Structural Threshold Regression*

Andros Kourtellos
Department of Economics
University of Cyprus†

Thanasis Stengos
Department of Economics
University of Guelph‡

Chih Ming Tan
Department of Economics
Clark University§

PRELIMINARY DRAFT
PLEASE DO NOT CITE OR CIRCULATE WITHOUT PERMISSION

October 15, 2011

Abstract

This paper introduces the structural threshold regression model that allows for an endogeneous threshold variable as well as for endogenous regressors. This model provides a parsimonious way of modeling nonlinearities and has many potential applications in economics and finance. Our framework can be viewed as a generalization of the simple threshold regression framework of Hansen (2000) and Caner and Hansen (2004) to allow for the endogeneity of the threshold variable and regime specific heteroskedasticity. Our estimation of the threshold parameter is based on a concentrated least squares method that involves an inverse Mills ratio bias correction term in each regime. We derive the asymptotic distribution of our estimator and propose a method to construct bootstrap confidence intervals. Finally, we investigate the performance of the asymptotic approximations and the bootstrap using a Monte Carlo simulation that indicates the applicability of the method in finite samples.

JEL Classifications: C13, C51

*Acknowledgements: We would like to thank Bruce Hansen for helpful comments and seminar participants at the University of Athens, University of Cambridge, Ryerson University, University of Waterloo, Simon Fraser University, the University of Western Ontario, 10th World Congress of the Econometric Society in Shanghai, and 27th Annual Meeting of the Canadian Econometrics Study Group in Vancouver.

†P.O. Box 537, CY 1678 Nicosia, Cyprus, e-mail: andros@ucy.ac.cy.

‡Guelph, Ontario N1G 2W1, Canada, email: tstengos@uoguelph.ca

§Department of Economics, Clark University 222 Jonas Clark Hall, 950 Main Street, Worcester, MA 01610, email: CTan@clarku.edu.

1 Introduction

One of the most interesting forms of nonlinear regression models with wide applications in economics is the threshold regression model. The attractiveness of this model stems from the fact that it treats the sample split value (threshold parameter) as unknown. That is, it internally sorts the data, on the basis of some threshold determinant, into groups of observations each of which obeys the same model. While threshold regression is parsimonious it also allows for increased flexibility in functional form and at the same time is not as susceptible to curse of dimensionality problems as nonparametric methods.

A crucial assumption in all the studies of the current literature is that the threshold variable is exogenous. This assumption severely limits the usefulness of threshold regression models in practice, since in economics many plausible threshold variables are endogenous. For example, Papageorgiou (2002) organized countries into multiple growth regimes using the trade share, defined as the ratio of imports plus exports to real GDP in 1985, as a threshold variable. Similarly, Tan (2010) classified countries into development clubs using the average expropriation risk from 1984-97 as the threshold variable. In each of these cases, there is strong evidence in the growth literature; see, Frankel and Romer (1999) and Acemoglu, Johnson, and Robinson (2001), respectively, that the proposed threshold variable is endogenous.

In this paper we introduce the Structural Threshold Regression (STR) model that allows for endogeneity in the threshold variable as well as in the slope regressors. Our research is related to several recent papers in the literature; see for example Hansen (2000) and Caner and Hansen (2004), Seo and Linton (2007), Gonzalo and Wolf (2005), and Yu (2010, 2011). The main difference of all these papers with our work is that they maintain the assumption that the threshold variable is exogenous. If the threshold variable is endogenous, the above approaches will yield inconsistent slope coefficients for the two regimes. The reason for the bias is that, just as in the limited dependent variable framework, a set of inverse Mills ratio bias correction terms is required to restore the orthogonality of the errors.

Intuitively, the main strategy of this paper is to exploit the insight obtained from the limited dependent variable literature (e.g., Heckman (1979)), and to relate the problem of having an endogenous threshold variable with the analogous problem of having an endogenous dummy variable or sample selection in the limited dependent variable framework. However, there is one important difference. While in sample selection models, we observe the assignment of observations into regimes but the (threshold) variable that drives this assignment is taken to be latent, here, it is the opposite; we do not know which observations belong to which regime (i.e., we do not know the threshold value), but we can observe the threshold variable. To put it differently, while endogenous dummy models treat the threshold variable as unobserved and the sample split as observed (dummy), here we treat the sample split value as unknown and we estimate it.

Specifically, we propose to estimate the threshold parameter using a concentrated least squares method and the slope estimates using 2SLS or GMM. We show the consistency of our estimators and derive the corresponding asymptotic distributions. To do so, we cast STR as a threshold regression model that is subject to cross-regime restrictions. Specifically, it imposes the restriction of having a different inverse Mills ratio for each regime. Analyzing such a restricted threshold regression model is nontrivial for two reasons. First, the estimates cannot be analyzed using results obtained regime by regime in the presence of restrictions across regimes, and, second, the orthogonalized errors of the structural model are regime specific heteroskedastic.

To overcome these problems we explore the relationship between the restricted and unrestricted sum of squared errors. We show that the threshold estimate has the same properties with or without restrictions, which implies that ignoring the restrictions will result in the same estimates and inference for the threshold. Our finding is similar to the result of Perron and Qu (2006) who consider structural change models with restrictions across regimes. To put it differently, if we were to impose the implausible assumption of regime specific homoskedasticity, then if one ignores the endogeneity in threshold and employ existing methods as in Hansen (2000) and Caner and Hansen (2004) the result will be the same as far as the properties of the threshold are concerned. Of course, the story is different for the estimates of the slope parameters, which suffer from bias when one ignores the endogeneity in the threshold and omits the inverse Mills ratio terms.

With regards to inference, the asymptotic distribution of the threshold estimate is nonstandard because the threshold parameter is not identified under the null. STR employs the framework of Hansen (2000) and Caner and Hansen (2004) who assume that the threshold effect diminishes as the sample increases. This assumption is the key to overcoming a problem that was first pointed out by Chan (1993). Chan shows that while the threshold estimate is superconsistent, the asymptotic distribution of the threshold estimate turns out to be too complicated for inference as it depends on nuisance parameters, including the marginal distribution of the regressors and all the regression coefficients.

Under regime specific heteroskedastcity, the asymptotic distribution is further characterized by parameters associated with regime specific heteroskedasticity as in the case of structural change models; see Bai (1997). More precisely, it involves two independent Brownian motions with two different scales and two different drifts. While these parameters are in principle estimable, inverting the likelihood ratio to obtain a confidence interval is not trivial as it involves a nonlinear algorithm. Instead, we employ a bootstrap inverted likelihood ratio approach. To examine the finite sample properties of our estimators we provide a Monte Carlo analysis.

In terms of the broader literature, our paper is related to Seo and Linton (2007) who allow the threshold variable to be a linear index of observed variables. They avoid the assumption of the shrinking threshold by proposing a smoothed least squares estimation strategy based on smoothing the objective function in the sense of Horowitz's smoothed maximum scored estimator. While they show that their estimator exhibits asymptotic normality it depends on the choice of bandwidth. Gonzalo and Wolf (2005) proposed subsampling to conduct inference in the context of threshold autoregressive models. Yu (2010) explores bootstrap methods for the threshold regression. He shows that while the nonparametric bootstrap is inconsistent the parametric bootstrap is consistent for inference on the threshold point in discontinuous threshold regression. He also finds that the asymptotic nonparametric bootstrap distribution of the threshold estimate depends on the sampling path of the original data. Finally, Yu (2011) proposes a semiparametric empirical Bayes estimator of the threshold parameter and shows that it is semiparametrically efficient.

The paper is organized as follows. Section 2 describes the model and the setup. Section 3 derives results for inference. Section 4 presents our Monte Carlo experiments. Section 5 concludes. In the appendix we collect the proofs of the main results.

2 The model

We assume weakly dependent data $\{y_i, x_i, q_i, z_i, u_i\}_{i=1}^n$ where y_i is real valued, x_i is a $p \times 1$ vector of covariates, q_i is a threshold variable, and z_i is a $l \times 1$ vector of instruments with $l \geq p$. Consider the following structural threshold regression model,

$$y_i = \boldsymbol{\beta}_1' \mathbf{x}_i + u_i, \quad q_i \leq \gamma \tag{2.1}$$

$$y_i = \boldsymbol{\beta}_2' \mathbf{x}_i + u_i, \quad q_i > \gamma \tag{2.2}$$

where $E(u_i|\mathbf{z}_i) = 0$. Equations (2.1) and (2.2) describe the relationship between the variables of interest in each of the two regimes and q_i is the threshold variable with γ being the sample split (threshold) value. The reduced form equation that determines the threshold variable is analogous to a selection equation that appears in the literature on limited dependent variable models; see Heckman (1979). The main difference is that while limited dependent variable models treat q_i as latent and the sample split as observed, here we treat the sample split value as unknown and we estimate it. The selection equation that determines which regime applies takes the form

$$q_i = \boldsymbol{\pi}_q' \mathbf{z}_i + v_{qi} \tag{2.3}$$

where $E(v_{qi}|\mathbf{z}_i) = 0$.

Let us consider the following partition $\mathbf{x}_i = (\mathbf{x}_{1i}, \mathbf{x}_{2i})$ where $\mathbf{x}_{1i}$ are endogenous and $\mathbf{x}_{2i}$ are exogenous and the $l \times 1$ vector of instrumental variables $\mathbf{z}_i = (\mathbf{z}_{1i}, \mathbf{z}_{2i})$ where $\mathbf{x}_{2i} \in \mathbf{z}_i$. If both q_i and $\mathbf{x}_i$ are exogenous then we get the threshold regression (TR) model studied by Hansen (2000). If q_i and $\mathbf{x}_{2i}$ are exogenous and $\mathbf{x}_{1i}$ is not a null set, then we get the instrumental variable threshold (IVTR) model studied by Caner and Hansen (2004). If $v_{qi} = 0$ then we get the smoothed exogenous threshold model as in Seo and Linton (2005), which allows the threshold variable to be a linear index of observed variables. In this paper we focus on the case where q_i is endogenous and the general case where $\mathbf{x}_{1i}$ is not a null set[1].

By defining the indicator function

$$I(q_i \leq \gamma) = \begin{cases} 1 & \text{iff } q_i \leq \gamma \Leftrightarrow v_{qi} \leq \gamma - \mathbf{z}_i'\boldsymbol{\pi}_q : \text{Regime 1} \\ 0 & \text{iff } q_i > \gamma \Leftrightarrow v_{qi} > \gamma - \mathbf{z}_i'\boldsymbol{\pi}_q : \text{Regime 2} \end{cases} \tag{2.4}$$

and $I(q_i > \gamma) = 1 - I(q_i \leq \gamma)$, we can rewrite the structural model (2.1)-(2.2) as

$$y_i = \boldsymbol{\beta}_{\mathbf{x}1}'\mathbf{x}_i I(q_i \leq \gamma) + \boldsymbol{\beta}_{\mathbf{x}2}'\mathbf{x}_i I(q_i > \gamma) + u_i \tag{2.5}$$

The reduced form model[2], $\mathbf{g}_{\mathbf{x}i} \equiv \mathbf{g}_{\mathbf{x}}(\mathbf{z}_i; \boldsymbol{\pi}) = E(\mathbf{x}_i|\mathbf{z}_i) = \boldsymbol{\Pi}'\mathbf{z}_i$, is given by

$$\mathbf{x}_i = \boldsymbol{\Pi}'\mathbf{z}_i + \mathbf{v}_{\mathbf{x}i} \tag{2.6}$$

$$E(\mathbf{v}_{\mathbf{x}i}|\mathbf{z}_i) = 0 \tag{2.7}$$

We assume joint normality of the errors such that

$$\begin{pmatrix} u_i \\ v_{q,i} \\ \mathbf{v}_{\mathbf{x}i} \end{pmatrix} |\mathbf{z}_i \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & \sigma_{uv_q} & \boldsymbol{\sigma}_{uv} \\ \sigma_{uv_q} & 1 & \mathbf{0} \\ \boldsymbol{\sigma}_{uv}' & \mathbf{0}' & \boldsymbol{\Sigma}_v \end{pmatrix} \right). \tag{2.8}$$

and obtain the following conditional expectations

$$E(y_i|\mathbf{z}_i, q_i \leq \gamma) = \boldsymbol{\beta}_{\mathbf{x}1}'\mathbf{g}_{\mathbf{x}i} + E(u_i|\mathbf{z}_i, q_i \leq \gamma) = \boldsymbol{\beta}_{\mathbf{x}1}'\mathbf{g}_{\mathbf{x}i} + \kappa\lambda_{1i}(\gamma) \tag{2.9}$$

$$E(y_i|\mathbf{z}_i, q_i > \gamma) = \boldsymbol{\beta}_{\mathbf{x}2}'\mathbf{g}_{\mathbf{x}i} + E(u_i|\mathbf{z}_i, q_i > \gamma) = \boldsymbol{\beta}_{\mathbf{x}2}'\mathbf{g}_{\mathbf{x}i} + \kappa\lambda_{2i}(\gamma) \tag{2.10}$$

where $\lambda_{1i}(\gamma) = \lambda_1(\gamma - z_i'\pi_q) = -\frac{\phi(\gamma - \mathbf{z}_i'\pi_q)}{\Phi(\gamma - \mathbf{z}_i'\pi_q)}$ and $\lambda_{2i}(\gamma) = \lambda_2(\gamma - z_i'\pi_q) = \frac{\phi(\gamma - \mathbf{z}_i'\pi_q)}{1 - \Phi(\gamma - \mathbf{z}_i'\pi_q)}$ are the inverse

[1] Note that we exclude (i) the special case of a continuous threshold model; see Hansen (2000) and Chan and Tsay (1998) and (ii) the case that $q_i \in \mathbf{x}_{1i}$. Our framework can be extended to consider these cases.

[2] One may easily consider alternative reduced form models, such as a threshold model; see Caner and Hansen (2004).

Mills ratio bias correction terms and $\phi(\cdot)$ and $\Phi(\cdot)$ are the normal pdf and cdf, respectively[3].

Then we can write the STR model as follows

$$y_i = \boldsymbol{\beta}'_{\mathbf{x}1}\mathbf{g}_{\mathbf{x}i} + \kappa\lambda_{1i}(\gamma) + e_{1i}, \quad q_i \le \gamma \tag{2.11}$$

$$y_i = \boldsymbol{\beta}'_{\mathbf{x}2}\mathbf{g}_{\mathbf{x}i} + \kappa\lambda_{2i}(\gamma) + e_{2i}, \quad q_i > \gamma \tag{2.12}$$

where $e_{1i} = \boldsymbol{\beta}'_{\mathbf{x}1}\mathbf{v}_{\mathbf{x}i} - \kappa\lambda_{1i}(\gamma) + u_i$ and $e_{2i} = \boldsymbol{\beta}'_{\mathbf{x}2}\mathbf{v}_{\mathbf{x}i} - \kappa\lambda_{2i}(\gamma) + u_i$.

Let $\lambda_i(\gamma) = \lambda_{1i}(\gamma) I(q_i \le \gamma) + \lambda_{2i}(\gamma) I(q_i > \gamma)$ and define $\boldsymbol{\delta}_{\mathbf{x},n} = \boldsymbol{\beta}_{\mathbf{x}1} - \boldsymbol{\beta}_{\mathbf{x}2}$. We can then express (2.11) and (2.12) as

$$y_i = \mathbf{g}'_{\mathbf{x}i}\boldsymbol{\beta}_{\mathbf{x}} + \mathbf{g}'_{\mathbf{x}i} I(q_i \le \gamma)\boldsymbol{\delta}_{\mathbf{x}n} + \lambda_i(\gamma)\kappa_n + e_i \tag{2.13}$$

or

$$y_i = \mathbf{g}'_{\mathbf{x}i}\boldsymbol{\beta}_{\mathbf{x}} + \mathbf{g}'_i(\gamma)\boldsymbol{\delta}_n + e_i \tag{2.14}$$

with regression parameters $\mathbf{g}_i(\gamma) = (\mathbf{g}'_{\mathbf{x}i} I(q_i \le \gamma), \lambda_i(\gamma))'$, $\boldsymbol{\beta}_{\mathbf{x}} = \boldsymbol{\beta}_{\mathbf{x}2}$, and $\boldsymbol{\delta}_n = (\boldsymbol{\delta}_{\mathbf{x},n}, \kappa_n)'$ and $E(e_i|\mathbf{z}_i) = 0$.

A few remarks are in order. First, note that when the error structure in the two regimes (2.2) and (2.1) is different $u_1 \neq u_2$ then the slope coefficient of the inverse Mills ratio terms κ_1 and κ_2 can be different across the two regimes $\kappa_1 \neq \kappa_2$. Here, for simplicity we assume $\kappa_1 = \kappa_2$ but our results carry over to the more general case. Second, when $\kappa = 0$, this model nests Caner and Hansen's TR model and if additionally $\mathbf{x}_i$ is exogenous then it coincides with Hansen (2000)'s TR model. In general, the main difference with both of the above cases is that in the latter the inverse Mills ratio bias correction term is omitted and as we will be arguing below this yields inconsistent estimates of the slope parameters $\boldsymbol{\beta}_{\mathbf{x}1}$ and $\boldsymbol{\beta}_{\mathbf{x}2}$.

In the following section we propose a consistent profile estimation procedure for STR that takes into account the inverse Mills ratio bias correction.

2.1 Estimation

We proceed in three steps. First, we estimate by LS the reduced models (2.3) and (2.6) to obtain $\widehat{\boldsymbol{\Pi}}$ and $\widehat{\boldsymbol{\pi}}_q$. The fitted values are then given by $\widehat{q}_i = \boldsymbol{\pi}'_q\mathbf{z}_i$ and $\widehat{\mathbf{x}}_i = \widehat{\mathbf{g}}_{\mathbf{x}i} = \widehat{\boldsymbol{\Pi}}'\mathbf{z}_i$ along with first stage residuals as $\widehat{\mathbf{v}}_{\mathbf{x}i} = \mathbf{x}_i - \widehat{\mathbf{x}}_i$and $\widehat{v}_{qi} = q_i - \widehat{q}$, respectively. We can also define the following functions of γ, $\widehat{\lambda}_{1i}(\gamma) = \lambda_1(\gamma - z'_i\widehat{\pi}_q)$ and $\widehat{\lambda}_{2i}(\gamma) = \lambda_2(\gamma - z'_i\widehat{\pi}_q)$, and $\widehat{\lambda}_i(\gamma) = \widehat{\lambda}_{1,i}(\gamma) I(q_i \le \gamma) + \widehat{\lambda}_{2,i}(\gamma) I(q_i > \gamma)$.

[3]Note that equations (2.9) and (2.10) hold even when one relaxes the assumption of Normality but with the correction terms being unknown functions (depending on the error distributions). These functions can be estimated by using a series approximation, or by using Robinson's two-step partially linear estimator; see Li and Wooldridge (2002).

Second, we estimate the threshold parameter γ by minimizing a Concentrated Least Squares (CLS) criterion

$$\widehat{\gamma} = \arg\min_{\gamma} S_n(\gamma) \tag{2.15}$$

where

$$S_n(\gamma) = \sum_{i=1}^{n}(y_i - \widehat{\mathbf{g}}'_{\mathbf{x}i}\boldsymbol{\beta}_{\mathbf{x}} - \widehat{\mathbf{g}}'_{\mathbf{x}i}I(q_i \leq \gamma)\boldsymbol{\delta}_{\mathbf{x}n} - \widehat{\lambda}_i(\gamma)\kappa_n)^2 = \sum_{i=1}^{n}(y_i - \widehat{\mathbf{g}}'_{\mathbf{x}i}\boldsymbol{\beta}_{\mathbf{x}} - \widehat{\mathbf{g}}'_i(\gamma)\boldsymbol{\delta}_n)^2. \tag{2.16}$$

Finally, once we obtain the split samples implied by $\widehat{\gamma}$, we estimate the slope parameters by 2SLS or GMM. This estimation strategy using concentration is exactly the same as in Hansen (2000) and Caner and Hansen (2004). Notice that conditional on γ, estimation in each regime mirrors the Heckman (1979) sample selection bias correction model, the Heckit model.

