi}_s' v_{q,t}\varepsilon_s\right) =$

$O(1)$ under Assumption 1, and ϕ is the s^{th} element of $\left(\boldsymbol{X}_0 - \boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{\Phi}'_{Ln},\gamma_0}\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\Phi}_{Ln}\boldsymbol{\Phi}'_{Ln},\gamma_0}\boldsymbol{\Delta}_0\right)$

$$\begin{aligned}\boldsymbol{\Omega}_{22} &= \boldsymbol{\Sigma}_{\varepsilon,\mathbf{x}\mathbf{x}',\gamma_0} - \boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{\Phi}'_{Ln},\gamma_0}\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\Phi}_{Ln}\boldsymbol{\Phi}'_{Ln},\gamma_0}\boldsymbol{\Sigma}_{\varepsilon,\boldsymbol{\Phi}_{Ln}\mathbf{x}',\gamma_0} - \boldsymbol{\Sigma}_{\varepsilon,\mathbf{x}\boldsymbol{\Phi}'_{Ln},\gamma_0}\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\Phi}_{Ln}\boldsymbol{\Phi}'_{Ln},\gamma_0}\boldsymbol{\Sigma}_{\boldsymbol{\Phi}_{Ln}\mathbf{x}',\gamma_0} \\ &\quad + \boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{\Phi}'_{Ln},\gamma_0}\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\Phi}_{Ln}\boldsymbol{\Phi}'_{Ln},\gamma_0}\boldsymbol{\Sigma}_{\varepsilon,\boldsymbol{\Phi}_{Ln}\boldsymbol{\Phi}'_{Ln},\gamma_0}\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\Phi}_{Ln}\boldsymbol{\Phi}'_{Ln},\gamma_0}\boldsymbol{\Sigma}_{\boldsymbol{\Phi}_{Ln}\mathbf{x},\gamma_0}.\end{aligned}$$

Therefore, we obtain $\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \overset{d}{\to} N(\mathbf{0}, \boldsymbol{\Omega})$, where

$$\boldsymbol{\Omega} = \boldsymbol{B}\left[E\left(\mathbf{z}_1\mathbf{z}_1'\right)\right]^{-1}\boldsymbol{\Omega}_{11}\left[E\left(\mathbf{z}_1\mathbf{z}_1'\right)\right]^{-1}\boldsymbol{B} - 2\boldsymbol{B}\boldsymbol{\Omega}_{12} + \boldsymbol{\Omega}_{22}. \tag{A.23}$$

This completes the proof of this theorem.

Proof of Theorem 4: Denoting $\mathcal{K}_h(\omega) = K\left(\left(\hat{q}_t - \omega\right)/h\right)$, $\mathcal{W}_t = \left[1, \left(\hat{q}_t - \omega\right)/h\right]^T$, and $\mathcal{X}_{t,\gamma} = [1, I(q_t \le \gamma)]'$, $\mathbf{x}_{t,\gamma} = \mathbf{x}_t I(q_t \le \gamma)$, and $\eta_{0,\gamma}(\omega) = \eta_0(\omega) I(q_t \le \gamma)$, we have

$$\begin{aligned}\begin{bmatrix}\tilde{\psi}(\omega) \\ \dot{\psi}(\omega)\end{bmatrix} &\equiv \left[\frac{1}{nh}\sum_{t=1}^{n} K\left(\frac{\hat{q}_t - w}{h}\right)\left(\mathcal{W}_t\mathcal{W}_t'\right) \otimes \left(\mathcal{X}_{t,\hat{\gamma}}\mathcal{X}'_{t,\hat{\gamma}}\right)\right]^{-1}\frac{1}{nh}\sum_{t=1}^{n} K\left(\frac{\hat{q}_t - w}{h}\right)\left(\mathcal{W}_t \otimes \mathcal{X}_{t,\hat{\gamma}}\right)\tilde{y}_t \\ &= \begin{bmatrix}\psi(\omega) \\ h\psi^{(1)}(\omega)\end{bmatrix} + \boldsymbol{A}_{n1}^{-1}\left(\boldsymbol{A}_{n2}/2 + \boldsymbol{A}_{n3} + \boldsymbol{A}_{n4}\right),\end{aligned}$$

where $\tilde{y}_t = y_t - \boldsymbol{x}'_{-,t}\hat{\boldsymbol{\beta}}_1 - \boldsymbol{x}'_{+,t}\hat{\boldsymbol{\beta}}_2$ and $\psi^{(s)}(\omega) = \partial^s\psi(\omega)/\partial w^s$ for an integer $s > 0$, $\tilde{\psi}(\omega)$ and

$\dot{\psi}(\omega)$ are the estimator for $\psi(\omega)$ and $h\psi^{(1)}(\omega)$, respectively, and

$$\begin{aligned}
\boldsymbol{A}_{n1} &= \frac{1}{nh}\sum_{t=1}^{n} K\left(\frac{\hat{q}_t - w}{h}\right)\left(\mathcal{W}_t\mathcal{W}_t'\right)\otimes\left(\mathcal{X}_{t,\hat{\gamma}}\mathcal{X}_{t,\hat{\gamma}}'\right) \\
\boldsymbol{A}_{n2} &= \frac{1}{nh}\sum_{t=1}^{n}\left(\hat{q}_t-\omega\right)^2 K\left(\frac{\hat{q}_t - w}{h}\right)\left(\mathcal{W}_t\otimes\mathcal{X}_{t,\hat{\gamma}}\right)\mathcal{X}_{t,\gamma}'\psi^{(2)}\left(\tilde{q}_t\right) \\
\boldsymbol{A}_{n3} &= \frac{1}{nh}\sum_{t=1}^{n} K\left(\frac{\hat{q}_t - w}{h}\right)\left(\mathcal{W}_t\otimes\mathcal{X}_{t,\hat{\gamma}}\right)\lambda_t \\
\boldsymbol{A}_{n4} &= \frac{1}{nh}\sum_{t=1}^{n} K\left(\frac{\hat{q}_t - w}{h}\right)\left(\mathcal{W}_t\otimes\mathcal{X}_{t,\hat{\gamma}}\right)\varepsilon_t,
\end{aligned}$$

$\lambda_t = \mathbf{x}_t'\left(\boldsymbol{\beta}_2-\hat{\boldsymbol{\beta}}_2\right)+\boldsymbol{\delta}_n'\mathbf{x}_{t,\gamma_0}-\hat{\boldsymbol{\delta}}_n'\mathbf{x}_{t,\hat{\gamma}}+h_2\left(\mathbf{z}_t'\boldsymbol{\pi}_q\right)-h_2\left(\hat{q}_t\right)+n^{-\varrho}\left[\eta_{0,\gamma_0}\left(\mathbf{z}_t'\boldsymbol{\pi}_q\right)-\eta_{0,\hat{\gamma}}\left(\mathbf{z}_t'\hat{\boldsymbol{\pi}}_q\right)\right]$, and $\tilde{q}_t$ lies between $\hat{q}_t$ and w.

Firstly, we calculate

$$\begin{aligned}
K\left(\frac{\hat{q}_t - w}{h}\right) &= K\left(\frac{\mathbf{z}_t'\boldsymbol{\pi}_q - w}{h}\right)+K'\left(\frac{\mathbf{z}_t'\boldsymbol{\pi}_q - w}{h}\right)\frac{\mathbf{z}_t'\left(\hat{\boldsymbol{\pi}}_q-\boldsymbol{\pi}_q\right)}{h} \\
&\quad +\ldots+\frac{1}{r!}K^{(r_k)}\left(\frac{\zeta_t\mathbf{z}_t'\left(\hat{\boldsymbol{\pi}}_q-\boldsymbol{\pi}_q\right)+\mathbf{z}_t'\boldsymbol{\pi}_q-w}{h}\right)\left(\frac{\mathbf{z}_t'\left(\hat{\boldsymbol{\pi}}_q-\boldsymbol{\pi}_q\right)}{h}\right)^{r_k}
\end{aligned} \tag{A.24}$$

for some $\zeta_t\in(0,1)$ uniformly for all t and some $r_k>2$. Therefore, we obtain

$$\begin{aligned}
\boldsymbol{A}_{n1} &= \frac{1}{nh}\sum_{t=1}^{n} K\left(\frac{\mathbf{z}_t'\boldsymbol{\pi}_q - w}{h}\right)\left(\mathcal{W}_t\mathcal{W}_t'\right)\otimes\left(\mathcal{X}_{t,\hat{\gamma}}\mathcal{X}_{t,\hat{\gamma}}'\right) \\
&\quad +\frac{1}{nhj!}\sum_{j=1}^{r_k-1}\sum_{t=1}^{n} K^{(j)}\left(\frac{\mathbf{z}_t'\boldsymbol{\pi}_q - w}{h}\right)\left(\frac{\mathbf{z}_t'\left(\hat{\boldsymbol{\pi}}_q-\boldsymbol{\pi}_q\right)}{h}\right)^{j}\left(\mathcal{W}_t\mathcal{W}_t'\right)\otimes\left(\mathcal{X}_{t,\hat{\gamma}}\mathcal{X}_{t,\hat{\gamma}}'\right) \\
&\quad +\frac{1}{nhr_k!}\sum_{t=1}^{n} K^{(r_k)}\left(\frac{\zeta_t\mathbf{z}_t'\left(\hat{\boldsymbol{\pi}}_q-\boldsymbol{\pi}_q\right)+\mathbf{z}_t'\boldsymbol{\pi}_q-w}{h}\right)\left(\frac{\mathbf{z}_t'\left(\hat{\boldsymbol{\pi}}_q-\boldsymbol{\pi}_q\right)}{h}\right)^{r_k}\left(\mathcal{W}_t\mathcal{W}_t'\right)\otimes\left(\mathcal{X}_{t,\hat{\gamma}}\mathcal{X}_{t,\hat{\gamma}}'\right) \\
&= \boldsymbol{A}_{n1,1}+\boldsymbol{A}_{n1,2}+\boldsymbol{A}_{n1,3},
\end{aligned}$$

where as $\hat{\boldsymbol{\pi}}_q - \boldsymbol{\pi}_q = O_p\left(n^{-1/2}\right)$ and $\hat{\gamma} - \gamma_0 = O_p\left(n^{-1+2\min(\varsigma,\varrho)}\right)$, we obtain

$$\begin{aligned} \boldsymbol{A}_{n1,1} &= f_{z\pi}(\omega) \begin{bmatrix} 1 & 0 \\ 0 & \kappa_{1,2} \end{bmatrix} \otimes E\left(\mathcal{X}_{1,\gamma_0}\mathcal{X}'_{1,\gamma_0} | \mathbf{z}'_1\boldsymbol{\pi}_q = \omega\right) \\ &\quad + O_p\left(h^2 + (nh)^{-1/2} h^{(2-r')/(2r')} + \left(\sqrt{n}h\right)^{-1} + n^{2\min(\varsigma,\varrho)-1}\right) \end{aligned}$$

