*Econometric Theory*, **38**, 2022, 562–595.
doi:10.1017/S0266466621000256

# ENDOGENEITY IN SEMIPARAMETRIC THRESHOLD REGRESSION

Andros Kourtellos
*University of Cyprus*

Thanasis Stengos
*University of Guelph*

Yiguo Sun
*University of Guelph*

This paper estimates threshold regression models with an endogenous threshold variable using a nonparametric control function approach. Assuming diminishing threshold effects, we derive the consistency and limiting distribution of our proposed estimator constructed from the series approximation method for weakly dependent data. In addition, we propose a test for the endogeneity of the threshold variable, which is valid regardless of whether the threshold effects exist. We assess the performance of our methods using Monte Carlo simulations.

## 1. INTRODUCTION

Several economic hypotheses suggest the presence of threshold-like structures. For example, debt levels that are above a particular threshold parameter 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 may result in heterogeneous transmission effects. These types of nonlinearities can be modeled by threshold regression models, which sort 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 the nonlinear time series literature. For example, the threshold-autoregressive (TAR) model allows the autoregressive coefficients to change between the regimes, while the

The first author acknowledges that this research has received funding from Marie Skłodowska-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. The authors would like to thank the editor, the co-editor, and the four anonymous referees for their patience and excellent comments that significantly improved the manuscript. Address correspondence to Yiguo Sun, Department of Economics and Finance, University of Guelph, Guelph, Ontario N1G 2W1, Canada; E-mail: yisun@uoguelph.ca.

© The Author(s), 2021. Published by Cambridge University Press. 

https://doi.org/10.1017/S0266466621000256 Published online by Cambridge University Press

coefficients keep constant within 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 predetermined threshold variables. Notable examples include the works of Chan (1993), Hansen (2000), Caner and Hansen (2004), Gonzalo and Wolf (2005), Seo and Linton (2007), and Yu (2012, 2015). Recently, there is a growing interest in threshold regression models that accommodate endogenous threshold variables. Kourtellos, Stengos, and Tan (2016) propose estimation and make inference for a threshold regression model that allows for an endogenous threshold variable as well as for endogenous regressors under certain parametric assumptions. In particular, in the spirit of Heckman's sample selection method, they account for the threshold endogeneity bias by including regime specific inverse Mills ratio bias correction terms in the threshold regression. Closely related to Kourtellos et al. (2016), Christopoulos, McAdam, and Tzavalis (2021) apply copula theory to deal with the threshold endogeneity. Seo and Shin (2016) study a dynamic threshold panel data model, which allows both regressors and threshold variable to be endogenous. In particular, they propose a first-differenced GMM estimator with lagged variables as instruments, as well as a Hausman-type test statistic for testing the null hypothesis of threshold variable exogeneity. Alternatively, without resorting to instruments, Yu and Phillips (2018) construct an integrated difference kernel (IDK) estimator to estimate the threshold parameter in the presence of an endogenous threshold variable and endogenous regressors within the fixed threshold effect framework for an i.i.d. sample. The IDK estimator is shown to enjoy the rate $n$ convergence for a parametric endogenous threshold regression.

In this paper, we propose a semiparametric approach to deal with the endogeneity of threshold variable and regressors that avoids the challenges of pure nonparametric models and at the same time relaxes parametric assumptions. Specifically, we are able to transfer a linear threshold model with endogenous threshold effects into a partially linear threshold model via a control function approach in the spirit of Yu et al. (2020). Compared with the parametric model considered in Kourtellos et al. (2016), our model introduces nonparametric threshold bias correction terms to avoid possible misspecification of the normality assumption. In doing so, we propose to estimate the threshold parameter using a concentrated least-squares estimation method, which includes a regime specific control function estimated by a series estimation method.[2] We derive the limiting distribution of our proposed estimators for both the threshold and slope parameters as well as for the smooth

[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 the TAR model with parametric smoothed transition functions such as the logistic and the exponential functions.

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

https://doi.org/10.1017/S0266466621000256 Published online by Cambridge University Press

control functions. Finally, we propose a test for the endogeneity of the threshold variable.

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 for the case of exogenous regressors. Section 3 presents a test for the endogeneity of the threshold variable. Section 4 reports some Monte Carlo simulation results to assess the finite sample performance of our estimators and test statistic. Section 5 concludes. We delay all the mathematical proofs to the Appendix. Supplementary proofs can be found in the supplementary section of this paper; we will refer to this as the Online Appendix.

## 2. THE MODEL AND ESTIMATOR

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

$$y_t = \mathbf{x}_t'\beta_1 + u_{1t}, \text{if } q_t \leq \gamma_0 \tag{2.1}$$

$$y_t = \mathbf{x}_t'\beta_2 + u_{2t}, \text{if } q_t > \gamma_0 \tag{2.2}$$

for $t = 1, 2, \ldots, n$, where $y_t$ is the dependent variable, $\mathbf{x}_t$ is a $d_{\mathbf{x}} \times 1$ vector of exogenous regressors, $\beta_1$ and $\beta_2$ are regime-specific coefficients, and $u_{jt}$ is a regime-specific error with zero mean and variance $\sigma_j^2 > 0$ for $j = 1, 2$. Furthermore, $q_t$ is a scalar endogenous threshold variable with $\gamma_0$ being the threshold parameter. A reduced form equation for $q_t$ is given by

$$q_t = \mathbf{z}_t'\pi_q + v_t,\ t = 1, 2, \ldots, n, \tag{2.3}$$

where $\mathbf{z}_t$ is a $d_{\mathbf{z}} \times 1$ vector of instrumental variables for $q_t$ and $\mathbf{z}_t$ is allowed to contain $\mathbf{x}_t$. Also, $E(v_t|\mathcal{F}_{n,t-1}) = 0$ for all $t$, where $\mathcal{F}_{n,t}$ is the smallest sigma-field generated from $\left\{\left(\mathbf{x}_{s+1}', \mathbf{z}_{s+1}', q_s, u_s\right) : 1 \leq s \leq t \leq n\right\}$ with $u_t = u_{1t}I(q_t \leq \gamma_0) + u_{2t}I(q_t > \gamma_0)$. The endogeneity of the threshold variable $q_t$ comes from the contemporaneous correlation between $u_{jt}$ and $v_t$ for $j = 1, 2$.

Assuming that $E(u_{jt}|\mathcal{F}_{n,t-1}, q_t) = E(u_{jt}|v_t) = h_j(v_t)$ holds almost surely for $j = 1, 2$ and all $t$, we have

$$E\left(y_t|\mathcal{F}_{n,t-1}, q_t\right) = \left[\mathbf{x}_t'\beta_1 + h_1(v_t)\right] I(q_t \leq \gamma_0) + \left[\mathbf{x}_t'\beta_2 + h_2(v_t)\right] I(q_t > \gamma_0).$$

Therefore, our semiparametric threshold regression model is given by

$$y_t = \mathbf{x}_t'\beta_1 I(q_t \leq \gamma_0) + \mathbf{x}_t'\beta_2 I(q_t > \gamma_0) + h_1(v_t) I(q_t \leq \gamma_0) + h_2(v_t) I(q_t > \gamma_0) + \varepsilon_t, \tag{2.4}$$

where $\varepsilon_t = \varepsilon_{1t}I(q_t \leq \gamma_0) + \varepsilon_{2t}I(q_t > \gamma_0)$ with $\varepsilon_{jt} = u_{jt} - h_j(v_t)$ and $E\left(h_j(v_t)\right) = 0$ as $E(u_{jt}) = 0$ for $j = 1, 2$. Model (2.4) becomes a partially linear regression model with

https://doi.org/10.1017/S0266466621000256 Published online by Cambridge University Press

exogenous threshold effects.[3] In the presence of threshold variable endogeneity, the unknown control functions, $h_1(\cdot)$ and $h_2(\cdot)$, as the endogenous threshold bias correction terms, will be nonzero over at least one nonempty interval. If the errors are homogeneous across regimes in models (2.1) and (2.2), $h_1(\cdot)$ and $h_2(\cdot)$ will be the same. If the threshold variable $q_t$ is exogenous, we have $h_1(\cdot) \equiv h_2(\cdot) \equiv 0$ and model (2.4) is the standard linear threshold regression. Yu et al. (2020) consider the parametric case that $h_1(\cdot)$ and $h_2(\cdot)$ are known functions up to a finite number of unknown parameters, while we extend their work to a semiparametric case with unknown $h_1(\cdot)$ and $h_2(\cdot)$.

Denoting $\delta = \beta_1 - \beta_2$ and $\eta(\cdot) = h_1(\cdot) - h_2(\cdot)$, we can rewrite (2.4) as

$$y_t = \mathbf{x}_t'\beta_2 + \mathbf{x}_t'\delta I(q_t \le \gamma_0) + h_2(v_t) + \eta(v_t) I(q_t \le \gamma_0) + \varepsilon_t. \tag{2.5}$$

As in Hansen (2000), our asymptotic results are derived in the framework of "small threshold" effects; i.e., we assume that the threshold effects, $\delta = \delta_n$ and $\eta(\cdot) = \eta_n(\cdot)$, both approach to zero slowly as $n$ diverges. It means that regime changes temporally exist around a threshold parameter. Below, we summarize the assumptions that support our model (2.5).

**Assumption 1.**

(i) $\left\{\left(\mathbf{x}_t', q_t, \mathbf{z}_t', u_t\right)\right\}$ is a strictly stationary, ergodic, and $\rho$-mixing sequence with $\rho$-mixing coefficients $\{\rho_m\}$ satisfying $\sum_m^{\infty} \rho_m^{1/2} < \infty$, $y_t$ is generated from model (2.1) and (2.2), and $\gamma_0$ is an interior point of a bounded set $[\underline{\gamma}, \overline{\gamma}]$.

(ii) $E\left(\|\mathbf{z}_t v_t\|^4\right) < \infty$, $E\left(\|\mathbf{z}_t\|^4\right) < \infty$ and $E\left(\mathbf{z}_t\mathbf{z}_t'\right)$ is nonsingular.

(iii) $E\left(v_t|\mathcal{F}_{n,t-1}\right) = 0$, $E\left(u_{jt}|\mathcal{F}_{n,t-1}\right) = 0$, $E\left(u_{jt}^2|\mathcal{F}_{n,t-1}\right) = \sigma_j^2 > 0$, and $E(u_{jt}|\mathcal{F}_{n,t-1}, v_t = v) = h_j(v)$ for any for $j = 1, 2$ and any $v$.

(iv) For any $\lambda \neq 0$, there is no measurable function $m(\cdot)$ such that $\mathbf{x}'\lambda = m(v)$.

(v) $\eta_n(\cdot) = n^{-\varrho}\eta_0(\cdot)$ and $\delta_n = n^{-\varsigma}\delta_0$ for some $0 < \varsigma, \varrho < 1/2$, $\delta_0 \neq 0$, and $\eta_0(\cdot) \neq 0$ over at least one nonempty interval.

Assumption 1(i) allows for weakly dependent data but excludes time trends and integrated processes; see the discussion given in Hansen (2000). Assumption 1(ii) contains unconditional moment conditions. Assumption 1(iii) defines threshold variable endogeneity and ensures the specification of model (2.4). Applying the law of iterated expectation gives $E\left(u_{jt}|\mathcal{F}_{n,t-1}\right) = E\left[E\left(u_{jt}|\mathcal{F}_{n,t-1}, v_t\right)|\mathcal{F}_{n,t-1}\right] = E\left[h_j(v_t)|\mathcal{F}_{n,t-1}\right]$. Hence, assuming that $v_t$ is independent of $\mathcal{F}_{n,t-1}$ is sufficient to ensure $E\left(u_{jt}|\mathcal{F}_{n,t-1}\right) = 0$ in Assumption 1(iii). 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 limiting distribution of our proposed estimator implicitly incur unbounded moments beyond the second order, we therefore assume that $\left\{\left(\varepsilon_t, \mathcal{F}_{n,t}^*\right)\right\}_{t=1}^n$ is a martingale difference sequence ensured under Assumption 1(iii), where $\mathcal{F}_{n,t}^*$ is the smallest sigma-field generated from $\left\{\left(\mathbf{x}_{s+1}', \mathbf{z}_{s+1}', q_{s+1}, \varepsilon_s\right) : 1 \le s \le t \le n\right\}$.

[3] We thank an anonymous referee for pointing this out to us.

https://doi.org/10.1017/S0266466621000256 Published online by Cambridge University Press

Assumption 1(iv) is an identification condition similar to Assumption 2.1 in Newey (2009).

Assumption 1(v) allows for the threshold effects to vanish at a different rate as the sample size increases. The reason for the diminishing threshold effects assumption is that in the case of fixed threshold effects the limiting distribution of the estimator for the threshold parameter is too complicated for inference purpose. Specifically, Chan (1993) shows that the asymptotic distribution of the threshold parameter estimator depends on a host of nuisance parameters, including the marginal distribution of the regressors. To overcome this difficulty, Hansen (2000) considers the case where 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 estimator of the threshold parameter based on the simulation of a related compound Poisson process. An alternative approach is to introduce smoothness in the objective criterion to achieve asymptotic normality but with a slower convergence rate in the threshold parameter estimation. For example, Seo and Linton (2007) propose a smoothed least-squares estimation strategy that leads to an asymptotically normally distributed estimator of the threshold parameter by replacing the indicator function in the objective function with an integrated kernel. Seo and Shin (2016) exploit the smoothness of the GMM criterion to achieve the asymptotic normality of the estimator for the threshold parameter, regardless of whether the threshold effects are fixed or diminishing.[4]

## 2.1. Estimation

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

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

respectively, where $\Phi_{L_n}(v) = \left[\phi_1(v), \ldots, \phi_{L_n}(v)\right]'$ denotes an $L_n \times 1$ vector.[5] As $\eta_n(v_t) = h_1(v_t) - h_2(v_t) = n^{-\varrho}\eta_0(v_t)$, we have $n^{-\varrho}\alpha_{L_n,0} = \alpha_{L_n,1} - \alpha_{L_n,2}$ and $\eta_n(v_t)$ is approximated by $\eta_n^*(v_t) = n^{-\varrho}\alpha'_{L_n,0}\Phi_{L_n}(v_t)$. Below, we explain our proposed estimation procedure.

**Step 1.** Given instruments $\mathbf{z}_t$, the OLS estimator from model (2.3) is, $\hat{\pi}_q = \left(\sum_{t=1}^n \mathbf{z}_t\mathbf{z}_t'\right)^{-1}\sum_{t=1}^n \mathbf{z}_t q_t$. Assumptions 1(i)–(iii) imply that $\hat{\pi}_q$ exists and ensure

[4]The control function approach used in the paper relies on the conditional moment restrictions defined in Assumption 1(iii) to identify the unknown parameters in models (2.1) and (2.2). In contrast, the GMM estimation used in Seo and Shin (2016) employs unconditional orthogonal moment conditions between errors and instruments to identify unknown parameters.

[5]We suppress the dependence of $h_j^*(\cdot)$ and $\eta_0^*(\cdot)$ on $L_n$ to simplify the notation.

https://doi.org/10.1017/S0266466621000256 Published online by Cambridge University Press

consistency with $\hat{\pi}_q = \pi_q + O_p\left(n^{-1/2}\right)$. We then denote the residuals as $\hat{v}_t = q_t - \mathbf{z}_t'\hat{\pi}_q$ for all $t$ throughout the rest of this paper.

**Step 2.** For a given $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, we estimate $\theta = \left(\beta_1', \alpha_{L_n,1}', \beta_2', \alpha_{L_n,2}'\right)'$ from the objective function

$$\hat{\theta}(\gamma) = \arg\min_{\theta} \sum_{t=1}^{n} \Big[ y_t - \beta_1' \mathbf{x}_t I(q_t \le \gamma) - \alpha_{L_n,1}' \Phi_{L_n,\gamma}^{-}\left(\hat{v}_t\right) - \beta_2' \mathbf{x}_t I(q_t > \gamma) - \alpha_{L_n,2}' \Phi_{L_n}^{+}\left(\hat{v}_t\right) \Big]^2, \tag{2.6}$$

where we denote $\Phi_{L_n,\gamma}^{-}\left(\hat{v}_t\right) = \Phi_{L_n}\left(\hat{v}_t\right) I(q_t \le \gamma)$ and $\Phi_{L_n,\gamma}^{+}\left(\hat{v}_t\right) = \Phi_{L_n}\left(\hat{v}_t\right) I(q_t > \gamma)$. Denoting an $n \times [2(d_{\mathbf{x}} + L_n)]$ matrix $\mathbf{X}_\gamma = \left[\mathbf{X}_{-,\gamma}, \mathbf{X}_{+,\gamma}\right]$, where $\mathbf{X}_{-,\gamma}$ stacks up $\left[\mathbf{x}_t' I(q_t \le \gamma), \Phi_{L_n,\gamma}^{-\prime}\left(\hat{v}_t\right)\right]$ and $\mathbf{X}_{+,\gamma}$ stacks up $\left[\mathbf{x}_t' I(q_t > \gamma), \Phi_{L_n,\gamma}^{+\prime}\left(\hat{v}_t\right)\right]$, and solving (2.6) gives

$$\hat{\theta}(\gamma) = \left(\mathbf{X}_\gamma' \mathbf{X}_\gamma\right)^{-1} \mathbf{X}_\gamma' \mathbf{y}, \tag{2.7}$$

where $\mathbf{y} = [y_1, \ldots, y_n]'$. We can estimate the threshold parameter $\gamma_0$ by minimizing the concentrated-squares criterion

$$\widehat{\gamma} = \arg \min_{\gamma \in \left[\underline{\gamma}, \overline{\gamma}\right]} \sum_{t=1}^{n} \left[ y_t - \mathbf{x}_{t,\gamma}' \hat{\theta}(\gamma) \right]^2 \tag{2.8}$$

and then estimate $\theta$ by $\hat{\theta} = \hat{\theta}\left(\hat{\gamma}\right)$, where $\mathbf{x}_{t,\gamma}$ in (2.8) is the $t^{th}$ row vector of $\mathbf{X}_\gamma$.

**Step 3**. Calculating $\tilde{y}_t = y_t - \hat{\beta}_1' \mathbf{x}_t I(q_t \le \hat{\gamma}) - \hat{\beta}_2' \mathbf{x}_t I(q_t > \hat{\gamma})$, we can re-estimate $h_2(\cdot)$ and $\eta_n(\cdot)$ at an interior point $v$ by the local linear regression approach from $\tilde{y}_t = h_2(v_t) + \eta_n(v_t) I(q_t \le \hat{\gamma}) + \tilde{\varepsilon}_t$, $t = 1, 2, \ldots, n$.

Of course, one can obtain the series estimator for $h_1(\cdot)$, $h_2(\cdot)$, 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, and as such, we add Step 3 in the paper. Moreover, if one is interested in estimating a partially linear regression model with an exogenous threshold variable, $q_t$, and an observable exogenous regressor, $v_t$, defined in equation (2.4), ignoring Step 1 and applying Step 2 with $\hat{v}_t$ being replaced with $v_t$ can be used to estimate all the slope parameters and threshold parameter.

For independent data, Lee, Seo, and Shin (2016) study the oracle efficiency property of the lasso estimation for high-dimensional regression with possible fixed threshold effects, where the slope regressors and threshold variable are all exogenous. After replacing the unknown bias correction terms with series approximation, model (2.4) is approximated by a linear threshold regression with an exogenous threshold variable, but with an increasing number of parameters $\theta$, which fits into Lee et al. (2016) framework. Adding a penalty term similar to what is used in Lee et al. (2016) to the objective function (2.6) allows for a lasso

https://doi.org/10.1017/S0266466621000256 Published online by Cambridge University Press

estimation of $\theta$. However, pursuing this penalty estimation method is beyond the scope of this paper. We leave it for our future research.