Let $\widetilde{\mathbf{X}}_\gamma$ and $\widetilde{\mathbf{X}}_\perp$ be the matrices of stacked vectors $\widetilde{\mathbf{x}}_{\gamma i}(\widehat{\gamma})=(\mathbf{x}'_i I(q_i \leq \widehat{\gamma}), \lambda_{1,i}(\widehat{\gamma})\, I(q_i \leq \widehat{\gamma}))'$ and $\widetilde{\mathbf{x}}_i(\widehat{\gamma})=(\mathbf{x}'_i I(q_i > \widehat{\gamma}), \lambda_{2,i}(\widehat{\gamma})\, I(q_i > \widehat{\gamma}))'$. Similarly, $\widetilde{\mathbf{Z}}_\gamma$, and $\widetilde{\mathbf{Z}}_\perp$ denote the matrices of stacked vectors $\widetilde{\mathbf{z}}_\gamma = \mathbf{z}_i I(q_i \leq \widehat{\gamma})$ and $\widetilde{\mathbf{z}}_\perp = \mathbf{z}_i I(q_i > \widehat{\gamma})$. Next, define the following matrices $\widetilde{\mathbf{X}} = (\widetilde{\mathbf{X}}_\gamma, \widetilde{\mathbf{X}}_\perp)$ and $\widetilde{\mathbf{Z}} = (\widetilde{\mathbf{Z}}_\gamma, \widetilde{\mathbf{Z}}_\perp)$ and the weight matrix $\widetilde{\mathbf{W}}$. Note that these matrices have a block diagonal form due to the indicator function.

Then, we can define the class of GMM estimators $\widetilde{\boldsymbol{\beta}} = (\widetilde{\boldsymbol{\beta}}'_1, \widetilde{\boldsymbol{\beta}}'_2)'$for $\boldsymbol{\beta}_1 = (\boldsymbol{\beta}'_{\mathbf{x}1}, \kappa_1)'$ and $\boldsymbol{\beta}_2 = (\boldsymbol{\beta}'_{\mathbf{x}2}, \kappa_2)'$

$$\widetilde{\boldsymbol{\beta}} = \left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{Z}}\widetilde{\mathbf{W}}\widetilde{\mathbf{Z}}'\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}'\widetilde{\mathbf{Z}}\widetilde{\mathbf{W}}\widetilde{\mathbf{Z}}'\mathbf{Y}. \tag{2.17}$$

When $\widetilde{\mathbf{W}} = (\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}})^{-1}$ we obtain the 2SLS estimator $\widetilde{\boldsymbol{\beta}}_{2SLS} = (\widetilde{\boldsymbol{\beta}}'_{1,2SLS}, \widetilde{\boldsymbol{\beta}}'_{2,2SLS})'$. The 2SLS residual is given by $\widetilde{e}_{i,2SLS} = y_i - \mathbf{x}_{1i}(\widehat{\gamma})'\widetilde{\boldsymbol{\beta}}_{1,2SLS} - \mathbf{x}_{2i}(\widehat{\gamma})'\widetilde{\boldsymbol{\beta}}_{2,2SLS}$. Define $\widetilde{\boldsymbol{\Sigma}}_j = \sum_{i=1}^{n}\mathbf{z}_{ji}\mathbf{z}'_{ji}\widetilde{e}_{i,2SLS}$, $j = 1, 2$. When $\widetilde{\mathbf{W}} = \widetilde{\boldsymbol{\Sigma}}^{-1} = diag(\widetilde{\boldsymbol{\Sigma}}_1^{-1}, \widetilde{\boldsymbol{\Sigma}}_2^{-1})$ then we obtain the efficient GMM estimator, $\widetilde{\boldsymbol{\beta}}_{GMM}$.

3 Threshold Regression with Restrictions

In this section we rewrite the STR model in equation (2.14) as a threshold regression subject to restrictions. We do so because the restricted problem above cannot be analyzed using results obtained regime by regime. Therefore we first analyze the unrestricted threshold regression and then relate to the restricted problem by using the relationship between the restricted and unrestricted problems. In particular, the unrestricted problem generalizes Caner and Hansen (2004) by including both inverse mills ratio terms in both regimes.

Define $\boldsymbol{\lambda}_i(\gamma) = (\lambda_{1i}(\gamma), \lambda_{1i}(\gamma))'$, $\boldsymbol{\beta}_{\boldsymbol{\lambda}} = \boldsymbol{\beta}_{\boldsymbol{\lambda}2} = (\kappa_{21}, \kappa_{22})'$, $\boldsymbol{\beta}_{\boldsymbol{\lambda}1} = (\kappa_{11}, \kappa_{12})'$,and $\boldsymbol{\delta}_{\boldsymbol{\lambda}n} = \boldsymbol{\beta}_{\boldsymbol{\lambda}1} - \boldsymbol{\beta}_{\boldsymbol{\lambda}2}$.

Then the unrestricted model takes the form

$$y_i = \mathbf{g}'_{\mathbf{x}i}\boldsymbol{\beta}_{\mathbf{x}} + \boldsymbol{\lambda}_i(\gamma)'\boldsymbol{\beta}_{\boldsymbol{\lambda}} + \mathbf{g}'_{\mathbf{x}i}I(q_i \leq \gamma)\boldsymbol{\delta}_{\mathbf{x}n} + \boldsymbol{\lambda}_i(\gamma)'I(q_i \leq \gamma)\boldsymbol{\delta}_{\boldsymbol{\lambda}n} + e_i, \tag{3.18}$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}'_{\mathbf{x}}, \boldsymbol{\beta}'_{\boldsymbol{\lambda}})'$ and $\boldsymbol{\delta}_n = \left(\boldsymbol{\delta}'_{\mathbf{x}n}, \boldsymbol{\delta}'_{\boldsymbol{\lambda}n}\right)'$ are the slope coefficients and e_i is the error of the unrestricted threshold model

$$e_i = \left(\mathbf{v}'_{\mathbf{x}i}\boldsymbol{\beta}_{\mathbf{x}1} - \boldsymbol{\lambda}_i(\gamma)'\boldsymbol{\beta}_{\boldsymbol{\lambda}1}\right) I(q_i \leq \gamma) + \left(\mathbf{v}'_{\mathbf{x}i}\boldsymbol{\beta}_{\mathbf{x}2} - \boldsymbol{\lambda}_i(\gamma)'\boldsymbol{\beta}_{\boldsymbol{\lambda}2}\right) I(q_i > \gamma) + u_i. \tag{3.19}$$

It is easy to verify that the STR model in equation (2.14) is a special case of (3.18) under the following restrictions

$$\kappa_{12} = \kappa_{21} = 0 \tag{3.20}$$

and

$$\kappa_{11} = \kappa_{22} = \kappa. \tag{3.21}$$

In general, let $\mathbf{g}_i(\gamma) = (\mathbf{g}'_{\mathbf{x}i}, \boldsymbol{\lambda}_i(\gamma)')', \boldsymbol{\theta} = \left(\boldsymbol{\beta}', \boldsymbol{\delta}'_n\right)'$. Then we can write the model as

$$y_i = \mathbf{g}'_i(\gamma)\boldsymbol{\beta} + \mathbf{g}'_i(\gamma)I(q_i \leq \gamma)\boldsymbol{\delta}_n + e_i, \tag{3.22}$$

subject to the restriction

$$\mathbf{R}'\boldsymbol{\theta} = \boldsymbol{\vartheta} \tag{3.23}$$

with $\mathbf{R}$ a $2q \times r$ matrix of rank r and $\boldsymbol{\vartheta}$ a r dimensional vector of constants and $e_i = e_{1i}I(q_i \leq \gamma) + e_{2i}I(q_i > \gamma)$. Then, the estimate of the threshold parameter γ can be viewed as the minimizer of the unconstrained Concentrated Least Squares (CLS) problem subject to the constraint in equation (3.23).

Consider the unrestricted GMM estimator $\widehat{\boldsymbol{\beta}}$ and a consistent weight matrix $\widehat{\mathbf{W}}$. Then, the relationship between the unrestricted and restricted slope coefficients, $\widetilde{\boldsymbol{\beta}}$, is given by

$$\widetilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}} - \widehat{\mathbf{W}}\mathbf{R}\left(\mathbf{R}'\widehat{\mathbf{W}}\mathbf{R}\right)^{-1}\left(\mathbf{R}'\widehat{\boldsymbol{\beta}} - \boldsymbol{\vartheta}\right). \tag{3.24}$$

We show in Lemma 4 of the Appendix that inference for the threshold estimator is the same with or without restrictions. We note that Perron and Qu (2006) obtained a similar finding in the context of structural change models. Therefore, we proceed by presenting the assumptions for the unrestricted threshold regression.

4 Inference

Define $\overline{\mathbf{g}}_i = \sup_{\gamma\in\Gamma}|\mathbf{g}_i(\gamma)|$ and $\overline{\mathbf{g}}_i|e_i| = \sup_{\gamma\in\Gamma}|\mathbf{g}_i(\gamma)e_i|$. Then define the moment functionals

$$\begin{aligned}
\mathbf{M}(\gamma) &= E(\mathbf{g}_i(\gamma)\mathbf{g}_i(\gamma)'), \\
\mathbf{D}(\gamma) &= E\left(\mathbf{g}_i(\gamma)\mathbf{g}_i(\gamma)'|q_i=\gamma\right), \\
\mathbf{\Omega}(\gamma) &= E\left(\mathbf{g}_i(\gamma)\mathbf{g}_i(\gamma)'e_i^2|q_i=\gamma\right).
\end{aligned}$$

Let $f_q(q)$ be the density function of q and γ_0 denotes the true value of γ. Let $\lim_{\gamma\nearrow\gamma_0}$ and $\lim_{\gamma\searrow\gamma_0}$, denote the limits from below and above the threshold γ_0, respectively. Then, we can define the following limits:

$\mathbf{D}_1 = \lim_{\gamma\nearrow\gamma_0}\mathbf{D}(\gamma) = \lim_{\gamma\nearrow\gamma_0}E\left(\mathbf{g}_i(\gamma)\mathbf{g}_i(\gamma)'|q_i=\gamma\right)$

$\mathbf{D}_2 = \lim_{\gamma\searrow\gamma_0}\mathbf{D}(\gamma) = \lim_{\gamma\searrow\gamma_0}E\left(\mathbf{g}_i(\gamma)\mathbf{g}_i(\gamma)'|q_i=\gamma\right)$

$\mathbf{\Omega}_1 = \lim_{\gamma\nearrow\gamma_0}\mathbf{\Omega}(\gamma) = \lim_{\gamma\nearrow\gamma_0}E\left(\mathbf{g}_i(\gamma)\mathbf{g}_i(\gamma)'e_i^2|q_i=\gamma\right)$

$\mathbf{\Omega}_2 = \lim_{\gamma\searrow\gamma_0}\mathbf{\Omega}(\gamma) = \lim_{\gamma\searrow\gamma_0}E\left(\mathbf{g}_i(\gamma)\mathbf{g}_i(\gamma)'e_i^2|q_i=\gamma\right)$

Assumption 1

(1.1) $\{\mathbf{z}_i, \mathbf{g}_i(\gamma), u_i, \mathbf{v}_i, v_{qi}\}$ is strictly stationary and ergodic with ρ mixing coefficients $\sum_{m=1}^{\infty}\rho_m^{1/2} < \infty$,

(1.2) $E(u_i|F_{i-1}) = 0$,

(1.3) $E(\mathbf{v}_i|F_{i-1}) = 0$,

(1.4) $E|\overline{\mathbf{g}}_i|^4 < \infty$ and $E|\overline{\mathbf{g}}_i e_i|^4 < \infty$,

(1.5) for all $\gamma\in\Gamma, E(|\overline{\mathbf{g}}_i|^4|q_i=\gamma) \le C$, $\lim_{\gamma\searrow\gamma_0}E(|\mathbf{g}_i(\gamma)|^4 e_i^4|q_i=\gamma) \le C$, $\lim_{\gamma\nearrow\gamma_0}E(|\mathbf{g}_i(\gamma)|^4 e_i^4|q_i=\gamma) \le C$, and for some $C<\infty$,

(1.6) for all $\gamma\in\Gamma$, the marginal distribution of the threshold variable, $f_q(\gamma) \le \overline{f} < \infty$ and it is continuous at $\gamma=\gamma_0$.

(1.7) $\mathbf{D}(\gamma)$ and $\mathbf{\Omega}(\gamma)$ are semi-continuous at $\gamma=\gamma_0$.

(1.8) $\boldsymbol{\delta}_n = \boldsymbol{\beta}_1 - \boldsymbol{\beta}_2 = \mathbf{c}n^{-\alpha} \to 0$, $\mathbf{c}\neq 0$, $\alpha\in(0,1/2)$,

(1.9) $f_q(\gamma) > 0$, $\mathbf{c}'\mathbf{D}(\gamma)\mathbf{c} > 0$, $\mathbf{c}'\boldsymbol{\Omega}(\gamma)\mathbf{c} > 0$.

(1.10) for all $\gamma \in \Gamma$, $\overline{\mathbf{M}} > \mathbf{M}(\gamma) > 0$.

(1.11) for all $\gamma \in \Gamma$,$\widehat{\gamma} = \underset{\gamma\in\Gamma}{\arg\min} \sum_{i=1}^{n}(y_i - \mathbf{g}_i'(\gamma)\boldsymbol{\beta} - \mathbf{g}_i'(\gamma)I(q_i \leq \gamma)\boldsymbol{\delta}_n)^2$ exists and it is unique. Furthermore, $\widehat{\gamma}$ lies in the interior of Γ, with Γ compact and convex. This set of assumptions is similar to Hansen (2000) and Caner and Hansen (2004).

While most assumptions are rather standard, Assumption 1.8 is not. Assumption 1.8 assumes that a "small threshold" asymptotic framework applies in the sense that $\boldsymbol{\delta}_n = \left(\boldsymbol{\delta}'_{\mathbf{x}n}, \boldsymbol{\delta}'_{\boldsymbol{\lambda}n}\right)'$will tend to go to zero as $n \to \infty$. Under this assumption, Hansen (2000) showed in the case without regime specific heteroskedasticity that the threshold estimate has an asymptotic distribution free of nuisance parameters.

In our context, this assumption allows us to derive an asymptotic distribution of the threshold estimate that only depends on parameters associated with regime specific heteroskedasticity that are, in principle, estimable. This result is related to the asymptotic distribution of Bai (1997) in the case of structural change models. However, structural change models (i.e., $q_i = i$) assume that the stochastic process of $\sum_{i=1}^{n} g_i e_i I\{q_i < \gamma\}$ is a martingale in γ, but this may not be true for the case of STR unless the data are independent across i.

In the case of the STR model in equation (2.14), this assumption also implies that κ_n vanishes as $n \to \infty$ at the same rate as $\boldsymbol{\delta}_{\boldsymbol{\lambda}n}$ to ensures that the bias correction (i.e. the inverse Mills ration terms) to the endogeneity of the threshold will not be present when the model is linear.

Assumption 1.11 is satisfied given the monotonicity of the inverse Mills ratios. The above assumptions are also sufficient to guarantee that the first stage regressions are consistent for the true conditional means i.e. $\widehat{\mathbf{r}} = (\widehat{\mathbf{r}}_{\mathbf{x}i}, \widehat{\mathbf{r}}_{u_1 i}, \widehat{\mathbf{r}}_{u_2 i})= \mathbf{g}_i(\gamma) - \widehat{\mathbf{g}}_i(\gamma) = o_p(1)$.

4.1 Threshold Estimate

Proposition 4.1 *Consistency of* $\widehat{\gamma}$

Under Assumption 1, the estimator for γ *obtained by minimizing the CLS criterion (2.16),* $\widehat{\gamma}$*, is consistent. That is,*

$$\widehat{\gamma} \underset{p}{\rightarrow} \gamma_0$$

The proof is given in the appendix.

Corollary 4.1 *Consider the shorter (mispecified) STR model based on a subset of regressors that* *belongs in the span of the columns of the true STR model. Then, under Assumption 1, the estimator for γ obtained by minimizing the CLS based on a restricted projection is also consistent. The proof is immediate from the proof of Proposition 4.1.*

Remark 1 When we ignore the endogeneity in the threshold we would still get a consistent estimator for γ_0. This means that Hansen's TR and Caner-Hansen's IVTR will both yield consistent estimators for γ_0, regardless of whether there is endogeneity in the slope.

Remark 2 Although the endogeneity in the threshold does not generate bias in the threshold estimate, it does yield a bias for the estimation of the slope coefficients. As in the standard omitted variable case, the bias will depend on the degree of correlation between the omitted inverse Mills ratio term and the included regressors.

To obtain the asymptotic distribution let us first define two independent standard Wiener processes $W_1(s)$ and $W_2(s)$ defined on $[0, \infty)$.

Let

$$T(s) = \begin{cases} -\frac{1}{2}|s| + W_1(-s), & \text{if } v \leq 0 \\ -\frac{1}{2}\xi|s| + \sqrt{\phi}W_2(s) & \text{if } v > 0 \end{cases},$$

where $\xi = \frac{\mathbf{c}'\mathbf{D}_2\mathbf{c}}{\mathbf{c}'\mathbf{D}_1\mathbf{c}}$, and $\varphi = \frac{\mathbf{c}'\mathbf{\Omega}_2\mathbf{c}}{\mathbf{c}'\mathbf{\Omega}_1\mathbf{c}}$.[4]

Theorem 4.1 *Asymptotic Distribution of* $\widehat{\gamma}$

Under Assumption 1

$$n^{1-2\alpha}(\widehat{\gamma} - \gamma_0) \xrightarrow{d} \omega T \tag{4.25}$$

where $\omega = \frac{\mathbf{c}'\mathbf{\Omega}_1\mathbf{c}}{(\mathbf{c}'\mathbf{D}_1\mathbf{c})^2 f}$ *and* $T \equiv \underset{-\infty<s<\infty}{\arg\max}\, T(s)$. *The proof is given in the appendix.*

The distribution function of T is given by Bai (1997). For $x < 0$, the cdf of T is given by

$$P(T \leq x) = -\sqrt{\frac{|x|}{2\pi}}\exp(-\frac{|x|}{8}) - c\exp(a|x|)\Phi(-b\sqrt{|x|}) + (d - 2 + \frac{|x|}{2})\Phi(-\frac{\sqrt{|x|}}{2}), \tag{4.26}$$

where $a = \frac{1}{2}\frac{\xi}{\varphi}(1 + \frac{\xi}{\varphi})$, $b = \frac{1}{2} + \frac{\xi}{\varphi}, c = \frac{\varphi(\varphi+2\xi)}{\xi(\varphi+\xi)}$, and $d = \frac{(\varphi+2\xi)^2}{\xi(\varphi+\xi)}$.