under Assumption **4**, where we apply Davydov's inequality to obtain the stochastic order of the variance of each term in $A_{n1,1}$. In addition, under Assumption **5**(ii), we have

$$\begin{aligned} \boldsymbol{A}_{n1,2} &= \frac{1}{nhj!}\sum_{j=1}^{r-1}\sum_{t=1}^{n} K^{(j)}\left(\frac{\mathbf{z}'_t\boldsymbol{\pi}_q - w}{h}\right)\left(\frac{\mathbf{z}'_t\left(\hat{\boldsymbol{\pi}}_q - \boldsymbol{\pi}_q\right)}{h}\right)^j (\mathcal{W}_t\mathcal{W}'_t) \otimes \left(\mathcal{X}_{t,\hat{\gamma}}\mathcal{X}'_{t,\hat{\gamma}}\right) \\ &\leq \frac{1}{nhj!}\sum_{j=1}^{r-1}\sum_{t=1}^{n} \left| K^{(j)}\left(\frac{\mathbf{z}'_t\boldsymbol{\pi}_q - w}{h}\right)\left(\frac{\mathbf{z}'_t\left(\hat{\boldsymbol{\pi}}_q - \boldsymbol{\pi}_q\right)}{h}\right)^j \right| \left\|(\mathcal{W}_t\mathcal{W}'_t) \otimes \left(\mathcal{X}_{t,\hat{\gamma}}\mathcal{X}'_{t,\hat{\gamma}}\right)\right\| \\ &= O_p\left(\frac{1}{(\sqrt{n}h)^j}\right) \end{aligned}$$

and

$$\begin{aligned} \boldsymbol{A}_{n1,3} &= \frac{1}{nhr_k!}\sum_{t=1}^{n} K^{(r_k)}\left(\frac{\zeta_t\mathbf{z}'_t\left(\hat{\boldsymbol{\pi}}_q - \boldsymbol{\pi}_q\right) + \mathbf{z}'_t\boldsymbol{\pi}_q - w}{h}\right)\left(\frac{\mathbf{z}'_t\left(\hat{\boldsymbol{\pi}}_q - \boldsymbol{\pi}_q\right)}{h}\right)^{r_k} (\mathcal{W}_t\mathcal{W}'_t) \otimes \left(\mathcal{X}_{t,\hat{\gamma}}\mathcal{X}'_{t,\hat{\gamma}}\right) \\ &\leq \frac{M}{nhr_k!}\sum_{t=1}^{n} \left| \left(\frac{\mathbf{z}'_t\left(\hat{\boldsymbol{\pi}}_q - \boldsymbol{\pi}_q\right)}{h}\right)^{r_k} \right| \left\|(\mathcal{W}_t\mathcal{W}'_t) \otimes \left(\mathcal{X}_{t,\hat{\gamma}}\mathcal{X}'_{t,\hat{\gamma}}\right)\right\| = O_p\left(n^{-r_k/2}h^{-r_k-3}\right). \end{aligned}$$

Therefore, under Assumption **5**, we have $A_{n1,2} = o_p(1)$ and $A_{n1,3} = o_p(1)$.

Similarly, we can show that

$$\begin{aligned}
&\frac{\boldsymbol{A}_{n2}}{h^2} \\
= \; & f_{z\pi}(\omega)\begin{bmatrix}\kappa_{1,2}\\ 0\end{bmatrix}\otimes\begin{bmatrix} h_2^{(2)}(\omega)+n^{-\varrho}\eta_0^{(2)}(\omega)\,E\left[I\left(q_t\le\gamma_0\right)|w\right]\\ \left[h_2^{(2)}(\omega)+n^{-\varrho}\eta_0^{(2)}(\omega)\right]E\left[I\left(q_t\le\gamma_0\right)|w\right]\end{bmatrix} \\
&+O_p\left(h^2+(nh)^{-1/2}h^{(2-r')/(2r')}+\left(\sqrt{n}h\right)^{-1}+n^{2\min(\varsigma,\varrho)-1}\right) \\
&+O_p\left(n^{-(r_k+2)/2}\left[h^{-(r_k+3)},h^{-(r_k+4)}\right]'\right)
\end{aligned}$$

and by Theorem 3 and under Assumptions **4** and **5**, we have

$$\begin{aligned}
\boldsymbol{A}_{n3} \;=\; & O_p\left(\|\boldsymbol{\Phi}_{L_n}\|_1 n^{-1/2}\left(h^2+(nh)^{-1/2}h^{(2-r')/(2r')}+\left(\sqrt{n}h\right)^{-1}+n^{2\min(\varsigma,\varrho)-1}\right)\right) \\
&+O_p\left(\|\boldsymbol{\Phi}_{L_n}\|_1 n^{-(r_k+1)/2}\left[h^{-(r_k+1)},h^{-(r_k+3)}\right]\right).
\end{aligned}$$

Now, we consider $\boldsymbol{A}_{n4}$. Applying similar method used above, we have

$$\boldsymbol{A}_{n4}=\frac{1}{nh}\sum_{t=1}^{n}K\left(\frac{\hat{q}_t-w}{h}\right)\left(\mathcal{W}_t\otimes\mathcal{X}_{t,\hat{\gamma}}\right)\varepsilon_t=\boldsymbol{A}_{n4,1}+\boldsymbol{A}_{n4,2},$$

where we have

$$\begin{aligned}
\sqrt{nh}\boldsymbol{A}_{n4,1} \;=\; & \frac{1}{\sqrt{nh}}\sum_{t=1}^{n}K\left(\frac{\hat{q}_t-w}{h}\right)\left(\mathcal{W}_t\otimes\mathcal{X}_{t,\gamma_0}\right)\varepsilon_t \\
=\; & \frac{1}{\sqrt{nh}}\sum_{t=1}^{n}\varepsilon_t K\left(\frac{\mathbf{z}_t'\boldsymbol{\pi}_q-w}{h}\right)\left(\begin{bmatrix}1\\ \left(\mathbf{z}_t'\boldsymbol{\pi}_q-\omega\right)/h\end{bmatrix}\otimes\mathcal{X}_{t,\gamma_0}\right) \\
&+O_p\left(\left(nh^{3/2}\right)^{-1}\right)+O_p\left(n^{-r_k/2}\left[h^{-(r_k+1)},h^{-(r_k+2)}\right]'\right) \\
\xrightarrow{d}\; & N\left(0,f_{z\pi}(\omega)\begin{bmatrix}\kappa_{2,0} & 0\\ 0 & \kappa_{2,2}\end{bmatrix}\otimes E\left(\varepsilon_t^2\mathcal{X}_{t,\gamma_0}\mathcal{X}_{t,\gamma_0}'|\mathbf{z}_1'\boldsymbol{\pi}_q=\omega\right)\right)
\end{aligned}$$

by (A.24) and the central limit theorem for martingale difference sequence (e.g., Theorem 5.24 in White (2001)) under Assumptions **4** and **5,** and

$$\sqrt{nh}\boldsymbol{A}_{n4,2} = \frac{1}{\sqrt{nh}}\sum_{t=1}^{n} K\left(\frac{\hat{q}_t - w}{h}\right)\left(\mathcal{W}_t \otimes \begin{bmatrix} 0 \\ d_t(\hat{\gamma}, \gamma_0) \end{bmatrix}\right)\varepsilon_t = O_p\left(n^{2\min(\varsigma,\varrho)-1}\right).$$

Therefore we obtain $\sqrt{nh}\boldsymbol{A}_{n4,1}$ is the leading term of $\sqrt{nh}\boldsymbol{A}_{n4}$. Taking together all the results above completes the proof of this theorem.

Proof of Theorem 5: Given that the proof of this theorem closely follows the proof of Theorem 1, we only provide detailed proofs where they differ. Also, we borrow the same notation used in the proof of Theorem 1 unless defined differently. Our objective function is rewritten as

$$\begin{aligned}
S_n(\gamma) &= \sum_{t=1}^{n}\left[y_t - \mathcal{X}_{t,\gamma}'\hat{\boldsymbol{\theta}}(\gamma)\right]^2 = \boldsymbol{y}'(\boldsymbol{I}_n - \boldsymbol{P}_\gamma)\boldsymbol{y} \\
&= \left[\mathbf{z}(\boldsymbol{\Pi}_{\boldsymbol{x}} - \hat{\boldsymbol{\pi}}_x)'\boldsymbol{\beta}_2 + \boldsymbol{X}_-\boldsymbol{\delta}_n + \boldsymbol{\eta}_-(\mathbf{z}'\boldsymbol{\pi}_q) + \hat{\boldsymbol{\varepsilon}}\right]'(\boldsymbol{I}_n - \boldsymbol{P}_\gamma) \\
&\quad \cdot\left[\boldsymbol{z}(\boldsymbol{\Pi}_x - \hat{\boldsymbol{\pi}}_{\boldsymbol{x}})'\boldsymbol{\beta}_2 + \boldsymbol{X}_-\boldsymbol{\delta}_n + \boldsymbol{\eta}_-(\mathbf{z}'\boldsymbol{\pi}_q) + \hat{\boldsymbol{\varepsilon}}\right] \\
&= \boldsymbol{\beta}_2'(\boldsymbol{\Pi}_{\boldsymbol{x}} - \hat{\boldsymbol{\pi}}_{\boldsymbol{x}})\boldsymbol{z}'(\boldsymbol{I}_n - \boldsymbol{P}_\gamma)\boldsymbol{z}(\boldsymbol{\Pi}_{\boldsymbol{x}} - \hat{\boldsymbol{\pi}}_{\boldsymbol{x}})'\boldsymbol{\beta}_2 + \boldsymbol{\delta}_n'\boldsymbol{X}_-'(\boldsymbol{I}_n - \boldsymbol{P}_\gamma)\boldsymbol{X}_-\boldsymbol{\delta}_n \\
&\quad +\boldsymbol{\eta}_-(\boldsymbol{z}\boldsymbol{\pi}_q)'(\boldsymbol{I}_n - \boldsymbol{P}_\gamma)\boldsymbol{\eta}_-(\boldsymbol{z}'\boldsymbol{\pi}_q) + \hat{\boldsymbol{\varepsilon}}'(\boldsymbol{I}_n - \boldsymbol{P}_\gamma)\hat{\boldsymbol{\varepsilon}} \\
&\quad +2\boldsymbol{\delta}_n'\boldsymbol{X}_-'(\boldsymbol{I}_n - \boldsymbol{P}_\gamma)\boldsymbol{\eta}_-(\boldsymbol{z}\boldsymbol{\pi}_q) + 2\boldsymbol{\delta}_n'X_-'(\boldsymbol{I}_n - \boldsymbol{P}_\gamma)\hat{\boldsymbol{\varepsilon}} + 2\boldsymbol{\eta}_-(\boldsymbol{z}\boldsymbol{\pi}_q)'(\boldsymbol{I}_n - \boldsymbol{P}_\gamma)\hat{\boldsymbol{\varepsilon}} \\
&\quad +2\boldsymbol{\beta}_2'(\boldsymbol{\Pi}_{\boldsymbol{x}} - \hat{\boldsymbol{\pi}}_{\boldsymbol{x}})\boldsymbol{z}'(\boldsymbol{I}_n - \boldsymbol{P}_\gamma)\boldsymbol{\eta}_-(\boldsymbol{z}\boldsymbol{\pi}_q) + 2\boldsymbol{\beta}_2'(\boldsymbol{\Pi}_{\boldsymbol{x}} - \hat{\boldsymbol{\pi}}_{\boldsymbol{x}})\mathbf{z}'(\boldsymbol{I}_n - \boldsymbol{P}_\gamma)\boldsymbol{X}_-\boldsymbol{\delta}_n \\
&\quad +2\boldsymbol{\beta}_2'(\boldsymbol{\Pi}_{\boldsymbol{x}} - \hat{\boldsymbol{\pi}}_{\boldsymbol{x}})\boldsymbol{z}'(\boldsymbol{I}_n - \boldsymbol{P}_\gamma)\hat{\boldsymbol{\varepsilon}} \\
&= S_{n0} + S_{n1} + S_{n2} + S_{n3} + 2(S_{n4} + S_{n5} + S_{n6}) + 2(S_{n7} + S_{n8} + S_{n9}),
\end{aligned}$$