## 2.2. Limiting Results

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

**Assumption 2.**

(i) $E\|\mathbf{x}_t\|^4 < \infty$, $E\|\varepsilon_t\mathbf{x}_t\|^4 < \infty$, $\max_{\gamma\in[\underline{\gamma},\overline{\gamma}]}E(\|\mathbf{x}_t\|^4|q_t=\gamma) < \infty$, and $\max_{\gamma\in[\underline{\gamma},\overline{\gamma}]}E(\|\mathbf{x}_t\|^4\varepsilon_t^4|q_t=\gamma) < \infty$, where $||\cdot||$ denotes the Euclidean norm;

(ii) for every $L_n$, there exist constants $\underline{c}$ and $\overline{c}$ such that $0 < \underline{c} \leq \lambda_{\min}\left(\Sigma_\gamma\right) \leq \lambda_{\max}\left(\Sigma_\gamma\right) \leq \overline{c} < \infty$ over $\gamma \in [\underline{\gamma},\overline{\gamma}]$, where $\Sigma_\gamma$ can take the form of either $E\left(\mathbf{x}^*_{t,\gamma}\mathbf{x}^{*\prime}_{t,\gamma}\right)$ or $E\left(\varepsilon_t^2\mathbf{x}^*_{t,\gamma}\mathbf{x}^{*\prime}_{t,\gamma}\right)$, and $\mathbf{x}^*_{t,\gamma}$ equals $\mathbf{x}_{t,\gamma}$ with $\hat{v}_t$ being replaced with $v_t$;

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

(iv) $h_2(\cdot)$ and $\eta_0(\cdot)$ are squared integrable, and there exist $\alpha_{L_n,2}$ and $\alpha_{L_n,0}$ such that $\sup_{v\in\mathcal{R}}\left|h_2(v)-\alpha'_{L_n,2}\Phi_{L_n}(v)\right| \leq ML_n^{-\xi}$ and $\sup_{v\in\mathcal{R}}\left|\eta_0(v)-\alpha'_{L_n,0}\Phi_{L_n}(v)\right| \leq ML_n^{-\xi}$;

(v) $\{\phi_l(\cdot), l=1,2,...\}$ is a sequence of orthonormal basis functions in $\Lambda^\xi(\mathcal{R})$ and uniformly bounded over $\mathcal{R}$;

(vi) $E\left(\chi_t^*\chi_t^{*\prime}|q_t=\gamma\right)$ and $E\left(\varepsilon_{jt}^2\chi_t^*\chi_t^{*\prime}|q_t=\gamma\right)$ are continuous at $\gamma=\gamma_0$ for $j=1,2$, where $\chi_t^* = \left[\mathbf{x}_t', \Phi'_{L_n}(v_t)\right]'$.

Assumption 2(i) contains unconditional and conditional moment conditions. Assumption 2(ii) ensures the existence of $\hat{\theta}(\gamma)$, which is standard in the literature; see, e.g., Newey (1997) and Delgado et al. (2020). 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 for estimating nonparametric models. Assumption 2(iv) restricts the sieve approximation error, and it holds by Theorem 1.1 in Dzyadyk and Shevchuk (2008) if $v_t$ has compact support and $\eta_0(\cdot)$ and $h_j(\cdot)$ are all $\xi$-smooth. If $v_t$ has unbounded support, $(-\infty,\infty)$, Xiang (2012) showed that Assumption 2(iv) holds for the normalized Hermite orthonormal basis functions if $\eta_0(\cdot)$ and $h_2(\cdot)$ are all $p$-smooth for some $p > 2(\xi+1)$. Let $\sup_{j\leq s}\sup_{v\in\mathcal{R}}\sum_{l=1}^{L_n}\left[\phi_l^{(j)}(v)\right]^2 = \left\|\Phi_{L_n}\right\|_s^2$ for $s \geq 0$, where $\phi_l^{(j)}(\cdot)$ is the $j$th derivative of $\phi_l(\cdot)$. Assumption 2(v) describes the properties of the basis functions and implies

https://doi.org/10.1017/S0266466621000256 Published online by Cambridge University Press

$\left\|\Phi_{L_n}\right\|_0 = O(L_n^{1/2})$ and $\left\|\Phi_{L_n}\right\|_1 = O(L_n^{3/2})$; see, e.g., the normalized Hermite functions and wavelet functions defined in Blundell et al. (2007). Assumption 2(vi) implies that $E\left(\chi_t^*\chi_t^{*\prime}|q_t=\gamma\right)$ and $E\left(\varepsilon_{jt}^2\chi_t^*\chi_t^{*\prime}|q_t=\gamma\right)$ are well defined in the neighborhood of $\gamma_0$ for $j=1,2$.

**Assumption 3.** Denote $\vartheta_n = L_n^{-\xi} + \sqrt{L_n/n}$. (i) $\vartheta_n = o(1)$ and $\left\|\Phi_{L_n}\right\|_1^2 L_n/n = o(1)$; (ii) $n^{\min(\varsigma,\varrho)}\vartheta_n = o(1)$; and (iii) $\sqrt{n}L_n^{-\xi} = o(1)$.

Assumption 3(i) is used to derive the consistency result of $\widehat{\gamma}$ and $\hat{\theta}$ in Lemma 1. Assumption 3(ii) is imposed to remove the asymptotic impact of the estimation of the unknown parameter $\theta$ on the limiting result of $\hat{\gamma}$ in Theorem 1; see the discussion given below Lemma 1. Assumptions 3(iii) is used to remove the asymptotic impact of estimating the unknown threshold parameter, $\gamma_0$, on the limiting distribution for $\hat{\beta}$ in Theorem 3.

Let $L_n = [c_0 n^r]$ for some constant $c_0 > 0$ and $r \in (0,1)$. Then, Assumption 3(i) implies $r \in (0,1/4)$, and Assumptions 3(i)–(iii) imply $r \in (1/(2\xi), \min(1/4, 1-2\min(\varsigma,\varrho)))$ for some $\xi > 2$. It follows that $\min(\varsigma,\varrho) < 1/2 - 1/(4\xi)$. Thus, the semiparametric threshold model requires a relatively slower diminishing threshold effects than a parametric threshold model, although this requirement is nonrestrictive for sufficiently smooth endogenous bias correction terms. Below, we give the limiting results for $\widehat{\gamma}$ and $\hat{\theta}$.

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

Lemma 1 shows the consistency of $\widehat{\gamma}$ and $\hat{\theta}$, where $\vartheta_n$ is the conventional convergence rate of the series estimator and is not affected by the parametric estimation of the unknown errors, $v_t$, from the reduced-form model of the endogenous threshold variable, $q_t$.

To remove the asymptotic impact of the estimation of the unknown parameter $\theta$ on the limiting result of $\hat{\gamma}$ in Theorem 1, we need to show that $n^{\min(\varsigma,\varrho)}\left\|\hat{\theta} - \theta\right\| = o_p(1)$ in parallel to Lemma A.6 in Hansen (2000). This leads to Assumption 3(ii) by the result given in Lemma 1. This assumption is necessary for the estimation of model (2.4) in the presence of unknown curves, but it is not required for the estimation of a parametric model as in Hansen (2000).

2.2.1. *Threshold parameter.* We begin by deriving the asymptotic distribution of the threshold estimator $\widehat{\gamma}$.

THEOREM 1. *Under Assumptions 1–3(i)(ii), we have*

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

https://doi.org/10.1017/S0266466621000256 Published online by Cambridge University Press

*where we denote* $\tau_j^2 = E\left\{\varepsilon_{jt}^2\left[I(\varsigma \le \varrho)\,\delta_0'\mathbf{x}_t + I(\varsigma \ge \varrho)\,\eta_0(v_t)\right]^2 | q_t = \gamma_0\right\}$ *for* $j = 1, 2$, $\omega_0 = E\left\{\left[I(\varsigma \le \varrho)\,\delta_0'\mathbf{x}_t + I(\varsigma \ge \varrho)\,\eta_0(v_t)\right]^2 | q_t = \gamma_0\right\}$,

$$\varpi = \frac{\tau_1^2}{\omega_0^2 f_q(\gamma_0)} \text{ and } T = \underset{-\infty<r<\infty}{argmax}\, T(r).$$

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

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

*where* $\varphi = \tau_2^2/\tau_1^2$ *and* $W_1(r)$ *and* $W_2(r)$ *are two independent standard Brownian motion processes defined on* $[0, \infty)$.

Theorem 1 shows that the asymptotic distribution of the threshold estimator is a Chernoff-type distribution featuring unequal scales for each regime. Although the definition of $\tau_j^2$ and $\omega_0$ depend on the diminishing speed of the threshold effects, we do not need to know $\varsigma$ and $\varrho$ for inference purposes as explained below Theorem 2.

Making inference on the threshold parameter is generally difficult. Following Hansen (2000) we obtain the confidence interval of $\gamma_0$ by inverting a likelihood ratio test statistic under the very restrictive assumption of normally distributed errors. The likelihood ratio test of the null hypothesis $H_0 : \gamma = \gamma_0$ is defined as

$$LR_n(\gamma) = n\frac{S_n(\gamma) - S_n(\hat{\gamma})}{S_n(\hat{\gamma})},$$

where $S_n(\gamma) = \sum_{t=1}^n \left[y_t - \mathbf{x}_{t,\gamma}'\hat{\theta}(\gamma)\right]^2$. Then, the following theorem provides the asymptotic distribution of $LR_n(\gamma_0)$.

THEOREM 2. *Under* $H_0 : \gamma = \gamma_0$ *and Assumptions 1–3(i)–(ii), we have*

$$LR_n(\gamma_0) \xrightarrow{d} \eta^2 \zeta,$$

*where*

$$\eta^2 = \frac{\tau_1^2}{\bar{\sigma}_\varepsilon^2 \mu},$$

*and*

$$\zeta = \sup_s [-|s| + 2W_1(s)] I(s \le 0) + \left[-|s| + 2\sqrt{\varphi} W_2(s)\right] I(s > 0)$$

*with* $\mu = E\left\{\left[I(\varsigma \le \varrho)\,\delta_0'\mathbf{x}_t + I(\varsigma \ge \varrho)\,\eta_0(v_t)\right]^2 | q_t = \gamma_0\right\} f_q(\gamma_0)$ *and* $\bar{\sigma}_\varepsilon^2 = \lim_{n\to\infty} n^{-1} \sum_{t=1}^n E\left(\varepsilon_t^2\right)$. *Also, the distribution of* $\zeta$ *is* $\Pr(\zeta \le x) = \left(1 - e^{-x/2}\right)\left(1 - e^{-x/(2\varphi)}\right)$.

https://doi.org/10.1017/S0266466621000256 Published online by Cambridge University Press

Theorem 2 states that the asymptotic $p$-values for observed test statistics are given as follows

$$p_n = p_n(\varphi, \eta^2) = 1 - \left[1 - exp\left(-LR_n/(2\eta^2)\right)\right]\left[1 - exp\left(LR_n/(2\varphi\eta^2)\right)\right].$$

Then, the null hypothesis is rejected asymptotically at the significance level $\alpha$ if $p_n < \alpha$ for some $\alpha \in (0,1)$. To obtain a feasible test statistic, we need to obtain a consistent estimator for the nuisance parameters $\varphi$ and $\eta^2$. Note that we have $\tau_j^2 = n^{2min(\varsigma,\varrho)} E\left\{\varepsilon_{jt}^2\left[\delta_n'\mathbf{x}_t + \eta_n(v_t)\right]^2 | q_t = \gamma_0\right\}$ for $j = 1, 2$ by Theorem 1 and $\mu = n^{2min(\varsigma,\varrho)} E\left\{\left[\delta_n'\mathbf{x}_t + \eta_n(v_t)\right]^2 | q_t = \gamma_0\right\} f_q(\gamma_0)$ by Theorem 2. Hence, $n^{2min(\varsigma,\varrho)}$ is cancelled out from the formula of $\eta^2$ and $\varphi$, and $\eta^2$ and $\varphi$ can be consistently estimated once we consistently estimate $\delta_n$ and $\eta_n(\cdot)$ without the knowledge of $\varsigma$ and $\varrho$.

Specifically, we have

$$\varphi = \frac{b_2(\gamma_0)}{b_1(\gamma_0)} = \frac{E\left\{\varepsilon_{2t}^2\left[\delta_n'\mathbf{x}_t + \eta_n(v_t)\right]^2 | q_t = \gamma_0\right\}}{E\left\{\varepsilon_{1t}^2\left[\delta_n'\mathbf{x}_t + \eta_n(v_t)\right]^2 | q_t = \gamma_0\right\}}$$

and

$$\eta^2 = \frac{b_1(\gamma_0)}{\bar{\sigma}_\varepsilon^2 b_3(\gamma_0)} = \frac{E\left\{\varepsilon_{1t}^2\left[\delta_n'\mathbf{x}_t + \eta_n(v_t)\right]^2 | q_t = \gamma_0\right\}}{\bar{\sigma}_\varepsilon^2 E\left\{\left[\delta_n'\mathbf{x}_t + \eta_n(v_t)\right]^2 | q_t = \gamma_0\right\} f_q(\gamma_0)},$$

where $\bar{\sigma}_\varepsilon^2$ can be estimated by $\hat{\sigma}_\varepsilon^2 = n^{-1}\sum_{t=1}^n \hat{\varepsilon}_t^2$ and $\hat{\varepsilon}_t = y_t - \mathbf{x}_{t,\hat{\gamma}}'\hat{\theta}$ for $t = 1, \ldots, n$.

Next, replacing $\varepsilon_{1t} = \varepsilon_t I(q_t \le \gamma_0)$, $\varepsilon_{2t} = \varepsilon_t I(q_t > \gamma_0)$, $\delta_n$, and $\eta_n(v_t)$ in $b_j(\gamma_0)$ with $\hat{\varepsilon}_{1t} = \hat{\varepsilon}_t I\left(q_t \le \hat{\gamma}\right)$, $\hat{\varepsilon}_{2t} = \hat{\varepsilon}_t I\left(q_t > \hat{\gamma}\right)$, $\hat{\delta}_n$, and $\hat{\eta}_n\left(\hat{v}_t\right)$, respectively, we obtain $b_j(\hat{\gamma})$ for $j = 1, 2, 3$. We then estimate the conditional expectations $b_j\left(\hat{\gamma}\right)$ by kernel method as suggested in Hansen (2000) from which we obtain the consistent estimates of the nuisance parameters $\eta^2$ and $\varphi$. The null hypothesis is rejected at the significance level $\alpha$ if $p_n(\hat{\varphi}, \hat{\eta}^2) < \alpha$, and the $100(1-\alpha)\%$ confidence interval for $\gamma_0$ is given by $\widehat{\Gamma} = \{\gamma : LR_n(\gamma) \le c_\zeta(1-\alpha, \hat{\varphi}, \hat{\eta}^2)\}$, where $c_\zeta(1-\alpha, \varphi, \eta^2)$ equals the $LR_n$ value that solves $p_n(\varphi, \eta^2) = \alpha$. The asymptotic validity of the confidence interval results from $P(\gamma_0 \in \widehat{\Gamma}) \to 1-\alpha$ under $H_0$ by Theorem 2. Note, however, that the confidence intervals are asymptotically valid under the assumption of shrinking threshold effects. This suggests that the actual coverage may differ from the desired level for large values of $\delta_n$ and $\eta_n(\cdot)$.[6]

Our proposed method can also be applied to the case where both $q_t$ and some components of $\mathbf{x}_t$ are endogenous. Suppose that $\mathbf{x}_{1,t}$ and $\mathbf{x}_{2,t}$ are the endogenous and exogenous components of $\mathbf{x}_t = \left[\mathbf{x}_{1,t}', \mathbf{x}_{2,t}'\right]'$. Let $\mathbf{r}_t = \left[\mathbf{x}_{1,t}', q_t\right]'$ contain all the

[6] Under additional restrictive assumptions as in (Hansen (2000), Thm. 3) one can show that these confidence intervals are also asymptotically valid in the case of fixed threshold effects.

https://doi.org/10.1017/S0266466621000256 Published online by Cambridge University Press

endogenous variables, and $\mathbf{z}_t$ contains all the instrumental variables and is allowed to include $\mathbf{x}_{2,t}$. Then, we extend model (2.3) to

$$\mathbf{r}_t = \pi \mathbf{z}_t + \mathbf{v}_t,\ t = 1, 2, \ldots, n, \tag{2.9}$$

where $\mathbf{z}_t$ is a $d_{\mathbf{z}} \times 1$ vector of instrumental variables for $\mathbf{r}_t$ satisfying $E(\mathbf{v}_t \mathbf{z}_t') = \mathbf{0}$ for all $t$. The endogeneity of variable $\mathbf{r}_t$ comes from the contemporaneous correlation between $u_{jt}$ and $\mathbf{v}_t$ for $j = 1, 2$.

Assuming that $E(u_{jt}|\mathcal{F}_{n,t-1}, \mathbf{r}_t) = E(u_{jt}|\mathbf{v}_t) = h_j(\mathbf{v}_t)$ holds almost surely and $h_j(\cdot)$ is nonzero over at least one nonempty interval for $j = 1, 2$, where $\mathcal{F}_{n,t}$ is the smallest sigma-field generated from $\left\{\left(\mathbf{x}_{1,s}', \mathbf{x}_{2,s+1}', \mathbf{z}_{s+1}', q_s, u_s\right) : 1 \le s \le t \le n\right\}$, we have

$$\begin{aligned} E\left(y_t|\mathcal{F}_{n,t-1}, \mathbf{r}_t\right) = & \left[\mathbf{z}_t' \pi_{\mathbf{x}}' \beta_{1,1} + \mathbf{x}_{2,t}' \beta_{1,2} + h_1(\mathbf{v}_t)\right] I(q_t \le \gamma_0) + \\ & \left[\mathbf{z}_t' \pi_q' \beta_{2,1} + \mathbf{x}_{2,t}' \beta_{2,2} + h_2(\mathbf{v}_t)\right] I(q_t > \gamma_0), \end{aligned} \tag{2.10}$$

where $\beta_j = \left[\beta_{1,j}', \beta_{2,j}'\right]'$ are conformable with the decomposition of $\mathbf{x}_t = \left[\mathbf{x}_{1,t}', \mathbf{x}_{2,t}'\right]'$ for $j = 1, 2$, and $\pi = \left[\pi_{\mathbf{x}}', \pi_q'\right]'$ is conformable with the decomposition of $\mathbf{r}_t = \left[\mathbf{x}_{1,t}', q_t\right]'$. Once we estimate $\pi$ by the OLS estimator, the rest of the estimation procedure is the same as that presented in Section 2.1. We will not give a further detailed discussion as it will not shed additional light on our methodology.