For $x > 0$,

$$P(T \leq x) = 1 + \xi\sqrt{\frac{x}{2\pi\varphi}}\exp(-\frac{\xi^2 x}{8\varphi}) - c\exp(ax)\Phi(-b\sqrt{x}) + (-d + 2 - \frac{\xi^2 x}{2\varphi})\Phi(-\frac{\xi}{2}\sqrt{\frac{x}{\varphi}}), \tag{4.27}$$

[4] The case of the asymmetric two sided Brownian motion argmax distribution with unequal variances was first examined by Stryhn (1996).

where $a = \frac{\varphi+\xi}{2}$, $b = \frac{2\varphi+\xi}{2\sqrt{\varphi}}$, $c = \frac{\xi(\xi+2\varphi)}{\varphi(\varphi+\xi)}$, and $d = \frac{(\xi+2\varphi)^2}{\varphi(\varphi+\xi)}$. The distribution is not symmetric when $\varphi \neq 1$ or $\xi \neq 1$. In the case of $\varphi = \xi = 1$, we get the symmetric case; see for example Hansen (2000).

Note that a simpler case occurs when we assume regime specific heteroskedasticity but homoskedasticity within each regime. In this case we get $\mathbf{\Omega}_1 = \sigma_{e_1}^2 \mathbf{D}_1$, $\mathbf{\Omega}_2 = \sigma_{e_2}^2 \mathbf{D}_2$, where $\sigma_{e_1}^2 = E(e_{1i}^2|q = \gamma), \sigma_{e_2}^2 = E(e_{2i}^2|q = \gamma)$. This implies that $\omega = \frac{\sigma_{e_1}^2}{(\mathbf{c}'\mathbf{D}_1\mathbf{c})f}$, and $\varphi = \frac{\sigma_{e_2}^2}{\sigma_{e_1}^2}\xi$. Furthermore, note that when $\mathbf{D}_1 = \mathbf{D}_2 = \mathbf{D}$ and $\mathbf{\Omega}_1 = \mathbf{\Omega}_2 = \mathbf{\Omega}$ we obtain the case that excludes regime specific heteroskedasticity. In this case we obtain $\xi = 1$, $\varphi = 1$, $\omega = \frac{\mathbf{c}'\mathbf{\Omega}\mathbf{c}}{(\mathbf{c}'\mathbf{D}\mathbf{c})^2 f}$. Hence, when we define $W(s) = W_1(s)$ for $v \leq 0$ and $W(s) = W_2(s)$ for $v > 0$, we can easily see that the distribution coincides with the two sided Wiener distribution established in Hansen (2000) and Caner and Hansen (2004).

Next using the distributional result in Theorem 4.1 we can construct confidence intervals for γ_0. We consider the pseudo Likelihood Ratio (LR) statistic

$$LR_n(\gamma) = n\frac{S_n(\gamma) - S_n(\widehat{\gamma})}{S_n(\widehat{\gamma})}.$$

Define

$$\eta^2 = \frac{\mathbf{c}'\mathbf{\Omega}_1\mathbf{c}}{(\mathbf{c}'\mathbf{D}_1\mathbf{c})\sigma_e^2}$$

and

$$\psi = \sup_{-\infty<s<\infty}\left(\left(-\frac{1}{2}|s| + W_1(-s)\right) I(q < \gamma_0) + \left(-\frac{1}{2}\xi|s| + \sqrt{\phi}W_2(s)\right) I(q > \gamma_0)\right)$$

Then we have the following theorem.

Theorem 4.2 *Asymptotic Distribution of* $LR(\gamma_0)$

Under Assumption 1, the asymptotic distribution of the likelihood ratio test under H_0 *is given by*

$$LR_n(\gamma_0) \xrightarrow{d} \eta^2\psi \tag{4.28}$$

where the distribution of ψ *is* $P(\psi \leq x) = (1 - e^{-x/2})(1 - e^{-\xi x/2})^{\sqrt{\varphi}}$

The proof is given in the appendix.

Note that when we exclude regime specific heteroskedasticity we obtain $\xi = \varphi = 1$ and the distribution is identical to the distribution of Hansen (2000) and Caner and Hansen (2004). Under homoskedasticity within each regime the distribution of the asymptotic distribution of the LR statistic is free of nuisance parameters and simplifies to $LR_n(\gamma) = n\frac{S_n(\gamma) - S_n(\widehat{\gamma})}{S_n(\widehat{\gamma})} \xrightarrow{d} \psi$ since $\eta^2 = 1$.

Define $\widehat{\Gamma} = \{\gamma : LR_n(\gamma) \leq c\}$ and let $1 - a$ denote the desired asymptotic confidence level and let $c = c_\psi(1 - a)$ be the critical value for ψ. Assuming $\alpha = 1$, $\xi = \varphi = 1$, $\eta^2 = 1$ and Gaussian errors we can invoke Theorem 3 of Hansen (2000) to show that the likelihood ratio test is asymptotically conservative. This implies that at least in this special case inferences based on the confidence region $\widehat{\Gamma}$ are asymptotically valid.

The nuisance parameters, η^2, ξ, and φ, are in principle estimable. They can be estimated for each regime separately as in Section 3.4 of Hansen (2000). However, it is quite difficult to apply the test-inversion method of Hansen (2000) to construct an asymptotic confidence interval for γ_0 because there is no closed form solution for $1 - a = (1 - e^{-x/2})(1 - e^{-\xi x/2})^{\sqrt{\varphi}}$. Therefore we propose to use a bootstrap inverted likelihood ratio approach that we describe next.

The bootstrap for the threshold regression model has been studied by Yu (2010) who shows that the parametric bootstrap is consistent while the nonparametric bootstrap is inconsistent for inference on the threshold estimate. The problem is that typically the parametric bootstrap is not feasible as one needs to specify a complete likelihood. To overcome this problem we follow Hansen (2000), who under the framework of an asymptotically diminishing threshold effect, shows that the confidence interval constructed by inverting the likelihood ratio statistic is asymptotically valid. In this framework the bootstrap is valid.[5]

Given consistent estimates for $(\widehat{\boldsymbol{\delta}}_{\mathbf{x}n}, \widehat{\boldsymbol{\beta}}_{\mathbf{x}}, \widehat{\kappa}_n, \widehat{\mathbf{g}}_{\mathbf{x}i}, \widehat{\lambda}_i(\widehat{\gamma}))$ we define the residuals of the STR model

$$\widehat{e}_i = y_i - \widehat{\mathbf{g}}'_{\mathbf{x}i}\widehat{\boldsymbol{\beta}}_{\mathbf{x}} - \widehat{\mathbf{g}}'_{\mathbf{x}i}I(q_i \leq \widehat{\gamma})\widehat{\boldsymbol{\delta}}_{\mathbf{x}n} - \widehat{\lambda}_i(\widehat{\gamma})\widehat{\kappa}_n$$

Then following Hansen (1996) we fix the regressors and define the bootstrap dependent variable $y_i^* = \widetilde{e}_i(\gamma)\zeta_i$, where ζ_i is Normal *i.i.d.* and $\widetilde{e}_i$ is the recentered residual $\widehat{e}_i$.

To construct bootstrap confidence intervals for γ we follow the test-inversion method of Hansen (2000) and then obtain the bootstrap distribution of the likelihood ratio statistic using the bootstrap estimates

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

We store likelihood ratio values from bootstraps $\{LR_n^{*(1)}(\gamma), ..., LR_n^{*(B)}(\gamma)\}$ and sort them to determine the $a(B+1)^{th}$ LR value, $LR_n^*(c_a^*)$ as the critical value for $1-a$ confidence level. Then we construct the bootstrapped inverted LR confidence region for γ_0 , $\widehat{\Gamma}^* = \{\gamma : LR_n(\gamma) \leq LR_n^*(c_a^*)\}$, where $LR_n(\gamma)$ is computed from the data.

[5] Antoch et al (1995) established the validity of the nonparametric bootstrap under the assumptions of an asymptotically diminishing threshold and *i.i.d* errors in the context of structural change models.

One problem with the bootstrap is that it heavily relies on the assumptions of the underlying model and in particular on the assumption of the diminishing threshold effect. In practice, however, it is not clear how one can distinguish whether a given dataset follows the STR model or that in Chan (1993). This is a problem because the bootstrap is invalid in the framework of Chan (1993) as shown in Yu (2010). This problem is beyond the scope of the paper and is left for future research.

4.2 Slope Parameters

Consider the unrestricted vector of covariates $\mathbf{x}_i(\gamma_0) = (\mathbf{x}_i, \lambda_{1i}(\gamma_0), \lambda_{2i}(\gamma_0))'$. Then , the inference on the slope parameters follows a restricted version of Caner and Hansen (2004). Let us define the following matrices

$\mathbf{Q}_1 = E(\mathbf{z}_i\mathbf{z}_i' I(q_i \leq \gamma_0), \mathbf{Q}_2 = E(\mathbf{z}_i\mathbf{z}_i' I(q_i > \gamma_0)$

$\mathbf{S}_1 = E(\mathbf{z}_i\mathbf{x}_i(\gamma_0)' I(q_i \leq \gamma_0), \mathbf{S}_2 = E(\mathbf{z}_i\mathbf{x}_i(\gamma_0)' I(q_i > \gamma_0)$

$\mathbf{\Sigma}_1 = E(\mathbf{z}_i\mathbf{z}_i' u_i^2 I(q_i \leq \gamma_0), \mathbf{\Sigma}_2 = E(\mathbf{z}_i\mathbf{z}_i' u_i^2 I(q_i > \gamma_0)$

$\mathbf{V}_{1,} = (\mathbf{S}_1'\mathbf{Q}_1^{-1}\mathbf{S}_1)^{-1}\mathbf{S}_1'\mathbf{Q}_1^{-1}\mathbf{\Sigma}_1\mathbf{Q}_1^{-1}\mathbf{S}_1(\mathbf{S}_1'\mathbf{Q}_1\mathbf{S}_1)^{-1}$

$\mathbf{V}_2 = (\mathbf{S}_2'\mathbf{Q}_2^{-1}\mathbf{S}_2)^{-1}\mathbf{S}_2'\mathbf{Q}_2^{-1}\mathbf{\Sigma}_2\mathbf{Q}_2^{-1}\mathbf{S}_2(\mathbf{S}_2'\mathbf{Q}_2^{-1}\mathbf{S}_2)^{-1}$

$\mathbf{V} = diag(\mathbf{V}_{1,2SLS}, \mathbf{V}_{2,2SLS})$

$\mathbf{Q} = diag(\mathbf{Q}_1, \mathbf{Q}_2)$

$\overline{\mathbf{V}}_1 = (\mathbf{S}_1'\mathbf{\Sigma}_1^{-1}\mathbf{S}_1)^{-1}, \overline{\mathbf{V}}_2 = (\mathbf{S}_2'\mathbf{\Sigma}_2^{-1}\mathbf{S}_2)^{-1}$

$\overline{\mathbf{V}} = diag(\overline{\mathbf{V}}_1, \overline{\mathbf{V}}_2)$

Then the following theorem establishes the asymptotic distributions of the 2SLS and GMM slope estimators.

Theorem 4.3 *Under Assumption 1,*

(a)

$$\sqrt{n}(\widetilde{\boldsymbol{\beta}}_{2SLS} - \boldsymbol{\beta}) \Longrightarrow N(0, \widetilde{\mathbf{V}}_{2SLS}) \tag{4.29}$$

where

$$\begin{aligned} \widetilde{\mathbf{V}}_{2SLS} = \; & \mathbf{V} - \mathbf{Q}^{-1}\mathbf{R}\left(\mathbf{R}'\mathbf{Q}^{-1}\mathbf{R}\right)^{-1}\mathbf{R}'\mathbf{V} - \mathbf{V}\mathbf{R}\left(\mathbf{R}'\mathbf{Q}^{-1}\mathbf{R}\right)^{-1}\mathbf{R}'\mathbf{Q}^{-1} \\ & + \mathbf{Q}^{-1}\mathbf{R}\left(\mathbf{R}'\mathbf{Q}^{-1}\mathbf{R}\right)^{-1}\mathbf{R}'\mathbf{V}\mathbf{R}\left(\mathbf{R}'\mathbf{Q}^{-1}\mathbf{R}\right)^{-1}\mathbf{R}'\mathbf{Q}^{-1}. \end{aligned} \tag{4.30}$$

(b)

$$\sqrt{n}(\widetilde{\boldsymbol{\beta}}_{GMM} - \boldsymbol{\beta}) \Longrightarrow N(0, \widetilde{\mathbf{V}}_{GMM}) \tag{4.31}$$

where

$$\widetilde{\mathbf{V}}_{GMM} = \overline{\mathbf{V}} - \overline{\mathbf{V}}\mathbf{R}\left(\mathbf{R}'\overline{\mathbf{V}}\mathbf{R}\right)^{-1}\mathbf{R}'\overline{\mathbf{V}} \tag{4.32}$$

5 Monte Carlo

We proceed below with an exhaustive simulation study that compares the finite sample performance of our estimator with that of Hansen (2000) and Caner and Hansen (2004). We explore two designs. First, we focus on the endogeneity of the threshold variable and assume that the slope variable is exogenous. Second, we assume that both the threshold and the slope variables are endogenous.

The Monte Carlo design is based on the following threshold regression

$$y_i = \begin{cases} \beta_{1,1} + \beta_{1,2}x_i + u_i, & q_i \leq 2 \\ \beta_{2,1} + \beta_{2,2}x_i + u_i, & q_i > 2 \end{cases} \tag{5.33}$$

where

$$q_i = 2 + z_{1,i} + v_{q,i} \tag{5.34}$$

with $z_{1,i}, v_{q,i}, \varepsilon_i \sim NIID(0,1)$ and $u_i = v_x + \kappa v_{q,i} + (0.1)\, N(0,1)$.The degree of endogeneity of the threshold variable is controlled by κ, where $\kappa = 0.01\sqrt{\widetilde{\kappa}^2/(1-\widetilde{\kappa}^2)}$. We sampled v_x and v_q independently. We fix $\widetilde{\kappa} = 0.95$ and set $\beta_{2,1} = \beta_{2,2} = \beta_2 = 1$ and $\beta_{1,1} = \beta_{1,2} = \beta_1$, and vary β_1 by examining various $\delta = \beta_1 - \beta_2$. We report three values of $\delta = \{0.5, 1, 2\}$, that correspond to a small, medium, and large threshold[6]. In the case of endogenous threshold and endogenous slope variable we assume that $x_i = z_{2,i} + v_i$, where $z_{2,i} \sim NIID(0,1)$ and $v_i = 0.5u_i$. Finally we consider sample sizes of $100, 200$, and 500 using 1000 Monte Carlo simulations. We also investigated what happened when we varied the degree of the correlation between the instrumental variables z and the exogenous slope variables x_2. As in the case of Heckman's estimator, our estimator becomes more efficient as this correlation decreases and the degree of multicollinearity between $\Pi' z$ and x is small.

First, we consider the estimation of the threshold value γ. Table 1 presents the 5th, 50th, and 95th quantiles for the distribution of the threshold estimate $\hat{\gamma}$ under STR, TR, and IVTR. Specifically,

[6]We have conducted a large number of expirements and the results are similar. Specifically, our experiments investigated a broader range of values of δ, different degrees of threshold endogeneity (σ_{uv_q}), and different degrees of correlation between the instrumental variables z and the included exogenous slope variable x_2. We investigated different degrees of threshold endogeneity between the threshold and the errors of two regimes. All results are available from the authors on request.

columns (1)-(6) of Table 1 consider the case where the threshold variable is endogenous but the slope variable is exogenous and compare the distribution of the TR estimates with those of STR. Columns (7)-(12) of Table 1 consider the case where both the threshold variable and slope variable are endogenous and compare the distribution of the IVTR estimates with those of STR.

Figures 1 and 2 present the corresponding Gaussian kernel density estimates for $\hat{\gamma}$ for the case where the slope variable is exogenous or endogenous, respectively. The kernel density estimates are obtained using Silverman's bandwidth parameter for various values of δ and sample sizes. Specifically, Figures 1(a)-(c) present the density estimates for various sample sizes for $\delta = 1$ while Figures 1(d)-(f) present the density estimates for various values of δ for $n = 500$. We present the results for STR in solid line in Figure 1 while the results for TR or IVTR are given by the dotted line.

We see that the performance of the threshold estimator of STR improves as δ and/or n increases. We also find that the threshold estimates of STR vis-a-vis those of Hansen (2000) and Caner and Hansen (2004) behave similarly. All three estimators appear to be consistent; as δ and/or n increases all three estimators appear to converge upon the true value of $\gamma = 2$. STR appears to be relatively more efficient for the case where the threshold variable is endogenous, while the opposite is true for the case where the threshold variable is exogenous.

Table 2 presents the results for the slope coefficient β_2 As in the case of the threshold estimates we find that the performance of the slope coefficient estimate of STR improves as δ and/or n increases. In sharp contrast to the results for the threshold estimate, however, we do not find, in this case, that the results for TR and IVTR are similar to STR. Table 2 suggests that the distribution of $\widehat{\beta}_2$ for STR converges to the true value of $\beta_2 = 1$. However, this is not the case for either TR or IVTR. In both cases, the median of the distribution centers away from the true value of $\beta_2 = 1$; specifically, the median for TR coverges to around 0.918 while that for IVTR converges to around 1.17. More revealingly, for the case of TR, the true value of $\beta_2 = 1$ is actually getting further away from the interval covered by the 5th to 95th quantiles as the sample size gets large. These findings suggest that, consistent with the theory, the omission of the inverse Mills ratio bias correction terms results in the estimators for the slope parameters of TR and IVTR to be inconsistent.

Finally, Table 3 presents bootstrap coverage probabilities of a nominal 95% interval $\widehat{\Gamma}$ using 300 bootstrap replications. We report results where δ varies from 0.5, 1, and 2 for sample sizes 50, 100, 250, and 500. Table 3 shows that the coverage probability increases for all the values of δ as n increases. We find that the coverage becomes more conservative for larger sample sizes. Similarly, for fixed sample size, n, the coverage probability increases as δ increases. Our bootstrap results are consistent with the simulation findings of Caner and Hansen (2004), which are based on the distribution theory.

6 Conclusion

In this paper we introduce the Structural Threshold Regression (STR) model that allows for the endogeneity of the threshold variable as well as the slope regressors. We develop a concentrated least squares estimator that deals with the problem of endogeneity in the threshold variable by including a correction term based on the inverse Mills ratios in each regime. We show that our estimators are consistent and derive their asymptotic distributions. Our monte carlo simulation experiments demonstrate the good finite sample properties of our estimators.

References

[1] Antoch, J., M. Huskova, and N. Veraverbeke, (1995), Change-point Problem and Bootstrap, *Journal of Nonparametric Statistics*, 5, 123-144.

[2] Acemoglu, D. Johnson, S. and J. A. Robinson, (2001), "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review,* 91, 1369-1401.

[3] Bai, M., (1997), "Estimation of a Change Point in Multiple Regression Models," *The Review of Economics and Statistics*, 79, 551-563.

[4] Caner, M. and B. Hansen, (2004), "Instrumental Variable Estimation of a Threshold Model," *Econometric Theory*, 20, 813-843.

[5] 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.

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

[7] Gonzalo, J. and M. Wolf, (2005), "Subsampling Inference in Threshold Autoregressive Models," *Journal of Econometrics*, 127, 2, 201-224.

[8] Hansen, B. E., (2000), "Sample Splitting and Threshold Estimation," *Econometrica*, 68, 3, 575-604.

[9] Heckman, J., (1979), "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 1, 153-161.

[10] Li , Q. and J. Wooldridge, (2002) "Semiparametric estimation of partially linear models for dependent data with generated regressors," *Econometric Theory,* 18, 625-645.

[11] Papageorgiou, C., (2002), "Trade as a Threshold Variable for Multiple Regimes," *Economics Letters*, 77, 85-91.