where S_{nj} for j=0,1,...,9 are named accoroding to the sequence of appearance, $\boldsymbol{X}_-$ is an $n \times d_x$ matrix with its t^{th} row equal to $\hat{\mathbf{x}}_t' I(q_t \leq \gamma_0)$, $\hat{\boldsymbol{\varepsilon}}_t = h_2(\mathbf{z}_t'\boldsymbol{\pi}_q) - h_2^*(\mathbf{z}_t'\hat{\boldsymbol{\pi}}_q) + \varepsilon_t$, and

$\varepsilon_t = \varepsilon_{1t} I(q_t \leq \gamma_0) + \varepsilon_{2t} I(q_t \leq \gamma_0)$ with $\varepsilon_{jt} = \boldsymbol{\beta}_j' \boldsymbol{v}_{x,t} + \sigma_j u_t - h_j(\mathbf{z}_t' \boldsymbol{\pi}_q)$ for j=1,2.

Firstly, we consider $S_{n0} = \boldsymbol{\beta}_2' (\boldsymbol{\Pi_x} - \hat{\boldsymbol{\pi}}_{\boldsymbol{x}}) \boldsymbol{z}' (\boldsymbol{I}_n - \boldsymbol{P}_\gamma) \mathbf{z} (\boldsymbol{\Pi_x} - \hat{\boldsymbol{\pi}}_x)' \boldsymbol{\beta}_2$. As $\boldsymbol{\pi}_x - \hat{\boldsymbol{\pi}}_x = O_p(n^{-1/2})$ and $\boldsymbol{I}_n - \boldsymbol{P}_\gamma$ is an idemponent matrix, we have

$$\begin{aligned}\max_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} |S_{n0}| &\leq \left|\boldsymbol{\beta}_2' (\boldsymbol{\Pi_x} - \hat{\boldsymbol{\pi}}_{\boldsymbol{x}}) \boldsymbol{z}' \boldsymbol{z} (\boldsymbol{\Pi_x} - \hat{\boldsymbol{\pi}}_{\boldsymbol{x}})' \boldsymbol{\beta}_2\right| \\ &\leq \lambda_{\max}(\boldsymbol{z}'\boldsymbol{z}) \left|\boldsymbol{\beta}_2' (\boldsymbol{\Pi_x} - \hat{\boldsymbol{\pi}}_{\boldsymbol{x}}) (\boldsymbol{\Pi_x} - \boldsymbol{\pi_x})' \boldsymbol{\beta}_2\right| = O_p(1).\end{aligned}$$

Then, under Assumption **1'**(vi) and by Lemmas 6 and 7 of the Online Appendix, we have (A.3)-(A.8) hold for S_{n1}, S_{n2} and S_{n4} with newly defined $\mathcal{X}_{t,\gamma}^*$ in Assumption **2'**, $\boldsymbol{g}_1(\gamma) = E\left[\mathcal{X}_{t,\gamma}^* \mathbf{z}_t' I(q_t \leq \gamma_0)\right] \boldsymbol{\pi}_{\boldsymbol{x}}'$, $\boldsymbol{m}_1(\gamma_0) = \boldsymbol{\Pi_x} E[\mathbf{z}_t \mathbf{z}_t' I(q_t \leq \gamma_0)] \boldsymbol{\Pi}_{\boldsymbol{x}}'$, and $\boldsymbol{m}_3(\gamma_0) = \boldsymbol{\Pi_x} E[\mathbf{z}_t \eta_0(\mathbf{z}_t' \boldsymbol{\pi}_q) I(q_t \leq \gamma_0)]$. Also, (A.11), (A.14) and (A.15) continue to hold for $S_{n,j}(\gamma)$ for j=3,5 and 6 by Lemmas 7 and 8 of the Online Appendix. In addition, for $S_{n7}(\gamma) = \boldsymbol{\beta}_2' (\boldsymbol{\Pi_x} - \hat{\boldsymbol{\pi}}_{\boldsymbol{x}}) \boldsymbol{z}' (\boldsymbol{I}_n - \boldsymbol{P}_\gamma) \boldsymbol{\eta}_-(\boldsymbol{z}\boldsymbol{\pi}_q)$, $S_{n8}(\gamma) = \boldsymbol{\beta}_2' (\boldsymbol{\Pi_x} - \hat{\boldsymbol{\pi}}_{\boldsymbol{x}}) \boldsymbol{z}' (\boldsymbol{I}_n - \boldsymbol{P}_\gamma) \boldsymbol{X}_- \boldsymbol{\delta}_n$, and $S_{n9}(\gamma) = \boldsymbol{\beta}_2' (\boldsymbol{\Pi_x} - \hat{\boldsymbol{\pi}}_{\boldsymbol{x}}) \boldsymbol{z}' (\boldsymbol{I}_n - \boldsymbol{P}_\gamma) \hat{\boldsymbol{\varepsilon}}$, applying Lemma 8 of the Online Appendix and $\hat{\boldsymbol{\pi}}_{\boldsymbol{x}} - \boldsymbol{\pi}_{\boldsymbol{x}} = O_p(n^{-1/2})$, we can show that

$$\max_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} \left|n^{-1} S_{n,7}(\gamma)\right| = O_p\left(n^{-\varrho} \vartheta_n\right), \tag{A.25}$$

$$\max_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} \left|n^{-1} S_{n,8}(\gamma)\right| O_p\left(n^{-\varrho - 1/2} \vartheta_n\right), \tag{A.26}$$

$$\max_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} \left|n^{-1} S_{n9}(\gamma)\right| = O_p\left(n^{-1/2} \vartheta_n\right). \tag{A.27}$$

Therefore, taking together all these results gives

$$\max_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} \left|n^{-1} S_n(\gamma) - S_1(\gamma)\right| = o_p(1) \tag{A.28}$$

where $S_1(\gamma)$ has the same formula as in the proof of Theorem 1 with newly defined $\boldsymbol{g}_1(\gamma)$,

$\boldsymbol{g}_2(\gamma)$, and $m_j(\gamma_0)$ for j=1,2,3.

Secondly, denote $\boldsymbol{\chi}_t^* = \left[\mathbf{z}_t'\boldsymbol{\pi}_x', \boldsymbol{\Phi}_{L_n}'(\mathbf{z}_t'\boldsymbol{\pi}_q)\right]'$ and $\boldsymbol{D}_1(\gamma) = E\left[\boldsymbol{\chi}_t^*\mathbf{z}_t'\boldsymbol{\Pi}_{\boldsymbol{x}}' I(q_t \le \gamma)\right]$. Closely following the proof of Theorem 1, we obtain $\hat{\gamma} \xrightarrow{p} \gamma_0$. Taking together this result with Lemma 9 of the Online Appendix completes the proof of this theorem.

Proof of Theorem 6: In matrix form, we have

$$\boldsymbol{y} = \boldsymbol{Z}\boldsymbol{\pi}_{\boldsymbol{x}}'\boldsymbol{\beta}_2 + \boldsymbol{Z}_{\gamma_0}\boldsymbol{\Pi}_{\mathbf{x}}'\boldsymbol{\delta}_{\mathbf{n}} + \mathbf{h_2}(\mathbf{z}\boldsymbol{\pi}_{\mathbf{q}}) + \boldsymbol{\eta}_{\mathbf{n},\gamma_0}(\mathbf{z}\boldsymbol{\pi}_{\mathbf{q}}) + \boldsymbol{\varepsilon},$$

and

$$\begin{aligned} & \boldsymbol{y} - \mathcal{X}_\gamma\hat{\boldsymbol{\theta}} \\ = \; & \boldsymbol{Z}\boldsymbol{\Pi}_{\mathbf{x}}'\boldsymbol{\beta}_2 + \mathbf{Z}_{\gamma_0}\boldsymbol{\pi}_{\mathbf{x}}'\boldsymbol{\delta}_{\mathbf{n}} + \mathbf{h_2}(\mathbf{z}\boldsymbol{\pi}_{\mathbf{q}}) + \boldsymbol{\eta}_{\mathbf{n},\gamma_0}(\mathbf{z}\boldsymbol{\pi}_{\mathbf{q}}) + \boldsymbol{\varepsilon} - \mathbf{Z}\hat{\boldsymbol{\pi}}_{\mathbf{x}}'\hat{\boldsymbol{\beta}}_2 - \mathbf{Z}_\gamma\hat{\boldsymbol{\pi}}_{\mathbf{x}}'\hat{\boldsymbol{\delta}}_{\mathbf{n}} - \hat{\mathbf{h}}_2^*(\mathbf{z}\hat{\boldsymbol{\pi}}_{\mathbf{q}}) - \hat{\boldsymbol{\eta}}_{\mathbf{n},\gamma}^*(\mathbf{z}\hat{\boldsymbol{\pi}}_{\mathbf{q}}) \\ = \; & \boldsymbol{\varepsilon} + \boldsymbol{\Delta}_n - \Delta\boldsymbol{Z}_\gamma\boldsymbol{\Pi}_{\mathbf{x}}'\hat{\boldsymbol{\delta}}_{\mathbf{n}} - \boldsymbol{\Delta}\hat{\boldsymbol{\eta}}_{\mathbf{n},\gamma}^*(\mathbf{z}\hat{\boldsymbol{\pi}}_{\mathbf{q}}) \end{aligned}$$