2.2.2. *Slope parameters.* Let $\hat{\beta} = \left(\hat{\beta}_1', \hat{\beta}_2'\right)'$ and $\beta = \left(\beta_1', \beta_2'\right)'$, and we define the following moments

$$\Sigma_{\mathbf{x}\mathbf{x}', \gamma_0} \equiv \begin{bmatrix} E\left[\mathbf{x}_t \mathbf{x}_t' I(q_t \le \gamma_0)\right] & \mathbf{0} \\ \mathbf{0}' & E\left[\mathbf{x}_t \mathbf{x}_t' I(q_t > \gamma_0)\right] \end{bmatrix},$$

$$\Sigma_{\mathbf{x}\Phi_{L_n}', \gamma_0} \equiv \begin{bmatrix} E\left[\mathbf{x}_t \Phi_{L_n}(v_t)' I(q_t \le \gamma_0)\right] & \mathbf{0} \\ \mathbf{0}' & E\left[\mathbf{x}_t \Phi_{L_n}(v_t)' I(q_t > \gamma_0)\right] \end{bmatrix},$$

$$\Sigma_{\Phi_{L_n}\Phi_{L_n}', \gamma_0} \equiv \begin{bmatrix} E\left[\Phi_{L_n}(v_t) \Phi_{L_n}(v_t)' I(q_t \le \gamma_0)\right] & \mathbf{0} \\ \mathbf{0}' & E\left[\Phi_{L_n}(v_t) \Phi_{L_n}'(v_t) I(q_t > \gamma_0)\right] \end{bmatrix},$$

$$\Gamma_{\mathbf{x},1} \equiv E\left[\mathbf{x}_t h_1^{(1)}(v_t) \mathbf{z}_t' I(q_t \le \gamma_0)\right],$$

$$\Gamma_{\mathbf{x},2} \equiv E\left[\mathbf{x}_t h_2^{(1)}(v_t) \mathbf{z}_t' I(q_t > \gamma_0)\right],$$

$$\Gamma_{\Phi_{L_n},1} \equiv E\left[\Phi_{L_n}(v_t) h_1^{(1)}(v_t) \mathbf{z}_t' I(q_t \le \gamma_0)\right],$$

$$\Gamma_{\Phi_{L_n},2} \equiv E\left[\Phi_{L_n}(v_t) h_2^{(1)}(v_t) \mathbf{z}_t' I(q_t > \gamma_0)\right],$$

$$\Omega_{11} \equiv \lim_{n\to\infty} Var\left(n^{-1/2} \sum_{t=1}^{n} \mathbf{z}_t v_t\right),$$

https://doi.org/10.1017/S0266466621000256 Published online by Cambridge University Press

$$\Omega_{22} \equiv \lim_{n\to\infty} n^{-1} Var\left(\left(\mathbf{X}_0 - \Sigma_{\mathbf{x}\Phi'_{L_n},\gamma_0}\Sigma^{-1}_{\Phi_{L_n}\Phi'_{L_n},\gamma_0}\Delta_0\right)'\varepsilon\right),$$

$$\Omega_{12} \equiv \lim_{n\to\infty} n^{-1}\sum_{t=2}^{n}\sum_{s=1}^{t-1} E\left(\mathbf{z}_t\varphi'_s v_t\varepsilon_s\right),$$

where $\varepsilon = (\varepsilon_1, ..., \varepsilon_n)'$, $h_j^{(1)}(\cdot)$ is the first-order derivative of $h_j(\cdot)$, $\Omega_{11}$, $\Omega_{22}$, and $\Omega_{12}$ are all finite matrices under Assumptions 1 and 2, the $t$th row of $\mathbf{X}_0$ and $\Delta_0$ are $\left[\mathbf{x}'_t I(q_t \le \gamma_0), \mathbf{x}'_t I(q_t > \gamma_0)\right]$ and $\left[\Phi_{L_n}(v_t)' I(q_t \le \gamma_0), \Phi_{L_n}(v_t)' I(q_t > \gamma_0)\right]$, respectively, and $\varphi_t$ is the $t$th row vector of $\mathbf{X}_0 - \Sigma_{\mathbf{x}'\Phi_{L_n},\gamma_0}\Sigma^{-1}_{\Phi_{L_n}\Phi'_{L_n},\gamma_0}\Delta_0$.

Then, the following theorem provides the asymptotic distribution of $\hat{\beta}$.

THEOREM 3. *Under Assumptions 1–3, we have*

$$\sqrt{n}(\mathbf{J}_n^{-1}\Omega_n\mathbf{J}_n^{-1})^{-1/2}\left(\hat{\beta} - \beta\right) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{I}_{2d_{\mathbf{x}}}\right), \tag{2.11}$$

*where the nonsingular matrix* $\mathbf{J}_n$ *and* $\Omega_n$ *are defined by*

$$\mathbf{J}_n \equiv \Sigma_{\mathbf{x}\mathbf{x}',\gamma_0} - \Sigma_{\mathbf{x}\Phi'_{L_n},\gamma_0}\Sigma^{-1}_{\Phi_{L_n}\Phi'_{L_n},\gamma_0}\Sigma_{\Phi'_{L_n}\mathbf{x},\gamma_0} \tag{2.12}$$

*and*

$$\Omega_n \equiv \mathbf{B}_n\left[E\left(\mathbf{z}_1\mathbf{z}'_1\right)\right]^{-1}\Omega_{11}\left[E\left(\mathbf{z}_1\mathbf{z}'_1\right)\right]^{-1}\mathbf{B}'_n - \Omega'_{12}\left[E\left(\mathbf{z}_1\mathbf{z}'_1\right)\right]^{-1}\mathbf{B}'_n$$
$$-\mathbf{B}_n\left[E\left(\mathbf{z}_1\mathbf{z}'_1\right)\right]^{-1}\Omega_{12} + \Omega_{22} \tag{2.13}$$

*with* $\mathbf{B}_n \equiv \left[\Gamma'_{\mathbf{x},1}, \Gamma'_{\mathbf{x},2}\right]' - \Sigma_{\mathbf{x}\Phi'_{L_n},\gamma_0}\Sigma^{-1}_{\Phi_{L_n}\Phi'_{L_n},\gamma_0}\left[\Gamma'_{\Phi_{L_n},1}, \Gamma'_{\Phi_{L_n},2}\right]'$.

Theorem 3 shows that the estimator of $\beta$ is root-n consistent and asymptotically normally distributed, where $\mathbf{J}_n$ is nonsingular under Assumption 2(ii). Statistical inference on $\hat{\beta}$ is obtained using consistent estimators of $\mathbf{J}_n$ and $\Omega_n$. Specifically, by (2.12), we can estimate $\mathbf{J}_n$ by $\hat{\mathbf{J}}_n$, which is obtained from $\mathbf{J}_n$ by replacing $E(\cdot)$ with their corresponding sample analogs, $\gamma_0$ with $\hat{\gamma}$, and $\pi_q$ with $\hat{\pi}_q$. Likewise, by (2.13), we can estimate $\Omega_n$ by $\hat{\Omega}_n$. Specifically, replacing $E(\cdot)$ in $\mathbf{B}_n$ with their corresponding sample analogs and all unknown parameters and unknown curves in $\mathbf{B}_n$ by their estimates gives $\hat{\mathbf{B}}_n$. For the calculation of $\hat{\Omega}_n$, we first replace $E(\cdot)$, $\gamma_0$ and $(\varepsilon_t, v_t)$ in $\begin{pmatrix} n^{-1/2}\sum_{t=1}^{n}\mathbf{z}_t v_t \\ n^{-1/2}\left(\mathbf{X}_0 - \sum_{\mathbf{x}\Phi'_{L_n},\gamma_0}\sum^{-1}_{\Phi_{L_n}\Phi'_{L_n},\gamma_0}\Delta_0\right)'\varepsilon \end{pmatrix}$ by the sample analogs, $\hat{\gamma}$ and $\left(\hat{\varepsilon}_t, \hat{v}_t\right)$, respectively, and then calculate the sample variance-covariance matrix using a HAC-type estimator. Despite the tedious calculations the standard error of $\hat{\beta}$ is then readily given.

2.2.3. *Control functions.* The following conditions are required for the derivation of the limiting distribution of the local linear estimator of $\eta_n(\cdot)$ and $h_2(\cdot)$ explained in Step 3.

https://doi.org/10.1017/S0266466621000256 Published online by Cambridge University Press

**Assumption 4.** (i) $f_v(v)$, $E[I(q_t \le \gamma_0)|v_t = v]$, and $E\left(\varepsilon_t^2 \varkappa_{t,\gamma_0}|v_t = v\right)$ all belong to the Hölder space $\Lambda^{\psi}(\mathcal{R})$ for some $\psi > 2$, the functions and their first- and second-order derivatives are uniformly bounded in the neighborhood of $v$, where $f_v(\cdot)$ is the probability density function of $v_t$ and $\varkappa_{t,\gamma_0} = [1, I(q_t \le \gamma_0)]'$; (ii) $E\left(\left\|\mathbf{z}_t\mathbf{x}_t'\right\||v_t = v\right)$, $E\left(\|\mathbf{z}_t\|^j|v_t = v\right)$, $E\left(\|\varepsilon_t\mathbf{z}_t\|^j|v_t = v\right)$, and $E\left(|\varepsilon_t|^{2+r}|v_t = v\right)$ are all bounded in the neighborhood of $v$ for some $r > 0$ and $1 \le j \le r_k$ with $r_k$ being defined in Assumption 5; (iii) $\Omega(v) = E\left(\varkappa_{1,\gamma_0}\varkappa_{1,\gamma_0}'|v_1 = v\right)$ is nonsingular in the neighborhood of $v$.

**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 $r_k > 2$; and (iii) as $n \to \infty$, $h \to 0$, $nh^{2(r_k+3)/r_k} \to \infty$, and $\lim_{n\to\infty} nh^5 = c_0 > 0$. Also, we denote $\kappa_{i,j} = \int K^i(u)\, u^j du$.

Assumption 4 contains regularity smoothness conditions used in the local linear estimation literature, while the conditional moment conditions with an exponent $j$ in Assumption 4(ii) are technical requirements to remove the asymptotic impact of the estimation of $\pi_q$; see the proofs given in the Appendix. Assumption 4(iii) holds if $E[I(q_1 \le \gamma_0)|v_1 = v]$ equals neither 0 nor 1 in the neighborhood of $v$. 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 of $\pi_q$ on $\tilde{\psi}(v) = \left[\tilde{h}_2(v), \tilde{\eta}_n(v)\right]'$, which is the second-step estimator of $\psi(v) = [h_2(v), \eta_n(v)]'$. Moreover, Assumption 5(iii) implies that the conventional optimal bandwidth of order $n^{-1/5}$ can be used to calculate $\tilde{\psi}(v)$. Below, we give our limiting result of $\tilde{\psi}(v)$.

THEOREM 4. *Under Assumptions 1–5, we have at an interior point v*

$$\sqrt{nh}\left[\tilde{\psi}(v) - \psi(v) - \frac{\kappa_{1,2}}{2}h^2\mathbf{B}(v)\right] \xrightarrow{d} N\left(0, \frac{\kappa_{2,0}}{f_v(v)}\Omega(v)^{-1}E\left(\varepsilon_1^2\varkappa_{1,\gamma_0}\varkappa_{1,\gamma_0}'|v_1 = v\right)\Omega(v)^{-1}\right).$$

*where* $\mathbf{B}(v) = \left[h_2^{(2)}(v), \eta_n^{(2)}(v)\right]'$.

Theorem 4 implies that the limiting distribution of the local linear estimator, $\tilde{\psi}(v)$, is not affected by the estimation of the unknown parameters $\left(\gamma_0, \beta_1', \beta_2'\right)$. Finally, the bandwidth can be chosen by conventional cross-validation methods.

## 3. MODEL SPECIFICATION TESTING

In this section, we are interested in testing whether the threshold variable, $q_t$, is endogenous in the linear threshold models (2.1) and (2.2). Note that Kapetanios (2010) constructs a Hausman-type test statistic to test for potential regressor

https://doi.org/10.1017/S0266466621000256 Published online by Cambridge University Press

endogeneity. In contrast, we are interested in testing whether the threshold variable, $q_t$, is endogenous. We formulate the following hypotheses

$$\mathrm{H}_0 : h_1(\cdot) \equiv h_2(\cdot) \equiv 0 \text{ vs. } \mathrm{H}_1 : \text{not } \mathrm{H}_0, \tag{3.1}$$

and the working null and alternative hypotheses can be written as, $\mathrm{H}_0' : \alpha_{L_n,1} = \alpha_{L_n,2} = \mathbf{0}_{L_n}$ against $\mathrm{H}_1' :$ not $\mathrm{H}_0'$.

Denoting $\mathbf{Q}_\gamma$ as an $n \times (2d_{\mathbf{x}})$ matrix with its $t$th row equal to $\left[\mathbf{x}_t' I(q_t \le \gamma), \mathbf{x}_t' I(q_t > \gamma)\right]$ and $\mathbf{M}_{\gamma,\mathbf{Q}} = \mathbf{I}_n - \mathbf{Q}_\gamma \left(\mathbf{Q}_\gamma' \mathbf{Q}_\gamma\right)^{-1} \mathbf{Q}_\gamma'$ and solving (2.6) by the partitioned OLS estimator gives

$$\hat{\alpha}_{L_n}(\gamma) = \left[\hat{\alpha}_{L_n,1}'(\gamma), \hat{\alpha}_{L_n,2}'(\gamma)\right]' = \left(\hat{\Phi}_{L_n,\gamma}' \mathbf{M}_{\gamma,\mathbf{Q}} \hat{\Phi}_{L_n,\gamma}\right)^{-1} \hat{\Phi}_{L_n,\gamma}' \mathbf{M}_{\gamma,\mathbf{Q}} \mathbf{y} \tag{3.2}$$

for any given $\gamma \in \left[\underline{\gamma}, \bar{\gamma}\right]$, where we denote $\hat{\Phi}_{L_n,\gamma} = \left[\hat{\Phi}_{L_n,\gamma}^-(\mathbf{v}), \hat{\Phi}_{L_n,\gamma}^+(\mathbf{v})\right]$. We then construct a Wald statistic using the covariance estimator introduced in Andrews (1991):

$$\begin{aligned} W_n(\gamma) = {} & \hat{\alpha}_{L_n}(\gamma)' \hat{\Phi}_{L_n,\gamma}' \mathbf{M}_{\gamma,\mathbf{Q}} \hat{\Phi}_{L_n,\gamma} \left(\hat{\Phi}_{L_n,\gamma}' \mathbf{M}_{\gamma,\mathbf{Q}} \hat{\Omega}_{n,\gamma} \mathbf{M}_{\gamma,\mathbf{Q}} \hat{\Phi}_{L_n,\gamma}\right)^{-1} \\ & \times \hat{\Phi}_{L_n,\gamma}' \mathbf{M}_{\gamma,\mathbf{Q}} \hat{\Phi}_{L_n,\gamma} \hat{\alpha}_{L_n}(\gamma), \end{aligned} \tag{3.3}$$

where $\hat{\Omega}_{n,\gamma} = \mathrm{diag}\left\{\hat{\varepsilon}_{\gamma,1}^2 / \left(1 - \hat{h}_{\gamma,11}\right), \ldots, \hat{\varepsilon}_{\gamma,n}^2 / \left(1 - \hat{h}_{\gamma,nn}\right)\right\}$ with $\hat{\varepsilon}_{\gamma,t} = y_t - \hat{\beta}_1'(\gamma)\mathbf{x}_t I(q_t \le \gamma) - \hat{\beta}_2'(\gamma)\mathbf{x}_t I(q_t > \gamma) - \hat{\alpha}_{L_n,1}'(\gamma)\Phi_{L_n,\gamma}^-(\hat{v}_t) - \hat{\alpha}_{L_n,2}'(\gamma)\Phi_{L_n}^+(\hat{v}_t)$ being the $t$th element of $\hat{\varepsilon}_\gamma = \left[\hat{\varepsilon}_{\gamma,1}, \ldots, \hat{\varepsilon}_{\gamma,n}\right]'$, an $n \times 1$ vector of residuals calculated from the alternative hypothesis, and $\hat{h}_{\gamma,tt}$ is the $(t,t)$th element of $\mathbf{M}_{\gamma,\mathbf{Q}} \hat{\Phi}_{L_n,\gamma} \left(\hat{\Phi}_{L_n,\gamma}' \mathbf{M}_{\gamma,\mathbf{Q}} \hat{\Omega}_{n,\gamma} \mathbf{M}_{\gamma,\mathbf{Q}} \hat{\Phi}_{L_n,\gamma}\right)^{-1} \hat{\Phi}_{L_n,\gamma}' \mathbf{M}_{\gamma,\mathbf{Q}}$.

Next, motivated by Gonzalo and Pitarakis (2007), we replace $\gamma$ in (3.3) by $\hat{\gamma}$, the estimate of $\gamma_0$ under the null hypothesis. That is,

$$\hat{\gamma} = \arg \min_{\gamma \in \left[\underline{\gamma}, \bar{\gamma}\right]} \mathbf{y}' \mathbf{M}_{\gamma,\mathbf{Q}} \mathbf{y}. \tag{3.4}$$

Our final test statistic is defined as $W_n(\hat{\gamma})$. The test statistic in Kapetanios (2010) needs to estimate all the regression coefficients under both the null and alternative hypotheses, while our test statistic estimates the threshold parameter only under the null hypothesis, and the regression coefficients are estimated only under the alternative hypothesis. Evidently, the limiting result of $\hat{\gamma}$ will depend on whether the threshold effects exist. In the presence of the threshold effects, $\beta_1 \neq \beta_2$ and $\gamma_0$ is identified. However, $\gamma_0$ vanishes when $\beta_1 = \beta_2$. Below, we give the limiting results for $\hat{\gamma}$.

https://doi.org/10.1017/S0266466621000256 Published online by Cambridge University Press

THEOREM 5. *Under Assumptions 1 and 2 and* $H_0$ *and* $E\left(u_{jt}^4\right) < M < \infty$ *for j=1,2, we have (i)* $n^{1-2\varsigma}\left|\hat{\gamma} - \gamma_0\right| = O_p(1)$ *when* $\beta_1 \neq \beta_2$*; (ii)* $\hat{\lambda} \xrightarrow{d} \lambda^*$ *when* $\beta_1 = \beta_2$*, where* $\lambda = E[I(q_t \leq \gamma)] = F_q(\gamma)$*,* $\hat{\lambda} = F_q(\hat{\gamma})$ *and* $\lambda^*$ *is defined in (A.23).*

Under the null hypothesis and $\beta_1 \neq \beta_2$, Theorem 5(i) results from Theorem 1 in Hansen (2000). Under the null hypothesis and $\beta_1 = \beta_2$, Theorem 5(ii) shows that $\hat{\lambda}$ converges in distribution to a random variable $\lambda^*$, where $F_q(\cdot)$ is the cumulative distribution function of $q_t$. Based on the limiting results in Theorem 5, we derive the limiting distribution of $W_n\left(\hat{\gamma}\right)$ under the null hypothesis as follows.

THEOREM 6. *Under the conditions in Theorem 5, Assumption 3(i),* $E(u_{jt}^4|\mathcal{F}_{n,t-1}) < \infty$ *and* $H_0$*, we have* $J_n\left(\hat{\gamma}\right) \xrightarrow{d} N(0,1)$ *holds, where* $J_n\left(\hat{\gamma}\right) = \left(W_n\left(\hat{\gamma}\right) - 2L_n\right)/(2\sqrt{L_n})$*, regardless of whether* $\beta_1$ *equals* $\beta_2$*.*

Theorem 6 constructs a test statistic that follows a standard normal distribution in large samples under the null hypothesis of exogenous threshold variable, regardless of whether the threshold effects exist in models (2.1) and (2.2). In addition, when the sample size is not large, $L_n$ can be very small, in which case the distribution of $W_n(\hat{\gamma})$ can be well approximated by the chi-squared distribution with $2L_n$ degrees of freedom, so that the critical value from the chi-squared distribution can be used to improve the finite sample performance of the proposed test.