[12] Perron, P. and Z. Qu, (2006), "Estimating Restricted Structural Change Models," *Journal of Econometrics*, 134, 372-399.

[13] Seo, M. H. and O. Linton, (2007), "A Smoothed Least Squares Estimator For Threshold Regression Models," *Journal of Econometrics*, 141, 2, 704-735.

[14] Stryhn, H., (1996), "The Location of the Maximum of Asymmetric Two-Sided Brownian Motion with Triangular Drift," *Statistics and Probability Letters*, 29, 279-284.

[15] Tan, C. M., (2009), "No One True Path: Uncovering the Interplay Between Geography, Institutions, and Fractionalization in Economic Development," *Journal of Applied Econometrics*, 25, 7, 1100-1127.

[16] Van der Vaart, A. W., (1998), "Asymptotic Statistics," Cambridge University Press.

[17] Van der Vaart, A. W. and J. A. Wellner, (1996), "Weak Convergence and Empirical Processes," Springer.

[18] Yu P., (2010a), "Bootstrap in Threshold Regression," University of Auckland, mimeo.

[19] Yu P., (2010b), "Threshold Regression with Weak and Strong Identifications," University of Auckland, mimeo.

[20] Yu P., (2011a), "Likelihood Estimation and Inference in Threshold Regression," University of Auckland, mimeo.

[21] Yu P., (2011b), "Adaptive Estimation of the Threshold Point in Threshold Regression," University of Auckland, mimeo.

[22]

A Appendix

The model in matrix notation

Recall that $\mathbf{g}_i(\gamma) = (\mathbf{g}_{\mathbf{x}i}, \lambda_{1i}(\gamma), \lambda_{2i}(\gamma))'$. Define the regime specific matrix $\mathbf{G}_\gamma(\gamma) = (\mathbf{G}_{\mathbf{x},\gamma}, \mathbf{\Lambda}_{1,\gamma}(\gamma), \mathbf{\Lambda}_{2,\gamma}(\gamma))$ by stacking $\mathbf{g}_{\gamma i}(\gamma) = (\mathbf{g}_{\mathbf{x}i} I(q_i \leq \gamma), \lambda_{1i}(\gamma) I(q_i \leq \gamma), \lambda_{2i}(\gamma) I(q_i \leq \gamma))'$. Similarly, we can define its orthogonal matrix, $\mathbf{G}_\perp(\gamma) = (\mathbf{G}_{\mathbf{x},\perp}, \mathbf{\Lambda}_{1,\perp}(\gamma), \mathbf{\Lambda}_{2,\perp}(\gamma))$. Let $\mathbf{Y}$ and $\mathbf{e}$ be the stacked vectors of y_i and e_i, respectively. Then we can write (2.14) as follows.

$$\mathbf{Y} = \mathbf{G}(\gamma_0)\boldsymbol{\beta} + \mathbf{G}_0(\gamma_0)\boldsymbol{\delta}_n + \mathbf{e} \tag{A.1}$$

or

$$\mathbf{Y} = \mathbf{G}^*(\gamma)\boldsymbol{\beta}^* + \mathbf{e} \tag{A.2}$$

where $\mathbf{G}^*(\gamma) = (\mathbf{G}_\gamma(\gamma), \mathbf{G}_\perp(\gamma))$ and $\boldsymbol{\beta}^* = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$.

Let us now define the projection matrices by first noting that $\widehat{\mathbf{x}}_i = \widehat{\mathbf{g}}_{\mathbf{x}i}$ so that $\widehat{\mathbf{G}}_{\mathbf{x}} = \widehat{\mathbf{X}}$. Let $\widehat{\mathbf{X}}_\gamma(\gamma) = (\widehat{\mathbf{X}}_\gamma, \widehat{\mathbf{\Lambda}}_{1,\gamma}(\gamma), \widehat{\mathbf{\Lambda}}_{2,\gamma}(\gamma))$ be the stacked vector of $\widehat{\mathbf{x}}_{\gamma i}(\gamma) = (\widehat{\mathbf{x}}_i' I(q_i \leq \gamma), \widehat{\lambda}_{1,i}(\gamma) I(q_i \leq \gamma), \widehat{\lambda}_{2,i}(\gamma) I(q_i \leq \gamma))'$ and similarly define its orthogonal matrix $\widehat{\mathbf{X}}_\perp(\gamma) = (\widehat{\mathbf{X}}_\perp, \widehat{\mathbf{\Lambda}}_{1,\perp}(\gamma), \widehat{\mathbf{\Lambda}}_{2,\perp}(\gamma))$.

We can then define the projections $\mathbf{P}_\gamma(\gamma) = \widehat{\mathbf{X}}_\gamma(\gamma)(\widehat{\mathbf{X}}_\gamma(\gamma)'\widehat{\mathbf{X}}_\gamma(\gamma))^{-1}\widehat{\mathbf{X}}_\gamma(\gamma)$, $\mathbf{P}_\perp(\gamma) = \widehat{\mathbf{X}}_\perp(\gamma)(\widehat{\mathbf{X}}_\perp(\gamma)'\widehat{\mathbf{X}}_\perp(\gamma))^{-1}\widehat{\mathbf{X}}_\perp(\gamma)'$, and $\mathbf{P}^*(\gamma)=\widehat{\mathbf{X}}^*(\gamma)(\widehat{\mathbf{X}}^*(\gamma)'\widehat{\mathbf{X}}^*(\gamma))^{-1}\widehat{\mathbf{X}}^*(\gamma)'$ where $\widehat{\mathbf{X}}^*(\gamma) = (\widehat{\mathbf{X}}_\gamma(\gamma), \widehat{\mathbf{X}}_\perp(\gamma))$ such that $\mathbf{P}^*(\gamma) = \mathbf{P}_\gamma(\gamma) + \mathbf{P}_\perp(\gamma)$.

Finally, let us also define the second stage residual $\widehat{e}_i = \widehat{\mathbf{r}}'_{\mathbf{x}}\boldsymbol{\beta} + e_i$ and its vector form $\widehat{\mathbf{e}} = \widehat{\mathbf{r}}_{\mathbf{x}}\boldsymbol{\beta} + \mathbf{e}$.

■

LEMMA 1. For some $B < \infty$ and $\underline{\gamma} \leq \gamma' \leq \gamma \leq \overline{\gamma}$ and $r \leq 4$, uniformly in γ

$$
\begin{aligned}
Eh_i^r(\gamma, \gamma') &\leq B|\gamma - \gamma'| \quad (A.3)\\
Ek_i^r(\gamma, \gamma') &\leq B|\gamma - \gamma'| \quad (A.4)
\end{aligned}
$$

Proof of Lemma 1.

Define $d_i(\gamma) = I_{\{q_i \leq \gamma\}}$ and $d_i^{\perp}(\gamma) = I_{\{q_i > \gamma\}}$. Define $h_i(\gamma, \gamma') = |(h_i(\gamma) - h_i(\gamma'))\, e_i|$ and $k_i(\gamma, \gamma') = |(h_i(\gamma) - h_i(\gamma'))|$. In the case of the STR model in equation (2.13) $h_i(\gamma) = (\mathbf{g}_i d_i(\gamma), \lambda_i(\gamma))$ and thus $h_i(\gamma, \gamma')$ takes the form

$$
h_i(\gamma, \gamma') = \begin{pmatrix} |\mathbf{g}_i\varepsilon_i||d_i(\gamma) - d_i(\gamma')| \\ |\lambda_i(\gamma)\varepsilon_i - \lambda_i(\gamma')\varepsilon_i| \end{pmatrix}
$$

The first argument in our $h_i(\gamma, \gamma')$ is the same as Hansen (2000) and Caner and Hansen (2004) so it is sufficient to show that

$$
E|\lambda_i(\gamma)\varepsilon_i - \lambda_i(\gamma')\varepsilon_i|^r \leq B|\gamma - \gamma'|^\lambda
$$

$E|\lambda_i(\gamma)\varepsilon_i - \lambda_i(\gamma')\varepsilon_i|^r =$

$E|\left((\lambda_{2i}(\gamma) - \lambda_{2i}(\gamma')) + (\lambda_{1i}(\gamma)d_i(\gamma) - \lambda_{1i}(\gamma')d_i(\gamma')) - (\lambda_{2i}(\gamma)d_i(\gamma) - \lambda_{2i}(\gamma')d_i(\gamma'))\right)\varepsilon_i|^r \leq$

$(E|(\lambda_{2i}(\gamma) - \lambda_{2i}(\gamma'))\,\varepsilon_i|^r)^{1/r} + (E|(\lambda_{1i}(\gamma)d_i(\gamma) - \lambda_{1i}(\gamma')d_i(\gamma'))\,\varepsilon_i|^r)^{1/r}$

$+ (E|(\lambda_{2i}(\gamma)d_i(\gamma) - \lambda_{2i}(\gamma')d_i(\gamma'))\varepsilon_i|^r)^{1/r} \leq$

$\left(E|\left(\overline{\lambda}_{2i} - \underline{\lambda}_{2i}\right)\varepsilon_i|^r\right)^{1/r} + \left(E|(\overline{\lambda}_{1i} - \underline{\lambda}_{1i})\varepsilon_i(d_i(\gamma) - d_i(\gamma')|^r\right)^{1/r} + \left(E|(\overline{\lambda}_{2i} - \underline{\lambda}_{2i})\varepsilon_i(d_i(\gamma) - d_i(\gamma')|^r\right)^{1/r}.$

The last inequality is due to the monotonicity of $\lambda_{1i}(\gamma)$ and $\lambda_{2i}(\gamma)$. Then by Lemma A1 of Hansen (2000) it follows that

$E|\lambda_i(\gamma)\varepsilon_i - \lambda_i(\gamma')\varepsilon_i|^r \leq C_1 + C_2\,|\gamma - \gamma'| + C_3\,|\gamma - \gamma'| \leq B\,|\gamma - \gamma'|\,.$

■

LEMMA 2. Uniformly in $\gamma \in \Gamma$ as $n \longrightarrow \infty$

$$\frac{1}{n}\widehat{\mathbf{X}}^{*}(\gamma)'\widehat{\mathbf{X}}^{*}(\gamma) = \frac{1}{n}\sum_{i=1}^{n}\widehat{\mathbf{x}}_i^{*}(\gamma)\widehat{\mathbf{x}}_i^{*}(\gamma)' \xrightarrow{p} \mathbf{M}(\gamma) \tag{A.5}$$

$$\frac{1}{n}\widehat{\mathbf{X}}^{*}(\gamma_0)'\mathbf{G}^{*}(\gamma_0) = \frac{1}{n}\sum_{i=1}^{n}\widehat{\mathbf{x}}_i^{*}(\gamma)\widehat{\mathbf{x}}_i^{*}(\gamma)' \xrightarrow{p} \mathbf{M}(\gamma_0) \tag{A.6}$$

$$\frac{1}{\sqrt{n}}\widehat{\mathbf{X}}^{*}(\gamma)'\widehat{\mathbf{e}} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{\mathbf{x}}_i^{*}(\gamma)\widehat{\mathbf{e}}_i = O_p(1) \tag{A.7}$$

Proof of Lemma 2.

To show (A.5) note that

$\frac{1}{n}\widehat{\mathbf{X}}_{\gamma}(\gamma)'\widehat{\mathbf{X}}_{\gamma}(\gamma)$

$$= \begin{pmatrix} \frac{1}{n}\widehat{\mathbf{X}}_{\gamma}'\widehat{\mathbf{X}}_{\gamma} & \frac{1}{n}\widehat{\mathbf{X}}_{\gamma}'\widehat{\mathbf{\Lambda}}_{1\gamma}(\gamma) & \frac{1}{n}\widehat{\mathbf{X}}_{\gamma}'\widehat{\mathbf{\Lambda}}_{2\gamma}(\gamma) \\ \frac{1}{n}\widehat{\mathbf{\Lambda}}_{1\gamma}(\gamma)'\widehat{\mathbf{X}}_{\gamma} & \frac{1}{n}\widehat{\mathbf{\Lambda}}_{1\gamma}(\gamma)'\widehat{\mathbf{\Lambda}}_{1\gamma}(\gamma) & \frac{1}{n}\widehat{\mathbf{\Lambda}}_{1\gamma}(\gamma)'\widehat{\mathbf{\Lambda}}_{2\gamma}(\gamma) \\ \frac{1}{n}\widehat{\mathbf{\Lambda}}_{2\gamma}(\gamma)'\widehat{\mathbf{X}}_{\gamma} & \frac{1}{n}\widehat{\mathbf{\Lambda}}_{2\gamma}(\gamma)'\widehat{\mathbf{\Lambda}}_{1\gamma}(\gamma) & \frac{1}{n}\widehat{\mathbf{\Lambda}}_{2\gamma}(\gamma)'\widehat{\mathbf{\Lambda}}_{2\gamma}(\gamma) \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{n}\sum_i(\widehat{\mathbf{x}}_i\widehat{\mathbf{x}}_i' I(q_i \le \gamma)) & \frac{1}{n}\sum_i\widehat{\lambda}_{1i}(\gamma)\widehat{\mathbf{x}}_i I(q_i \le \gamma) & \frac{1}{n}\sum_i\widehat{\lambda}_{2i}(\gamma)\widehat{\mathbf{x}}_i I(q_i \le \gamma) \\ \frac{1}{n}\sum_i\widehat{\lambda}_{1i}(\gamma)\widehat{\mathbf{x}}_i' I(q_i \le \gamma)) & \frac{1}{n}\sum_i(\widehat{\lambda}_{1i}(\gamma))^2 I(q_i \le \gamma) & \frac{1}{n}\sum_i\widehat{\lambda}_{1i}(\gamma)\widehat{\lambda}_{2i}(\gamma) I(q_i \le \gamma) \\ \frac{1}{n}\sum_i\widehat{\lambda}_{2i}(\gamma)\widehat{\mathbf{x}}_i' I(q_i \le \gamma)) & \frac{1}{n}\sum_i\widehat{\lambda}_{1i}(\gamma)\widehat{\lambda}_{2i}(\gamma) I(q_i \le \gamma) & \frac{1}{n}\sum_i(\widehat{\lambda}_{2i}(\gamma))^2 I(q_i \le \gamma) \end{pmatrix} \text{ and}$$

recall that $\widehat{\mathbf{x}}_i = \widehat{\mathbf{g}}_{\mathbf{x}i} = \mathbf{g}_{\mathbf{x}i} - \widehat{\mathbf{r}}_{\mathbf{x}i}$, $\widehat{\lambda}_{1i}(\gamma) = \lambda_{1i}(\gamma) - \widehat{r}_{u_1 i}$, and $\widehat{\lambda}_{2i}(\gamma) = \lambda_{2i}(\gamma) - \widehat{r}_{u_2 i}$.

First note that $\frac{1}{n}\sum_i(\widehat{\mathbf{x}}_i\widehat{\mathbf{x}}_i' I(q_i \le \gamma)) \xrightarrow{p} E\left(\mathbf{g}_i\mathbf{g}_i' I(q_i \le \gamma)\right)$ follows from Caner and Hansen (2004) and Lemma 1 of Hansen (1996). Since the first stage regressions are consistently estimated, from Lemma 1 of Hansen (1996) we get for $j = 1, 2$

$$\begin{aligned} \frac{1}{n}\sum_i\widehat{\lambda}_{ji}(\gamma)\widehat{\mathbf{x}}_i I(q_i \le \gamma) &= \frac{1}{n}\sum_i\widehat{\lambda}_{ji}(\gamma)\mathbf{g}_i' I(q_i \le \gamma) - \frac{1}{n}\sum_i\widehat{\mathbf{r}}_{\mathbf{x}i}\widehat{\lambda}_{ji}(\gamma) I(q_i \le \gamma) \\ &= \frac{1}{n}\sum_i\lambda_{ji}(\gamma)\mathbf{g}_i' I(q_i \le \gamma) - \frac{1}{n}\sum_i\widehat{r}_{u_j i}\mathbf{g}_i' I(q_i \le \gamma) \\ &\quad -\frac{1}{n}\sum_i\widehat{\mathbf{r}}_{\mathbf{x}i}\lambda_{ji}(\gamma) I(q_i \le \gamma) + \frac{1}{n}\sum_i\widehat{\mathbf{r}}_{\mathbf{x}i}\widehat{r}_{u_j i} I(q_i \le \gamma) \end{aligned}$$

$$\frac{1}{n}\sum_i(\widehat{\lambda}_{ji}(\gamma))^2 I(q_i \le \gamma) = \frac{1}{n}\sum_i\left(\lambda_{ji}(\gamma)\right)^2 I(q_i \le \gamma) - 2\frac{1}{n}\sum_i\lambda_{ji}(\gamma)\widehat{r}_{u_j i} I(q_i \le \gamma) + \frac{1}{n}\sum_i\widehat{r}_{u_j i}^2 I(q_i \le \gamma)$$

Similarly, we can show that

$$\begin{aligned} \frac{1}{n}\sum_i\widehat{\lambda}_{1i}(\gamma)\widehat{\lambda}_{2i}(\gamma) I(q_i \le \gamma) &= \frac{1}{n}\sum_i\lambda_{1i}(\gamma)\lambda_{2i}(\gamma) I(q_i \le \gamma) - \frac{1}{n}\sum_i\lambda_{1i}(\gamma)\widehat{r}_{u_1 i} I(q_i \le \gamma) \\ &\quad -\frac{1}{n}\sum_i\lambda_{2i}(\gamma)\widehat{r}_{u_2 i} I(q_i \le \gamma) + \frac{1}{n}\sum_i\widehat{r}_{u_1 i}\widehat{r}_{u_2 i} I(q_i \le \gamma) \end{aligned}$$

Therefore, uniformly in $\gamma \in \Gamma$, $\frac{1}{n}\widehat{\mathbf{X}}_{\gamma}(\gamma)'\widehat{\mathbf{X}}_{\gamma}(\gamma) \xrightarrow{p} E(\mathbf{g}_{\gamma i}(\gamma)\mathbf{g}_{\gamma i}(\gamma)') = \mathbf{M}_{\gamma}(\gamma)$, where

$$\mathbf{M}_{\gamma}(\gamma) = \begin{pmatrix} E\left(\mathbf{g}_i\mathbf{g}_i' I(q_i \leq \gamma)\right) & E(\lambda_{1i}\left(\gamma\right)\mathbf{g}_i I(q_i \leq \gamma)) & E(\lambda_{2i}\left(\gamma\right)\mathbf{g}_i I(q_i \leq \gamma)) \\ E(\lambda_{1i}\left(\gamma\right)\mathbf{g}_i' I(q_i \leq \gamma)) & E\left(\lambda_{1i}\left(\gamma\right)\right)^2 I(q_i \leq \gamma) & E\lambda_{1i}\left(\gamma\right)\lambda_{2i}\left(\gamma\right) I(q_i \leq \gamma) \\ E(\lambda_{2i}\left(\gamma\right)\mathbf{g}_i' I(q_i \leq \gamma)) & E(\lambda_{2i}\left(\gamma\right)\lambda_{1i}\left(\gamma\right) I(q_i \leq \gamma)) & E\left(\lambda_{2i}\left(\gamma\right)\right)^2 I(q_i \leq \gamma) \end{pmatrix}$$

Similarly we can show that, $\frac{1}{n}\widehat{\mathbf{X}}_{\perp}(\gamma)'\widehat{\mathbf{X}}_{\perp}(\gamma) \xrightarrow{p} E(\mathbf{g}_{\perp i}(\gamma)\mathbf{g}_{\perp i}(\gamma)') = \mathbf{M}_{\perp}(\gamma)$. Then, we get (A.5)

$$\frac{1}{n}\widehat{\mathbf{X}}^*(\gamma)'\widehat{\mathbf{X}}^*(\gamma) \xrightarrow{p} \mathbf{M}(\gamma) = \begin{pmatrix} \mathbf{M}_{\gamma}(\gamma) & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_{\perp}(\gamma) \end{pmatrix}$$

(A.6) follows similarly. We now show (A.7).