where $\Delta\boldsymbol{Z}_\gamma = \boldsymbol{Z}_\gamma - \boldsymbol{Z}_{\gamma_0}$, $\Delta\hat{\boldsymbol{\eta}}_{n,\gamma}^* = \hat{\boldsymbol{\eta}}_{n,\gamma}^*(\boldsymbol{z}\hat{\boldsymbol{\pi}}_q) - \hat{\boldsymbol{\eta}}_{n,\gamma_0}^*(\boldsymbol{z}\hat{\boldsymbol{\pi}}_q)$, $\boldsymbol{\Delta}_n = \boldsymbol{Z}\boldsymbol{\Pi}_{\mathbf{x}}'\left(\boldsymbol{\beta}_2 - \hat{\boldsymbol{\beta}}_2\right) + \mathbf{Z}(\boldsymbol{\Pi}_{\mathbf{x}} - \hat{\boldsymbol{\pi}}_{\mathbf{x}})'\hat{\boldsymbol{\beta}}_2 + \mathbf{Z}_{\gamma_0}\boldsymbol{\pi}_{\mathbf{x}}'\left(\boldsymbol{\delta}_{\mathbf{n}} - \hat{\boldsymbol{\delta}}_{\mathbf{n}}\right) + \mathbf{Z}_{\gamma_0}(\boldsymbol{\Pi}_{\mathbf{x}} - \hat{\boldsymbol{\pi}}_{\mathbf{x}})'\hat{\boldsymbol{\delta}}_{\mathbf{n}} + \boldsymbol{h}_2(\boldsymbol{z}\boldsymbol{\pi}_q) - \hat{\boldsymbol{h}}_2^*(\boldsymbol{z}\hat{\boldsymbol{\pi}}_q) + \boldsymbol{\eta}_{n,\gamma_0}(\boldsymbol{z}\boldsymbol{\pi}_q) - \hat{\boldsymbol{\eta}}_{n,\gamma_0}^*(\boldsymbol{z}\hat{\boldsymbol{\pi}}_q)$, and the typical element of $\boldsymbol{Z}$, $\boldsymbol{Z}_\gamma$, $\boldsymbol{h}_2(\boldsymbol{z}\boldsymbol{\pi}_q)$, $\hat{\boldsymbol{h}}_2^*(\boldsymbol{z}\hat{\boldsymbol{\pi}}_q)$, $\boldsymbol{\eta}_{n,\gamma}(\boldsymbol{z}\boldsymbol{\pi}_q)$, and $\hat{\boldsymbol{\eta}}_{n,\gamma}^*(\boldsymbol{z}\hat{\boldsymbol{\pi}}_q)$ are $\mathbf{z}_t'$, $\mathbf{z}_t' I(q_t \le \gamma)$, $h_2(\mathbf{z}_t'\boldsymbol{\pi}_q)$, $\hat{h}_2^*(\boldsymbol{z}_t'\hat{\boldsymbol{\pi}}_q)$, $\eta_n(\mathbf{z}_t'\boldsymbol{\pi}_q) I(q_t \le \gamma)$, and $\hat{\eta}_{n,\gamma}^*(\mathbf{z}_t'\hat{\boldsymbol{\pi}}_q) I(q_t \le \gamma)$, respectively. Denote $S_n(\gamma) = \left(\boldsymbol{y} - \mathcal{X}_\gamma \hat{\boldsymbol{\theta}}\right)' \left(\boldsymbol{y} - \mathcal{X}_\gamma \hat{\boldsymbol{\theta}}\right)$. Then $\hat{\gamma}$ minimizes

$$
\begin{aligned}
& S_n(\gamma) - S_n(\gamma_0) \\
= & \left(\Delta \hat{\boldsymbol{\eta}}^*_{n,\gamma}(\boldsymbol{z}\hat{\boldsymbol{\pi}}_q) \; + \Delta \boldsymbol{Z}_\gamma \boldsymbol{\Pi}'_{\mathbf{x}} \hat{\boldsymbol{\delta}}_{\mathbf{n}}\right)' \left(\Delta \hat{\boldsymbol{\eta}}^*_{n,\gamma}(\boldsymbol{z}\hat{\boldsymbol{\pi}}_q) \; + \Delta \boldsymbol{Z}_\gamma \boldsymbol{\pi}'_x \hat{\boldsymbol{\delta}}_n\right) \\
& -2(\varepsilon + \boldsymbol{\Delta}_n)' \left(\Delta \hat{\boldsymbol{\eta}}^*_{n,\gamma}(\boldsymbol{z}\hat{\boldsymbol{\pi}}_q) \; + \Delta \boldsymbol{Z}_\gamma \boldsymbol{\pi}'_x \hat{\boldsymbol{\delta}}_n\right) \\
= & \boldsymbol{\kappa}'_n \sum_{t=1}^n \boldsymbol{\chi}_t \boldsymbol{\chi}'_t d_t^2(\gamma, \gamma_0) \boldsymbol{\kappa}_n - 2\hat{\boldsymbol{\kappa}}'_n \sum_{t=1}^n \varepsilon_t \boldsymbol{\chi}_t d_t(\gamma, \gamma_0) \\
& -2\hat{\boldsymbol{\kappa}}'_n \sum_{t=1}^n \Bigg[\left(\boldsymbol{\beta}_2 - \hat{\boldsymbol{\beta}}_2\right)' \boldsymbol{\Pi}_{\mathbf{x}} \mathbf{z}_{\mathbf{t}} + \left(\boldsymbol{\delta}_{\mathbf{n}} - \hat{\boldsymbol{\delta}}_{\mathbf{n}}\right)' \boldsymbol{\Pi}_{\mathbf{x}} \mathbf{z}_{\mathbf{t}} \mathbf{I}(\mathbf{q}_{\mathbf{t}} \leq \boldsymbol{\gamma}_{\mathbf{0}}) \\
& + h_2(\boldsymbol{z}'_t \boldsymbol{\pi}_q) - \hat{h}^*_2(\boldsymbol{z}'_t \hat{\boldsymbol{\pi}}_q) + n^{-\varrho}\left(\eta_0(\boldsymbol{z}'_t \boldsymbol{\pi}_q) - \hat{\eta}^*_0(\boldsymbol{z}'_t \hat{\boldsymbol{\pi}}_q)\right) I(q_t \leq \gamma_0)\Bigg] \boldsymbol{\chi}_t d_t(\gamma, \gamma_0) \\
& + (\hat{\boldsymbol{\kappa}}_n + \boldsymbol{\kappa}_n)' \sum_{t=1}^n \boldsymbol{\chi}_t \boldsymbol{\chi}'_t d_t^2(\gamma, \gamma_0)(\hat{\boldsymbol{\kappa}}_n - \boldsymbol{\kappa}_n) \\
= & S^*_{n,1}(\gamma) - 2S^*_{n,2}(\gamma) - 2S^*_{n,3}(\gamma) + S^*_{n,4}(\gamma) \qquad \text{(A.29)}
\end{aligned}
$$

where we denote $\boldsymbol{\chi}_t = \left[\boldsymbol{z}'_t \boldsymbol{\pi}'_x, \boldsymbol{\Phi}'_{L_n}(\boldsymbol{z}'_t \hat{\boldsymbol{\pi}}_q)\right]'$, and $S^*_{n,j}(\gamma)$'s are defined the same as in the proof of Theorem 2 with newly defined χ_t. Closely following the proof of Theorem 2 and applying Lemma 10 of the Online Appendix complete the proof of this theorem.

Proof of Theorem 7: The notation is defined the same as in the proof of Theorem 3 unless defined differently. Throughout this proof, we replace $\boldsymbol{x}_t$ in $\boldsymbol{X}^-_{v,t}$ and $\boldsymbol{X}^+_{v,t}$ with $\hat{\boldsymbol{\pi}}_x \mathbf{z}_t$. This notation replacement only affects $\boldsymbol{X}_{\hat{v}}$ in $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left[\boldsymbol{X}'_{\hat{v}}(\boldsymbol{I}_n - \boldsymbol{P}_{\hat{v}})\boldsymbol{X}_{\hat{v}}\right]^{-1} \boldsymbol{X}'_{\hat{v}}(\boldsymbol{I}_n - \boldsymbol{P}_{\hat{v}})(\boldsymbol{y} - \boldsymbol{X}_{\hat{v}}\boldsymbol{\beta})$.

Firstly, closely following the proof of Theorem 3 and applying Lemmas 6 and 7 of the Online Appendix, we obtain $n^{-1}\boldsymbol{X}'_{\hat{v}}\boldsymbol{X}_{\hat{v}} \xrightarrow{p} \boldsymbol{\Sigma}_{\boldsymbol{\Pi}_{\boldsymbol{x}}\mathbf{z}\mathbf{z}'\boldsymbol{\Pi}'_{\boldsymbol{x}},\gamma_0}$, $n^{-1}\boldsymbol{X}'_{\hat{v}}\boldsymbol{\Delta}_{\hat{v}} \xrightarrow{p} \boldsymbol{\Sigma}_{\boldsymbol{\pi}_x\mathbf{z}\boldsymbol{\Phi}'_{L_n},\gamma_0}$, and

$n^{-1}\boldsymbol{\Delta}_{\hat{v}}'\boldsymbol{\Delta}_{\hat{v}} \xrightarrow{p} \boldsymbol{\Sigma}_{\boldsymbol{\Phi}_{L_n}\boldsymbol{\Phi}_{L_n}',\gamma_0}$, where

$$\boldsymbol{\Sigma}_{\boldsymbol{\Pi_x}\mathbf{z}\mathbf{z}'\boldsymbol{\Pi}_{\boldsymbol{x}}',\gamma_0} \equiv \begin{bmatrix} \boldsymbol{\Pi_x} E\left[\mathbf{z}_t\mathbf{z}_t' I\left(q_t \leq \gamma_0\right)\right]\boldsymbol{\Pi}_{\boldsymbol{x}}' & \mathbf{0} \\ \mathbf{0}' & \boldsymbol{\Pi_x} E\left[\mathbf{z}_t\mathbf{z}_t' I\left(q_t > \gamma_0\right)\right]\boldsymbol{\pi}_x' \end{bmatrix},$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\Pi_x}\mathbf{z}\boldsymbol{\Phi}_{L_n}',\gamma_0} \equiv \begin{bmatrix} \boldsymbol{\Pi_x} E\left[\mathbf{z}_t\boldsymbol{\Phi}_{L_n}\left(\mathbf{z}_t'\boldsymbol{\pi}_q\right)' I\left(q_t \leq \gamma_0\right)\right] & \mathbf{0} \\ \mathbf{0}' & E\left[\boldsymbol{\Phi}_{L_n}\left(\mathbf{z}_t'\boldsymbol{\pi}_q\right)\mathbf{z}_t' I\left(q_t > \gamma_0\right)\right]\boldsymbol{\Pi}_{\boldsymbol{x}}' \end{bmatrix}.$$

It then follows

$$n^{-1}\boldsymbol{X}_{\hat{v}}'\left(\boldsymbol{I}_n - \boldsymbol{P}_{\hat{v}}\right)\boldsymbol{X}_{\hat{v}} \xrightarrow{p} \boldsymbol{\Sigma}_{\boldsymbol{\Pi_x}\mathbf{z}\mathbf{z}'\boldsymbol{\Pi}_{\boldsymbol{x}}',\gamma_0} - \boldsymbol{\Sigma}_{\boldsymbol{\Pi_x}\mathbf{z}\boldsymbol{\Phi}_{L_n}',\gamma_0}\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\Phi}_{L_n}\boldsymbol{\Phi}_{L_n}',\gamma_0}\boldsymbol{\Sigma}_{\boldsymbol{\Phi}_{L_n}'\mathbf{z}'\boldsymbol{\Pi}_{\boldsymbol{x}}',\gamma_0} \equiv \boldsymbol{J}. \tag{A.30}$$