## 4. MONTE CARLO SIMULATIONS

4.1. *Threshold and Slope Parameters.* We generate our data from the following model:

$$y_t = \beta_1 + \beta_2 x_t + (\delta_1 + \delta_2 x_t) I\left(q_t \leq \gamma\right) + u_t, \tag{4.1}$$

where

$$q_t = 2 + z_{qt} + v_{qt}. \tag{4.2}$$

The threshold parameter is set at the center of the distribution of $q_t$; i.e., $\gamma = 2$. The instrumental variable $z_{qt}$ is given by

$$z_{qt} = \left(x_t + \varsigma_{zt}\right)/2\sqrt{2} \tag{4.3}$$

and

$$u_t = 0.1\varsigma_{ut} + \kappa v_{qt}, \tag{4.4}$$

where $x_t$, $v_{qt}$, $\varsigma_{zt}$, and $\varsigma_{ut}$ are independent i.i.d. $N(0,1)$ random variables. The degree of endogeneity of the threshold variable is controlled by $\kappa$. We fix $\kappa = 0.95$,

https://doi.org/10.1017/S0266466621000256 Published online by Cambridge University Press

**TABLE 1.** Threshold parameter and threshold effect

| | Threshold parameter | | | Threshold effect | | |
|---|---|---|---|---|---|---|
| Quantile | 5th | 50th | 95th | 5th | 50th | 95th |
| $n$ | | | | | | |
| | | | $\delta_2 = 1$ | | | |
| 100 | 1.880 | 1.979 | 2.039 | 0.943 | 1.000 | 1.045 |
| 250 | 1.953 | 1.991 | 2.011 | 0.970 | 1.001 | 1.028 |
| 500 | 1.977 | 1.996 | 2.004 | 0.983 | 1.000 | 1.019 |
| 1,000 | 1.990 | 1.998 | 2.002 | 0.988 | 1.000 | 1.013 |
| | | | $\delta_2 = 2$ | | | |
| 100 | 1.887 | 1.978 | 2.007 | 1.943 | 2.000 | 2.045 |
| 250 | 1.954 | 1.991 | 2.000 | 1.970 | 2.001 | 2.028 |
| 500 | 1.978 | 1.996 | 2.001 | 1.983 | 2.000 | 2.019 |
| 1,000 | 1.990 | 1.998 | 2.001 | 1.988 | 2.000 | 2.012 |
| | | | $\delta_2 = 3$ | | | |
| 100 | 1.888 | 1.978 | 2.000 | 2.942 | 3.000 | 3.045 |
| 250 | 1.954 | 1.991 | 2.000 | 2.970 | 3.001 | 3.028 |
| 500 | 1.978 | 1.996 | 2.000 | 2.983 | 3.000 | 3.019 |
| 1,000 | 1.990 | 1.998 | 2.001 | 2.988 | 3.000 | 3.012 |
| | | | $\delta_2 = 4$ | | | |
| 100 | 1.893 | 1.978 | 2.000 | 3.942 | 4.000 | 4.046 |
| 250 | 1.954 | 1.991 | 2.000 | 3.970 | 4.001 | 4.028 |
| 500 | 1.978 | 1.995 | 2.000 | 3.983 | 4.000 | 4.019 |
| 1,000 | 1.990 | 1.998 | 2.000 | 3.988 | 4.000 | 4.012 |

*Note:* This table presents Monte Carlo quantiles of the estimates of the true threshold parameter $\gamma = 2$ and true threshold effect $\delta_2 = 1, 2, 3, 4$ using a 6th order Hermite basis function and sample sizes $n$ for the DGP in (4.1)–(4.4).

$\beta_1 = \beta_2 = 1$, and $\delta_1 = 0$ and vary $\delta_2$ over the values of $1, 2, 3, 4$, which correspond to a range of small to large threshold effects.[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 $1,000$ using $1,000$ Monte Carlo replications. Table 1 presents the quantiles from the empirical distribution of $\hat{\gamma}$ and $\hat{\delta}_2$ by varying the threshold effect $\delta_2$ over the values 1, 2, 3, and 4 using a 6th order Hermite basis function for the DGP in (4.1)–(4.4).[8] We see that the estimators' performance for both the threshold parameter and the threshold effect improves as the threshold effect and the sample

[7] In the Online Appendix, we also present results based on an alternative DGP that adds an endogenous regressor to model (4.1) with similar findings.

[8] We also investigated the sensitivity of our results to the choice of the order of the Hermite basis function using $L_n = 1, 2..., 6$ and found that our results are similar but slightly better for $L_n = 6$. These results are presented in the Online Appendix.

https://doi.org/10.1017/S0266466621000256 Published online by Cambridge University Press

**TABLE 2.** Confidence interval coverage of the threshold parameter

| $\delta_2$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $n$ | | | | |
| | | $L_n = 2$ | | |
| 100 | 0.92 | 0.98 | 0.98 | 0.99 |
| 250 | 0.90 | 0.97 | 0.97 | 0.97 |
| 500 | 0.91 | 0.96 | 0.96 | 0.96 |
| 1,000 | 0.92 | 0.95 | 0.93 | 0.92 |
| | | $L_n = 3$ | | |
| 100 | 0.96 | 0.98 | 0.98 | 0.98 |
| 250 | 0.96 | 0.97 | 0.97 | 0.97 |
| 500 | 0.94 | 0.95 | 0.95 | 0.94 |
| 1,000 | 0.93 | 0.92 | 0.91 | 0.90 |
| | | $L_n = 4$ | | |
| 100 | 0.96 | 0.98 | 0.98 | 0.98 |
| 250 | 0.96 | 0.96 | 0.97 | 0.97 |
| 500 | 0.93 | 0.94 | 0.94 | 0.94 |
| 1,000 | 0.89 | 0.89 | 0.89 | 0.89 |
| | | $L_n = 5$ | | |
| 100 | 0.96 | 0.97 | 0.98 | 0.98 |
| 250 | 0.96 | 0.96 | 0.97 | 0.97 |
| 500 | 0.94 | 0.93 | 0.93 | 0.93 |
| 1,000 | 0.89 | 0.89 | 0.88 | 0.88 |
| | | $L_n = 6$ | | |
| 100 | 0.94 | 0.97 | 0.97 | 0.97 |
| 250 | 0.96 | 0.96 | 0.97 | 0.97 |
| 500 | 0.94 | 0.93 | 0.93 | 0.93 |
| 1,000 | 0.89 | 0.89 | 0.88 | 0.88 |

*Note:* This table presents Monte Carlo results about the nominal 90% confidence interval coverage of the threshold parameter for true threshold effect $\delta_2 = 1, 2, 3, 4$, order of Hermite basis function $L_n = 2, 3, 4, 5, 6$, and sample sizes $n$ for the DGP in (4.1)–(4.4).

size increases. Specifically, the 50th percentile approaches the true threshold parameter, $\gamma_0 = 2$, as the sample size $n$ increases and the width of the empirical distribution of $\hat{\gamma}$ becomes narrower as $\delta_2$ increases.

Finally, Table 2 shows the coverage rates of a nominal 90% confidence interval on the threshold parameter for different values of the threshold effect $\delta_2$ and different orders of Hermite basis functions. The coverage rate is relatively robust to the choice of the smoothing parameter and is closer to the nominal level as the sample size increases.

It is worth noting that the convergence rate of $\hat{\beta}$ is root-n in Theorem 3 and the convergence rate of $\hat{\gamma}$ in Theorem 1 depends on the diminishing rate of

https://doi.org/10.1017/S0266466621000256 Published online by Cambridge University Press

**TABLE 3.** Size and power

| $\kappa$ | 0 | 0.1 | 0.2 | 0.3 | 0.4 |
|---|---|---|---|---|---|
| $n$ | | | | | |
| | | | $L_n = 2$ | | |
| 100 | 0.13 | 0.35 | 0.82 | 0.99 | 1.00 |
| 250 | 0.08 | 0.28 | 0.82 | 0.99 | 1.00 |
| 500 | 0.06 | 0.26 | 0.80 | 0.99 | 1.00 |
| 1,000 | 0.06 | 0.25 | 0.82 | 1.00 | 1.00 |
| | | | $L_n = 3$ | | |
| 100 | 0.16 | 0.37 | 0.80 | 0.99 | 1.00 |
| 250 | 0.10 | 0.31 | 0.80 | 0.99 | 1.00 |
| 500 | 0.07 | 0.25 | 0.79 | 0.99 | 1.00 |
| 1,000 | 0.07 | 0.25 | 0.80 | 1.00 | 1.00 |
| | | | $L_n = 4$ | | |
| 100 | 0.23 | 0.42 | 0.80 | 0.99 | 1.00 |
| 250 | 0.15 | 0.35 | 0.79 | 0.99 | 1.00 |
| 500 | 0.10 | 0.28 | 0.78 | 0.99 | 1.00 |
| 1,000 | 0.08 | 0.24 | 0.79 | 1.00 | 1.00 |
| | | | $L_n = 5$ | | |
| 100 | 0.28 | 0.44 | 0.80 | 0.98 | 1.00 |
| 250 | 0.19 | 0.35 | 0.78 | 0.98 | 1.00 |
| 500 | 0.14 | 0.30 | 0.77 | 0.99 | 1.00 |
| 1,000 | 0.12 | 0.25 | 0.78 | 0.99 | 1.00 |
| | | | $L_n = 6$ | | |
| 100 | 0.39 | 0.52 | 0.83 | 0.97 | 1.00 |
| 250 | 0.19 | 0.34 | 0.75 | 0.98 | 1.00 |
| 500 | 0.16 | 0.33 | 0.75 | 0.98 | 1.00 |
| 1,000 | 0.14 | 0.29 | 0.74 | 0.99 | 1.00 |

*Note:* This table presents the size ($\kappa = 0$) and the power ($\kappa > 0$) of the test for endogeneity defined in equation (3.3) using the Andrews (1991) covariance estimator for various orders of Hermite basis functions $L_n$ and sample sizes $n$ for the DGP in (4.1)–(4.4). The significance level is 0.05.

the threshold effect. Therefore, the choice of $L_n$ only affects the second-order asymptotics of our estimator. Consequently, the estimation results of $\hat{\gamma}$ and $\hat{\beta}$ should be relatively robust to the choice of $L_n$ as long as Assumption 3 holds. It is well known in semiparametric estimation literature that the choice of optimal smoothing parameter is of no big concern to the estimation of the parametric part of parameters due to the fast parametric convergence rate.

4.2. *Size and Power of the Wald Statistic.* We assess the size and power of the Wald statistic in equation (3.3) at the 5% significance level, where $\gamma_0$ is estimated by the objective function defined by (3.4). Table 3 provides the results for the

https://doi.org/10.1017/S0266466621000256 Published online by Cambridge University Press

DGP defined in equation (4.1) for various orders of Hermite basis functions $L_n$ and sample sizes. We present the size ($\kappa = 0$) and the power ($\kappa > 0$) for various orders, $L_n$, of Hermite basis functions and sample sizes.

While our findings reveal that our test performs quite well in terms of power, they suggest a more nuanced and complex characterization of the size of the test. In particular, size distortions appear to be increasing with $L_n$ for any given sample size, and these distortions can be substantial, especially for relatively small sample sizes. Interestingly, our findings suggest that these size distortions are minimal when $L_n = 2$ and the sample size is greater than 250.[9]

## 5. CONCLUSION

In this paper, we propose to estimate a threshold regression model with an endogenous threshold variable based on a nonparametric control function approach. Specifically, the estimation of the threshold parameter is based on a two-stage concentrated least squares method and the regime-specific control functions are estimated by the series estimation method. We develop estimation and inference for weakly dependent data for the estimators of both the threshold and slope parameters, including a test for the endogeneity of the threshold variable. Finally, we assess the performance of the proposed estimator and test statistic using Monte Carlo simulations.

Our method can be further extended in several directions. One useful extension to our approach is to apply the penalized method to simultaneously select and estimate the threshold endogeneity bias correction terms. Lee et al. (2016) study high-dimensional regression with possible fixed threshold effects. It is interesting to extend the penalized method to study the diminishing threshold effects case, as well as to derive the limiting distribution of a post-lasso estimator. Another extension to our approach is to relax the linear specification in the structural models (2.1) and (2.2), as well as in the reduced-form model (2.3) of the threshold variable, to a more flexible semi-/nonparametric specification.

# APPENDIX

Throughout the Appendix, capital boldface letters denote matrices and lowercase boldface letters are reserved for vectors, unless otherwise specified differently. Also, $\mathbf{a}(\mathbf{v}) = [a(v_1), \ldots, a(v_n)]'$ denotes an $n \times 1$ vector of function $a(\cdot)$ at point $\mathbf{v} = [v_1, \ldots, v_n]'$.

**Proof of Lemma 1:** Denote $\mathbf{P}_\gamma = \mathbf{X}_\gamma \left(\mathbf{X}_\gamma'\mathbf{X}_\gamma\right)^{-1}\mathbf{X}_\gamma'$, $\mathbf{P}_\gamma^- = \mathbf{X}_{-,\gamma}\left(\mathbf{X}_{-,\gamma}'\mathbf{X}_{-,\gamma}\right)^{-1}\mathbf{X}_{-,\gamma}'$, and $\mathbf{P}_\gamma^+ = \mathbf{X}_{+,\gamma}\left(\mathbf{X}_{+,\gamma}'\mathbf{X}_{+,\gamma}\right)^{-1}\mathbf{X}_{+,\gamma}'$, where $\mathbf{X}_{-,\gamma}$ and $\mathbf{X}_{+,\gamma}$ are defined in

[9] We also investigated a finite-sample correction of the White estimator as well as the Horn, Horn, and Duncan (1975) estimator but the results were no better than the Andrews estimator.

https://doi.org/10.1017/S0266466621000256 Published online by Cambridge University Press

Section 2.1. Applying simple calculation gives $\mathbf{P}_\gamma = \mathbf{P}_\gamma^- + \mathbf{P}_\gamma^+$. As $\left(\mathbf{I}_n - \mathbf{P}_\gamma\right)\mathbf{X}_\gamma = \mathbf{0}$, by (2.7), we have

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

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

Specifically, applying simple algebra gives

$$\begin{aligned}
S_n(\gamma) &= \left(\mathbf{X}_-\delta_n + \eta_-(\mathbf{v}) + \hat{\varepsilon}\right)'\left(\mathbf{I}_n - \mathbf{P}_\gamma\right)\left(\mathbf{X}_-\delta_n + \eta_-(\mathbf{v}) + \hat{\varepsilon}\right) \\
&= \delta_n'\mathbf{X}_-'\left(\mathbf{I}_n - \mathbf{P}_\gamma\right)\mathbf{X}_-\delta_n + \eta_-(\mathbf{v})'\left(\mathbf{I}_n - \mathbf{P}_\gamma\right)\eta_-(\mathbf{v}) + \hat{\varepsilon}'\left(\mathbf{I}_n - \mathbf{P}_\gamma\right)\hat{\varepsilon} \\
&\quad + 2\delta_n'\mathbf{X}_-'\left(\mathbf{I}_n - \mathbf{P}_\gamma\right)\eta_-(\mathbf{v}) + 2\delta_n'\mathbf{X}_-'\left(\mathbf{I}_n - \mathbf{P}_\gamma\right)\hat{\varepsilon} + 2\eta_-(\mathbf{v})'\left(\mathbf{I}_n - \mathbf{P}_\gamma\right)\hat{\varepsilon} \\
&= S_{n1} + S_{n2} + S_{n3} + 2\left(S_{n4} + S_{n5} + S_{n6}\right),
\end{aligned}$$

where $\mathbf{X}_-$ is an $n \times d_{\mathbf{x}}$ matrix with its $t$th row equal to $\mathbf{x}_t' I(q_t \le \gamma_0)$, and $\eta_-(\mathbf{v})$ is an $n \times 1$ vector with its $t$th element equal to $\eta_n(v_t) I(q_t \le \gamma_0)$, $\hat{\varepsilon} = \left[\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n\right]'$ with $\hat{\varepsilon}_t = h_2(v_t) - h_2^*\left(\hat{v}_t\right) + \varepsilon_t$, and the subscript $j$ in $S_{nj}$ $(j = 1,\ldots,6)$ is labeled according to the ordering of appearance.

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

$$\begin{aligned}
S_{n1} &= \delta_n'\mathbf{X}_-'\left(\mathbf{I}_n - \mathbf{P}_\gamma\right)\mathbf{X}_-\delta_n = n^{-2\varsigma}\delta_0'\mathbf{X}_-'\left(\mathbf{I}_n - \mathbf{P}_\gamma\right)\mathbf{X}_-\delta_0, \\
S_{n2} &= \eta_-(\mathbf{v})'\left(\mathbf{I}_n - \mathbf{P}_\gamma\right)\eta_-(\mathbf{v}) = n^{-2\varrho}\eta_{0,-}(\mathbf{v})\left(\mathbf{I}_n - \mathbf{P}_\gamma\right)\eta_{0,-}(\mathbf{v}), \\
S_{n4} &= \delta_n'\mathbf{X}_-'\left(\mathbf{I}_n - \mathbf{P}_\gamma\right)\eta_-(\mathbf{v}) = n^{-\varsigma-\varrho}\delta_0'\mathbf{X}_-'\left(\mathbf{I}_n - \mathbf{P}_\gamma\right)\eta_{0,-}(\mathbf{v}),
\end{aligned}$$

where $\eta_{0,-}(\mathbf{v})$ is an $n \times 1$ vector with its $t$th element equal to $\eta_0(v_t) I(q_t \le \gamma_0)$. By Lemmas B.1 and B.2 of the Online Appendix, we have

$$\max_{\gamma \in \left[\underline{\gamma}, \overline{\gamma}\right]}\left|n^{-1}\delta_0'\mathbf{X}_-'\mathbf{P}_\gamma\mathbf{X}_-\delta_0 - \delta_0'\mathbf{g}_1'(\gamma)\Sigma_{\mathbf{x}^*\mathbf{x}^{*\prime},\gamma}^{-1}\mathbf{g}_1(\gamma)\delta_0\right| = o_p(1), \tag{A.3}$$

$$\max_{\gamma \in \left[\underline{\gamma}, \overline{\gamma}\right]}\left|n^{-1}\eta_{0,-}'(\mathbf{v})\mathbf{P}_\gamma\eta_{0,-}(\mathbf{v}) - \mathbf{g}_2'(\gamma)\Sigma_{\mathbf{x}^*\mathbf{x}^{*\prime},\gamma}^{-1}\mathbf{g}_2(\gamma)\right| = o_p(1), \tag{A.4}$$

$$\max_{\gamma \in \left[\underline{\gamma}, \overline{\gamma}\right]}\left|n^{-1}\delta_0'\mathbf{X}_-'\mathbf{P}_\gamma\eta_{0,-}(\mathbf{v}) - \delta_0'\mathbf{g}_1'(\gamma)\Sigma_{\mathbf{x}^*\mathbf{x}^{*\prime},\gamma}^{-1}\mathbf{g}_2(\gamma)\right| = o_p(1), \tag{A.5}$$

where $\Sigma_{\mathbf{x}^*\mathbf{x}^{*\prime},\gamma} = E\left(\mathbf{x}_{t,\gamma}^*\mathbf{x}_{t,\gamma}^{*\prime}\right)$, $\mathbf{g}_1(\gamma) = E\left[\mathbf{x}_{t,\gamma}^*\mathbf{x}_t' I(q_t \le \gamma_0)\right]$ and $\mathbf{g}_2(\gamma) = E\left[\mathbf{x}_{t,\gamma}^*\eta_0(v_t) I(q_t \le \gamma_0)\right]$. As $\left\{\left(\mathbf{x}_t', q_t, \mathbf{z}_t', u_t\right)\right\}$ is ergodic under Assumption 1(i),

https://doi.org/10.1017/S0266466621000256 Published online by Cambridge University Press