First note that $\frac{1}{n}\sum_i(\mathbf{x}_i\widehat{\mathbf{e}}_i' I(q_i \leq \gamma)) = O_p(1)$ follows from Caner and Hansen (2004). Second, from Lemma A.4 of Hansen (2000) and Theorem 1 of Hansen (1996) we can obtain for $j = 1, 2$

$$\begin{aligned} \frac{1}{n}\sum_i \widehat{\lambda}_{ji}\left(\gamma\right)\widehat{e}_i' I(q_i \leq \gamma)) &= \frac{1}{n}\sum_i \lambda_{ji}\left(\gamma\right)\widehat{e}_i' I(q_i \leq \gamma) - \frac{1}{n}\sum_i \widehat{r}_{u_j i}\widehat{e}_i' I(q_i \leq \gamma) \\ &= \frac{1}{n}\sum_i \lambda_{ji}\left(\gamma\right)\boldsymbol{\beta}'\widehat{\mathbf{r}}_{\mathbf{x}i} I(q_i \leq \gamma) + \frac{1}{n}\sum_i \lambda_{ji}\left(\gamma\right) e_i' I(q_i \leq \gamma) \\ &\quad -\frac{1}{n}\sum_i \widehat{r}_{u_j i}\boldsymbol{\beta}'\widehat{\mathbf{r}}_{\mathbf{x}i} I(q_i \leq \gamma) - \frac{1}{n}\sum_i \widehat{r}_{u_j i} e_i I(q_i \leq \gamma) \\ &= O_p(1) \end{aligned}$$

Then,

$$\frac{1}{\sqrt{n}}\widehat{\mathbf{X}}_{\gamma}(\gamma)'\widehat{\mathbf{e}} = \begin{pmatrix} \frac{1}{n}\sum_i \widehat{\mathbf{x}}_i\widehat{e}_i I(q_i \leq \gamma) \\ \frac{1}{n}\sum_i(\widehat{\lambda}_{1i}\left(\gamma\right)\widehat{e}_i' I(q_i \leq \gamma)) \\ \frac{1}{n}\sum_i(\widehat{\lambda}_{2i}\left(\gamma\right)\widehat{e}_i' I(q_i \leq \gamma)) \end{pmatrix} \xrightarrow{p} O_p(1)$$

Similarly, we can show that $\frac{1}{\sqrt{n}}\widehat{\mathbf{X}}_{\perp}(\gamma)'\widehat{\mathbf{e}} \xrightarrow{p} O_p(1)$ and hence $\frac{1}{\sqrt{n}}\widehat{\mathbf{X}}^*(\gamma)'\widehat{\mathbf{e}} \xrightarrow{p} O_p(1)$.

■

Proof of Proposition 1.

The proof proceeds as follows. First, we show that $\widehat{\gamma}$ is consistent for the unrestricted problem following the proof strategy of Caner and Hansen (2004). Then, we show that the same estimator has to be consistent for the restricted problem.

Define $\widehat{\mathbf{e}} = \widehat{\mathbf{r}}\boldsymbol{\beta} + \mathbf{e}$. Given that $\mathbf{G}(\gamma) = \widehat{\mathbf{G}}(\gamma) + \widehat{\mathbf{V}}$ and $\widehat{\mathbf{G}}(\gamma) = \widehat{\mathbf{X}}(\gamma)$ is in the span of $\widehat{\mathbf{X}}^*(\gamma)$ then $(\mathbf{I} - \mathbf{P}^*(\gamma))\mathbf{G}(\gamma) = (\mathbf{I} - \mathbf{P}^*(\gamma))\widehat{\mathbf{r}}$ and

$$(\mathbf{I} - \mathbf{P}^*(\gamma))\mathbf{Y} = (\mathbf{I} - \mathbf{P}^*(\gamma))(\mathbf{G}(\gamma_0)\boldsymbol{\beta} + \mathbf{G}_0(\gamma_0)\boldsymbol{\delta}_n + \widehat{\mathbf{e}})$$

Then

$$\begin{aligned} S_n^U(\gamma) &= \mathbf{Y}'(\mathbf{I} - \mathbf{P}^*(\gamma))\mathbf{Y} & \text{(A.8)} \\ &= (n^{-\boldsymbol{\alpha}}\mathbf{c}'\mathbf{G}_0(\gamma_0)' + \widehat{\mathbf{e}}')(\mathbf{I} - \mathbf{P}^*(\gamma))(\mathbf{G}_0(\gamma_0)n^{-\boldsymbol{\alpha}}\mathbf{c} + \widehat{\mathbf{e}}) & \text{(A.9)} \\ &= (n^{-\boldsymbol{\alpha}}\mathbf{c}'\mathbf{G}_0(\gamma_0)' + \widehat{\mathbf{e}}')(\mathbf{G}_0(\gamma_0)n^{-\boldsymbol{\alpha}}\mathbf{c} + \widehat{\mathbf{e}}) \\ &\quad -(n^{-\boldsymbol{\alpha}}\mathbf{c}'\mathbf{G}_0(\gamma_0)' + \widehat{\mathbf{e}}')\mathbf{P}^*(\gamma)(\mathbf{G}_0(\gamma_0)n^{-\boldsymbol{\alpha}}\mathbf{c} + \widehat{\mathbf{e}}) & \text{(A.10)} \end{aligned}$$

Because the first term in the last equality does not depend on γ, and $\widehat{\gamma}$ minimizes $S_n^U(\gamma)$, we can equivalently write that $\widehat{\gamma}$ maximizes $S_n^*(\gamma)$ where

$$\begin{aligned} S_n^{*U}(\gamma) &= n^{-1+2\alpha}(n^{-\boldsymbol{\alpha}}\mathbf{c}'\mathbf{G}_0(\gamma_0)' + \widehat{\mathbf{e}}')\mathbf{P}^*(\gamma)(\mathbf{G}_0(\gamma_0)n^{-\boldsymbol{\alpha}}\mathbf{c} + \widehat{\mathbf{e}}) \\ &= n^{-1+2\alpha}\widehat{\mathbf{e}}'\mathbf{P}^*(\gamma)\widehat{\mathbf{e}} + 2n^{-1+\alpha}\mathbf{c}'\mathbf{G}_0(\gamma_0)'\mathbf{P}^*(\gamma)\widehat{\mathbf{e}} + n^{-1}\mathbf{c}'\mathbf{G}_0(\gamma_0)'\mathbf{P}^*(\gamma)\mathbf{G}_0(\gamma_0)\mathbf{c} \end{aligned}$$

Let us now examine $S_n^{*U}(\gamma)$ for $\gamma \in (\gamma_0, \overline{\gamma}]$. Note that $\mathbf{G}_0(\gamma_0)'\mathbf{P}_\perp(\gamma) = 0$

From Lemma 2 we can show that for all $\gamma \in \Gamma$,

$n^{-1+2\alpha}\widehat{\mathbf{e}}'\mathbf{P}_\gamma(\gamma)\widehat{\mathbf{e}} = n^{-1+2\alpha}(\frac{1}{\sqrt{n}}\widehat{\mathbf{e}}'\widehat{\mathbf{X}}_\gamma(\gamma))(\frac{1}{n}\widehat{\mathbf{X}}_\gamma(\gamma)'\widehat{\mathbf{X}}_\gamma(\gamma))^{-1}(\frac{1}{\sqrt{n}}\widehat{\mathbf{X}}_\gamma(\gamma)'\widehat{\mathbf{e}}) \xrightarrow{p} 0$

$n^{-1+2\alpha}\widehat{\mathbf{e}}'\mathbf{P}_\perp(\gamma)\widehat{\mathbf{e}} = n^{2\alpha-1}(\frac{1}{\sqrt{n}}\widehat{\mathbf{e}}'\widehat{\mathbf{X}}_\perp(\gamma))(\frac{1}{n}\widehat{\mathbf{X}}_\perp(\gamma)'\widehat{\mathbf{X}}_\perp(\gamma))^{-1}(\frac{1}{\sqrt{n}}\widehat{\mathbf{X}}_\perp(\gamma)'\widehat{\mathbf{e}}) \xrightarrow{p} 0$

and

$n^{-1+\alpha}\mathbf{c}_{\boldsymbol{\delta}}'\mathbf{G}_0(\gamma_0)'\mathbf{P}_\gamma(\gamma)\widehat{\mathbf{e}} = n^{\alpha-1/2}(\frac{1}{n}\mathbf{G}_0(\gamma_0)'\widehat{\mathbf{X}}_0(\gamma))(\frac{1}{n}\widehat{\mathbf{X}}_\gamma(\gamma)'\widehat{\mathbf{X}}_\gamma(\gamma))^{-1}(\frac{1}{\sqrt{n}}\widehat{\mathbf{X}}_\gamma(\gamma)'\widehat{\mathbf{e}}) \xrightarrow{p} 0$

So

$S_n^{*U}(\gamma) = n^{-1+2\alpha}\widehat{\mathbf{e}}'\mathbf{P}_\gamma(\gamma)\widehat{\mathbf{e}} + n^{-1+2\alpha}\widehat{\mathbf{e}}'\mathbf{P}_\perp(\gamma)\widehat{\mathbf{e}} + 2n^{-1+\alpha}\mathbf{c}'\mathbf{G}_0(\gamma_0)'\mathbf{P}_\gamma(\gamma)\widehat{\mathbf{e}}$

$+n^{-1}\mathbf{c}'\mathbf{G}_0(\gamma_0)'\mathbf{P}_\gamma(\gamma)\mathbf{G}_0(\gamma_0)\mathbf{c}.$

Before examining the last two terms let us calculate $\frac{1}{n}\widehat{\mathbf{X}}_1(\gamma)'\mathbf{G}(\gamma_0)$ and $\frac{1}{n}\widehat{\mathbf{X}}_2(\gamma)'\mathbf{G}(\gamma_0)$

$$\frac{1}{n}\widehat{\mathbf{X}}_{\gamma}(\gamma)'\mathbf{G}_0(\gamma_0) = \begin{pmatrix} \frac{1}{n}\widehat{\mathbf{X}}_{\gamma}'\mathbf{G}_{\mathbf{x},0} & \frac{1}{n}\widehat{\mathbf{X}}_{\gamma}'\mathbf{\Lambda}_{1,0}(\gamma_0) & \frac{1}{n}\widehat{\mathbf{X}}_{\gamma}'\mathbf{\Lambda}_{2,0}(\gamma_0) \\ \frac{1}{n}\widehat{\mathbf{\Lambda}}_{1,\gamma}(\gamma)'\mathbf{G}_{\mathbf{x},0} & \frac{1}{n}\widehat{\mathbf{\Lambda}}_{1,\gamma}(\gamma)'\mathbf{\Lambda}_{1,0}(\gamma_0) & \frac{1}{n}\widehat{\mathbf{\Lambda}}_{1,\gamma}(\gamma)'\mathbf{\Lambda}_{2,0}(\gamma_0) \\ \frac{1}{n}\widehat{\mathbf{\Lambda}}_{2,\gamma}(\gamma)'\mathbf{G}_{\mathbf{x},0} & \frac{1}{n}\widehat{\mathbf{\Lambda}}_{2,\gamma}(\gamma)'\mathbf{\Lambda}_{1,0}(\gamma_0) & \frac{1}{n}\widehat{\mathbf{\Lambda}}_{2,\gamma}(\gamma)'\mathbf{\Lambda}_{2,0}(\gamma_0) \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{n}\sum_i \mathbf{g}_{\mathbf{x}i}'\widehat{\mathbf{x}}_i I(q_i \le \gamma_0) & \frac{1}{n}\sum_i \lambda_{1,i}(\gamma_0)\,\widehat{\mathbf{x}}_i I(q_i \le \gamma_0) & \frac{1}{n}\sum_i \lambda_{2,i}(\gamma_0)\,\widehat{\mathbf{x}}_i I(q_i \le \gamma_0) \\ \frac{1}{n}\sum_i \mathbf{g}_{\mathbf{x}i}'\widehat{\lambda}_{1i}(\gamma)\,I(q_i \le \gamma_0) & \frac{1}{n}\sum_i \lambda_{1i}(\gamma_0)\,\widehat{\lambda}_{1i}(\gamma)\,I(q_i \le \gamma) & \frac{1}{n}\sum_i \lambda_{2i}(\gamma_0)\,\widehat{\lambda}_{1i}(\gamma)\,I(q_i \le \gamma_0) \\ \frac{1}{n}\sum_i \mathbf{g}_{\mathbf{x}i}'\widehat{\lambda}_{2i}(\gamma)\,I(q_i \le \gamma_0) & \frac{1}{n}\sum_i \lambda_{1i}(\gamma_0)\,\widehat{\lambda}_{2i}(\gamma)\,I(q_i \le \gamma_0) & \frac{1}{n}\sum_i \lambda_{2i}(\gamma_0)\,\widehat{\lambda}_{2i}(\gamma)\,I(q_i \le \gamma_0) \end{pmatrix}$$

$$\to \begin{pmatrix} E(\mathbf{g}_{\mathbf{x}i}\mathbf{g}_{\mathbf{x}i}'I(q_i \le \gamma_0) & E(\mathbf{g}_{\mathbf{x}i}\lambda_{1i}(\gamma_0)\,I(q_i \le \gamma_0)) & E(\lambda_{2,i}(\gamma_0)\,\mathbf{g}_{\mathbf{x}i}I(q_i \le \gamma_0)) \\ E(\lambda_{1i}(\gamma)\,\mathbf{g}_{\mathbf{x}i}'I(q_i \le \gamma_0)) & E(\lambda_{1i}(\gamma_0)\,\lambda_{1i}(\gamma)\,I(q_i \le \gamma_0)) & E(\lambda_{2i}(\gamma_0)\,\lambda_{1i}(\gamma)\,I(q_i \le \gamma_0)) \\ E(\lambda_{2i}(\gamma)\,\mathbf{g}_{\mathbf{x}i}'I(q_i \le \gamma_0)) & E(\lambda_{1i}(\gamma_0)\,\lambda_{2i}(\gamma)\,I(q_i \le \gamma_0)) & E(\lambda_{2i}(\gamma_0)\,\lambda_{2i}(\gamma)\,I(q_i \le \gamma_0)) \end{pmatrix}$$

$$\equiv \mathbf{M}_0(\gamma_0, \gamma).$$

Note that when $\gamma = \gamma_0$, $\mathbf{M}_0(\gamma_0, \gamma_0) = \mathbf{M}_0(\gamma_0)$ as it is in the case of Hansen (2000) and Caner and Hansen (2004).

Therefore,

$$\frac{1}{n}\mathbf{G}_0(\gamma_0)'\mathbf{P}_{\gamma}(\gamma)\,\mathbf{G}_0(\gamma_0) \to \mathbf{M}_0(\gamma_0, \gamma)'\mathbf{M}_{\gamma}(\gamma)^{-1}\mathbf{M}_0(\gamma_0, \gamma)$$

Then, uniformly for $\gamma \in (\gamma_0, \overline{\gamma}]$ we get

$$S_n^{*U}(\gamma) \to \mathbf{c}'\mathbf{M}_0(\gamma_0, \gamma)'\mathbf{M}_{\gamma}(\gamma)^{-1}\mathbf{M}_0(\gamma_0, \gamma)\mathbf{c} \tag{A.11}$$

by a Glivenko-Cantelli theorem for stationary ergodic processes.

Given the monotonicity of the inverse Mills ratio, $\mathbf{M}_0(\gamma_0, \gamma_0 + \epsilon) \ge \mathbf{M}_0(\gamma_0)$ for any $\epsilon > 0$ with equality at $\gamma = \gamma_0$. To see this note that for $\epsilon > 0$, $\lambda_{1i}(\gamma_0 + \epsilon) > \lambda_{1i}(\gamma_0)$ and $\lambda_{2i}(\gamma_0 + \epsilon) > \lambda_{2i}(\gamma_0)$. Therefore, we need to show that $S_n^{*U}(\gamma) < \mathbf{M}_0(\gamma_0)$ for any $\gamma \in (\gamma_0, \overline{\gamma}]$. It is sufficient to show that $\mathbf{M}_0(\gamma_0)'\mathbf{M}_{\gamma}(\gamma)^{-1}\mathbf{M}_0(\gamma_0) < \mathbf{M}_0(\gamma_0)$, which reduces to $\mathbf{M}_{\gamma}(\gamma) > \mathbf{M}_0(\gamma_0)$ for any $\gamma \in (\gamma_0, \overline{\gamma}]$.

To see this recall that $\mathbf{M}_\gamma(\gamma) = E\left(\mathbf{g}_{\gamma i}(\gamma)\mathbf{g}'_{\gamma i}(\gamma)\right)$. Then,

$$\begin{aligned}
\mathbf{M}_\varepsilon(\gamma_0+\epsilon) - \mathbf{M}_0(\gamma_0) &= \int\limits_{\gamma_0}^{\gamma_0+\epsilon} E(\mathbf{g}_i(t)\mathbf{g}_i(t)'|q=t)f_q(t)dt \\
&> \inf_{\gamma_0<\gamma\leq\gamma_0+\epsilon} E\mathbf{g}_i(\gamma)\mathbf{g}_i'(\gamma))|q=\gamma)\left(\int\limits_{\gamma_0}^{\gamma_0+\epsilon} f(\nu)d\nu\right) \\
&= \inf_{\gamma_0<\gamma\leq\gamma_0+\epsilon} \mathbf{D}_1(\gamma)\left(\int\limits_{\gamma_0}^{\gamma_0+\epsilon} f(\nu)d\nu\right) > 0
\end{aligned}$$

Therefore, $S^*(\gamma)$ is uniquely maximized at γ_0, for $\gamma \in (\gamma_0, \overline{\gamma}]$. The case of $\gamma \in [\underline{\gamma}, \gamma_0]$ can be proved using symmetric arguments.

Thus, the conditions of Theorem 5.7 by Van der Vaart (1998) are satisfied. Given the uniform convergence of $S_n^*(\gamma)$, i.e. $\sup\limits_{\gamma\in\Gamma}|S_n^{*U}(\gamma) - S_n^{*U}(\gamma_0)| \xrightarrow{p} 0$ as $n \longrightarrow \infty$, the compactness of Γ, and the fact that $S_n^{*U}(\gamma)$ is uniquely maximized at γ_0, we can have $\sup\limits_{|\gamma-\gamma_0|\geq\epsilon} S_n^{*U}(\gamma) < S_n^{*U}(\gamma_0)$ for every $\epsilon > 0$. Therefore, it follows that $\widehat{\gamma} \xrightarrow{p} \gamma_0$ for the unrestricted problem.

Assuming the restrictions in equation (3.23) hold we have

$$S_n^R(\widehat{\gamma}) \leq S_n^R(\gamma_0) \leq S_n^U(\gamma) \tag{A.12}$$

When $\widehat{\gamma}$ is not consistent it must be the case that $S_n^R(\widehat{\gamma}) \geq S_n^U(\gamma) + C||\boldsymbol{\beta}_{10} - \boldsymbol{\beta}_1||^2 + ||\boldsymbol{\beta}_{20} - \boldsymbol{\beta}_2||^2 + o_p(1)$, where $\boldsymbol{\beta}_{10}$ and $\boldsymbol{\beta}_{20}$ are the true slope coefficients for the two regimes. But since $S_n^U(\widehat{\gamma}) \leq S_n^R(\widehat{\gamma})$ we also have $S_n^R(\widehat{\gamma}) \geq S_n^U(\gamma) + C||\boldsymbol{\beta}_{10} - \boldsymbol{\beta}_1||^2 + ||\boldsymbol{\beta}_{20} - \boldsymbol{\beta}_2||^2 + o_p(1)$, which yields a contradition with (A.12). This completes the proof.