Secondly, we consider $B_n\left(\hat{v}\right) = n^{-1}\boldsymbol{X}_v'\left(\boldsymbol{I}_n - \boldsymbol{P}_v\right)\left(\boldsymbol{y} - \boldsymbol{X}_v\boldsymbol{\beta}\right)$, where $y_t - X_{v,t}'\boldsymbol{\beta} = \eta_{v,0,t} + h_1\left(\mathbf{z}_t'\boldsymbol{\pi}_q\right)I\left(q_t \leq \gamma_0\right) + h_2\left(\mathbf{z}_t'\boldsymbol{\pi}_q\right)I\left(q_t > \gamma_0\right) + \varepsilon_t$ and $\eta_{v,0,t} = \boldsymbol{\beta}_2'\left(\boldsymbol{\Pi_x} - \hat{\boldsymbol{\pi}}_x\right)\mathbf{z}_t + \boldsymbol{\delta}_n'\left(\boldsymbol{\Pi_x} - \hat{\boldsymbol{\pi}}_x\right)\mathbf{z}_t I\left(q_t \leq \gamma_0\right) - \boldsymbol{\delta}_n'\boldsymbol{\Pi_x}\mathbf{z}_t d_t\left(\gamma_0 + v/a_n, \gamma_0\right)$. By $\boldsymbol{\pi}_x - \hat{\boldsymbol{\pi}}_x = O_p\left(n^{-1/2}\right)$ and Lemma 6 of the Online Appendix, we obtain $n^{-1}\boldsymbol{X}_{\hat{v}}'\left(\boldsymbol{I}_n - \boldsymbol{P}_{\hat{v}}\right)\boldsymbol{\eta}_{\hat{v},0} = O_p\left(a_n^{-1/2}n^{-\varsigma}\right) = o_p\left(n^{-1/2}\right)$. In addition, we have

$$\begin{aligned} n^{-1}\boldsymbol{X}_{\hat{v}}'\left(\boldsymbol{I}_n - \boldsymbol{P}_{\hat{v}}\right)\boldsymbol{h}_{1,\gamma_0}\left(\boldsymbol{z}\boldsymbol{\pi}_q\right) &= \boldsymbol{B}_1\left(\boldsymbol{\pi}_q - \hat{\boldsymbol{\pi}}_q\right) + o_p\left(n^{-1/2}\right) \\ n^{-1}\boldsymbol{X}_{\hat{v}}'\left(\boldsymbol{I}_n - \boldsymbol{P}_{\hat{v}}\right)\boldsymbol{h}_{2,\gamma_0}\left(\boldsymbol{z}\boldsymbol{\pi}_q\right) &= \boldsymbol{B}_2\left(\boldsymbol{\pi}_q - \hat{\boldsymbol{\pi}}_q\right) + o_p\left(n^{-1/2}\right) \end{aligned}$$

where

$$\boldsymbol{B}_1 = \left[\Gamma_{\boldsymbol{\Pi_x}\mathbf{z},1}', \mathbf{0}'\right]' - \boldsymbol{\Sigma}_{\boldsymbol{\Pi_x}\mathbf{z}\boldsymbol{\Phi}_{L_n}',\gamma_0}\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\Phi}_{L_n}\boldsymbol{\Phi}_{L_n}',\gamma_0}\left[\Gamma_{\boldsymbol{\Phi}_{L_n},1}', \mathbf{0}'\right]' \tag{A.31}$$

$$\boldsymbol{B}_2 = \left[\mathbf{0}', \Gamma_{\boldsymbol{\Pi_x}\mathbf{z},2}'\right]' - \boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{\Phi}_{L_n}',\gamma_0}\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\Phi}_{L_n}\boldsymbol{\Phi}_{L_n}',\gamma_0}\left[\mathbf{0}', \Gamma_{\boldsymbol{\Phi}_{L_n},2}'\right]' \tag{A.32}$$

where

$\Gamma_{\mathbf{\Pi}_{\boldsymbol{x}}\mathbf{z},1} = E\left[\mathbf{\Pi}_{\boldsymbol{x}}\mathbf{z}_t h_2^{(1)}\left(\mathbf{z}_t'\boldsymbol{\pi}_q\right)\mathbf{z}_t' I\left(q_t \le \gamma_0\right)\right]$ and $\Gamma_{\mathbf{\Pi}_{\boldsymbol{x}}\mathbf{z},2} = E\left[\mathbf{\Pi}_{\boldsymbol{x}}\mathbf{z}_t h_2^{(1)}\left(\mathbf{z}_t'\boldsymbol{\pi}_q\right)\mathbf{z}_t' I\left(q_t > \gamma_0\right)\right]$. It follows that

$$\begin{aligned}\boldsymbol{B}_n\left(\hat{v}\right) &= n^{-1}\boldsymbol{X}_{\hat{v}}'\left(\boldsymbol{I}_n - \boldsymbol{P}_{\hat{v}}\right)\left(\boldsymbol{y} - \boldsymbol{X}_{\hat{v}}\boldsymbol{\beta}\right)\\ &= -\boldsymbol{B}\left(\hat{\boldsymbol{\pi}}_q - \boldsymbol{\pi}_q\right) + n^{-1}\left(\boldsymbol{Z}_0\mathbf{\Pi}_{\boldsymbol{x}}' - \mathbf{\Sigma}_{\mathbf{\Pi}_{\boldsymbol{x}}\mathbf{z}\mathbf{\Phi}_{L_n}',\gamma_0}\mathbf{\Sigma}^{-1}_{\mathbf{\Phi}_{L_n}\mathbf{\Phi}_{L_n}',\gamma_0}\mathbf{\Delta}_0\right)\boldsymbol{\varepsilon} + o_p\left(n^{-1/2}\right)\end{aligned}$$

where $\boldsymbol{B} = \boldsymbol{B}_1 + \boldsymbol{B}_2 = \left[\mathbf{\Gamma}_{\mathbf{\Pi}_{\boldsymbol{x}}\mathbf{z},1}', \mathbf{\Gamma}_{\mathbf{\Pi}_{\boldsymbol{x}}\mathbf{z},2}'\right]' - \mathbf{\Sigma}_{\mathbf{\Pi}_{\boldsymbol{x}}\mathbf{z}\mathbf{\Phi}_{L_n}',\gamma_0}\mathbf{\Sigma}^{-1}_{\mathbf{\Phi}_{L_n}\mathbf{\Phi}_{L_n}',\gamma_0}\left[\mathbf{\Gamma}_{\mathbf{\Phi}_{L_n},1}', \mathbf{\Gamma}_{\mathbf{\Phi}_{L_n},2}'\right]'$, and that $\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{\Omega}\right)$ where

$$\mathbf{\Omega} = \boldsymbol{B}\left[E\left(\mathbf{z}_1\mathbf{z}_1'\right)\right]^{-1}\mathbf{\Omega}_{11}\left[E\left(\mathbf{z}_1\mathbf{z}_1'\right)\right]^{-1}\boldsymbol{B} - 2\boldsymbol{B}\mathbf{\Omega}_{12} + \mathbf{\Omega}_{22} \tag{A.33}$$

and

$$\begin{aligned}\mathbf{\Omega}_{22} &= \mathbf{\Sigma}_{\varepsilon,\mathbf{\Pi}_{\boldsymbol{x}}\mathbf{z}\mathbf{z}'\mathbf{\Pi}_{\boldsymbol{x}}',\gamma_0} - \mathbf{\Sigma}_{\mathbf{\Pi}_{\boldsymbol{x}}\mathbf{z}\mathbf{\Phi}_{L_n}',\gamma_0}\mathbf{\Sigma}^{-1}_{\mathbf{\Phi}_{L_n}\mathbf{\Phi}_{L_n}',\gamma_0}\mathbf{\Sigma}_{\varepsilon,\mathbf{\Phi}_{L_n}'\mathbf{\Pi}_{\boldsymbol{x}}'\mathbf{z}',\gamma_0} - \mathbf{\Sigma}_{\varepsilon,\mathbf{\Pi}_{\boldsymbol{x}}\mathbf{z}\mathbf{\Phi}_{L_n}',\gamma_0}\mathbf{\Sigma}^{-1}_{\mathbf{\Phi}_{L_n}\mathbf{\Phi}_{L_n}',\gamma_0}\mathbf{\Sigma}_{\mathbf{\Phi}_{L_n}\mathbf{\Pi}_{\boldsymbol{x}}'\mathbf{z}',\gamma_0}\\ &\quad + \mathbf{\Sigma}_{\mathbf{\Pi}_{\boldsymbol{x}}\mathbf{z}\mathbf{\Phi}_{L_n}',\gamma_0}\mathbf{\Sigma}^{-1}_{\mathbf{\Phi}_{L_n}\mathbf{\Phi}_{L_n}',\gamma_0}\mathbf{\Sigma}_{\varepsilon,\mathbf{\Phi}_{L_n}\mathbf{\Phi}_{L_n}',\gamma_0}\mathbf{\Sigma}^{-1}_{\mathbf{\Phi}_{L_n}\mathbf{\Phi}_{L_n}',\gamma_0}\mathbf{\Sigma}_{\mathbf{\Phi}_{L_n}\boldsymbol{\pi}_x'\mathbf{z}',\gamma_0}.\end{aligned}$$

This completes the proof of this theorem.