$E\left(\left\|\mathbf{x}_t\mathbf{x}_t'\right\|\right) < \infty$ 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

$$n^{-1}\delta_0'\mathbf{X}_-'\mathbf{X}_-\delta_0 \xrightarrow{a.s.} \delta_0'E\left[\mathbf{x}_t\mathbf{x}_t'I(q_t \le \gamma_0)\right]\delta_0 \equiv \delta_0'\mathbf{m}_1(\gamma_0)\delta_0, \tag{A.6}$$

$$n^{-1}\eta_{0,-}'(\mathbf{v})\eta_{0,-}(\mathbf{v}) \xrightarrow{a.s.} E\left[\eta_0^2(v_t)I(q_t \le \gamma_0)\right] \equiv m_2(\gamma_0), \tag{A.7}$$

$$n^{-1}\delta_0'\mathbf{X}_-'\eta_{0,-}(\mathbf{v}) \xrightarrow{a.s.} \delta_0'E\left[\mathbf{x}_t\eta_0(v_t)I(q_t \le \gamma_0)\right] \equiv \delta_0'\mathbf{m}_3(\gamma_0). \tag{A.8}$$

(ii) We consider $S_{n,3}(\gamma) = \hat{\varepsilon}'\left(\mathbf{I}_n - \mathbf{P}_\gamma\right)\hat{\varepsilon} = \hat{\varepsilon}'\hat{\varepsilon} - \hat{\varepsilon}'\mathbf{P}_\gamma\hat{\varepsilon}$. Note that $h_2(v_t) - h_2^*\left(\hat{v}_t\right) = \left[h_2(v_t) - h_2\left(\hat{v}_t\right)\right] + \left[h_2\left(\hat{v}_t\right) - h_2^*\left(\hat{v}_t\right)\right]$, where the second term is uniformly bounded by $O\left(L_n^{-\xi}\right)$ for all $t$ by Assumption 2(iv), and the first term $h_2(v_t) - h_2\left(\hat{v}_t\right) = h_2^{(1)}(\bar{v}_t)\mathbf{z}_t'\left(\hat{\pi}_q - \pi_q\right)$ with $\bar{v}_t$ lying between $v_t$ and $\hat{v}_t$ as $v_t - \hat{v}_t = \mathbf{z}_t'\left(\hat{\pi}_q - \pi_q\right)$. It therefore follows

$$\frac{1}{n}\sum_{t=1}^{n}\left[h_2(v_t) - h_2^*\left(\hat{v}_t\right)\right]^2 = O_p\left(L_n^{-2\xi} + n^{-1}\right) \tag{A.9}$$

under Assumptions 1(i) and (ii) and 2(iii) and (iv). Therefore, we have

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

In addition, we have $\hat{\varepsilon}'\mathbf{P}_\gamma\hat{\varepsilon} = \left\|\mathbf{P}_\gamma\hat{\varepsilon}\right\|^2 \le 2\left(\left\|\mathbf{h}_2(\mathbf{v}) - \mathbf{h}_2^*\left(\hat{\mathbf{v}}\right)\right\|^2 + \left\|\mathbf{P}_\gamma\varepsilon\right\|^2\right)$ since $\left|\mathbf{x}'\mathbf{A}\mathbf{x}\right| \le \lambda_{\max}(\mathbf{A})\mathbf{x}'\mathbf{x}$ and an idempotent matrix's eigenvalues equal either zeros or ones. Applying Lemma B.3 of the Online Appendix gives $\max_{\gamma\in[\underline{\gamma},\overline{\gamma}]} n^{-1}\varepsilon'\mathbf{P}_\gamma\varepsilon = O_p(L_n/n)$. Therefore, we obtain

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

where we denote $\vartheta_n = L_n^{-\xi} + \sqrt{L_n/n}$.

(iii) We consider $S_{n,5}(\gamma) = \delta_n'\mathbf{X}_-'\left(\mathbf{I}_n - \mathbf{P}_\gamma\right)\hat{\varepsilon} = n^{-\varsigma}\delta_0'\mathbf{X}_-'\left(\mathbf{I}_n - \mathbf{P}_\gamma\right)\hat{\varepsilon}$, where

$$\begin{aligned} n^{-1}\delta_0'\mathbf{X}_-'\hat{\varepsilon} &= n^{-1}\sum_{t=1}^{n}\delta_0'\mathbf{x}_t'I(q_t \le \gamma_0)\left[h_2(v_t) - h_2^*\left(\hat{v}_t\right) + \varepsilon_t\right] \\ &= O_p\left(L_n^{-\xi} + n^{-1/2}\right) \end{aligned} \tag{A.12}$$

and

$$\max_{\gamma\in\left[\underline{\gamma},\overline{\gamma}\right]}\frac{\left|\delta_0'\mathbf{X}_-'\mathbf{P}_\gamma\hat{\varepsilon}\right|}{n} \le \frac{1}{n}\max_{\gamma\in\left[\underline{\gamma},\overline{\gamma}\right]}\left\|\mathbf{X}_-\delta_0\right\|\left\|\mathbf{P}_\gamma\hat{\varepsilon}\right\| = O_p(\vartheta_n). \tag{A.13}$$

It follows that

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

https://doi.org/10.1017/S0266466621000256 Published online by Cambridge University Press

Similarly, for $S_{n,6}(\gamma) = \eta_-(\mathbf{v})'(\mathbf{I}_n - \mathbf{P}_\gamma)\hat{\varepsilon} = n^{-\varrho}\eta_{0,-}(\mathbf{v})'(\mathbf{I}_n - \mathbf{P}_\gamma)\hat{\varepsilon}$, 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 $S_1(\gamma) = \bar{\sigma}_\varepsilon^2 + n^{-2\varsigma}\delta_0'\mathbf{m}_1(\gamma_0)\delta_0 + n^{-2\varrho}m_2(\gamma_0) + 2n^{-\varsigma-\varrho}\delta_0'\mathbf{m}_3(\gamma_0) - \mu(\gamma)$ and $\mu(\gamma) = \left[n^{-\varsigma}\mathbf{g}_1(\gamma)\delta_0 + n^{-\varrho}\mathbf{g}_2(\gamma)\right]'\Sigma^{-1}_{\mathbf{x}^*\mathbf{x}^{*\prime},\gamma}\left[n^{-\varsigma}\mathbf{g}_1(\gamma)\delta_0 + n^{-\varrho}\mathbf{g}_2(\gamma)\right]$. Evidently, $S_1(\gamma)$ is continuous in $\gamma$.

(v) Denote $\mathbf{D}(\gamma) = E\left[\chi_t^*\chi_t^{*\prime}I(q_t \le \gamma)\right]$ and $\mathbf{M}(\gamma) = E\left[\chi_t^*\eta_0(v_t)I(q_t \le \gamma)\right]$, where $\chi_t^* = \left[\mathbf{x}_t', \Phi_{L_n}'(v_t)\right]'$. Then, we have

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

and for $\gamma \in [\gamma_0, \overline{\gamma}]$, $\mathbf{g}_1(\gamma) = [\mathbf{D}_1(\gamma), \mathbf{0}]$ and $\mathbf{g}_2(\gamma) = [\mathbf{M}(\gamma), \mathbf{0}]$, where $\mathbf{D}_1(\gamma) = E\left[\chi_t^*\mathbf{x}_t'I(q_t \le \gamma)\right]$ and $\Sigma_{\chi^{*\prime}\chi^*} = E\left(\chi_t^*\chi_t^{*\prime}\right)$. As for any random variable $z$, $d\{E[zI(q \le \gamma)]\}/d\gamma = E(z|q = \gamma)f_q(\gamma)$, we have $\mathbf{D}^{(1)}(\gamma) = E\left(\chi_t^*\chi_t^{*\prime}|q_t = \gamma\right)f_q(\gamma)$, $M^{(1)}(\gamma) = E\left(\chi_t^*\eta_0(v_t)|q_t = \gamma\right)f_q(\gamma)$, and $\partial\Sigma_{\mathbf{x}^*\mathbf{x}^{*\prime},\gamma}/\partial\gamma = \text{diag}\{1,-1\} \otimes \mathbf{D}^{(1)}(\gamma)$. Moreover, applying Propositions 17.3(a) and 17.25 in Seber (2008), for any differentiable function $\mathbf{a}(\gamma)$, we have

$$\begin{aligned}&\frac{d\left(\mathbf{a}(\gamma)\Sigma^{-1}_{\mathbf{x}^*\mathbf{x}^{*\prime},\gamma}\mathbf{a}(\gamma)\right)}{d\gamma} \\ &\quad = 2\mathbf{a}'(\gamma)\Sigma^{-1}_{\mathbf{x}^*\mathbf{x}^{*\prime},\gamma}\frac{d(\mathbf{a}(\gamma))}{d\gamma} - \mathbf{a}'(\gamma)\Sigma^{-1}_{\mathbf{x}^*\mathbf{x}^{*\prime},\gamma}\frac{d\Sigma_{\mathbf{x}^*\mathbf{x}^{*\prime},\gamma}}{d\gamma}\Sigma^{-1}_{\mathbf{x}^*\mathbf{x}^{*\prime},\gamma}\mathbf{a}(\gamma).\end{aligned}$$

Hence, we have

$$\begin{aligned}-\frac{d\mu(\gamma)}{d\gamma} &= \left[n^{-\varsigma}\mathbf{D}_1(\gamma)\delta_0 + n^{-\varrho}\mathbf{M}(\gamma)\right]' \\ &\quad \times \mathbf{D}^{-1}(\gamma)\mathbf{D}^{(1)}(\gamma)\mathbf{D}^{-1}(\gamma)\left[n^{-\varsigma}\mathbf{D}_1(\gamma)\delta_0 + n^{-\varrho}\mathbf{M}(\gamma)\right] > 0\end{aligned}$$

as $\mathbf{D}^{-1}(\gamma)\mathbf{D}^{(1)}(\gamma)\mathbf{D}^{-1}(\gamma)$ is a continuous and positively definite matrix over $\gamma$ under Assumption 2(ii). Therefore, $S_1(\gamma)$ is a strictly increasing function over $\gamma \in [\gamma_0, \overline{\gamma}]$.

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

$$S_n(\gamma) = \left[-\mathbf{X}_+\delta_n - \eta_+(\mathbf{v}) + \hat{\varepsilon}\right]'(\mathbf{I}_n - \mathbf{P}_\gamma)\left[-\mathbf{X}_+\delta_n - \eta_+(\mathbf{v}) + \hat{\varepsilon}\right],$$

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

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

https://doi.org/10.1017/S0266466621000256 Published online by Cambridge University Press

where $S_2(\gamma)$ equals $S_1(\gamma)$ with $I(q_t \leq \gamma_0)$ being replaced by $I(q_t > \gamma_0)$. For $\gamma \in [\underline{\gamma}, \gamma_0]$, we have $\mathbf{g}_1(\gamma) = \left[\mathbf{0}, \Sigma_{\chi^{*\prime}\chi^*} - \mathbf{D}(\gamma)\right]$ and $\mathbf{g}_2(\gamma) = \left[\mathbf{0}, E\left[\chi_t^* \eta_0(v_t) I(q_t > \gamma)\right]\right]$, so that

$$
\begin{aligned}
-\frac{d\mu(\gamma)}{d\gamma} &= -\left\{\left[\Sigma_{\chi^{*\prime}\chi^*} - \mathbf{D}(\gamma)\right]\delta_0 + E\left[\chi_t^* \eta_0(v_t) I(q_t > \gamma)\right]\right\}' \\
&\quad \times \left[\Sigma_{\chi^{*\prime}\chi^*} - \mathbf{D}(\gamma)\right]^{-1} \mathbf{D}^{(1)}(\gamma) \\
&\times \left[\Sigma_{\chi^{*\prime}\chi^*} - \mathbf{D}(\gamma)\right]^{-1} \left\{\left[\Sigma_{\chi^{*\prime}\chi^*} - \mathbf{D}(\gamma)\right]\delta_0 + E\left[\chi_t^* \eta_0(v_t) I(q_t > \gamma)\right]\right\} < 0.
\end{aligned}
$$

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

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 a continuous function of $\gamma$ and is uniquely minimized at $\gamma_0$. It then follows $\hat{\gamma} \xrightarrow{p} \gamma_0$.

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

**Proof of Theorem 1:** In matrix form, we have $\mathbf{y} = \mathbf{X}\beta_2 + \mathbf{X}_{-,\gamma_0}\mathbf{S}_{d_\mathbf{x}}\delta_n + \mathbf{h}_2(\mathbf{v}) + \eta_{-,\gamma_0}(\mathbf{v}) + \varepsilon$, and

$$
\begin{aligned}
&\mathbf{y} - \mathbf{X}_\gamma \hat{\theta} \\
&= \mathbf{X}\beta_2 + \mathbf{X}_{-,\gamma_0}\mathbf{S}_{d_\mathbf{x}}\delta_n + \mathbf{h}_2(\mathbf{v}) + \eta_{-,\gamma_0}(\mathbf{v}) + \varepsilon - \mathbf{X}\hat{\beta}_2 - \mathbf{X}_{-,\gamma}\mathbf{S}_{d_\mathbf{x}}\hat{\delta}_n - \hat{\mathbf{h}}_2^*(\hat{v}) - \hat{\eta}_{-,\gamma}^*(\hat{v}) \\
&= \varepsilon + \mathbf{X}\left(\beta_2 - \hat{\beta}_2\right) + \mathbf{X}_{-,\gamma_0}\mathbf{S}_{d_\mathbf{x}}\left(\delta_n - \hat{\delta}_n\right) + \Delta_n - \Delta\mathbf{X}_{-,\gamma}\mathbf{S}_{d_\mathbf{x}}\hat{\delta}_n - \Delta\hat{\eta}_{-,\gamma}^*,
\end{aligned} \tag{A.19}
$$

where $\mathbf{S}_{d_\mathbf{x}}$ equals the first $d_\mathbf{x}$ columns of the $(d_\mathbf{x} + L_n) \times (d_\mathbf{x} + L_n)$ identity matrix, $\Delta\mathbf{X}_{-,\gamma} = \mathbf{X}_{-,\gamma} - \mathbf{X}_{-,\gamma_0}$, $\Delta\hat{\eta}_{-,\gamma}^* = \hat{\eta}_{-,\gamma}^*(\hat{\mathbf{v}}) - \hat{\eta}_{-,\gamma_0}^*(\hat{\mathbf{v}})$, $\Delta_n = \mathbf{h}_2(\mathbf{v}) - \hat{\mathbf{h}}_2^*(\hat{\mathbf{v}}) + \eta_{-,\gamma_0}(\mathbf{v}) - \hat{\eta}_{-,\gamma_0}^*(\hat{\mathbf{v}})$, and the typical element of $\mathbf{X}$, $\mathbf{h}_2(\mathbf{v})$, $\hat{\mathbf{h}}_2^*(\hat{\mathbf{v}})$, $\eta_{-,\gamma}(\mathbf{v})$, and $\hat{\eta}_{-,\gamma}^*(\hat{\mathbf{v}})$ are $\mathbf{x}_t$, $h_2(v_t)$, $\hat{h}_2^*(\hat{v}_t)$, $\eta_n(v_t) I(q_t \leq \gamma)$, and $\hat{\eta}_n^*(\hat{v}_t) I(q_t \leq \gamma)$, respectively. Denoting $S_n(\gamma) = \left(\mathbf{y} - \mathbf{X}_\gamma\hat{\theta}\right)'\left(\mathbf{y} - \mathbf{X}_\gamma\hat{\theta}\right)$ we have

$$
\begin{aligned}
&S_n(\gamma) - S_n(\gamma_0) \\
&= \left(\Delta\hat{\eta}_{-,\gamma}^* + \Delta\mathbf{X}_{-,\gamma}\mathbf{S}_{d_\mathbf{x}}\hat{\delta}_n\right)'\left(\Delta\hat{\eta}_{-,\gamma}^* + \Delta\mathbf{X}_{-,\gamma}\mathbf{S}_{d_\mathbf{x}}\hat{\delta}_n\right) \\
&\quad - 2(\varepsilon + \Delta_n)'\left(\Delta\hat{\eta}_{-,\gamma}^* + \Delta\mathbf{X}_{-,\gamma}\mathbf{S}_{d_\mathbf{x}}\hat{\delta}_n\right) \\
&= \kappa_n' \sum_{t=1}^n \chi_t \chi_t' d_t^2(\gamma, \gamma_0)\kappa_n - 2\hat{\kappa}_n' \sum_{t=1}^n \varepsilon_t \chi_t d_t(\gamma, \gamma_0) \\
&\quad - 2\hat{\kappa}_n' \sum_{t=1}^n \left[\left(\beta_2 - \hat{\beta}_2\right)'\mathbf{x}_t + \left(\delta_n - \hat{\delta}_n\right)'\mathbf{x}_t I(q_t \leq \gamma_0)\right. \\
&\quad \left. + h_2(v_t) - \hat{h}_2^*(\hat{v}_t) + n^{-\varrho}\left(\eta_0(v_t) - \hat{\eta}_0^*(\hat{v}_t)\right) I(q_t \leq \gamma_0)\right] \chi_t d_t(\gamma, \gamma_0)
\end{aligned}
$$

https://doi.org/10.1017/S0266466621000256 Published online by Cambridge University Press

$$+\left(\hat{\kappa}_n+\kappa_n\right)'\sum_{t=1}^{n}\chi_t\chi_t' d_t^2\left(\gamma,\gamma_0\right)\left(\hat{\kappa}_n-\kappa_n\right)$$
$$=S_{n,1}^*\left(\gamma\right)-2S_{n,2}^*\left(\gamma\right)-2S_{n,3}^*\left(\gamma\right)+S_{n,4}^*\left(\gamma\right), \tag{A.20}$$

where we denote $d_t\left(\gamma,\gamma_0\right)=I(q_t\le\gamma)-I(q_t\le\gamma_0)$, $\chi_t=\left[\mathbf{x}_t',\Phi_{L_n}'\left(\hat{v}_t\right)\right]'$, $\hat{\kappa}_n=\left[\hat{\delta}_n',n^{-\varrho}\hat{\alpha}_{L_n,0}'\right]'$, and $\kappa_n=\left[n^{-\varsigma}\delta_0',n^{-\varrho}\alpha_{L_n,0}'\right]'$ with $\hat{\delta}_n=\hat{\beta}_1-\hat{\beta}_2$, $n^{-\varrho}\hat{\alpha}_{L_n,0}=\hat{\alpha}_{L_n,1}-\hat{\alpha}_{L_n,2}$, $n^{-\varsigma}\delta_0=\beta_1-\beta_2$, and $n^{-\varrho}\alpha_{L_n,0}=\alpha_{L_n,1}-\alpha_{L_n,2}$. Closely following the proof of Lemma A.9 in Hansen (2000), we can show that

$$a_n\left(\hat{\gamma}-\gamma_0\right)=\underset{\upsilon}{argmax}\,Q_n\left(\upsilon\right)=O_p\left(1\right),$$

where $a_n=n^{1-2\min(\varsigma,\varrho)}$ and $Q_n\left(\upsilon\right)=S_n\left(\gamma_0\right)-S_n\left(\gamma_0+\upsilon/a_n\right)$.