■

LEMMA 3. $a_n(\widehat{\gamma} - \gamma_0) = O_p(1)$.

Proof of Lemma 3.

Note that $S_n^R(\gamma) = S_n^U(\gamma) + (\boldsymbol{\vartheta} - \mathbf{R}'\widehat{\boldsymbol{\theta}})'(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma)'\widehat{\mathbf{X}}^*(\gamma))^{-1}\mathbf{R})^{-1}(\boldsymbol{\vartheta} - \mathbf{R}'\widehat{\boldsymbol{\theta}})$. The proof proceeds in steps. First we establish that the unrestricted and the restricted problems share the same rate of convergence.

$$\frac{S_n^R(\gamma) - S_n^R(\gamma_0)}{n^{2\alpha-1}(\gamma-\gamma_0)} \geq \frac{S_n^U(\gamma) - S_n^U(\gamma_0)}{n^{2\alpha-1}(\gamma-\gamma_0)} + o_p(1) \tag{A.13}$$

The key is to show that the second term is $o_p(1)$.

Using Lemma A.2 of Perron and Qu (2006) and the joint events A.24-A.32 of Caner and Hansen (2004) we can deduce that

$$(\widehat{\mathbf{X}}^*(\gamma)'\widehat{\mathbf{X}}^*(\gamma))^{-1} = (\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1} + O_p(\frac{|\gamma-\gamma_0|}{n^2}) \tag{A.14}$$

and

$$(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma)'\widehat{\mathbf{X}}^*(\gamma))^{-1}\mathbf{R})^{-1} = (\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\mathbf{R})^{-1} + O_p(|\gamma-\gamma_0|). \tag{A.15}$$

Consider $\Delta\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_0 = (\widehat{\mathbf{X}}^*(\gamma)'\widehat{\mathbf{X}}^*(\gamma))^{-1}\widehat{\mathbf{X}}^*(\gamma)'(\widehat{\mathbf{X}}^*(\gamma_0)\boldsymbol{\theta}_0 + \mathbf{e}) - \boldsymbol{\theta}_0 - (\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\widehat{\mathbf{X}}^*(\gamma_0)'\mathbf{e})$

$= (\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}((\widehat{\mathbf{X}}^*(\gamma) - \widehat{\mathbf{X}}^*(\gamma_0))'\widehat{\mathbf{X}}^*(\gamma_0)\boldsymbol{\theta}_0 + (\widehat{\mathbf{X}}^*(\gamma) - \widehat{\mathbf{X}}^*(\gamma_0))'\mathbf{e}) + |\gamma-\gamma_0|O_p(\frac{1}{n})$

$= (\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1/2}\mathbf{A}_n$

with

$\mathbf{A}_n = \widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1/2}(\widehat{\mathbf{X}}^*(\gamma) - \widehat{\mathbf{X}}^*(\gamma_0))'\widehat{\mathbf{X}}^*(\gamma_0)\boldsymbol{\theta}_0$

$+(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1/2}(\widehat{\mathbf{X}}^*(\gamma) - \widehat{\mathbf{X}}^*(\gamma_0))'\mathbf{e} + |\gamma-\gamma_0|O_p(\frac{1}{\sqrt{n}})$

$= |\gamma-\gamma_0|O_p(n^{-1/2}) + |\gamma-\gamma_0|O_p(1) + |\gamma-\gamma_0|O_p(\frac{1}{\sqrt{n}})$

$=|\gamma-\gamma_0|O_p(n^{-1/2})$ and hence $\Delta\widehat{\boldsymbol{\theta}} = |\gamma-\gamma_0|O_p(n^{-1})$. using (A.14) for the first equality and the fact that $(\widehat{\mathbf{X}}^*(\gamma) - \widehat{\mathbf{X}}^*(\gamma_0))'\widehat{\mathbf{X}}^*(\gamma_0)$ and $(\widehat{\mathbf{X}}^*(\gamma) - \widehat{\mathbf{X}}^*(\gamma_0))'\mathbf{e}$ are $|\gamma-\gamma_0|O_p(1)$ terms.

Then it is easy to see that $\Delta\widehat{\boldsymbol{\theta}}'\mathbf{R}$ and $(\boldsymbol{\vartheta} - \mathbf{R}'\widehat{\boldsymbol{\theta}})'$ are $|\gamma-\gamma_0|O_p(n^{-1})$, too.

$S_n^R(\gamma) - S_n^R(\gamma_0) =$

$[S_n^U(\gamma) - S_n^U(\gamma_0)]+$

$[(\boldsymbol{\vartheta} - \mathbf{R}'\widehat{\boldsymbol{\theta}})'(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma)'\widehat{\mathbf{X}}^*(\gamma))^{-1}\mathbf{R})^{-1}(\boldsymbol{\vartheta} - \mathbf{R}'\widehat{\boldsymbol{\theta}})-$

$(\boldsymbol{\vartheta} - \mathbf{R}'\widehat{\boldsymbol{\theta}}_0)'(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\mathbf{R})^{-1}(\boldsymbol{\vartheta} - \mathbf{R}'\widehat{\boldsymbol{\theta}}_0)] =$

$[S_n^U(\gamma) - S_n^U(\gamma_0)]+$

$[(\boldsymbol{\vartheta} - \mathbf{R}'\widehat{\boldsymbol{\theta}})'(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\mathbf{R})^{-1}(\boldsymbol{\vartheta} - \mathbf{R}'\widehat{\boldsymbol{\theta}})-$

$(\boldsymbol{\vartheta} - \mathbf{R}'\widehat{\boldsymbol{\theta}}_0)'(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\mathbf{R})^{-1}(\boldsymbol{\vartheta} - \mathbf{R}'\widehat{\boldsymbol{\theta}}_0)] + (\gamma-\gamma_0)^2O_p(n^{-1}) =$

$[S_n^U(\gamma) - S_n^U(\gamma_0)]+$

$(\boldsymbol{\theta}_0 + \Delta\widehat{\boldsymbol{\theta}})'\mathbf{R}(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\mathbf{R})^{-1}\mathbf{R}'(\boldsymbol{\theta}_0 + \Delta\widehat{\boldsymbol{\theta}})-$

$\widehat{\boldsymbol{\theta}}_0{}'\mathbf{R}(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\mathbf{R})^{-1}\mathbf{R}'\widehat{\boldsymbol{\theta}}_0-$

$2\boldsymbol{\vartheta}'\mathbf{R}(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\mathbf{R})^{-1}\mathbf{R}'(\widehat{\boldsymbol{\theta}}_1-\widehat{\boldsymbol{\theta}}_0)+$

$(\gamma-\gamma_0)^2O_p(n^{-1})$

$=[S_n^U(\gamma)-S_n^U(\gamma_0)]+$

$2\Delta\widehat{\boldsymbol{\theta}}'\mathbf{R}(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\mathbf{R})^{-1}\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\widehat{\mathbf{X}}^*(\gamma_0)'\mathbf{e}+$

$\Delta\widehat{\boldsymbol{\theta}}'\mathbf{R}(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\mathbf{R})^{-1}\mathbf{R}'\Delta\widehat{\boldsymbol{\theta}}+$

$2\Delta\widehat{\boldsymbol{\theta}}'\mathbf{R}(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\mathbf{R})^{-1}(\mathbf{R}\boldsymbol{\theta}_0-\boldsymbol{\vartheta})+$

$(\gamma-\gamma_0)^2O_p(n^{-1}).$

Note that the second and fourth term are $o_p(1)$ when divided by $(\gamma-\gamma_0)$ since $\Delta\widehat{\boldsymbol{\theta}}=|\gamma-\gamma_0|O_p(n^{-1})$. The third term is always positive. Therefore,we can now focus on the first term, which is the rate of convergence for the unrestricted problem.

Our proof follows in spirit Yu (2010b). In this lemma we use the notation for empirical processes in van der Vaart and Wellner (1996). Define $M_n(\theta)=\mathbb{P}_n m(\boldsymbol{\theta})$, where $\mathbb{P}_n$ denotes the empirical measure $\mathbb{P}_n=\frac{1}{n}\sum_{i=1}^n$, such that for any class of measurable function $f:x\to\mathbb{R}$, we denote $\mathbb{P}_n f=\frac{1}{n}\sum_{i=1}^n f(x_i)$. We also define $M(\theta)=\mathbb{P}m(\boldsymbol{\theta})$, where $\mathbb{P}m(\boldsymbol{\theta})=\int_x f(x)P(dx)$. Finally, define the empirical process $\mathbb{G}_n=\sqrt{n}(\mathbb{P}_n-\mathbb{P})$ so that $\mathbb{G}_n m(\boldsymbol{\theta})=\sqrt{n}(\mathbb{M}_n(\theta)-M)$.

Given that the theorem is for the maximization problem we will consider $m(\boldsymbol{\theta})=-(y_i-\mathbf{g}_i(\gamma)'\boldsymbol{\beta}_1 I(q_i\le\gamma)-\mathbf{g}_i(\gamma)'\boldsymbol{\beta}_2 I(q_i>\gamma))^2$ and let $\boldsymbol{\theta}=(\boldsymbol{\beta}_1',\boldsymbol{\beta}_2',\gamma)'$. Recall that $\gamma\in\Gamma=[\underline{\gamma},\overline{\gamma}]$, then we have $I(q_i\le\gamma)\le I(q\le\gamma\wedge\gamma_0)$ and $I(q_i>\gamma)\ge I(\gamma_0<q\le\gamma\vee\gamma_0)$, where "$\wedge$" and "$\vee$" denote the minimum and maximum, respectively.

We can derive the following formula.

$m(\boldsymbol{\theta})=-(y_i-\mathbf{g}_i(\gamma)'\boldsymbol{\beta}_1 I(q_i\le\gamma)-\mathbf{g}_i(\gamma)'\boldsymbol{\beta}_2 I(q_i>\gamma))^2$

$=$

$-\left[\mathbf{g}_i(\gamma_0)'\boldsymbol{\beta}_{10}-\mathbf{g}_i(\gamma)'\boldsymbol{\beta}_1+e_{1i}\right]^2 I(q_i\le\gamma\wedge\gamma_0)$

$-\left[\mathbf{g}_i(\gamma_0)'\boldsymbol{\beta}_{20}-\mathbf{g}_i(\gamma)'\boldsymbol{\beta}_2+e_{2i}\right]^2 I(q_i>\gamma\vee\gamma_0)$

$-\left[\mathbf{g}_i(\gamma_0)'\boldsymbol{\beta}_{10}-\mathbf{g}_i(\gamma)'\boldsymbol{\beta}_2+e_{1i}\right]^2 I(\gamma\wedge\gamma_0<q_i\le\gamma_0)$

$-\left[\mathbf{g}_i(\gamma_0)'\boldsymbol{\beta}_{20}-\mathbf{g}_i(\gamma)'\boldsymbol{\beta}_1+e_{2i}\right]^2 I(\gamma_0<q_i\le\gamma\vee\gamma_0)$

$=$

$-\left[\mathbf{g}_{\mathbf{x}i}'(\boldsymbol{\beta}_{\mathbf{x}10}-\boldsymbol{\beta}_{\mathbf{x}1})+\boldsymbol{\lambda}_i(\gamma_0)'(\boldsymbol{\beta}_{\lambda10}-\boldsymbol{\beta}_{\lambda1})+(\boldsymbol{\lambda}_i(\gamma)-\boldsymbol{\lambda}_i(\gamma_0))'\boldsymbol{\beta}_{\lambda1}+e_{1i}\right]^2 I(q\le\gamma\wedge\gamma_0)$

$-\left[\mathbf{g}_{\mathbf{x}i}'(\boldsymbol{\beta}_{\mathbf{x}20}-\boldsymbol{\beta}_{\mathbf{x}2})+\boldsymbol{\lambda}_i(\gamma_0)'(\boldsymbol{\beta}_{\lambda20}-\boldsymbol{\beta}_{\lambda2})+(\boldsymbol{\lambda}_i(\gamma)-\boldsymbol{\lambda}_i(\gamma_0))'\boldsymbol{\beta}_{\lambda2}+e_{2i}\right]^2 I(q>\gamma\vee\gamma_0)$

$$- \left[\mathbf{g}_{\mathbf{x}i}'(\boldsymbol{\beta}_{\mathbf{x}10} - \boldsymbol{\beta}_{\mathbf{x}2}) + \boldsymbol{\lambda}_i(\gamma_0)'(\boldsymbol{\beta}_{\lambda 10} - \boldsymbol{\beta}_{\lambda 2}) + (\boldsymbol{\lambda}_i(\gamma) - \boldsymbol{\lambda}_i(\gamma_0))'\boldsymbol{\beta}_{\lambda 2} + e_{1i}\right]^2 I(\gamma \wedge \gamma_0 < q \leq \gamma_0)$$

$$- \left[\mathbf{g}_{\mathbf{x}i}'(\boldsymbol{\beta}_{\mathbf{x}20} - \boldsymbol{\beta}_{\mathbf{x}1}) + \boldsymbol{\lambda}_i(\gamma_0)'(\boldsymbol{\beta}_{\lambda 20} - \boldsymbol{\beta}_{\lambda 1}) + (\boldsymbol{\lambda}_i(\gamma) - \boldsymbol{\lambda}_i(\gamma_0))'\boldsymbol{\beta}_{\lambda 1} + e_{2i}\right]^2 I(\gamma_0 < q \leq \gamma \vee \gamma_0)$$

$$=$$

$$- \left[\mathbf{g}_i(\gamma_0)'(\boldsymbol{\beta}_{\mathbf{x}10} - \boldsymbol{\beta}_{\mathbf{x}1}) + (\boldsymbol{\lambda}_i(\gamma) - \boldsymbol{\lambda}_i(\gamma_0))'\boldsymbol{\beta}_{\lambda 1} + e_{1i}\right]^2 I(q \leq \gamma \wedge \gamma_0)$$

$$- \left[\mathbf{g}_i(\gamma_0)'(\boldsymbol{\beta}_{\mathbf{x}20} - \boldsymbol{\beta}_{\mathbf{x}2}) + (\boldsymbol{\lambda}_i(\gamma) - \boldsymbol{\lambda}_i(\gamma_0))'\boldsymbol{\beta}_{\lambda 2} + e_{2i}\right]^2 I(q > \gamma \vee \gamma_0)$$

$$- \left[\mathbf{g}_i(\gamma_0)'(\boldsymbol{\beta}_{\mathbf{x}20} - \boldsymbol{\beta}_{\mathbf{x}2}) + (\boldsymbol{\lambda}_i(\gamma) - \boldsymbol{\lambda}_i(\gamma_0))'\boldsymbol{\beta}_{\lambda 2} + e_{2i}\right]^2 I(\gamma \wedge \gamma_0 < q \leq \gamma_0)$$

$$- \left[\mathbf{g}_i(\gamma_0)'(\boldsymbol{\beta}_{\mathbf{x}20} - \boldsymbol{\beta}_{\mathbf{x}1}) + (\boldsymbol{\lambda}_i(\gamma) - \boldsymbol{\lambda}_i(\gamma_0))'\boldsymbol{\beta}_{\lambda 1} + e_{2i}\right]^2 I(\gamma_0 < q \leq \gamma \vee \gamma_0)$$

Define

$$T(\boldsymbol{\theta}_{1,0}, \boldsymbol{\theta}_1) = \left(\mathbf{g}_i(\gamma_0)'(\boldsymbol{\beta}_{\mathbf{x}10} - \boldsymbol{\beta}_{\mathbf{x}1}) + (\boldsymbol{\lambda}_i(\gamma) - \boldsymbol{\lambda}_i(\gamma_0))'\boldsymbol{\beta}_{\lambda 1} + e_{1i}\right)^2 - e_{1i}^2$$

$$T(\boldsymbol{\theta}_{2,0}, \boldsymbol{\theta}_2) = \left(\mathbf{g}_i(\gamma_0)'(\boldsymbol{\beta}_{\mathbf{x}20} - \boldsymbol{\beta}_{\mathbf{x}2}) + (\boldsymbol{\lambda}_i(\gamma) - \boldsymbol{\lambda}_i(\gamma_0))'\boldsymbol{\beta}_{\lambda 2} + e_{2i}\right)^2 - e_{2i}^2$$

$$T(\boldsymbol{\theta}_{1,0}, \boldsymbol{\theta}_2) = \left(\mathbf{g}_i(\gamma_0)'(\boldsymbol{\beta}_{\mathbf{x}10} - \boldsymbol{\beta}_{\mathbf{x}2}) + (\boldsymbol{\lambda}_i(\gamma) - \boldsymbol{\lambda}_i(\gamma_0))'\boldsymbol{\beta}_{\lambda 2} + e_{1i}\right)^2 - e_{1i}^2$$

$$T(\boldsymbol{\theta}_{2,0}, \boldsymbol{\theta}_1) = \left(\mathbf{g}_i(\gamma_0)'(\boldsymbol{\beta}_{\mathbf{x}20} - \boldsymbol{\beta}_{\mathbf{x}1}) + (\boldsymbol{\lambda}_i(\gamma) - \boldsymbol{\lambda}_i(\gamma_0))'\boldsymbol{\beta}_{\lambda 1} + e_{2i}\right)^2 - e_{2i}^2$$

Define the discrepancy function $d(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = ||\boldsymbol{\beta} - \boldsymbol{\beta}_0|| + |\gamma_0 - \gamma| + \sqrt{F_q(\gamma) - F_q(\gamma_0)}$ for $\boldsymbol{\theta}$ in the neighborhood of $\boldsymbol{\theta}_0$. Note that $d(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \to 0$ if and only if $||\boldsymbol{\beta} - \boldsymbol{\beta}_0|| \to 0$ and $|\gamma - \gamma_0| \to 0$.

The proof of this lemma relies on two sufficient conditions. First, we need to show that $M(\theta) - M(\theta_0) \leq -Cd^2(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$ for $\boldsymbol{\theta}$ in a neighborhood of $\boldsymbol{\theta}_0$.