Proof of Theorem 8: We only need to give proofs when $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \boldsymbol{\beta}$ and $h_1\left(z\right) = h_2\left(z\right) \equiv 0$, under which model (2.1)-(2.2) becomes $y_t = \mathbf{x}_t'\boldsymbol{\beta} + e_t$, where $e_t = u_t\left[\sigma_1 I\left(q_t \le \gamma\right) + \sigma_2 I\left(q_t > \gamma\right)\right]$. It follows that $\boldsymbol{y}'\boldsymbol{M}_{\gamma,\boldsymbol{Q}}\boldsymbol{y} = e'\boldsymbol{M}_{\gamma,\boldsymbol{Q}}\boldsymbol{e} = \boldsymbol{e}'\boldsymbol{e} - \boldsymbol{e}'\boldsymbol{Q}_\gamma\left(\boldsymbol{Q}_\gamma'\boldsymbol{Q}_\gamma\right)^{-1}\boldsymbol{Q}_\gamma'\boldsymbol{e}$. Applying Lemma 1 in Hansen (1996, p.428), we obtain uniformly over $\gamma \in \left[\underline{\gamma}, \overline{\gamma}\right]$,

$$\begin{aligned}n^{-1}\boldsymbol{e}'\boldsymbol{e} &= \frac{\sigma_1^2}{n}\sum_{t=1}^n u_t^2 I\left(q_t \le \gamma\right) + \frac{\sigma_2^2}{n}\sum_{t=1}^n u_t^2 I\left(q_t > \gamma\right)\\ &\xrightarrow{a.s} \sigma_1^2 E\left[I\left(q_t \le \gamma\right)\right] + \sigma_2^2 E\left[I\left(q_t > \gamma\right)\right]\end{aligned}$$

and

$$
\begin{aligned}
& \boldsymbol{e}'\boldsymbol{Q}_\gamma\left(\boldsymbol{Q}_\gamma'\boldsymbol{Q}_\gamma\right)^{-1}\boldsymbol{Q}_\gamma'\boldsymbol{e} \\
= \; & \sigma_1^2\sum_{t=1}^n \mathbf{x}_t' u_t I\left(q_t\le\gamma\right)\left(\sum_{t=1}^n \mathbf{x}_t\mathbf{x}_t' I\left(q_t\le\gamma\right)\right)^{-1}\sum_{t=1}^n \mathbf{x}_t u_t I\left(q_t\le\gamma\right) \\
& +\sigma_2^2\sum_{t=1}^n \mathbf{x}_t' u_t I\left(q_t>\gamma\right)\left(\sum_{t=1}^n \mathbf{x}_t\mathbf{x}_t' I\left(q_t>\gamma\right)\right)^{-1}\sum_{t=1}^n \mathbf{x}_t u_t I\left(q_t>\gamma\right) \\
\Rightarrow \; & \sigma_1^2\boldsymbol{B}_u'\left(1,\lambda\right)\boldsymbol{B}_u\left(1,\lambda\right) \\
& +\sigma_2^2\left(\boldsymbol{\Sigma}_{xx'}^{1/2}\boldsymbol{B}_u\left(1\right)-\boldsymbol{\Sigma}_{xx',\gamma}^{1/2}\boldsymbol{B}_u\left(1,\lambda\right)\right)'\left(\boldsymbol{\Sigma}_{xx'}-\boldsymbol{\Sigma}_{xx',\gamma}\right)^{-1}\left(\boldsymbol{\Sigma}_{xx'}^{1/2}\boldsymbol{B}_u\left(1\right)-\boldsymbol{\Sigma}_{xx',\gamma}^{1/2}\boldsymbol{B}_u\left(1,\lambda\right)\right)
\end{aligned}
$$

because $\left[n^{-1/2}\boldsymbol{\Sigma}_{xx'}^{-1/2}\sum_{t=1}^{[ns]}\mathbf{x}_t u_t, n^{-1/2}\boldsymbol{\Sigma}_{xx',\gamma}^{-1/2}\sum_{t=1}^{[ns]}\mathbf{x}_t u_t I\left(q_t\le\gamma\right)\right]\Rightarrow\left[\boldsymbol{B}_u\left(s\right),\boldsymbol{B}_u\left(s,\lambda\right)\right]$ by the functional central limit theorem of Caner and Hansen (Theorem 1, 2001) for $s\in[0,1]$, where we denote $\lambda=F_q\left(\gamma\right)=E\left[I\left(q_t\le\gamma\right)\right]$, $\boldsymbol{\Sigma}_{xx'}=E\left(\mathbf{x}_t\mathbf{x}_t'\right)$, $\boldsymbol{\Sigma}_{xx',\gamma}=E\left[\mathbf{x}_t\mathbf{x}_t' I\left(q_t\le\gamma\right)\right]$, $\boldsymbol{B}_u\left(\lambda\right)$ is the d_x-dimensional standard multivariate Brownian motion, and "$\Longrightarrow$" denotes weak convergence on D[0,1] as $n\to\infty$ with D[0,1] being the space of cadlag functions on [0,1] equipped with Skorohod topology.

Denoting $\underline{\lambda}= F_q\left(\underline{\gamma}\right)$, $\bar{\lambda}=F_q\left(\bar{\gamma}\right)$, $\lambda^*=F_q\left(\gamma^*\right)$ and $\hat{\lambda}=F_q\left(\hat{\gamma}\right)$, we can rewritten the optimization problem (4.5) in terms of λ. If $\sigma_1^2=\sigma_2^2=\sigma^2$, we have

$$
\begin{aligned}
\hat{\lambda} \;\Rightarrow\; \lambda^* = & \arg\max_{\lambda\in\left[\underline{\lambda},\bar{\lambda}\right]}\boldsymbol{B}_u'\left(1,\lambda\right)\boldsymbol{B}_u\left(1,\lambda\right)+\left(\boldsymbol{\Sigma}_{xx'}^{1/2}\boldsymbol{B}_u\left(1\right)-\boldsymbol{\Sigma}_{xx',\gamma}^{1/2}\boldsymbol{B}_u\left(1,\lambda\right)\right)' \\
& \times\left(\boldsymbol{\Sigma}_{xx'}-\boldsymbol{\Sigma}_{xx',\gamma}\right)^{-1}\left(\boldsymbol{\Sigma}_{xx'}^{1/2}\boldsymbol{B}_u\left(1\right)-\boldsymbol{\Sigma}_{xx',\gamma}^{1/2}\boldsymbol{B}_u\left(1,\lambda\right)\right)
\end{aligned}
$$

so that $\hat{\gamma}\Rightarrow\gamma^*$ by the continuous mapping theorem, where

$$
\gamma^*=F_q^{-1}\left(\lambda^*\right). \tag{A.34}
$$

If $\sigma_1^2 \neq \sigma_2^2$, we have

$$\hat{\lambda} \xrightarrow{p} \lambda^* = \arg \min_{\lambda \in [\underline{\lambda}, \bar{\lambda}]} \sigma_1^2 \lambda + \sigma_2^2 (1 - \lambda) \tag{A.35}$$

so that $\lambda^* = \underline{\lambda} I (\sigma_1^2 > \sigma_2^2) + \bar{\lambda} I (\sigma_1^2 < \sigma_2^2)$ or equivalently, $\gamma^* = \underline{\gamma} I (\sigma_1^2 > \sigma_2^2) + \bar{\gamma} I (\sigma_1^2 < \sigma_2^2)$ because $F_q(\cdot)$ is strictly increasing. This completes the proof of this theorem.

Proof of Theorem 9: For a given γ, we have $\hat{\boldsymbol{\alpha}}_{L_n}(\gamma) = \left(\hat{\boldsymbol{\Phi}}_{L_n,\gamma}^{*\prime} \hat{\boldsymbol{\Phi}}_{L_n,\gamma}^{*} \right)^{-1} \hat{\boldsymbol{\Phi}}_{L_n,\gamma}^{*\prime} \boldsymbol{y}^*$, so that

$$\begin{aligned} W_n(\gamma) &= \hat{\boldsymbol{\alpha}}_{L_n}(\gamma)' \hat{\boldsymbol{\Phi}}_{L_n,\gamma}^{*\prime} \hat{\boldsymbol{\Phi}}_{L_n,\gamma}^{*} \left(\hat{\boldsymbol{\Phi}}_{L_n,\gamma}^{*\prime} \hat{\boldsymbol{\varepsilon}}_\gamma \hat{\boldsymbol{\varepsilon}}_\gamma' \hat{\boldsymbol{\Phi}}_{L_n,\gamma}^{*} \right)^{-1} \hat{\boldsymbol{\Phi}}_{L_n,\gamma}^{*\prime} \hat{\boldsymbol{\Phi}}_{L_n,\gamma}^{*} \hat{\boldsymbol{\alpha}}_{L_n}(\gamma) \\ &= \boldsymbol{y}^{*\prime} \hat{\boldsymbol{\Phi}}_{L_n,\gamma}^{*} \left(\hat{\boldsymbol{\Phi}}_{L_n,\gamma}^{*\prime} \hat{\boldsymbol{\varepsilon}}_\gamma \hat{\boldsymbol{\varepsilon}}_\gamma' \hat{\boldsymbol{\Phi}}_{L_n,\gamma}^{*} \right)^{-1} \hat{\boldsymbol{\Phi}}_{L_n,\gamma}^{*\prime} \boldsymbol{y}^*. \end{aligned}$$

Applying tedious but straightforward calculations, we obtain the respective tth row vector of $\hat{\boldsymbol{\Phi}}_{L_n,\gamma}^{-,*}$ and $\hat{\boldsymbol{\Phi}}_{L_n,\gamma}^{-,*}$ as follows

$$\boldsymbol{a}_{\gamma,t,-}' = [\boldsymbol{\Phi}_{L_n}(\hat{q}_t) - \hat{\boldsymbol{\pi}}_{\boldsymbol{\Phi},\gamma,-} \mathbf{x}_t]' I (q_t \leq \gamma) \text{ and } \boldsymbol{a}_{\gamma,t,+}' = [\boldsymbol{\Phi}_{L_n}(\hat{q}_t) - \hat{\boldsymbol{\pi}}_{\boldsymbol{\Phi},\gamma,+} \mathbf{x}_t]' I (q_t > \gamma),$$

where we denote $\boldsymbol{\Sigma}_{n,xx'} = n^{-1} \sum_{t=1}^n \mathbf{x}_t \mathbf{x}_t'$, $\boldsymbol{\Sigma}_{n,xx',\gamma} = n^{-1} \sum_{t=1}^n \mathbf{x}_t \mathbf{x}_t' I (\mathbf{q}_t \leq \gamma)$, $\hat{\boldsymbol{\pi}}_{\boldsymbol{\Phi},\gamma,-} = \sum_{s=1}^n \boldsymbol{\Phi}_{L_n}(\hat{q}_s) \mathbf{x}_s' I (q_s \leq \gamma) \boldsymbol{\Sigma}_{n,xx',\gamma}^{-1}$,
and $\hat{\boldsymbol{\pi}}_{\boldsymbol{\Phi},\gamma,+} = \sum_{s=1}^n \boldsymbol{\Phi}_{L_n}(\hat{q}_s) \mathbf{x}_s' I (q_s > \gamma) (\boldsymbol{\Sigma}_{n,xx'} - \boldsymbol{\Sigma}_{n,xx',\gamma})^{-1}$. It follows that $\hat{\boldsymbol{\Phi}}_{L_n,\gamma}^{*\prime} \boldsymbol{y}^* = \left[\left(\hat{\boldsymbol{\Phi}}_{L_n,\gamma}^{-,*\prime} \boldsymbol{y}^* \right)', \left(\hat{\boldsymbol{\Phi}}_{L_n,\gamma}^{+,*\prime} \boldsymbol{y}^* \right)' \right]' = \left[\sum_{t=1}^n y_t \boldsymbol{a}_{\gamma,t,-}', \sum_{t=1}^n y_t \boldsymbol{a}_{\gamma,t,+}' \right]'$, and $\hat{\boldsymbol{\Phi}}_{L_n,\gamma}^{*\prime} \hat{\boldsymbol{\varepsilon}}_\gamma \hat{\boldsymbol{\varepsilon}}_\gamma' \hat{\boldsymbol{\Phi}}_{L_n,\gamma}^{*}$ $= \begin{bmatrix} \Delta_{n,\gamma}^- & \mathbf{0}_{L_n \times L_n} \\ \mathbf{0}_{L_n \times L_n} & \Delta_{n,\gamma}^+ \end{bmatrix}$, where $\Delta_{n,\gamma}^- = \sum_{t=1}^n \sum_{t'=1}^n \hat{\varepsilon}_{\gamma,t}^- \hat{\varepsilon}_{\gamma,t'}^- \boldsymbol{a}_{\gamma,t,-}' \boldsymbol{a}_{\gamma,t',-}$, $\Delta_{n,\gamma}^+ =$