Now, we consider $S_{n,1}^*\left(\gamma\right)=\kappa_n'\sum_{t=1}^{n}\chi_t\chi_t' d_t^2\left(\gamma,\gamma_0\right)\kappa_n$. For any given $\upsilon\in\left[\underline{\upsilon},\overline{\upsilon}\right]$, a finite interval, we have

$$\begin{aligned}
S_{n,1}^*\left(\upsilon\right)&=\kappa_n'\sum_{t=1}^{n}\chi_t\chi_t' d_t^2\left(\gamma_0+\frac{\upsilon}{a_n},\gamma_0\right)\kappa_n\\
&=G_n\left(\upsilon\right)+2\sum_{t=1}^{n}\kappa_n'\left(\chi_t-\chi_t^*\right)\chi_t^{*\prime}\kappa_n d_t^2\left(\gamma_0+\frac{\upsilon}{a_n},\gamma_0\right)\\
&\quad+\sum_{t=1}^{n}\left[\kappa_n'\left(\chi_t-\chi_t^*\right)\right]^2 d_t^2\left(\gamma_0+\frac{\upsilon}{a_n},\gamma_0\right)\\
&=G_n\left(\upsilon\right)+2n^{-2\varrho}\sum_{t=1}^{n}\alpha_{L_n,0}'\left[\Phi_{L_n}\left(\hat{v}_t\right)-\Phi_{L_n}\left(v_t\right)\right]\eta_0^*\left(v_t\right)d_t^2\left(\gamma_0+\frac{\upsilon}{a_n},\gamma_0\right)\\
&\quad+n^{-2\varrho}\sum_{t=1}^{n}\left(\alpha_{L_n,0}'\left[\Phi_{L_n}\left(\hat{v}_t\right)-\Phi_{L_n}\left(v_t\right)\right]\right)^2 d_t^2\left(\gamma_0+\frac{\upsilon}{a_n},\gamma_0\right)\\
&=G_n\left(\upsilon\right)+A_n\left(\upsilon\right),
\end{aligned}$$

where $G_n\left(\upsilon\right)=\sum_{t=1}^{n}\left(\kappa_n'\chi_t^*\right)^2 d_t^2\left(\gamma_0+\upsilon/a_n,\gamma_0\right)$ is uniformly bounded in probability over $\upsilon\in\left[\underline{\upsilon},\overline{\upsilon}\right]$ by Lemma B.5 of the Online Appendix. As $\hat{\pi}_q-\pi_q=O_p\left(n^{-1/2}\right)$ and $\max_{1\le t\le n}\left|\eta_0^*\left(v_t\right)\right|<\infty$ under Assumptions 2(iii) and (iv), we have

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

https://doi.org/10.1017/S0266466621000256 Published online by Cambridge University Press

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

Secondly, we consider $S^*_{n,2}(\gamma) = \hat{\kappa}_n' \sum_{t=1}^n \varepsilon_t \chi_t d_t(\gamma, \gamma_0)$. For any given $\upsilon \in [\underline{\upsilon}, \overline{\upsilon}]$, we have

$$\begin{aligned} S^*_{n,2}(\upsilon) &= \hat{\kappa}_n' \sum_{t=1}^n \varepsilon_t \chi_t d_t\left(\gamma_0 + \frac{\upsilon}{a_n}, \gamma_0\right) \\ &= R_n(\upsilon) + \hat{\kappa}_n' \sum_{t=1}^n \varepsilon_t \left(\chi_t - \chi_t^*\right) d_t\left(\gamma_0 + \frac{\upsilon}{a_n}, \gamma_0\right) \\ &\quad + \left(\hat{\kappa}_n - \kappa_n\right)' \sum_{t=1}^n \varepsilon_t \chi_t^* d_t\left(\gamma_0 + \frac{\upsilon}{a_n}, \gamma_0\right) \\ &= R_n(\upsilon)\left[1 + o_p(1)\right] + O_p\left(n^{-1/2-\varrho+\min(\varrho,\varsigma)} \left\|\Phi_{L_n}\right\|_1\right) \end{aligned}$$

as $\left\|\hat{\kappa}_n - \kappa_n\right\| = o_p(\|\kappa_n\|)$ under Assumption 3(ii) by Lemma 1, where $R_n(\upsilon) = \sum_{t=1}^n \varepsilon_t \kappa_n' \chi_t^* d_t(\gamma_0 + \upsilon/a_n, \gamma_0) = O_e(1)$ holds uniformly over $\upsilon \in [\underline{\upsilon}, \overline{\upsilon}]$ by Lemma B.5 of the Online Appendix. Therefore, we show that the leading term of $S^*_{n,2}(\upsilon)$ is $R_n(\upsilon)$ under Assumption 3(ii).

Thirdly, we can show that $S^*_{n,4}(\upsilon) = (\hat{\kappa}_n + \kappa_n)' \sum_{t=1}^n \chi_t \chi_t' d_t^2(\gamma_0 + \upsilon/a_n, \gamma_0)(\hat{\kappa}_n - \kappa_n) = o_p(S^*_{n,1}(\upsilon))$ and $S^*_{n,3}(\upsilon) = o_p(S^*_{n,1}(\upsilon))$. And, taking above results together with Lemma B.5 of the Online Appendix, we have $Q_n(\upsilon) = -G_n(\upsilon) + 2R_n(\upsilon) + o_p(1)$ and

$$Q(\upsilon) = -\mu|\upsilon| + 2\sqrt{\tau_1^2} W_1(\upsilon) I(\upsilon \le 0) + 2\sqrt{\tau_2^2} W_2(\upsilon) I(\upsilon > 0).$$

Following the proof of Theorem 1 in Kourtellos et al. (2016), we complete the proof of this theorem. ■

**Proof of Theorem 2:** Under $H_0$, by Lemma B.5 of the Online Appendix and following the proof of Theorem 1, we have

$$\begin{aligned} &n^{-1} S_n\left(\hat{\theta}(\hat{\gamma}), \hat{\gamma}\right) LR_n(\gamma_0) - Q_n(\hat{\upsilon}) \\ &\quad = \left[S_n\left(\hat{\theta}(\gamma_0), \gamma_0\right) - S_n\left(\hat{\theta}(\hat{\gamma}), \hat{\gamma}\right)\right] - \left[S_n\left(\hat{\theta}, \gamma_0\right) - S_n\left(\hat{\theta}, \gamma_0 + \hat{\upsilon}/a_n\right)\right] \\ &\quad = S_n\left(\hat{\theta}(\gamma_0), \gamma_0\right) - S_n\left(\hat{\theta}, \gamma_0\right) \\ &\quad = \sum_{t=1}^n \left[y_t - \mathbf{x}_{t,\gamma_0}' \hat{\theta}(\gamma_0)\right]^2 - \sum_{t=1}^n \left(y_t - \mathbf{x}_{t,\gamma_0}' \hat{\theta}\right)^2 = o_p(1), \end{aligned}$$

where $S_n(\theta, \gamma) = \sum_{t=1}^n \left(y_t - \mathbf{x}_{t,\gamma}' \theta\right)^2$, $\hat{\gamma} = \gamma_0 + \hat{\upsilon}/a_n$ and $n^{-1} S_n\left(\hat{\theta}(\hat{\gamma}), \hat{\gamma}\right) \xrightarrow{p} \bar{\sigma}_\varepsilon^2$ following the proof of $S_{n3}$ in the proof of Lemma 1. Therefore, by Theorem 1 and the continuous mapping theorem we obtain

$$LR_n(\gamma_0) = \frac{\sup_\upsilon Q_n(\upsilon)}{\bar{\sigma}_\varepsilon^2} + o_p(1) \xrightarrow{d} \frac{\sup_\upsilon Q(\upsilon)}{\bar{\sigma}_\varepsilon^2},$$

https://doi.org/10.1017/S0266466621000256 Published online by Cambridge University Press

where denoting $\upsilon = \left(\tau_1^2/\mu^2\right)s$ and

$$\begin{aligned}
&\frac{\sup_{\upsilon} Q(\upsilon)}{\bar{\sigma}_\varepsilon^2} \\
&= \frac{1}{\bar{\sigma}_\varepsilon^2}\sup_{\upsilon} -\mu|\upsilon| + 2\sqrt{\tau_1^2}W_1(\upsilon)I(\upsilon \le 0) + 2\sqrt{\tau_2^2}W_2(\upsilon)I(\upsilon > 0) \\
&= \frac{1}{\bar{\sigma}_\varepsilon^2}\sup_{s} -\frac{\tau_1^2}{\mu}|s| + 2\sqrt{\tau_1^2}W_1\left(\frac{\tau_1^2}{\mu^2}s\right)I\left(\frac{\tau_1^2}{\mu^2}s \le 0\right) + 2\sqrt{\tau_2^2}W_2\left(\frac{\tau_1^2}{\mu^2}s\right)I\left(\frac{\tau_1^2}{\mu^2}s > 0\right) \\
&= \frac{\tau_1^2}{\bar{\sigma}_\varepsilon^2\mu}\sup_{s}[-|s| + 2W_1(s)]I(s \le 0) + \left[-|s| + 2\sqrt{\frac{\tau_2^2}{\tau_1^2}}W_2(s)\right]I(s > 0) \\
&= \eta^2\zeta
\end{aligned}$$

with $\eta^2 = \tau_1^2/\bar{\sigma}_\varepsilon^2\mu$. This completes the proof of this Theorem. ■

**Proof of Theorem 3:** Denote $\Delta_\upsilon = \left[\Delta_{-,\upsilon}, \Delta_{+,\upsilon}\right]$ and $\mathbf{X}_\upsilon = \left[\mathbf{X}_{-,\upsilon}, \mathbf{X}_{+,\upsilon}\right]$, where the $t$th row vector of $\Delta_{-,\upsilon}$, $\Delta_{+,\upsilon}$, $\mathbf{X}_{-,\upsilon}$ and $\mathbf{X}_{+,\upsilon}$ are $\Delta_{\upsilon,t}^- = \Phi_{L_n}(\hat{v}_t)I(q_t \le \gamma_0 + \upsilon/a_n)$, $\Delta_{\upsilon,t}^+ = \Phi_{L_n}(\hat{v}_t)I(q_t > \gamma_0 + \upsilon/a_n)$, $\mathbf{x}_{\upsilon,t}^- = \mathbf{x}_t I(q_t \le \gamma_0 + \upsilon/a_n)$, and $\mathbf{x}_{\upsilon,t}^+ = \mathbf{x}_t I(q_t > \gamma_0 + \upsilon/a_n)$, respectively. Also, denote $\mathbf{P}_\upsilon = \Delta_\upsilon\left(\Delta_\upsilon'\Delta_\upsilon\right)^{-1}\Delta_\upsilon'$ and $\hat{\upsilon} = a_n(\hat{\gamma} - \gamma_0)$. Applying the partitioned squares method gives $\hat{\beta} - \beta = \left[\mathbf{X}_{\hat{\upsilon}}'\left(\mathbf{I}_n - \mathbf{P}_{\hat{\upsilon}}\right)\mathbf{X}_{\hat{\upsilon}}\right]^{-1}\mathbf{X}_{\hat{\upsilon}}'\left(\mathbf{I}_n - \mathbf{P}_{\hat{\upsilon}}\right)\left(\mathbf{y} - \mathbf{X}_{\hat{\upsilon}}\beta\right)$.

Firstly, applying Lemmas B.1 and B.2 of the Online Appendix we obtain $n^{-1}\mathbf{X}_{\hat{\upsilon}}'\mathbf{X}_{\hat{\upsilon}} \xrightarrow{p} \Sigma_{\mathbf{x}\mathbf{x}',\gamma_0}$, $n^{-1}\mathbf{X}_{\hat{\upsilon}}'\Delta_{\hat{\upsilon}} = \Sigma_{\mathbf{x}\Phi_{L_n}',\gamma_0} + o_p(1)$ and $n^{-1}\Delta_{\hat{\upsilon}}'\Delta_{\hat{\upsilon}} = \Sigma_{\Phi_{L_n}\Phi_{L_n}',\gamma_0} + o_p(1)$. It then follows

$$n^{-1}\mathbf{X}_{\hat{\upsilon}}'\left(\mathbf{I}_n - \mathbf{P}_{\hat{\upsilon}}\right)\mathbf{X}_{\hat{\upsilon}} = \mathbf{J}_n + o_p(1), \tag{A.21}$$

where $\mathbf{J}_n$ is defined by (2.12).

Secondly, we consider $\mathbf{B}_n(\upsilon) = n^{-1}\mathbf{X}_\upsilon'\left(\mathbf{I}_n - \mathbf{P}_\upsilon\right)(\mathbf{y} - \mathbf{X}_\upsilon\beta)$, where $y_t - \mathbf{x}_{\upsilon,t}'\beta = \eta_{\upsilon,0,t} + h_2(v_t) + \eta_n(v_t)I(q_t \le \gamma_0) + \varepsilon_t$ and $\eta_{\upsilon,0,t} = -\delta_n'\mathbf{x}_t d_t(\gamma_0 + \upsilon/a_n, \gamma_0)$.

(i) We will show $n^{-1}\mathbf{X}_{\hat{\upsilon}}'\left(\mathbf{I}_n - \mathbf{P}_{\hat{\upsilon}}\right)\eta_{\hat{\upsilon},0} = o_p\left(n^{-1/2}\right)$, where $\eta_{\upsilon,0}$ is an $n \times 1$ vector with its $t$th element equal to $\eta_{\upsilon,0,t}$. By Lemma B.1 of the Online Appendix, $n^{-1}\sum_{t=1}^n \delta_n'\mathbf{x}_t\mathbf{x}_t' d_t(\gamma_0 + \upsilon/a_n, \gamma_0) = n^{-\varsigma}E\left[\delta_0'\mathbf{x}_t\mathbf{x}_t' d_t(\gamma_0 + \upsilon/a_n, \gamma_0)\right]\left[1 + o_p(1)\right] = O_p\left(a_n^{-1}n^{-\varsigma}\right)$ holds uniformly over $\upsilon \in \left[\underline{\upsilon}, \overline{\upsilon}\right]$. Hence, we obtain $n^{-1}\mathbf{X}_{\hat{\upsilon}}'\eta_{\hat{\upsilon},0} = O_p\left(a_n^{-1}n^{-\varsigma}\right)$. Moreover, we have

$$\begin{aligned}
&\left\|n^{-1}\mathbf{X}_{\hat{\upsilon}}'\mathbf{P}_{\hat{\upsilon}}\eta_{\hat{\upsilon},0}\right\| \le \left\|n^{-1}\mathbf{X}_{\hat{\upsilon}}'\Delta_{\hat{\upsilon}}\left(\Delta_{\hat{\upsilon}}'\Delta_{\hat{\upsilon}}/n\right)^{-1/2}\right\|_{sp}\left\|n^{-1}\left(\Delta_{\hat{\upsilon}}'\Delta_{\hat{\upsilon}}/n\right)^{-1/2}\Delta_{\hat{\upsilon}}'\eta_{\hat{\upsilon},0}\right\|_{sp} \\
&\le \lambda_{\max}^{1/2}\left(\Sigma_{\mathbf{x}\Phi_{L_n}',\gamma_0}\Sigma_{\Phi_{L_n}\Phi_{L_n}',\gamma_0}^{-1}\Sigma_{\Phi_{L_n}\mathbf{x}',\gamma_0}\right)\left[1 + o_p(1)\right]\sqrt{n^{-1}\eta_{\hat{\upsilon},0}'\eta_{\hat{\upsilon},0}} = O_p\left(a_n^{-1/2}n^{-\varsigma}\right).
\end{aligned}$$

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

https://doi.org/10.1017/S0266466621000256 Published online by Cambridge University Press

(ii) We consider $\left(\mathbf{I}_n - \mathbf{P}_{\hat{\upsilon}}\right)\mathbf{h}_2(\mathbf{v}) = \left(\mathbf{I}_n - \mathbf{P}_{\hat{\upsilon}}\right)\left[\mathbf{h}_2(\mathbf{v}) - \mathbf{h}_2^*(\hat{\mathbf{v}})\right]$, where $h_2(v_t) - h_2^*(\hat{v}_t) = \left[h_2(v_t) - h_2(\hat{v}_t)\right] + \left[h_2(\hat{v}_t) - h_2^*(\hat{v}_t)\right] = \eta_{\upsilon,1,t} + \eta_{\upsilon,2,t}$, where $\eta_{\upsilon,2,t} = h_2(\hat{v}_t) - h_2^*(\hat{v}_t)$. Let $\eta_{\upsilon,j} = \left[\eta_{\upsilon,j,1}, \ldots, \eta_{\upsilon,j,n}\right]'$ for $j = 1,2$. Applying again Lemma B.1 of the Online Appendix, we show that $n^{-1}\mathbf{X}_{\upsilon}'\eta_{\upsilon,2}$ and $n^{-1}\Delta_{\upsilon}'\eta_{\upsilon,2}$ are both of order $O_p\left(L_n^{-\xi}\right)$ uniformly over $\upsilon \in \left[\underline{\upsilon}, \overline{\upsilon}\right]$. Hence, we obtain $n^{-1}\mathbf{X}_{\hat{\upsilon}}'\left(\mathbf{I}_n - P_{\hat{\upsilon}}\right)\eta_{\hat{\upsilon},2} = O_p\left(L_n^{-\xi}\right) = o_p\left(n^{-1/2}\right)$ if $\sqrt{n}L_n^{-\xi} = o(1)$ or Assumption 3(iii) holds.

Next, as $h_2(v_t) - h_2(\hat{v}_t) = h_2^{(1)}(\bar{v}_t)\mathbf{z}_t'\left(\hat{\pi}_q - \pi_q\right)$, where $\bar{v}_t$ lies between $\hat{v}_t$ and $v_t$, closely following the proof of Lemma B.1 of the Online Appendix, we have

$$\max_{\upsilon\in[\underline{\upsilon},\overline{\upsilon}]}\left\| n^{-1}\sum_{t=1}^{n}\mathbf{w}_t e_t h_2^{(1)}(\bar{v}_t)\mathbf{z}_t' - E\left[\mathbf{w}_t e_t h_2^{(1)}(v_t)\mathbf{z}_t'\right]\right\| = o_p(1),$$

where $\mathbf{w}_t = \mathbf{x}_t$ or $\Phi_{L_n}(v_t)$ and $e_t = I(q_t \le \gamma_0 + \upsilon/a_n)$ or $e_t = I(q_t > \gamma_0 + \upsilon/a_n)$. It then follows $n^{-1}\mathbf{X}_{\hat{\upsilon}}'\eta_{\hat{\upsilon},1} = \Gamma_{\mathbf{x}}\left(\hat{\pi}_q - \pi_q\right)$ and $n^{-1}\Delta_{\hat{\upsilon}}'\eta_{\hat{\upsilon},1} = \Gamma_{\Phi_{L_n}}\left(\hat{\pi}_q - \pi_q\right)$. Hence, we have

$$n^{-1}\mathbf{X}_{\hat{\upsilon}}'\left(\mathbf{I}_n - \mathbf{P}_{\hat{\upsilon}}\right)\mathbf{h}_2(\mathbf{v}) = \mathbf{B}_{1n}\left(\hat{\pi}_q - \pi_q\right) + o_p\left(n^{-1/2}\right),$$

where $\mathbf{B}_{1n} = \Gamma_{\mathbf{x}} - \Sigma_{\mathbf{x}\Phi_{L_n}',\gamma_0}\Sigma_{\Phi_{L_n}\Phi_{L_n}',\gamma_0}^{-1}\Gamma_{\Phi_{L_n}}$ with $\Gamma_{\mathbf{x}} = [E(\mathbf{x}_t h_2^{(1)}(v_t)\mathbf{z}_t' I(q_t \le \gamma_0))', \Gamma_{\mathbf{x},2}']'$ and $\Gamma_{\Phi_{L_n}} = [E(\Phi_{L_n}(v_t)h_2^{(1)}(v_t)\mathbf{z}_t' I(q_t \le \gamma_0))', \Gamma_{\Phi_{L_n},2}']'$.