Consider

$$M(\theta) - M(\theta_0) =$$

$$-E\left[T(\boldsymbol{\theta}_{1,0}, \boldsymbol{\theta}_1) I(q_i \leq \gamma \wedge \gamma_0)\right]$$

$$-E\left[T(\boldsymbol{\theta}_{2,0}, \boldsymbol{\theta}_2) I(q_i > \gamma \vee \gamma_0)\right]$$

$$-E\left[T(\boldsymbol{\theta}_{1,0}, \boldsymbol{\theta}_2) I(\gamma \wedge \gamma_0 < q_i \leq \gamma_0)\right]$$

$$-E\left[T(\boldsymbol{\theta}_{2,0}, \boldsymbol{\theta}_1) I(\gamma_0 < q_i \leq \gamma \vee \gamma_0)\right]$$

$$\leq$$

$$-(\boldsymbol{\beta}_{10} - \boldsymbol{\beta}_1)' E\left(\mathbf{g}_i(\gamma_0)\, \mathbf{g}_i(\gamma_0)' I(q_i \leq \gamma \wedge \gamma_0)\right)(\boldsymbol{\beta}_{10} - \boldsymbol{\beta}_1)$$

$$-(\boldsymbol{\beta}_{20} - \boldsymbol{\beta}_2)'\left(E\mathbf{g}_i(\gamma_0)\, \mathbf{g}_i(\gamma_0)' I(q_i > \gamma \vee \gamma_0)\right)(\boldsymbol{\beta}_{20} - \boldsymbol{\beta}_2)$$

$$-(\boldsymbol{\beta}_{10} - \boldsymbol{\beta}_2)' E(\mathbf{g}_i(\gamma_0)\, \mathbf{g}_i(\gamma_0)' I(\gamma \wedge \gamma_0 < q_i \leq \gamma_0))(\boldsymbol{\beta}_{20} - \boldsymbol{\beta}_1)$$

$$-(\boldsymbol{\beta}_{20} - \boldsymbol{\beta}_1)' E(\mathbf{g}_i(\gamma_0)\, \mathbf{g}_i(\gamma_0)' I(\gamma_0 < q_i \leq \gamma \vee \gamma_0))(\boldsymbol{\beta}_{10} - \boldsymbol{\beta}_2) - C_\lambda |\gamma_0 - \gamma|^2$$

$\leq -C\left(||\boldsymbol{\beta}_{10}-\boldsymbol{\beta}_1||^2+||\boldsymbol{\beta}_{20}-\boldsymbol{\beta}_2||^2+|\gamma_0-\gamma|^2+|F_q(\gamma)-F_q(\gamma_0)|\right)=-Cd^2(\boldsymbol{\theta},\boldsymbol{\theta}_0)$, where the the first inequality is due to the monotonicity of $\lambda_1(\cdot)$ and $\lambda_2(\cdot)$, Assumption 1, and Lemma 1.

Let us now proceed to the second condition of this lemma, which requires that

$$E^*\left(\sup_{d(\theta,\theta_0)}|\mathbb{G}_n\left(m(w|\theta)-m(w|\theta_0)\right)|\right)\leq C\epsilon,$$

where E^* is the outer expectation and $\epsilon>0$.

To show this, let us first define the class of functions

$$\mathcal{M}_\epsilon=\{m(\boldsymbol{\theta})-m(\boldsymbol{\theta}_0):d(\boldsymbol{\theta},\boldsymbol{\theta}_0)<\epsilon\}$$

Let us also write $m(\theta)-m(\theta_0)$ as follows

$m(\theta)-m(\theta_0)=$

$-T(\boldsymbol{\theta}_{1,0},\boldsymbol{\theta}_1)I(q_i\leq\gamma\wedge\gamma_0)-T(\boldsymbol{\theta}_{2,0},\boldsymbol{\theta}_2)I(q_i>\gamma\vee\gamma_0)$

$-T(\boldsymbol{\theta}_{1,0},\boldsymbol{\theta}_2)I(\gamma\wedge\gamma_0<q_i\leq\gamma_0)-T(\boldsymbol{\theta}_{2,0},\boldsymbol{\theta}_1)I(\gamma_0<q_i\leq\gamma\vee\gamma_0)$

$=A+B+C+D$, where $A,B,C,$ and D are defined accordingly.

Note that $\{T(\boldsymbol{\theta}_{1,0},\boldsymbol{\theta}_1):d(\boldsymbol{\theta},\boldsymbol{\theta}_0)<\widetilde{\delta}\}$ is a finite-dimensional vector space of real valued functions. Then Lemma 2.4 of Pakes and Pollard (1989) implies that $\{I(q\leq\gamma\wedge\gamma_0):d(\boldsymbol{\theta},\boldsymbol{\theta}_0)<\widetilde{\delta}\}$ is a VC subgraph class of functions. Then it follows that $\{A_n:d(\boldsymbol{\theta},\boldsymbol{\theta}_0)<\widetilde{\delta}\}$ is also a VC subgraph by Lemma 2.14 (ii) of Pakes and Pollard (1989). Similarly, we can show that $\{B_n:d(\boldsymbol{\theta},\boldsymbol{\theta}_0)<\widetilde{\delta}\}$, $\{C_n:d(\boldsymbol{\theta},\boldsymbol{\theta}_0)<\widetilde{\delta}\}$, $\{D_n:d(\boldsymbol{\theta},\boldsymbol{\theta}_0)<\widetilde{\delta}\}$ are VC-classes.

Given these results we use Theorem 2.14.2 of Van der Vaart and Wellner (1996) to show that

$$E^*\left(\sup_{d(\theta,\theta_0)}|\mathbb{G}_n\left(m(w|\theta)-m(w|\theta_0)\right)|\right)\leq C\sqrt{PF^2},$$

where F is the envelope function of the class of functions defined by $\{m(w|\theta)-m(w|\theta_0):d(\boldsymbol{\theta},\boldsymbol{\theta}_0)<\widetilde{\delta}\}$. Given the functional form of $m(w|\theta)-m(w|\theta_0)$, $\sqrt{\mathbb{P}F^2}\leq C\widetilde{\delta}$ follows by Assumption 1.4 and 1.5.

Corollary 3.2.6 of van der Vaart and Wellner (1996) implies that $\phi(\widetilde{\delta})=\widetilde{\delta}$ and thus $\phi(\widetilde{\delta})/\widetilde{\delta}^{\alpha}=\delta^{1-\alpha}$ is decreasing for any $\alpha\in(1,2)$, hence Theorem 14.4 in Kosorok (2008) is satisfied. Since $r_n^2\phi(r_n^{-1})=r_n$ and hence $\sqrt{n}d(\widehat{\boldsymbol{\theta}},\boldsymbol{\theta}_0)=O_p(1)$. By the definition of d,we get that $||\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0||=O_p(n^{-1/2})$ and $|\widehat{\gamma}-\gamma_0|+|F\left(\widehat{\gamma}\right)-F\left(\gamma_0\right)|=O_p(n^{-1/2})+O_p(n^{-1})=O_p(n^{-1})$.

Therefore for any $\varepsilon > 0$, we can find M_ε such that $P(n(F(\widehat{\gamma}) - F(\gamma_0)) > M_\varepsilon)$=$P(n(F(\gamma_0 + a_n(\widehat{\gamma} - \gamma_0)/a_n) - F(\gamma_0)) > M_\varepsilon)$<$\varepsilon$, which implies that there exists a_n such that $P(a_n|\widehat{\gamma} - \gamma_0| > \overline{M}_\varepsilon) \leq \varepsilon$ for $n \geq \overline{n}$. This completes the proof.

■

LEMMA 4. $\underset{\overline{\nu}/a_n \leq |\gamma-\gamma_0| \leq B}{\arg\min} S_n^R(\gamma) - S_n^R(\gamma_0) = \underset{\overline{\nu}/a_n \leq |\gamma-\gamma_0| \leq B}{\arg\min} S_n^U(\gamma) - S_n^U(\gamma_0) + o_p(1)$

Proof of Lemma 4.

Recall that $S_n^R(\gamma) = S_n^U(\gamma) + (\boldsymbol{\vartheta} - \mathbf{R}'\widehat{\boldsymbol{\theta}})'(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma)'\widehat{\mathbf{X}}^*(\gamma))^{-1}\mathbf{R})^{-1}(\boldsymbol{\vartheta} - \mathbf{R}'\widehat{\boldsymbol{\theta}})$. Then

$$\begin{aligned} S_n^R(\gamma) - S_n^R(\gamma_0) &= [S_n^U(\gamma) - S_n^U(\gamma_0)] \\ &+[(\boldsymbol{\vartheta} - \mathbf{R}'\widehat{\boldsymbol{\theta}})'(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma)'\widehat{\mathbf{X}}^*(\gamma))^{-1}\mathbf{R})^{-1}(\boldsymbol{\vartheta} - \mathbf{R}'\widehat{\boldsymbol{\theta}}) \\ &-(\boldsymbol{\vartheta} - \mathbf{R}'\widehat{\boldsymbol{\theta}}_0)'(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\mathbf{R})^{-1}(\boldsymbol{\vartheta} - \mathbf{R}'\widehat{\boldsymbol{\theta}}_0)] \end{aligned}$$

The key is to show that the second term is $o_p(1)$.

Define $\Delta(\gamma) = I(q \leq \gamma) - I(q \leq \gamma_0)$ and $\widetilde{I} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

Let us consider the case of $\gamma > \gamma_0$,

$\frac{1}{n}||\widehat{\mathbf{X}}^*(\gamma)'\widehat{\mathbf{X}}^*(\gamma) - (\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0)|| =$

$\frac{1}{n}||(\sum_i \mathbf{g}_i(\gamma)\mathbf{g}_i(\gamma)'\Delta(\gamma) - \sum_i \mathbf{g}_i(\gamma)\widehat{\mathbf{r}}'\Delta(\gamma) - \sum_i \mathbf{g}_i(\gamma)\widehat{\mathbf{r}}'\Delta(\gamma) + \sum_i \widehat{\mathbf{r}}\widehat{\mathbf{r}}'\Delta(\gamma)) \otimes \widetilde{I}|| \leq$

$\frac{1}{n}||(\sum_i \mathbf{g}_i(\gamma)\mathbf{g}_i(\gamma)'\Delta(\gamma) \otimes \widetilde{I}|| + 2\frac{1}{n}||(\sum_i \mathbf{g}_i(\gamma)\widehat{\mathbf{r}}'\Delta(\gamma) \otimes \widetilde{I}|| + ||\sum_i \widehat{\mathbf{r}}\widehat{\mathbf{r}}'\Delta(\gamma)) \otimes \widetilde{I}|| \leq$

$\sqrt{2}\frac{1}{n}(tr(\sum_i \mathbf{g}_i(\gamma_0 + \epsilon)\mathbf{g}_i(\gamma_0 + \epsilon)'\Delta(\gamma))^2)^{1/2}+$

$\sqrt{2}\frac{2}{n}(tr(\sum_i \mathbf{g}_i(\gamma_0 + \epsilon)\widehat{\mathbf{r}}'\Delta(\gamma))^2)^{1/2}+$

$\sqrt{2}\frac{1}{n}(tr(\sum_i \widehat{\mathbf{r}}\widehat{\mathbf{r}}'\Delta(\gamma))^2)^{1/2} = o_p(1).$

So $\frac{1}{n}\widehat{\mathbf{X}}^*(\gamma)'\widehat{\mathbf{X}}^*(\gamma) = \frac{1}{n}\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0) + o_p(1)$. Then using Lemma A.2 of Perron and Qu (2006) we obtain

$$(\frac{1}{n}\widehat{\mathbf{X}}^*(\gamma)'\widehat{\mathbf{X}}^*(\gamma))^{-1} = (\frac{1}{n}\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1} + o_p(1). \quad \text{(A.16)}$$

and

$$\frac{1}{n}(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma)'\widehat{\mathbf{X}}^*(\gamma))^{-1}\mathbf{R})^{-1} = \frac{1}{n}(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\mathbf{R})^{-1} + o_p(1). \quad \text{(A.17)}$$

Note that $S_n^U(\gamma) - S_n^U(\gamma_0) = o_p(1)$. Then,

$S_n^R(\gamma) - S_n^R(\gamma_0) =$

$[S_n^U(\gamma) - S_n^U(\gamma_0)]+$

$[(\boldsymbol{\vartheta} - \mathbf{R}'\widehat{\boldsymbol{\theta}})'(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma)'\widehat{\mathbf{X}}^*(\gamma))^{-1}\mathbf{R})^{-1}(\boldsymbol{\vartheta} - \mathbf{R}'\widehat{\boldsymbol{\theta}})-$

$(\boldsymbol{\vartheta} - \mathbf{R}'\widehat{\boldsymbol{\theta}}_0)'(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\mathbf{R})^{-1}(\boldsymbol{\vartheta} - \mathbf{R}'\widehat{\boldsymbol{\theta}}_0)] =$

$[(\boldsymbol{\vartheta} - \mathbf{R}'\widehat{\boldsymbol{\theta}})'(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\mathbf{R})^{-1}(\boldsymbol{\vartheta} - \mathbf{R}'\widehat{\boldsymbol{\theta}})-$

$(\boldsymbol{\vartheta} - \mathbf{R}'\widehat{\boldsymbol{\theta}}_0)'(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\mathbf{R})^{-1}(\boldsymbol{\vartheta} - \mathbf{R}'\widehat{\boldsymbol{\theta}}_0)] + o_p(1) =$

$n^{1/2}(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}})'\mathbf{R}(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\mathbf{R})^{-1}\mathbf{R}'n^{1/2}(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}})-$

$n^{1/2}(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}}_0)'\mathbf{R}(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\mathbf{R})^{-1}\mathbf{R}'n^{1/2}(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}}_0) + o_p(1)$

$= o_p(1)$ since $n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = n^{1/2}(\widehat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_0) + o_p(1)$.

This completes the proof.

■

LEMMA 5. On $[-\overline{\upsilon}, \overline{\upsilon}]$,

$$Q_n(\upsilon) = S_n^U(\gamma_0) - S_n^U(\gamma_0 + \upsilon/a_n) \Longrightarrow \begin{cases} -\mu_1|\upsilon| + 2\zeta_1^{1/2}W_1(\upsilon), & \text{uniformly on } \upsilon \in [-\overline{\upsilon}, 0] \\ -\mu_2|\upsilon| + 2\zeta_2^{1/2}W_2(\upsilon), & \text{uniformly on } \upsilon \in [0, \overline{\upsilon}] \end{cases},$$

where $\mu_i = \mathbf{c}'\mathbf{D}_i\mathbf{c}f$ and $\zeta_i = \mathbf{c}'\boldsymbol{\Omega}_i\mathbf{c}f$, for $i = 1, 2$.

Proof of Lemma 5.

Proof: $S_n^{*U}(\gamma) = n^{-1+2\alpha}(n^{-\boldsymbol{\alpha}}\mathbf{c}'\mathbf{G}_0(\gamma_0)' + \widehat{\mathbf{e}}')\mathbf{P}^*(\gamma)(\mathbf{G}_0(\gamma_0)n^{-\boldsymbol{\alpha}}\mathbf{c} + \widehat{\mathbf{e}})$

Our proof strategy follows Caner and Hansen (2004). Let us reparameterize all functions of γ as functions of υ. For example, $\widehat{\mathbf{X}}_\upsilon = \widehat{\mathbf{X}}_{\gamma_0+\upsilon/a_n}$, $\mathbf{P}^*(\upsilon) = \mathbf{P}^*(\gamma_0 + \upsilon/a_n)$ and for $\Delta_i(\gamma) = I(q_i \leq \gamma) - I(q_i \leq \gamma_0)$ we have $\Delta_i(\upsilon) = \Delta_i(\gamma_0 + \upsilon/a_n)$. Then,

$$\begin{aligned} Q_n(\upsilon) &= S_n^U(\gamma_0) - S_n^U(\gamma_0 + \upsilon/a_n) \\ &= (n^{-\alpha}\mathbf{c}'\mathbf{G}(\gamma_0)' + \widehat{\mathbf{e}}')\mathbf{P}^*(\upsilon)(\mathbf{G}(\gamma_0)\mathbf{c}n^{-\alpha} + \widehat{\mathbf{e}}) - (n^{-\alpha}\mathbf{c}'\mathbf{G}(\gamma_0)' + \widehat{\mathbf{e}}')\mathbf{P}^*(\gamma_0)(\mathbf{G}(\gamma_0)\mathbf{c}n^{-\alpha} + \widehat{\mathbf{e}}) \\ &= n^{-2a}c'\mathbf{G}(\gamma_0)'(\mathbf{P}^*(\upsilon) - \mathbf{P}^*(\gamma_0))\mathbf{G}(\gamma_0)c + 2n^{-a}c'\mathbf{G}(\gamma_0)'(\mathbf{P}^*(\upsilon) - \mathbf{P}^*(\gamma_0))\widehat{\mathbf{e}} + \widehat{\mathbf{e}}'(\mathbf{P}^*(\upsilon) - \mathbf{P}^*(\gamma_0))\widehat{\mathbf{e}} \end{aligned}$$

We proceed by studying the behavior of each term: (i) $n^{-2a}\mathbf{c}'\mathbf{G}(\gamma_0)'(\mathbf{P}^*(\upsilon)-\mathbf{P}^*(\gamma_0))\mathbf{G}(\gamma_0)\mathbf{c}$; (ii) $2n^{-a}c'\mathbf{G}(\gamma_0)'(\mathbf{P}^*(\upsilon)-\mathbf{P}^*(\gamma_0))\widehat{\mathbf{e}}$; (iii) $\widehat{\mathbf{e}}'(\mathbf{P}^*(\upsilon)-\mathbf{P}^*(\gamma_0))\widehat{\mathbf{e}}$

(i)

Define $\widehat{\mathbf{X}}_\gamma(\gamma,\gamma_0)=(\widehat{\mathbf{X}}_\gamma,\widehat{\mathbf{\Lambda}}_{1,\gamma}(\gamma_0),\widehat{\mathbf{\Lambda}}_{2,\gamma}(\gamma_0))$ and $\widehat{\mathbf{X}}_\gamma(\gamma_0)=\widehat{\mathbf{X}}_\gamma(\gamma_0,\gamma_0)$. Furthermore, recall that

$\frac{1}{n}\widehat{\mathbf{X}}_\gamma(\gamma)'\widehat{\mathbf{X}}_\gamma(\gamma)=\frac{1}{n}\widehat{\mathbf{X}}_\gamma(\gamma,\gamma_0)'\widehat{\mathbf{X}}_\gamma(\gamma,\gamma_0)+o_p(1)$

$n^{-2\alpha}|\frac{1}{n}\widehat{\mathbf{X}}_\upsilon(\upsilon)'\widehat{\mathbf{X}}_\upsilon(\upsilon)-\frac{1}{n}\widehat{\mathbf{X}}_0(\gamma_0)'\widehat{\mathbf{X}}_0(\gamma_0)|$

$$\leq n^{-2\alpha}\sum_{i=1}^n|\mathbf{g}_i(\upsilon)|^2\Delta_i(\upsilon)+2n^{-2\alpha}\left|\sum_{i=1}^n\mathbf{g}_i(\upsilon)\widehat{\mathbf{e}}_i'\Delta_i(\upsilon)\right|+n^{-2\alpha}\left|\sum_{i=1}^n\widehat{\mathbf{e}}_i\widehat{\mathbf{e}}_i'\Delta_i(\upsilon)\right|\Longrightarrow$$

$$\begin{cases}|\mathbf{D}_1f|\,|\upsilon|, & \upsilon\in[-\overline{\upsilon},0]\\ |\mathbf{D}_2f|\,|\upsilon|, & \upsilon\in[0,\overline{\upsilon}]\end{cases}$$

Therefore, $n^{-2\alpha}\sup_{|\upsilon|\leq\overline{\upsilon}}|\widehat{\mathbf{X}}_\upsilon(\upsilon)'\widehat{\mathbf{X}}_\upsilon(\upsilon)-\widehat{\mathbf{X}}_0(\gamma_0)'\widehat{\mathbf{X}}_0(\gamma_0)|=O_p(1)$

We also know from Lemma 2 that

$$\frac{1}{n}\widehat{\mathbf{X}}_\upsilon(\upsilon)'\widehat{\mathbf{X}}_\upsilon(\upsilon)\Longrightarrow\mathbf{M}(\gamma_0)\tag{A.18}$$

Our analysis below will be restricted to the region $[\gamma_0+\overline{\upsilon}/a_n\leq\gamma\leq\gamma_0+B]$ for some constant $B>0$, which follows from Lemma 1. Note that this restriction implies that $\widehat{\mathbf{X}}_\gamma'\mathbf{G}_{\mathbf{x},0}=\widehat{\mathbf{X}}_0'\mathbf{G}_{\mathbf{x},0},\widehat{\mathbf{X}}_\gamma'\widehat{\mathbf{X}}_0=\widehat{\mathbf{X}}_0'\widehat{\mathbf{X}}_0$,

The analysis for the case $[\gamma_0-\overline{\upsilon}/a_n\geq\gamma\geq\gamma_0-B]$ is similar.