$\sum_{t=1}^{n}\sum_{t'=1}^{n}\hat{\boldsymbol{\varepsilon}}_{\gamma,t}^{+}\hat{\boldsymbol{\varepsilon}}_{\gamma,t'}^{+}\boldsymbol{a}_{\gamma,t,+}'\boldsymbol{a}_{\gamma,t',+}$,

$$\begin{aligned}\hat{\varepsilon}_{\gamma,t}^{-} &= \left[y_t - \mathbf{x}_t'\hat{\boldsymbol{\beta}}_1(\gamma) - \hat{\boldsymbol{\alpha}}_{L_n,1}'(\gamma)\boldsymbol{\Phi}_{L_n}(\hat{q}_t)\right] I(q_t \leq \gamma) \\ &= \left[\mathbf{x}_t'\left(\boldsymbol{\beta}_1 - \hat{\boldsymbol{\beta}}_1(\gamma)\right) + h_1(\mathbf{z}_t'\boldsymbol{\pi}_q) - \hat{\boldsymbol{\alpha}}_{L_n,1}'(\gamma)\boldsymbol{\Phi}_{L_n}(\hat{q}_t) + \sigma_1 u_t\right] I(q_t \leq \gamma)\end{aligned}$$

and

$$\begin{aligned}\hat{\varepsilon}_{\gamma,t}^{+} &= \left[y_t - \mathbf{x}_t'\hat{\boldsymbol{\beta}}_2(\gamma) - \hat{\boldsymbol{\alpha}}_{L_n,2}'(\gamma)\boldsymbol{\Phi}_{L_n}(\hat{q}_t)\right] I(q_t > \gamma) \\ &= \left[\mathbf{x}_t'\left(\boldsymbol{\beta}_2 - \hat{\boldsymbol{\beta}}_2(\gamma)\right) + h_2(\mathbf{z}_t'\boldsymbol{\pi}_q) - \hat{\boldsymbol{\alpha}}_{L_n,2}'(\gamma)\boldsymbol{\Phi}_{L_n}(\hat{q}_t) + \sigma_2 u_t\right] I(q_t > \gamma).\end{aligned}$$

It is readily seen that $n^{-1}\sum_{t=1}^{n}\hat{\boldsymbol{\varepsilon}}_{\gamma,t}^{-}\boldsymbol{a}_{\gamma,t,-} = \sigma_1 n^{-1}\sum_{t=1}^{n}u_t\boldsymbol{a}_{\gamma,t,-}$ and $n^{-1}\sum_{t=1}^{n}\hat{\boldsymbol{\varepsilon}}_{\gamma,t}^{+}\boldsymbol{a}_{\gamma,t,+} = \sigma_2 n^{-1}\sum_{t=1}^{n}u_t\boldsymbol{a}_{\gamma,t,+}$ under H_0.

Firstly, we consider the case that $\boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$ under which $\widehat{\gamma} = \gamma_0 + O_p\left(n^{-1+2\min(\varsigma,\varrho)}\right)$ by Theorem 2 and $y_t = \mathbf{x}_t'\boldsymbol{\beta}_1 I(q_t \leq \gamma_0) + \mathbf{x}_t'\boldsymbol{\beta}_2 I(q_t > \gamma_0) + \varepsilon_t$, where $\varepsilon_t = \sigma_1 u_t I(q_t \leq \gamma_0) + \sigma_2 u_t I(q_t > \gamma_0)$. In the proof of Theorem 2, we have $\left\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right\| = O_p(\vartheta_n)$ if $\hat{\gamma} = \gamma_0 + O_p\left(n^{-1+2\min(\varsigma,\varrho)}\right)$. For notation simplification, we denote $\boldsymbol{\chi}_{n,\omega,1} = n^{-1}\sum_{t=1}^{n}\boldsymbol{\omega}_t\mathbf{x}_t' I(q_t \leq \hat{\gamma}) I(q_t \leq \gamma_0)$, $\boldsymbol{\chi}_{n,\omega,2} = n^{-1}\sum_{t=1}^{n}\boldsymbol{\omega}_t\mathbf{x}_t' I(q_t \leq \hat{\gamma})$, $\boldsymbol{\lambda}_{n,\omega,1} = n^{-1}\sum_{t=1}^{n}u_t\boldsymbol{\omega}_t I(q_t \leq \hat{\gamma}) I(q_t \leq \gamma_0)$, and $\boldsymbol{\lambda}_{n,\omega,2} = n^{-1}\sum_{t=1}^{n}u_t\boldsymbol{\omega}_t I(q_t \leq \hat{\gamma})$ for $\boldsymbol{\omega}_t = \boldsymbol{\Phi}_{L_n}(\hat{q}_t)$

and $\mathbf{x}_t$. Then, we have $\hat{\boldsymbol{\pi}}_{\boldsymbol{\Phi},\gamma,-} = \boldsymbol{\chi}_{n,\boldsymbol{\Phi},2}\boldsymbol{\chi}_{n,x,2}^{-1}$ and

$$\begin{aligned}
& n^{-1}\hat{\boldsymbol{\Phi}}_{L_n,\hat{\gamma}}^{-,*\prime}\boldsymbol{y}^* \\
= & \left(\boldsymbol{\chi}_{n,\boldsymbol{\Phi},1} - \boldsymbol{\chi}_{n,\boldsymbol{\Phi},2}\boldsymbol{\chi}_{n,x,2}^{-1}\boldsymbol{\chi}_{n,x,1}\right)\boldsymbol{\delta}_n + \sigma_1\left(\boldsymbol{\lambda}_{n,\boldsymbol{\Phi},1} - \boldsymbol{\chi}_{n,\boldsymbol{\Phi},2}\boldsymbol{\chi}_{n,x,2}^{-1}\boldsymbol{\lambda}_{n,x,1}\right) \\
& +\sigma_2\left[\boldsymbol{\lambda}_{n,\boldsymbol{\Phi},2} - \boldsymbol{\lambda}_{n,\boldsymbol{\Phi},1} + \boldsymbol{\chi}_{n,\boldsymbol{\Phi},2}\boldsymbol{\chi}_{n,x,2}^{-1}\left(\boldsymbol{\lambda}_{n,x,1} - \boldsymbol{\lambda}_{n,x,2}\right)\right] \\
= & \sigma_1 n^{-1}\sum_{t=1}^{n} u_t\boldsymbol{\Phi}_{L_n}\left(\mathbf{z}_t'\boldsymbol{\pi}_q\right)I\left(q_t \leq \gamma_0\right) - \sigma_1 n^{-1}\sum_{t=1}^{n}\boldsymbol{\Phi}_{L_n}\left(\mathbf{z}_t'\boldsymbol{\pi}_q\right)\mathbf{x}_t' I\left(q_t \leq \gamma_0\right) \\
& \times\left[n^{-1}\sum_{t=1}^{n}\mathbf{x}_t\mathbf{x}_t' I\left(q_t \leq \gamma_0\right)\right]^{-1} n^{-1}\sum_{t=1}^{n} u_t\mathbf{x}_t I\left(q_t \leq \gamma_0\right) \\
& +O_p\left(n^{-1+2\min(\varsigma,\varrho)}\left(\left\|\boldsymbol{\Phi}_{L_n}\right\|_0 n^{-\varsigma} + \left\|\boldsymbol{\Phi}_{L_n}\right\|_0/\sqrt{n}\right)\right)
\end{aligned}$$

by Lemma 11 of the Online Appendix. By Lemma 1 of the Online Appendix, we have

$$n^{-1}\hat{\boldsymbol{\Phi}}_{L_n,\hat{\gamma}}^{-,*\prime}\boldsymbol{y}^* \approx \sigma_1 n^{-1}\sum_{t=1}^{n} u_t\boldsymbol{c}_{-,t}\left(\gamma_0\right) = O_p\left(\left\|\boldsymbol{\Phi}_{L_n}\right\|_0/\sqrt{n}\right) \tag{A.36}$$

where

we denote $\boldsymbol{c}_{-,t}\left(\gamma_0\right) = \left\{\boldsymbol{\Phi}_{L_n}\left(\mathbf{z}_t'\boldsymbol{\pi}_q\right) - E\left[\sum_{t=1}^{n}\boldsymbol{\Phi}_{L_n}\left(\mathbf{z}_t'\boldsymbol{\pi}_q\right)\mathbf{x}_t' I\left(q_t \leq \gamma_0\right)\right]\boldsymbol{\Sigma}_{xx',\gamma_0}^{-1}\mathbf{x}_t'\right\} I\left(q_t \leq \gamma_0\right)$.
Similarly, we can show that

$$n^{-1}\hat{\boldsymbol{\Phi}}_{L_n,\hat{\gamma}}^{+,*\prime}\boldsymbol{y}^* \approx \sigma_2 n^{-1}\sum_{t=1}^{n} u_t\boldsymbol{c}_{+,t}\left(\gamma_0\right) = O_p\left(\left\|\boldsymbol{\Phi}_{L_n}\right\|_0/\sqrt{n}\right) \tag{A.37}$$

where $\boldsymbol{c}_{+,t}\left(\gamma_0\right) =$
$\left\{\boldsymbol{\Phi}_{L_n}\left(\mathbf{z}_t'\boldsymbol{\pi}_q\right) - E\left[\sum_{t=1}^{n}\boldsymbol{\Phi}_{L_n}\left(\mathbf{z}_t'\boldsymbol{\pi}_q\right)\mathbf{x}_t' I\left(q_t > \gamma_0\right)\right]\left(\boldsymbol{\Sigma}_{xx',\gamma_0} - \boldsymbol{\Sigma}_{xx',\gamma_0}\right)^{-1}\mathbf{x}_t'\right\} I\left(q_t > \gamma_0\right)$.
Similarly, we obtain $n^{-1}\sum_{t=1}^{n}\hat{\boldsymbol{\varepsilon}}_{\hat{\gamma},t}^{-}\boldsymbol{a}_{\hat{\gamma},t,-} \approx \sigma_1 n^{-1}\sum_{t=1}^{n} u_t\boldsymbol{c}_{-,t}\left(\gamma_0\right)$ and $n^{-1}\sum_{t=1}^{n}\hat{\boldsymbol{\varepsilon}}_{\hat{\gamma},t}^{+}\boldsymbol{a}_{\hat{\gamma},t,+} \approx \sigma_2 n^{-1}\sum_{t=1}^{n} u_t\boldsymbol{c}_{+,t}\left(\gamma_0\right)$.