(iii) Let $\eta_{n,\gamma_0}(\mathbf{v})$ denote an $n \times 1$ vector and its $t$th element equals $\eta_n(v_t)I(q_t \le \gamma_0)$. Closely following the proofs for (i) and (ii) above, we show that $n^{-1}\mathbf{X}_{\hat{\upsilon}}'\left(\mathbf{I}_n - \mathbf{P}_{\hat{\upsilon}}\right)\eta_{n,\gamma_0}(\mathbf{v}) = n^{-1}\mathbf{X}_{\hat{\upsilon}}'\left(\mathbf{I}_n - \mathbf{P}_{\hat{\upsilon}}\right)\left[\eta_{n,\upsilon}(\mathbf{v}) - \eta_{n,\upsilon}(\hat{\mathbf{v}})\right] + O_p\left(L_n^{-\xi} + a_n^{-1}n^{-\varrho}\right)$, where $\eta_{n,\upsilon}(\mathbf{v})$ is an $n \times 1$ vector with its $t$th element equal to $\eta_n(v_t)I(q_t \le \gamma_0 + \upsilon/a_n)$. Following the proof for (ii) above, we have

$$n^{-1}\mathbf{X}_{\hat{\upsilon}}'\left(\mathbf{I}_n - \mathbf{P}_{\hat{\upsilon}}\right)\left[\eta_{n,\upsilon}(\mathbf{v}) - \eta_{n,\upsilon}(\hat{\mathbf{v}})\right] = \mathbf{B}_{2n}\left(\hat{\pi}_q - \pi_q\right) + o_p\left(n^{-1/2}\right),$$

where $\mathbf{B}_{2n} = \bar{\Gamma}_{\mathbf{x}} - \Sigma_{\mathbf{x}\Phi_{L_n}',\gamma_0}\Sigma_{\Phi_{L_n}\Phi_{L_n}',\gamma_0}^{-1}\bar{\Gamma}_{\Phi_{L_n}}$ with $\bar{\Gamma}_{\mathbf{x}} = \Big[E[\mathbf{x}_t(h_1^{(1)}(v_t) - h_2^{(1)}(v_t))\mathbf{z}_t' I(q_t \le \gamma_0)]', \mathbf{0}'\Big]'$, $\bar{\Gamma}_{\Phi_{L_n}} = \Big[E[\Phi_{L_n}(v_t)(h_1^{(1)}(v_t) - h_2^{(1)}(v_t))\mathbf{z}_t' I(q_t \le \gamma_0)]', \mathbf{0}'\Big]'$.

Taking together all the results above, we have

$$\begin{aligned}\mathbf{B}_n(\hat{\upsilon}) &= n^{-1}\mathbf{X}_{\hat{\upsilon}}'\left(\mathbf{I}_n - \mathbf{P}_{\hat{\upsilon}}\right)\left(\mathbf{y} - \mathbf{X}_{\hat{\upsilon}}\beta\right)\\ &= -\mathbf{B}_n\left(\hat{\pi}_q - \pi_q\right) + n^{-1}\mathbf{X}_{\hat{\upsilon}}'\left(\mathbf{I}_n - \mathbf{P}_{\hat{\upsilon}}\right)\varepsilon + o_p\left(n^{-1/2}\right)\\ &= -\mathbf{B}_n\left(\hat{\pi}_q - \pi_q\right) + n^{-1}\left(\mathbf{X}_0 - \Sigma_{\mathbf{x}\Phi_{L_n}',\gamma_0}\Sigma_{\Phi_{L_n}\Phi_{L_n}',\gamma_0}^{-1}\Delta_0\right)'\varepsilon + o_p\left(n^{-1/2}\right),\end{aligned}$$

where $\mathbf{B}_n$, $\mathbf{X}_0$ and $\Delta_0$ are defined in Section 2.2.2. Then, applying Wooldridge and White's central limit theorem for strong mixing process (White (2001), Thm. 5.20, p. 130) under Assumption 1(i), we obtain

$$\begin{pmatrix} n^{-1/2}\sum_{t=1}^{n}\mathbf{z}_t v_t \\ n^{-1/2}\left(\mathbf{X}_0 - \Sigma_{\mathbf{x}\Phi_{L_n}',\gamma_0}\Sigma_{\Phi_{L_n}\Phi_{L_n}',\gamma_0}^{-1}\Delta_0\right)'\varepsilon \end{pmatrix} \xrightarrow{d} N\left(\mathbf{0}, \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{12}' & \Omega_{22} \end{pmatrix}\right).$$

https://doi.org/10.1017/S0266466621000256 Published online by Cambridge University Press

Therefore, we obtain $\sqrt{n}(\mathbf{J}_n^{-1}\Omega_n\mathbf{J}_n^{-1})^{-1/2}\left(\hat{\beta}-\beta\right)\xrightarrow{d} N\left(\mathbf{0},\mathbf{I}_{2d_{\mathbf{x}}}\right)$, where $\mathbf{J}_n$ and $\Omega_n$ are defined by (2.12) and (2.13), respectively. This completes the proof of this theorem. ■

**Proof of Theorem 4:** Denoting $\mathcal{W}_t=\left[1,\left(\hat{v}_t-v\right)/h\right]'$, $\varkappa_{t,\gamma}=[1,I(q_t\le\gamma)]'$, and $\eta_{0,t,\gamma}(v)=\eta_0(v)I(q_t\le\gamma)$, we have

$$\begin{aligned}\left[\begin{array}{c}\tilde{\psi}(v)\\ \dot{\psi}(v)\end{array}\right]&\equiv\left[\frac{1}{nh}\sum_{t=1}^{n}K\left(\frac{\hat{v}_t-v}{h}\right)\left(\mathcal{W}_t\mathcal{W}_t'\right)\otimes\left(\varkappa_{t,\hat{\gamma}}\varkappa_{t,\hat{\gamma}}'\right)\right]^{-1}\\&\quad\times\frac{1}{nh}\sum_{t=1}^{n}K\left(\frac{\hat{v}_t-v}{h}\right)\left(\mathcal{W}_t\otimes\varkappa_{t,\hat{\gamma}}\right)\tilde{y}_t\\&=\left[\begin{array}{c}\psi(v)\\ h\psi^{(1)}(v)\end{array}\right]+\mathbf{A}_{n1}^{-1}\left(\mathbf{A}_{n2}/2+\mathbf{A}_{n3}+\mathbf{A}_{n4}\right),\end{aligned}$$

where $\tilde{y}_t=y_t-\hat{\beta}_1'\mathbf{x}_tI\left(q_t\le\hat{\gamma}\right)-\hat{\beta}_2'\mathbf{x}_tI\left(q_t>\hat{\gamma}\right)$, $\psi^{(s)}(v)=\partial^s\psi(v)/\partial v^s$ for an integer $s>0$, $\tilde{\psi}(v)$ and $\dot{\psi}(v)$ are the estimator for $\psi(v)$ and $h\psi^{(1)}(v)$, respectively, and

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

with $\lambda_t=\mathbf{x}_t'\left(\beta_2-\hat{\beta}_2\right)+\delta_n'\mathbf{x}_tI\left(q_t\le\gamma_0\right)-\hat{\delta}_n'\mathbf{x}_tI\left(q_t\le\hat{\gamma}\right)+h_2(v_t)-h_2\left(\hat{v}_t\right)-n^{-\varrho}\left[\eta_0(v_t)I(q_t\le\gamma_0)-\eta_0\left(\hat{v}_t\right)I\left(q_t\le\hat{\gamma}\right)\right]$ and $\bar{v}_t$ lying between $\hat{v}_t$ and $v$.

Firstly, we calculate

$$\begin{aligned}K\left(\frac{\hat{v}_t-v}{h}\right)&=K\left(\frac{v_t-v}{h}\right)+K^{(1)}\left(\frac{v_t-v}{h}\right)\frac{\mathbf{z}_t'\left(\pi_q-\hat{\pi}_q\right)}{h}\\&\quad+\cdots+\frac{1}{r!}K^{(r_k)}\left(\frac{\zeta_t\mathbf{z}_t'\left(\pi_q-\hat{\pi}_q\right)+v_t-v}{h}\right)\left(\frac{\mathbf{z}_t'\left(\pi_q-\hat{\pi}_q\right)}{h}\right)^{r_k},\end{aligned}\tag{A.22}$$

where $K^{(s)}(\cdot)$ denotes the $s$th order derivative of $K(\cdot)$, and $\zeta_t\in(0,1)$ over all $t$ for some $r_k>2$. Therefore, we obtain

$$\begin{aligned}\mathbf{A}_{n1}&=\frac{1}{nh}\sum_{t=1}^{n}K\left(\frac{v_t-v}{h}\right)\left(\mathcal{W}_t\mathcal{W}_t'\right)\otimes\left(\varkappa_{t,\hat{\gamma}}\varkappa_{t,\hat{\gamma}}'\right)\\&\quad+\frac{1}{nh}\sum_{j=1}^{r_k-1}\frac{1}{j!}\sum_{t=1}^{n}K^{(j)}\left(\frac{v_t-v}{h}\right)\left(\frac{\mathbf{z}_t'\left(\pi_q-\hat{\pi}_q\right)}{h}\right)^j\left(\mathcal{W}_t\mathcal{W}_t'\right)\otimes\left(\varkappa_{t,\hat{\gamma}}\varkappa_{t,\hat{\gamma}}'\right)\end{aligned}$$

https://doi.org/10.1017/S0266466621000256 Published online by Cambridge University Press

$$+\frac{1}{nhr_k!}\sum_{t=1}^{n}K^{(r_k)}\left(\frac{\zeta_t\mathbf{z}_t'\left(\pi_q-\hat{\pi}_q\right)+v_t-v}{h}\right)\left(\frac{\mathbf{z}_t'\left(\pi_q-\hat{\pi}_q\right)}{h}\right)^{r_k}$$
$$\times\left(\mathcal{W}_t\mathcal{W}_t'\right)\otimes\left(\varkappa_{t,\hat{\gamma}}\varkappa_{t,\hat{\gamma}}'\right)$$
$$=\mathbf{A}_{n1,1}+\mathbf{A}_{n1,2}+\mathbf{A}_{n1,3},$$

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

$$\mathbf{A}_{n1,1}=f_v(v)\begin{bmatrix}1&0\\0&\kappa_{1,2}\end{bmatrix}\otimes E\left(\varkappa_{t,\gamma_0}\varkappa_{t,\gamma_0}'|v_t=v\right)$$
$$+O_p\left(h^2+(nh)^{-1/2}+\left(\sqrt{n}h\right)^{-1}+n^{2\min(\varsigma,\varrho)-1}\right)$$

under Assumptions 1(i) and 4(ii). In addition, under Assumption 5, we have

$$\mathbf{A}_{n1,2}=\frac{1}{nh}\sum_{j=1}^{r_k-1}\frac{1}{j!}\sum_{t=1}^{n}K^{(j)}\left(\frac{v_t-v}{h}\right)\left(\frac{\mathbf{z}_t'\left(\pi_q-\hat{\pi}_q\right)}{h}\right)^{j}\left(\mathcal{W}_t\mathcal{W}_t'\right)\otimes\left(\varkappa_{t,\hat{\gamma}}\varkappa_{t,\hat{\gamma}}'\right)$$
$$\leq\frac{1}{nh}\sum_{j=1}^{r_k-1}\frac{1}{j!}\sum_{t=1}^{n}\left|K^{(j)}\left(\frac{v_t-v}{h}\right)\left(\frac{\mathbf{z}_t'\left(\pi_q-\hat{\pi}_q\right)}{h}\right)^{j}\right|\left\|\left(\mathcal{W}_t\mathcal{W}_t'\right)\otimes\left(\varkappa_{t,\hat{\gamma}}\varkappa_{t,\hat{\gamma}}'\right)\right\|$$
$$=O_p\left(\frac{1}{\sqrt{n}h}\right)$$

and

$$\mathbf{A}_{n1,3}=\frac{1}{nhr_k!}\sum_{t=1}^{n}K^{(r_k)}\left(\frac{\zeta_t\mathbf{z}_t'\left(\pi_q-\hat{\pi}_q\right)+v_t-v}{h}\right)\left(\frac{\mathbf{z}_t'\left(\pi_q-\hat{\pi}_q\right)}{h}\right)^{r_k}$$
$$\times\left(\mathcal{W}_t\mathcal{W}_t'\right)\otimes\left(\varkappa_{t,\hat{\gamma}}\varkappa_{t,\hat{\gamma}}'\right)$$
$$\leq\frac{M}{nhr_k!}\sum_{t=1}^{n}\left|\left(\frac{\mathbf{z}_t'\left(\pi_q-\hat{\pi}_q\right)}{h}\right)^{r_k}\right|\left\|\left(\mathcal{W}_t\mathcal{W}_t'\right)\otimes\left(\varkappa_{t,\hat{\gamma}}\varkappa_{t,\hat{\gamma}}'\right)\right\|=O_p\left(n^{-r_k/2}h^{-r_k-3}\right).$$

Therefore, under Assumption 5(ii), we have $\mathbf{A}_{n1,2}=o_p(1)$ and $\mathbf{A}_{n1,3}=o_p(1)$.

Similarly, we can show that

$$\frac{\mathbf{A}_{n2}}{h^2}=f_v(v)\begin{bmatrix}\kappa_{1,2}\\0\end{bmatrix}\otimes\begin{bmatrix}h_2^{(2)}(v)+n^{-\varrho}\eta_0^{(2)}(v)E\left[I\left(q_t\leq\gamma_0\right)|v\right]\\\left[h_2^{(2)}(v)+n^{-\varrho}\eta_0^{(2)}(v)\right]E\left[I\left(q_t\leq\gamma_0\right)|v\right]\end{bmatrix}$$
$$+O_p\left(h^2+(nh)^{-1/2}+\left(\sqrt{n}h\right)^{-1}+n^{2\min(\varsigma,\varrho)-1}\right)$$
$$+O_p\left(n^{-(r_k+2)/2}\left[h^{-(r_k+3)},h^{-(r_k+4)}\right]'\right)$$

and by Theorem 3 and under Assumptions 3(i), 4 and 5, we have

$$\mathbf{A}_{n3}=O_p\left(\left\|\Phi_{L_n}\right\|_1 n^{-1/2}\left(h^2+(nh)^{-1/2}+\left(\sqrt{n}h\right)^{-1}+n^{2\min(\varsigma,\varrho)-1}\right)\right)$$
$$+O_p\left(\left\|\Phi_{L_n}\right\|_1 n^{-(r_k+1)/2}\left[h^{-(r_k+1)},h^{-(r_k+3)}\right]\right)=o_p(1).$$

https://doi.org/10.1017/S0266466621000256 Published online by Cambridge University Press

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

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

where we have

$$\begin{aligned}
\sqrt{nh}\mathbf{A}_{n4,1} &= \frac{1}{\sqrt{nh}}\sum_{t=1}^{n} K\left(\frac{\hat{v}_t - v}{h}\right)\left(\boldsymbol{\mathcal{W}}_t \otimes \varkappa_{t,\gamma_0}\right)\varepsilon_t \\
&= \frac{1}{\sqrt{nh}}\sum_{t=1}^{n} \varepsilon_t K\left(\frac{v_t - v}{h}\right)\left(\begin{bmatrix} 1 \\ (v_t - v)/h \end{bmatrix} \otimes \varkappa_{t,\gamma_0}\right) \\
&\quad + O_p\left(\left(nh^{3/2}\right)^{-1}\right) + O_p\left(n^{-r_k/2}\left[h^{-(r_k+1)}, h^{-(r_k+2)}\right]'\right) \\
&\xrightarrow{d} N\left(0, f_v(v)\begin{bmatrix} \kappa_{2,0} & 0 \\ 0 & \kappa_{2,2} \end{bmatrix} \otimes E\left(\varepsilon_t^2 \varkappa_{t,\gamma_0}\varkappa_{t,\gamma_0}' | v_t = v\right)\right)
\end{aligned}$$

by (A.22), the central limit theorem of Theorem 5.20 in White (2001) and the Cramér–Wold device under Assumptions 1, 4, and 5, and

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

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

**Proof of Theorem 5:** We only need to give proofs when $\beta_1 = \beta_2 = \beta$ and $h_1(\cdot) \equiv h_2(\cdot) \equiv 0$, under which models (2.1) and (2.2) becomes $y_t = \mathbf{x}_t'\beta + u_t$, where $u_t = u_{1t}I(q_t \le \gamma_0) + u_{2t}I(q_t > \gamma_0)$. It follows that $\mathbf{y}'\mathbf{M}_{\gamma,\mathbf{Q}}\mathbf{y} = \mathbf{u}'M_{\gamma,\mathbf{Q}}\mathbf{u} = \mathbf{u}'\mathbf{u} - \mathbf{u}'\mathbf{Q}_\gamma\left(\mathbf{Q}_\gamma'\mathbf{Q}_\gamma\right)^{-1}\mathbf{Q}_\gamma'\mathbf{u}$, where the first term does not depend on $\gamma$. Denoting $\lambda = E[I(q_t \le \gamma)] = F_q(\gamma)$, $\underline{\lambda} = F_q\left(\underline{\gamma}\right)$, $\bar{\lambda} = F_q(\bar{\gamma})$, and $\hat{\lambda} = F_q\left(\hat{\gamma}\right)$, we can rewrite the optimization problem (3.4) in terms of $\lambda$ as follows:

$$\hat{\lambda} \Rightarrow \lambda^* = \arg\max_{\lambda \in [\underline{\lambda}, \bar{\lambda}]} \mathbf{u}'\mathbf{Q}_\lambda\left(\mathbf{Q}_\lambda'\mathbf{Q}_\lambda\right)^{-1}\mathbf{Q}_\lambda'\mathbf{u}.$$

By Cor. 4.2 in Wooldridge and White (1988) and closely following the proof of Thm. 1 in Caner and Hansen (2001) for $s \in [0,1]$, we obtain $[n^{-1/2}\Sigma_{u,\mathbf{x}\mathbf{x}'}^{-1/2}\sum_{t=1}^{[ns]}\mathbf{x}_t u_t, n^{-1/2}\Sigma_{u,\mathbf{x}\mathbf{x}',\gamma,-}^{-1/2}\sum_{t=1}^{[ns]}\mathbf{x}_t u_t I(q_t \le \gamma)] \Rightarrow [\mathbf{B}_u(s), \mathbf{B}_u(s,\lambda)]$, where we denote $\Sigma_{u,\mathbf{x}\mathbf{x}'} = E(u_t^2\mathbf{x}_t\mathbf{x}_t')$, $\Sigma_{u,\mathbf{x}\mathbf{x}',\gamma,-} = E[u_t^2\mathbf{x}_t\mathbf{x}_t'I(q_t \le \gamma)]$, $\mathbf{B}_u(\lambda)$ and $\mathbf{B}_u(s,\lambda)$ are the $d_\mathbf{x}$-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.

https://doi.org/10.1017/S0266466621000256 Published online by Cambridge University Press

It then follows that

$$\begin{aligned}
&\mathbf{u}'\mathbf{Q}_\gamma\left(\mathbf{Q}_\gamma'\mathbf{Q}_\gamma\right)^{-1}\mathbf{Q}_\gamma'\mathbf{u}\\
&\quad=\sum_{t=1}^{n}\mathbf{x}_t'u_tI\left(q_t\le\gamma\right)\left(\sum_{t=1}^{n}\mathbf{x}_t\mathbf{x}_t'I\left(q_t\le\gamma\right)\right)^{-1}\sum_{t=1}^{n}\mathbf{x}_tu_tI\left(q_t\le\gamma\right)\\
&\qquad+\sum_{t=1}^{n}\mathbf{x}_t'u_tI\left(q_t>\gamma\right)\left(\sum_{t=1}^{n}\mathbf{x}_t\mathbf{x}_t'I\left(q_t>\gamma\right)\right)^{-1}\sum_{t=1}^{n}\mathbf{x}_tu_tI\left(q_t>\gamma\right)\\
&\quad\Rightarrow\mathbf{B}_u'\left(1,\lambda\right)\Sigma_{u,\mathbf{xx}',\gamma,-}^{1/2}\Sigma_{\mathbf{xx}',\gamma,-}^{-1}\Sigma_{u,\mathbf{xx}',\gamma,-}^{1/2}\mathbf{B}_u\left(1,\lambda\right)\\
&\qquad+\left(\Sigma_{u,\mathbf{xx}'}^{1/2}\mathbf{B}_u\left(1\right)-\Sigma_{u,\mathbf{xx}',\gamma,-}^{1/2}\mathbf{B}_u\left(1,\lambda\right)\right)'\Sigma_{\mathbf{xx}',\gamma,+}^{-1}\\
&\qquad\left(\Sigma_{u,\mathbf{xx}'}^{1/2}\mathbf{B}_u\left(1\right)-\Sigma_{u,\mathbf{xx}',\gamma,-}^{1/2}\mathbf{B}_u\left(1,\lambda\right)\right)\equiv W_u(\lambda),
\end{aligned}$$

where $\gamma=F_q^{-1}(\lambda)$, $\Sigma_{\mathbf{xx}'}=E\left(\mathbf{x}_t\mathbf{x}_t'\right)$, $\Sigma_{\mathbf{xx}',\gamma,-}=E\left[\mathbf{x}_t\mathbf{x}_t'I\left(q_t\le\gamma\right)\right]$, and $\Sigma_{\mathbf{xx}',\gamma,+}=E\left[\mathbf{x}_t\mathbf{x}_t'I\left(q_t>\gamma\right)\right]$. We therefore obtain

$$\hat{\lambda}\Rightarrow\lambda^*=\arg\max_{\lambda\in\left[\underline{\lambda},\bar{\lambda}\right]}W_u(\lambda). \tag{A.23}$$

This completes the proof of this theorem. ■

**Proof of Theorem 6:** For a given $\gamma$, we have $\hat{\alpha}_{L_n}(\gamma)=\left(\hat{\Phi}_{L_n,\gamma}^{*\prime}\hat{\Phi}_{L_n,\gamma}^{*}\right)^{-1}\hat{\Phi}_{L_n,\gamma}^{*\prime}\mathbf{y}^*$, so that

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

where we denote $\hat{\Phi}_{L_n,\gamma}^{*}=\mathbf{M}_{\gamma,\mathbf{Q}}\hat{\Phi}_{L_n,\gamma}$ and $\mathbf{y}^*=\mathbf{M}_{\gamma,\mathbf{Q}}\mathbf{y}$.