Applying the Cramer-Wold device and Wooldrige and White's central limit theorem for strong mixing process gives

$$\sqrt{n}\boldsymbol{\Omega}^{-1/2}\left[\begin{array}{c}\sigma_1 n^{-1}\sum_{t=1}^{n}u_t\boldsymbol{c}_{-,t}\left(\gamma_0\right)\\ \sigma_2 n^{-1}\sum_{t=1}^{n}u_t\boldsymbol{c}_{+,t}\left(\gamma_0\right)\end{array}\right]\xrightarrow{d}N\left(\mathbf{0}_{2L_n},I_{2L_n}\right)$$

where we denote $\boldsymbol{\Omega}_{-}=n^{-1}\sum_{t=1}^{n}\sum_{t'=1}^{n}E\left[u_t u_{t'}\boldsymbol{c}_{-,t}\left(\gamma_0\right)\boldsymbol{c}'_{-,t'}\left(\gamma_0\right)\right]$ and $\boldsymbol{\Omega}_{+}=n^{-1}\sum_{t=1}^{n}\sum_{t'=1}^{n}E\left[u_t u_{t'}\boldsymbol{c}_{-,t}\left(\gamma_0\right)\boldsymbol{c}'_{-,t'}\left(\gamma_0\right)\right]$. Under Assumption 1, both $\boldsymbol{\Omega}_{-}$ and $\boldsymbol{\Omega}_{+}$ are finite nonsingular matrix. Taking the above results together gives $W_n\left(\hat{\gamma}\right)\xrightarrow{d}\boldsymbol{\chi}^2_{2L_n}$ as $n\to\infty$.

Next, we consider the case that $\boldsymbol{\beta}_1=\boldsymbol{\beta}_2=\boldsymbol{\beta}$ and $h_1\left(z\right)=h_2\left(z\right)\equiv 0$. We consider three cases.

(i) If $\sigma_1^2<\sigma_2^2$, $\hat{\gamma}\xrightarrow{p}\bar{\gamma}$ by Theorem 8. We have $n^{-1}\hat{\boldsymbol{\Phi}}^{-,*\prime}_{L_n,\hat{\gamma}}\boldsymbol{y}^*$ $=\sigma_1\left(\lambda_{n,\boldsymbol{\Phi},1}-\boldsymbol{\chi}_{n,\boldsymbol{\Phi},2}\boldsymbol{\chi}^{-1}_{n,x,2}\boldsymbol{\lambda}_{n,x,1}\right)$ $+\sigma_2\left[\boldsymbol{\lambda}_{n,\boldsymbol{\Phi},2}-\boldsymbol{\lambda}_{n,\boldsymbol{\Phi},1}+\boldsymbol{\chi}_{n,\boldsymbol{\Phi},2}\boldsymbol{\chi}^{-1}_{n,x,2}\left(\boldsymbol{\lambda}_{n,x,1}-\boldsymbol{\lambda}_{n,x,2}\right)\right]\approx n^{-1}\sum_{t=1}^{n}u_t\left\{\boldsymbol{c}_{-,t}\left(\gamma_0,\bar{\gamma}\right)\left(\sigma_1-\sigma_2\right)+\sigma_2\boldsymbol{c}_{-,t}\left(\bar{\gamma},\bar{\gamma}\right)\right\}$,
where $\boldsymbol{c}_{-,t}\left(\gamma_1,\gamma_2\right)=\left\{\boldsymbol{\Phi}_{L_n}\left(\mathbf{z}'_t\boldsymbol{\pi}_q\right)-E\left[\sum_{t=1}^{n}\boldsymbol{\Phi}_{L_n}\left(\mathbf{z}'_t\boldsymbol{\pi}_q\right)\mathbf{x}'_t I\left(q_t\leq\gamma_2\right)\right]\boldsymbol{\Sigma}^{-1}_{xx',\gamma_2}\mathbf{x}'_t\right\}I\left(q_t\leq\gamma_1\right)$, while $n^{-1}\hat{\boldsymbol{\Phi}}^{+,*\prime}_{L_n,\hat{\gamma}}\boldsymbol{y}^*\approx n^{-1}\sum_{t=1}^{n}u_t\boldsymbol{c}_{+,t}\left(\bar{\gamma}\right)$. In addition, we have $n^{-1}\sum_{t=1}^{n}\hat{\boldsymbol{\varepsilon}}^{-}_{\hat{\gamma},t}\boldsymbol{a}_{\hat{\gamma},t,-}=\sigma_1\left(\boldsymbol{\lambda}_{n,\boldsymbol{\Phi},2}-\boldsymbol{\chi}_{n,\boldsymbol{\Phi},2}\boldsymbol{\chi}^{-1}_{n,x,2}\boldsymbol{\lambda}_{n,x,2}\right)\approx n^{-1}\sum_{t=1}^{n}u_t\boldsymbol{c}_{-,t}\left(\bar{\gamma}\right)$ and $n^{-1}\sum_{t=1}^{n}\hat{\boldsymbol{\varepsilon}}^{+}_{\hat{\gamma},t}\boldsymbol{a}_{\hat{\gamma},t,+}\approx n^{-1}\sum_{t=1}^{n}u_t\boldsymbol{c}_{+,t}\left(\bar{\gamma}\right)$. As $n^{-1}\hat{\boldsymbol{\Phi}}^{-,*\prime}_{L_n,\hat{\gamma}}\boldsymbol{y}^*$ and $n^{-1}\sum_{t=1}^{n}\hat{\boldsymbol{\varepsilon}}^{-}_{\hat{\gamma},t}\boldsymbol{a}_{\hat{\gamma},t,-}$ converge to normal distribution with different variance, $W_n\left(\hat{\gamma}\right)\xrightarrow{d}\chi^2_{2L_n}$ fails to hold.

(ii) If $\sigma_1^2>\sigma_2^2$, $\hat{\gamma}\xrightarrow{p}\underline{\gamma}$ by Theorem 8, then $\boldsymbol{\chi}_{n,\omega,1}-\boldsymbol{\chi}_{n,\omega,2}=\mathbf{0}$ and $\boldsymbol{\lambda}_{n,\omega,1}-\boldsymbol{\lambda}_{n,\omega,2}=\mathbf{0}$. It is readily seen that $n^{-1}\hat{\boldsymbol{\Phi}}^{-,*\prime}_{L_n,\hat{\gamma}}\boldsymbol{y}^*=\sigma_1\left(\boldsymbol{\lambda}_{n,\boldsymbol{\Phi},1}-\boldsymbol{\chi}_{n,\boldsymbol{\Phi},2}\boldsymbol{\chi}^{-1}_{n,x,2}\lambda_{n,x,1}\right)$ $\approx\sigma_1 n^{-1}\sum_{t=1}^{n}u_t\boldsymbol{c}_{-,t}\left(\underline{\gamma}\right)$, $n^{-1}\hat{\Phi}^{+,*\prime}_{L_n,\hat{\gamma}}\boldsymbol{y}^*\approx\sigma_2 n^{-1}\sum_{t=1}^{n}u_t\boldsymbol{c}_{+,t}\left(\gamma_0,\underline{\gamma}\right)$, where $\boldsymbol{c}_{+,t}\left(\gamma_1,\gamma_2\right)=\left\{\Phi_{L_n}\left(\mathbf{z}'_t\boldsymbol{\pi}_q\right)-E\left[\sum_{t=1}^{n}\boldsymbol{\Phi}_{L_n}\left(\mathbf{z}'_t\boldsymbol{\pi}_q\right)\mathbf{x}'_t I\left(q_t>\gamma_2\right)\right]\boldsymbol{\Sigma}^{-1}_{xx',\gamma_2}\mathbf{x}'_t\right\}I\left(q_t>\gamma_1\right)$. In addition, $n^{-1}\sum_{t=1}^{n}\hat{\boldsymbol{\varepsilon}}^{-}_{\hat{\gamma},t}\boldsymbol{a}_{\hat{\gamma},t,-}\approx n^{-1}\sum_{t=1}^{n}u_t\boldsymbol{c}_{-,t}\left(\underline{\gamma}\right)$ and $n^{-1}\sum_{t=1}^{n}\hat{\boldsymbol{\varepsilon}}^{+}_{\hat{\gamma},t}\boldsymbol{a}_{\hat{\gamma},t,+}\approx n^{-1}\sum_{t=1}^{n}u_t\boldsymbol{c}_{+,t}\left(\underline{\gamma}\right)$. As $n^{-1}\hat{\Phi}^{+,*\prime}_{L_n,\hat{\gamma}}\boldsymbol{y}^*$ and $n^{-1}\sum_{t=1}^{n}\hat{\boldsymbol{\varepsilon}}^{+}_{\hat{\gamma},t}\boldsymbol{a}_{\hat{\gamma},t,+}$ converge to normal distribution with different variance, $W_n\left(\hat{\gamma}\right)\xrightarrow{d}\chi^2_{2L_n}$ fails to hold.

(iii) If $\sigma_1^2 = \sigma_2^2$, $\hat{\gamma} \xrightarrow{d} \gamma^*$ by Theorem 8. We then have $n^{-1}\hat{\Phi}_{L_n,\hat{\gamma}}^{-,*\prime}\boldsymbol{y}^* = \sigma\left(\boldsymbol{\lambda}_{n,\boldsymbol{\Phi},2} - \boldsymbol{\chi}_{n,\boldsymbol{\Phi},2}\boldsymbol{\chi}_{n,x,2}^{-1}\boldsymbol{\lambda}_{n,x,2}\right) \overset{d}{\approx} \sigma n^{-1}\sum_{t=1}^{n} u_t \boldsymbol{c}_{-,t}(\gamma^*)$ and $n^{-1}\hat{\Phi}_{L_n,\hat{\gamma}}^{+,*\prime}\boldsymbol{y}^* \overset{d}{\approx} \sigma n^{-1}\sum_{t=1}^{n} u_t \left\{\boldsymbol{c}_{+,t}(\gamma^*)\right\}$. In addition, $n^{-1}\sum_{t=1}^{n}\hat{\boldsymbol{\varepsilon}}_{\hat{\gamma},t}^{-}\boldsymbol{a}_{\hat{\gamma},t,-} \approx \sigma n^{-1}\sum_{t=1}^{n} u_t \boldsymbol{c}_{-,t}(\gamma^*)$ and $n^{-1}\sum_{t=1}^{n}\hat{\boldsymbol{\varepsilon}}_{\hat{\gamma},t}^{+}\boldsymbol{a}_{\hat{\gamma},t,+} \approx \sigma n^{-1}\sum_{t=1}^{n} u_t \boldsymbol{c}_{+,t}(\gamma^*)$. If $\{u_t\}$ is independent of $\{(\mathbf{x}_t, \mathbf{z}_t)\}$, all the four terms converges to mixed normal distribution with zero mean, therefore $W_n(\hat{\gamma}) \xrightarrow{d} \chi_{2L_n}^2$ continues to hold as $n \to \infty$.

This completes the proof of this theorem.