Applying tedious but straightforward calculations, we obtain the respective $t$th row vector of $\hat{\Phi}_{L_n,\gamma}^{-,*}=\mathbf{M}_{\gamma,\mathbf{Q}}\Phi_{L_n,\gamma}^{-}\left(\hat{\mathbf{v}}\right)$ and $\hat{\Phi}_{L_n,\gamma}^{-,*}=\mathbf{M}_{\gamma,\mathbf{Q}}\Phi_{L_n,\gamma}^{+}\left(\hat{\mathbf{v}}\right)$, respectively, as follows

$$\begin{aligned}
&\mathbf{a}_{\gamma,t,-}'=\left[\Phi_{L_n}\left(\hat{v}_t\right)-\hat{\pi}_{\Phi,\gamma,-}\mathbf{x}_t\right]'I\left(q_t\le\gamma\right)\text{ and}\\
&\mathbf{a}_{\gamma,t,+}'=\left[\Phi_{L_n}\left(\hat{v}_t\right)-\hat{\pi}_{\Phi,\gamma,+}\mathbf{x}_t\right]'I\left(q_t>\gamma\right),
\end{aligned}$$

where we denote

$$\begin{aligned}
&\Sigma_{n,\mathbf{xx}'}=n^{-1}\sum_{t=1}^{n}\mathbf{x}_t\mathbf{x}_t',\ \Sigma_{n,\mathbf{xx}',\gamma,-}=n^{-1}\sum_{t=1}^{n}\mathbf{x}_t\mathbf{x}_t'I\left(q_t\le\gamma\right),\\
&\hat{\pi}_{\Phi,\gamma,-}=n^{-1}\sum_{t=1}^{n}\Phi_{L_n}\left(\hat{v}_t\right)\mathbf{x}_t'I\left(q_t\le\gamma\right)\Sigma_{n,\mathbf{xx}',\gamma,-}^{-1},\\
&\hat{\pi}_{\Phi,\gamma,+}=n^{-1}\sum_{t=1}^{n}\Phi_{L_n}\left(\hat{v}_t\right)\mathbf{x}_t'I\left(q_t>\gamma\right)\left(\Sigma_{n,\mathbf{xx}'}-\Sigma_{n,\mathbf{xx}',\gamma,-}\right)^{-1}.
\end{aligned}$$

https://doi.org/10.1017/S0266466621000256 Published online by Cambridge University Press

It follows that

$$\hat{\Phi}_{L_n,\gamma}^{*\prime}\mathbf{y}^* = \left[\left(\hat{\Phi}_{L_n,\gamma}^{-,*\prime}\mathbf{y}^*\right)', \left(\hat{\Phi}_{L_n,\gamma}^{+,*\prime}\mathbf{y}^*\right)'\right]' = \left[\sum_{t=1}^{n} y_t \mathbf{a}'_{\gamma,t,-}, \sum_{t=1}^{n} y_t \mathbf{a}'_{\gamma,t,+}\right]'$$

and

$$\hat{\Phi}_{L_n,\gamma}^{*\prime}\hat{\Omega}_{n,\gamma}\hat{\Phi}_{L_n,\gamma}^{*} = \begin{bmatrix} \Delta_{n,\gamma}^{-} & \mathbf{0}_{L_n\times L_n} \\ \mathbf{0}_{L_n\times L_n} & \Delta_{n,\gamma}^{+} \end{bmatrix},$$

where $\hat{\varepsilon}_{\gamma,t} = \hat{\varepsilon}_{\gamma,t,0} + \hat{\varepsilon}_{\gamma,t}^{-} + \hat{\varepsilon}_{\gamma,t}^{+}$, $\hat{\varepsilon}_{\gamma,t,0} = -\mathbf{x}_t'\delta_n d_t(\gamma,\gamma_0) + u_t$, $\Delta_{n,\gamma}^{-} = \sum_{t=1}^{n}\left(\hat{\varepsilon}_{\gamma,t,0}+\hat{\varepsilon}_{\gamma,t}^{-}\right)^2/\hat{h}_{\gamma,tt}\mathbf{a}_{\gamma,t,-}\mathbf{a}'_{\gamma,t',-}$, $\Delta_{n,\gamma}^{+} = \sum_{t=1}^{n}\left(\hat{\varepsilon}_{\gamma,t,0}+\hat{\varepsilon}_{\gamma,t}^{+}\right)^2/\hat{h}_{\gamma,tt}\mathbf{a}_{\gamma,t,+}\mathbf{a}'_{\gamma,t',+}$, and

$$\hat{\varepsilon}_{\gamma,t}^{-} = \left[\mathbf{x}_t'\left(\beta_1 - \hat{\beta}_1(\gamma)\right) - \hat{\alpha}'_{L_n,1}(\gamma)\,\Phi_{L_n}(\hat{v}_t)\right] I(q_t \le \gamma),$$
$$\hat{\varepsilon}_{\gamma,t}^{+} = \left[\mathbf{x}_t'\left(\beta_2 - \hat{\beta}_2(\gamma)\right) - \hat{\alpha}'_{L_n,2}(\gamma)\,\Phi_{L_n}(\hat{v}_t)\right] I(q_t > \gamma).$$

It is readily seen that $n^{-1}\Delta_{n,\gamma}^{-} \approx n^{-1}\sum_{t=1}^{n}\varepsilon_t^2/h_{\gamma,tt}\mathbf{a}_{\gamma,t,-}\mathbf{a}'_{\gamma,t',-}$ and $n^{-1}\Delta_{n,\gamma}^{+} \approx n^{-1}\sum_{t=1}^{n}\varepsilon_t^2/h_{\gamma,tt}\mathbf{a}_{\gamma,t,+}\mathbf{a}'_{\gamma,t',+}$ under $\mathrm{H}_0$, where $h_{\gamma,tt} = 1 - a'_{\gamma,t,-}\left(\sum_{t=1}^{n}\mathbf{a}_{\gamma,t,-}\mathbf{a}'_{\gamma,t,-}\right)^{-1}\mathbf{a}_{\gamma,t,-} - \mathbf{a}'_{\gamma,t,+}\left(\sum_{t=1}^{n}\mathbf{a}_{\gamma,t,+}\mathbf{a}'_{\gamma,t,+}\right)^{-1}\mathbf{a}_{\gamma,t,+} = 1 + O_p(L_n/n)$ uniformly over $t$.

Firstly, when $\beta_1 \neq \beta_2$, $\widehat{\gamma} = \gamma_0 + O_p\left(n^{-1+2\varsigma}\right)$ by Theorem 5(i) and $y_t = \mathbf{x}_t'\beta_1 I(q_t \le \gamma_0) + \mathbf{x}_t'\beta_2 I(q_t > \gamma_0) + u_t$. In the proof of Theorem 3, we have $\left\|\hat{\beta} - \beta\right\| = O_p\left(n^{-1/2}\right)$. For simplification of exposition, we denote $\chi_{n,\omega,1} = n^{-1}\sum_{t=1}^{n}\omega_t\mathbf{x}_t' I\left(q_t \le \hat{\gamma}\right) I(q_t \le \gamma_0)$, $\chi_{n,\omega,2} = n^{-1}\sum_{t=1}^{n}\omega_t\mathbf{x}_t' I\left(q_t \le \hat{\gamma}\right)$, and $\lambda_{n,\omega} = n^{-1}\sum_{t=1}^{n}\varepsilon_t\omega_t I\left(q_t \le \hat{\gamma}\right)$ for $\omega_t = \Phi_{L_n}\left(\hat{v}_t\right)$ and $\mathbf{x}_t$. Then, we have $\hat{\pi}_{\Phi,\gamma,-} = \chi_{n,\Phi,2}\chi_{n,\mathbf{x},2}^{-1}$ and

$$\begin{aligned}
&n^{-1}\hat{\Phi}_{L_n,\hat{\gamma}}^{-,*\prime}\mathbf{y}^* \\
&\quad = \left(\chi_{n,\Phi,1} - \chi_{n,\Phi,2}\chi_{n,\mathbf{x},2}^{-1}\chi_{n,\mathbf{x},1}\right)\delta_n + \left(\lambda_{n,\Phi} - \chi_{n,\Phi,2}\chi_{n,\mathbf{x},2}^{-1}\lambda_{n,\mathbf{x}}\right) \\
&\quad = n^{-1}\sum_{t=1}^{n}\varepsilon_t\Phi_{L_n}(v_t)\,I(q_t \le \gamma_0) - n^{-1}\sum_{t=1}^{n}\Phi_{L_n}(v_t)\,\mathbf{x}_t' I(q_t \le \gamma_0) \\
&\qquad \times \left[n^{-1}\sum_{t=1}^{n}\mathbf{x}_t\mathbf{x}_t' I(q_t \le \gamma_0)\right]^{-1} n^{-1}\sum_{t=1}^{n}\varepsilon_t\mathbf{x}_t I(q_t \le \gamma_0) \\
&\qquad + O_p\left(n^{-1+2\varsigma}\left(\left\|\Phi_{L_n}\right\|_0 n^{-\varsigma} + \left\|\Phi_{L_n}\right\|_1/n\right)\right)
\end{aligned}$$

https://doi.org/10.1017/S0266466621000256 Published online by Cambridge University Press

by Lemma B.6 of the Online Appendix. By Lemma B.1 of the Online Appendix, we have

$$n^{-1}\hat{\Phi}^{-,*\prime}_{L_n,\hat{\gamma}}\mathbf{y}^* \approx n^{-1}\sum_{t=1}^{n}\varepsilon_t \mathbf{c}_{-,t}(\gamma_0) = O_p\left(\left\|\Phi_{L_n}\right\|_0/\sqrt{n}\right), \tag{A.24}$$

$$n^{-1}\hat{\Phi}^{+,*\prime}_{L_n,\hat{\gamma}}\mathbf{y}^* \approx n^{-1}\sum_{t=1}^{n}\varepsilon_t \mathbf{c}_{+,t}(\gamma_0) = O_p\left(\left\|\Phi_{L_n}\right\|_0/\sqrt{n}\right), \tag{A.25}$$

where we denote

$$\mathbf{c}_{-,t}(\gamma_0) = \left\{\Phi_{L_n}(v_t) - E\left[\Phi_{L_n}(v_t)\mathbf{x}_t' I(q_t \le \gamma_0)\right]\Sigma^{-1}_{\mathbf{xx}',\gamma_0,-}\mathbf{x}_t\right\} I(q_t \le \gamma_0),$$

$$\mathbf{c}_{+,t}(\gamma_0) = \left\{\Phi_{L_n}(v_t) - E\left[\Phi_{L_n}(v_t)\mathbf{x}_t' I(q_t > \gamma_0)\right]\left(\Sigma_{\mathbf{xx}'} - \Sigma_{\mathbf{xx}',\gamma_0,-}\right)^{-1}\mathbf{x}_t\right\} I(q_t > \gamma_0),$$

where $\Sigma_{\mathbf{xx}',\gamma_0,-}$ is given by $\Sigma_{\mathbf{xx}',\gamma,-}$ evaluated at $\gamma_0$. $n^{-1}\Delta^{-}_{n,\hat{\gamma}} \approx n^{-1}\sum_{t=1}^{n}\varepsilon_t^2 c_{-,t}(\gamma_0)c'_{-,t}(\gamma_0)$ and $n^{-1}\Delta^{+}_{n,\hat{\gamma}} \approx n^{-1}\sum_{t=1}^{n}\varepsilon_t^2 c_{+,t}(\gamma_0)c'_{+}(\gamma_0)$.

We denote $\Omega_{n,-} = n^{-1}\bar{\sigma}_\varepsilon^2\sum_{t=1}^{n}\mathbf{c}_{-,t}(\gamma_0)\mathbf{c}'_{-,t}(\gamma_0)$, and $\Omega_{n,+} = n^{-1}\bar{\sigma}_\varepsilon^2\sum_{t=1}^{n}\mathbf{c}_{+,t}(\gamma_0)\mathbf{c}'_{+,t}(\gamma_0)$. We then obtain $W_n(\hat{\gamma}) \approx n^{-1}\sum_{t=1}^{n}\varepsilon_t\mathbf{c}'_{-,t}(\gamma_0)\Omega^{-1}_{n,-}\sum_{t=1}^{n}\varepsilon_t\mathbf{c}_{-,t}(\gamma_0) + n^{-1}\sum_{t=1}^{n}\varepsilon_t\mathbf{c}'_{+,t}\Omega^{-1}_{n,-}\sum_{t=1}^{n}\varepsilon_t\mathbf{c}_{+,t}$, which is a U-statistic with an expectation equal to $2L_n$. Closely following the proof of Thm. 2.1 in Gao and Anh (2000), we show that $J_n(\hat{\gamma})$ converges to the standard normal distribution under $H_0$.

When $\beta_1 = \beta_2$ and $H_0$ holds true, following the similar proof method with $\gamma_0$ being replaced by the random variable $F_q^{-1}(\lambda^*)$. This completes the proof of this theorem. ■

## SUPPLEMENTARY MATERIAL

To view the supplementary material for this article, please visit: https://doi.org/10.1017/S0266466621000256

### *REFERENCES*


Andrews, D. W. K. (1991) Asymptotic normality of series estimators for nonparametric and semiparametric regression models. *Econometrica* 59, 307–345.

Blundell, R., X. Chen & D. Kristensen (2007) Semiparametric IV estimation of shape invariant engel curves. *Econometrica* 75, 1613–1669.

Caner, M. & B. Hansen (2004) Instrumental variable estimation of a threshold model. *Econometric Theory* 20, 813–843.

Caner, M. & B. E. Hansen (2001) Threshold autoregression with a unit root. *Econometrica* 69, 1555–1596.

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 & E. E. Leamer (eds.), *Handbook of Econometrics*, vol. 6B, pp. 5549–5632. Springer.

Christopoulos, D., P. McAdam & E. Tzavalis (2021) Dealing with endogeneity in threshold models using copulas. *Journal of Business and Economic Statistics* 39, 166–178.

Delgado, M. S., D. Ozabaci, Y. Sun & S. C. Kumbhakar (2020) Smooth coefficient models with endogenous environmental variables. *Econometric Reviews* 39, 158–180.

Durlauf, S. N. (1996) A theory of persistent income inequality. *Journal of Economic Growth* 1, 75–93.


https://doi.org/10.1017/S0266466621000256 Published online by Cambridge University Press

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

Galor, O. & J. Zeira (1993) Income distribution and macroeconomics. *Review of Economic Studies* 60, 35–52.

Gao, J. & V. Anh (2000) A central limit theorem for a random quadratic form of strictly stationary processes. *Statistics and Probability Letters* 49, 69–79.

Gonzalo, J. & J. Y. Pitarakis (2007) Inferring the predictability induced by a persistent regressor in a predictive threshold model. *Journal of Business and Economic Statistics* 35, 202–217.

Gonzalo, J. & 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. & R. Tibshirani (1993) Varying parameter models. *Journal of the Royal Statistical Society, Series B* 55, 757–796.

Horn, S. D., R. A. Horn & D. B. Duncan (1975) Estimating heteroscedastic variances in linear model. *Journal of the American Statistical Association* 70, 380–385.

Kapetanios, G. (2010) Testing for exogeneity in threshold models. *Econometric Theory* 26, 231–259.

Kourtellos, A., T. Stengos & C. M. Tan (2016) Structural threshold regression. *Econometric Theory* 32, 827–860.

Lee, S., M. H. Seo & Y. Shin (2016) The lasso for high dimensional regression with a possible change point. *Journal of the Royal Statistical Society. Series B* 78, 193–210.

Li, D. & 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, 147–168.

Newey, W. K. (2009) Two-step series estimation of sample selection models. *The Econometrics Journal* 12, 217–229.

Newey, W. K. & D. McFadden (1994) *Large Sample Estimation and Hypothesis Testing* (Vol. IV. Elsevier Science B.V., pp. 2111–2245).

Reinhart, C. M. & K. S. Rogoff (2010) Growth in time of debt. *American Economic Review* 100, 573–578.

Seber, G. A. F. (2008) *A Matrix Handbook for Statisticians*. Wiley.

Seo, M. H. & O. Linton (2007) A smoothed least squares estimator for threshold regression models. *Journal of Econometrics* 141, 704–735.

Seo, M. H. & Y. Shin (2016) Dynamic panels with threshold effect and endogeneity. *Journal of Econometrics* 195, 169–186.

Terasvirta, T., W. J. Granger & R. R. Nelson (1986) Aspects of modelling nonlinear time series. In T. Terasvirta, W. J. Granger, & R. R. Nelson (eds.), *Handbook of Econometrics*, 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.

Wooldridge, J. M. & H. White (1988) Some invariance principles and central limit theorems for dependent heterogeneous processes. *Econometrica* 4, 210–230.

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.

Yu, P. (2015) Adaptive estimation of the threshold point in threshold regression. *Journal of Econometrics* 189, 83–100.

Yu, P., Q. Liao, & P.C.B. Phillips (2020) New Control Function Approaches in Threshold Regression with Endogeneity, Working paper, University of Hong Kong. http://web.hku.hk/~pingyu/WorkingPapers/EndoTR_CF.pdf

Yu, P. & C. B. Phillips (2018) Threshold regression with endogeneity. *Journal of Econometrics* 203, 50–68.

https://doi.org/10.1017/S0266466621000256 Published online by Cambridge University Press