Then, by (A44), (A51), (A52), Lemma 2, (A40), 17, and Lemma A10 of Hansen (2000), we get

$n^{-2a}c'\mathbf{G}(\gamma_0)'(\mathbf{P}^*(\upsilon)-\mathbf{P}^*(\gamma_0))\mathbf{G}(\gamma_0)c=n^{-2a}c'\mathbf{G}(\gamma_0)'(\mathbf{P}_\upsilon(\upsilon)-\mathbf{P}_0(\gamma_0))\mathbf{G}(\gamma_0)c$

From equation A.44 of Caner and Hansen (2004) we can get

$n^{-2a}\mathbf{c}'\mathbf{G}(\gamma_0)'(\mathbf{P}^*(\upsilon)-\mathbf{P}^*(\gamma_0))\mathbf{G}(\gamma_0)\mathbf{c}$

$=n^{-2a}\mathbf{c}'\mathbf{G}(\gamma_0)'(\mathbf{P}_\upsilon(\upsilon)-\mathbf{P}_0(\gamma_0))\mathbf{G}(\gamma_0)\mathbf{c}$

$=n^{-2a}\mathbf{c}'\left(\widehat{\mathbf{X}}_\upsilon(\upsilon)'\widehat{\mathbf{X}}_\upsilon(\upsilon)-\widehat{\mathbf{X}}_0(\gamma_0)'\widehat{\mathbf{X}}_0(\gamma_0)\right)\mathbf{c}$

$-\mathbf{c}'\left(\widehat{\mathbf{X}}_\upsilon(\upsilon)'\widehat{\mathbf{X}}_\upsilon(\upsilon)-\widehat{\mathbf{X}}_0(\gamma_0)'\widehat{\mathbf{X}}_0(\gamma_0)\right)\left(\mathbf{I}-(\widehat{\mathbf{X}}_\upsilon(\upsilon)'\widehat{\mathbf{X}}_\upsilon(\upsilon))^{-1}\widehat{\mathbf{X}}_0(\gamma_0)'\widehat{\mathbf{X}}_0(\gamma_0)\right)\mathbf{c}$

$-\mathbf{c}\left(\mathbf{I}-\mathbf{G}_0(\gamma_0)'\widehat{\mathbf{X}}_0(\gamma_0)(\widehat{\mathbf{X}}_0(\gamma_0)'\widehat{\mathbf{X}}_0(\gamma_0))^{-1}\right)\left(\widehat{\mathbf{X}}_\upsilon(\upsilon)'\widehat{\mathbf{X}}_\upsilon(\upsilon)-\widehat{\mathbf{X}}_0(\gamma_0)'\widehat{\mathbf{X}}_0(\gamma_0)\right)(\widehat{\mathbf{X}}_\upsilon(\upsilon)'\widehat{\mathbf{X}}_\upsilon(\upsilon))^{-1}\widehat{\mathbf{X}}_0(\gamma_0)'\mathbf{G}_0(\gamma_0)\mathbf{c}+o_p(1)$

$$=n^{-2\alpha}\sum_{i=1}^{n}|\mathbf{g}_i(\upsilon)|^2\Delta_i(\upsilon)+o_p(1)\Longrightarrow\mu_2|\upsilon|.$$

This establishes that uniformly on $[\gamma_0+\overline{\upsilon}/a_n\le\gamma\le\gamma_0+B]$,

$$n^{-2a}c'\mathbf{G}(\gamma_0)'(\mathbf{P}^*(\gamma_0)-\mathbf{P}^*(\upsilon))\mathbf{G}(\gamma_0)c\Longrightarrow\mu_2|\upsilon| \tag{A.19}$$

(ii) From equation A.45 of Caner and Hansen (2004) we can get

$n^{-a}c'\mathbf{G}_0(\gamma_0)'(\mathbf{P}^*(\gamma_0)-\mathbf{P}^*(\upsilon))\widehat{\mathbf{e}}$

$=n^{-a}c'\mathbf{G}(\gamma_0)'(\mathbf{P}_0(\gamma_0)-\mathbf{P}_\upsilon(\upsilon))\widehat{\mathbf{e}}$

$=$

$\left[\mathbf{G}_0(\gamma_0)'\widehat{\mathbf{X}}_0(\gamma_0)(\widehat{\mathbf{X}}_0(\gamma_0)'\widehat{\mathbf{X}}_0(\gamma_0))^{-1}\right]\left[n^{-2\alpha}(\widehat{\mathbf{X}}_\upsilon(\upsilon)'\widehat{\mathbf{X}}_\upsilon(\upsilon)-\widehat{\mathbf{X}}_0(\gamma_0)'\widehat{\mathbf{X}}_0(\gamma_0))\right]\left[n^{\alpha}(\widehat{\mathbf{X}}_\upsilon(\upsilon)'\widehat{\mathbf{X}}_\upsilon(\upsilon))^{-1}\widehat{\mathbf{X}}_0(\gamma_0)'\widehat{\mathbf{e}}\right]$

$-\left[\mathbf{G}_0(\gamma_0)'\widehat{\mathbf{X}}_0(\gamma_0)(\widehat{\mathbf{X}}_\upsilon(\upsilon)'\widehat{\mathbf{X}}_\upsilon(\upsilon))^{-1}\right]\left[n^{-a}(\widehat{\mathbf{X}}_\upsilon(\upsilon)'-\widehat{\mathbf{X}}_0(\gamma_0)')\widehat{\mathbf{e}})\right]$

Note that by Lemma 2 and (A.18) we can get uniformly in $\upsilon\in[0,\overline{\upsilon}]$,

$$n^{\alpha}(\widehat{\mathbf{X}}_\upsilon(\upsilon)'\widehat{\mathbf{X}}_\upsilon(\upsilon))^{-1}\widehat{\mathbf{X}}_0(\gamma_0)'\widehat{\mathbf{e}}=(\frac{1}{n}\widehat{\mathbf{X}}_\upsilon(\upsilon)'\widehat{\mathbf{X}}_\upsilon(\upsilon))^{-1}(\frac{1}{n^{1-\alpha}}\widehat{\mathbf{X}}_0(\gamma_0)'\widehat{\mathbf{e}})=o_p(1) \tag{A.20}$$

and

$$\begin{aligned} n^{-a}(\widehat{\mathbf{X}}_\upsilon(\upsilon)'-\widehat{\mathbf{X}}_0(\gamma_0)')\widehat{\mathbf{e}}) &= n^{-a}\sum_{i=1}^{n}\widehat{\mathbf{g}}_i(\upsilon)\widehat{e}_i\Delta_i(\upsilon)\\ &= n^{-a}\sum_{i=1}^{n}\widehat{\mathbf{g}}_i(\upsilon)\widehat{\mathbf{r}}_i\boldsymbol{\beta}\Delta_i(\upsilon)+n^{-a}\sum_{i=1}^{n}\mathbf{g}_i(\upsilon)e_i\boldsymbol{\beta}\Delta_i(\upsilon)-n^{-a}\sum_{i=1}^{n}\widehat{\mathbf{r}}_ie_i\Delta_i(\upsilon)\\ &\Longrightarrow n^{-a}\sum_{i=1}^{n}\mathbf{g}_i(\upsilon)e_i\Delta_i(\upsilon)+o_p(1)=B_1(\upsilon). \end{aligned} \tag{A.21}$$

Then, it follows that

$$n^{-a}c'\mathbf{G}_0(\gamma_0)'(\mathbf{P}^*(\gamma_0)-\mathbf{P}^*(\upsilon))\widehat{\mathbf{e}}\Longrightarrow B_1(\upsilon).$$

where $B_1(\upsilon)$ a vector Brownian motion with covariance matrix $\mathbf{\Omega}_1 f$ and hence

$$n^{-a}c'\mathbf{G}(\gamma_0)'(\mathbf{P}^*(\gamma_0)-\mathbf{P}^*(\upsilon))\widehat{\mathbf{e}}\Longrightarrow\zeta_1^{1/2}W_1(\upsilon) \tag{A.22}$$

(iii)

$\widehat{\mathbf{e}}'(\mathbf{P}^*(\gamma_0)-\mathbf{P}^*(\upsilon))\widehat{\mathbf{e}}=$

$\left[n^{\alpha}\widehat{\mathbf{e}}'\widehat{\mathbf{X}}_0(\gamma_0)(\widehat{\mathbf{X}}_0(\gamma_0)'\widehat{\mathbf{X}}_0(\gamma_0))^{-1}\right]\left[n^{-2\alpha}(\widehat{\mathbf{X}}_\upsilon(\upsilon)'\widehat{\mathbf{X}}_\upsilon(\upsilon)-\widehat{\mathbf{X}}_0(\gamma_0)'\widehat{\mathbf{X}}_0(\gamma_0))\right]\left[n^{\alpha}(\widehat{\mathbf{X}}_\upsilon(\upsilon)'\widehat{\mathbf{X}}_\upsilon(\upsilon))^{-1}\widehat{\mathbf{X}}_0(\gamma_0)'\widehat{\mathbf{e}}\right]$ $= o_p(1)$. Hence,

$$\widehat{\mathbf{e}}'(\mathbf{P}^*(\gamma_0)-\mathbf{P}^*(\upsilon))\widehat{\mathbf{e}} \Longrightarrow 0. \tag{A.23}$$

Using equation (A.10) and (A.19)-(A.23) we get

$$\begin{aligned}
Q_n(\upsilon) &= S_n(\gamma_0)-S_n(\gamma_0+\upsilon/a_n)\\
&= (n^{-\alpha}\mathbf{c}'\mathbf{G}(\gamma_0)'+\widehat{\mathbf{e}}')\mathbf{P}^*(\upsilon)(\mathbf{G}(\gamma_0)\mathbf{c}n^{-\alpha}+\widehat{\mathbf{e}})-(n^{-\alpha}\mathbf{c}'\mathbf{G}(\gamma_0)'+\widehat{\mathbf{e}}')\mathbf{P}^*(\gamma_0)(\mathbf{G}(\gamma_0)\mathbf{c}n^{-\alpha}+\widehat{\mathbf{r}})\\
&= n^{-2a}c'\mathbf{G}(\gamma_0)'(\mathbf{P}^*(\gamma_0)-\mathbf{P}^*(\upsilon))\mathbf{G}(\gamma_0)c+2n^{-a}c'\mathbf{G}(\gamma_0)'(\mathbf{P}^*(\gamma_0)-\mathbf{P}^*(\upsilon))\widehat{\mathbf{e}}+\widehat{\mathbf{e}}'(\mathbf{P}^*(\gamma_0)-\mathbf{P}^*(\upsilon))\widehat{\mathbf{e}}\\
&\Longrightarrow -\mu_1|\upsilon|+2\zeta_1^{1/2}W_1(\upsilon),\ \text{uniformly on } \upsilon\in[-\bar{\varepsilon},0]
\end{aligned}$$

Similarly, we can show that uniformly on $\upsilon\in[0,\bar{\varepsilon}]$, $Q_n(\upsilon)\Longrightarrow -\mu_2|\upsilon|+2\zeta_2^{1/2}W_2(\upsilon)$, where W_2 is a Wiener process on $[0,\infty)$ independent of W_1.

■

Proof of Theorem 4.1

By Lemma 3, $a_n(\widehat{\gamma}-\gamma_0)=\arg\max_\upsilon Q_n(\upsilon)=O_p(1)$ and by Lemma 4,

$$Q_n^R(\upsilon)\Longrightarrow\begin{cases}-\mu_1|\upsilon|+2\zeta_1^{1/2}W_1(\upsilon), & \text{uniformly on } \upsilon\in[-\bar{\varepsilon},0]\\ -\mu_2|\upsilon|+2\zeta_2^{1/2}W_2(\upsilon), & \text{uniformly on } \upsilon\in[0,\bar{\varepsilon}]\end{cases}$$

Then, by Theorem 2.7 of Kim and Pollard (1990) and Theorem 1 of Hansen (2000) we can get $n^{1-2\alpha}(\widehat{\gamma}-\gamma_0)\xrightarrow{d}\underset{-\infty<\upsilon<\infty}{\arg\max}\,Q_n(\upsilon)$.

Set $\omega=\zeta_1/\mu_1^2$ and recall that $W_i(b^2\upsilon)=bW_i(\upsilon)$. By making the change of variables $\upsilon=(\zeta_1/\mu_1^2)s$ we can rewrite the asymptotic distribution as follows. For $s\in[-\bar{\varepsilon},0]$,

$$\begin{aligned}
&\underset{-\infty<\upsilon<\infty}{\arg\max}\,Q_n(\upsilon)\\
&=\begin{cases}\underset{-\infty<\upsilon<\infty}{\arg\max}\left(-\frac{\zeta_1}{\mu_1^2}\mu_1|s|+2\zeta_1^{1/2}W_1((\zeta_1/\mu_1^2)s)\right)=\omega\underset{-\infty<\upsilon<\infty}{\arg\max}\left(-\frac{1}{2}|s|+W_1(s)\right), & \text{if } s\in[-\bar{\varepsilon},0]\\ \underset{-\infty<\upsilon<\infty}{\arg\max}\left(-\frac{\zeta_1}{\mu_1^2}\mu_2|s|+2\zeta_2^{1/2}W_1((\zeta_1/\mu_1^2)s)\right)=\omega\underset{-\infty<\upsilon<\infty}{\arg\max}\left(-\frac{1}{2}\xi|s|+\sqrt{\varphi}W_2(s)\right), & \text{if } s\in[0,\bar{\varepsilon}]\end{cases}
\end{aligned}$$

where $\xi = \mu_2/\mu_1$ and $\varphi = \zeta_2/\zeta_1$. Hence, $n^{1-2\alpha}(\widehat{\gamma} - \gamma_0) \xrightarrow{d} \underset{-\infty<\upsilon<\infty}{\arg\max}\, \omega T(s)$, where

$$T(s) = \begin{cases} -\frac{1}{2}|s| + W_1(-s), & \text{if } s \in [-\overline{\varepsilon}, 0] \\ -\frac{1}{2}\xi|s| + \sqrt{\varphi} W_2(s), & \text{if } s \in [0, \overline{\varepsilon}] \end{cases}$$

■

Proof of Theorem 4.2

From Theorem 2 of Hansen (2000) we have $\widehat{\sigma}^2 LR_n(\gamma_0) - Q_n(\upsilon) \xrightarrow{p} 0$. Then,

$$LR_n(\gamma) = \frac{Q_n(\overline{\upsilon})}{\widehat{\sigma}^2} + o_p(1) = \frac{1}{\widehat{\sigma}^2} \sup_{-\infty<\upsilon<\infty} Q_n(\upsilon) + o_p(1) \xrightarrow{d} \frac{1}{\sigma^2} \sup_{-\infty<\upsilon<\infty} Q(\upsilon)$$

$$= \frac{1}{\sigma^2} \sup_{-\infty<\upsilon<\infty} \left(\left(-\mu_1|\upsilon| + 2\zeta_1^{1/2} W_1(\upsilon) \right) I(q < \gamma_0) + \left(-\mu_2|\upsilon| + 2\zeta_2^{1/2} W_2(\upsilon) \right) I(q > \gamma_0) \right)$$

By the change of variables $\upsilon = (\zeta_1/\mu_1^2)s$ the limiting distribution takes the form

$$\frac{1}{\sigma^2} \sup_{-\infty<\upsilon<\infty} Q(\upsilon) =$$

$$\frac{1}{\sigma^2} \sup_{-\infty<\upsilon<\infty} \left(\left(-\mu_1|\tfrac{\zeta_1}{\mu_1^2}s| + 2\zeta_1^{1/2} W_1(\tfrac{\zeta_1}{\mu_1^2}s) \right) I(q < \gamma_0) + \left(-\mu_2|\tfrac{\zeta_1}{\mu_1^2}s| + 2\zeta_2^{1/2} W_2(\tfrac{\zeta_1}{\mu_1^2}s) \right) I(q > \gamma_0) \right)$$

$$= \frac{\zeta_1}{\sigma^2\mu_1} \sup_{-\infty<\upsilon<\infty} \left((-|s| + 2W_1(s)) I(q < \gamma_0) + \left(-\xi|s| + 2\sqrt{\varphi} W_2(s) \right) I(q > \gamma_0) \right)$$

$$= \eta^2 \psi, \text{ where } \eta^2 = \frac{\zeta_1}{\sigma^2\mu_1}.$$

Note that $\psi = 2\max(\psi_1, \psi_2)$, where $\psi_1 = \sup_{s\le 0} (-|s| + 2W_1(s))$ and $\psi_2 = \sup_{s>0} \left(-\xi|s| + 2\sqrt{\varphi} W_2(s) \right)$. Note that while ψ_1 and ψ_2 are independent, they are not identical. ψ_1 is an exponential distribution while ψ_2 is a generalized distribution that depends on the parameters ξ and φ.

$$P(\psi \le x) = P(2\max(\psi_1, \psi_2) \le x) = P(\psi_1 \le x/2) P(\psi_2 \le x/2) = (1 - e^{-x/2})(1 - e^{-\xi x/2})^{\sqrt{\varphi}}.$$

■

Lemma 6 We prove the consistency of $\widehat{\boldsymbol{\beta}}_1$. The consistency of $\widehat{\boldsymbol{\beta}}_2$ can be shown similarly.

Proof of Lemma 6.

$$\widehat{\boldsymbol{\beta}}_1 = \left(\widehat{\mathbf{X}}_1' \widehat{\mathbf{Z}}_1 \widehat{\mathbf{W}}_1 \widehat{\mathbf{Z}}_1' \widehat{\mathbf{X}}_1 \right)^{-1} \widehat{\mathbf{X}}_1' \widehat{\mathbf{Z}}_1 \widehat{\mathbf{W}}_1 \widehat{\mathbf{Z}}_1' (\mathbf{X_1}\boldsymbol{\beta_{10}} + \mathbf{X_2}\boldsymbol{\beta_{20}} + \mathbf{e}) =$$

$\left((\frac{1}{n}\widehat{\mathbf{X}}_1'\widehat{\mathbf{Z}}_1)\widehat{\mathbf{W}}_1(\frac{1}{n}\widehat{\mathbf{Z}}_1'\widehat{\mathbf{X}}_1)\right)^{-1}(\frac{1}{n}\widehat{\mathbf{X}}_1'\widehat{\mathbf{Z}}_1)\widehat{\mathbf{W}}_1(\frac{1}{n}\widehat{\mathbf{Z}}_1'\mathbf{X_1})\boldsymbol{\beta}_{\mathbf{10}}+$

$\left((\frac{1}{n}\widehat{\mathbf{X}}_1'\widehat{\mathbf{Z}}_1)\widehat{\mathbf{W}}_1(\frac{1}{n}\widehat{\mathbf{Z}}_1'\widehat{\mathbf{X}}_1)\right)^{-1}(\frac{1}{n}\widehat{\mathbf{X}}_1'\widehat{\mathbf{Z}}_1)\widehat{\mathbf{W}}_1(\frac{1}{n}\widehat{\mathbf{Z}}_1'\mathbf{X_2})\boldsymbol{\beta}_{\mathbf{20}}+$

$\left((\frac{1}{n}\widehat{\mathbf{X}}_1'\widehat{\mathbf{Z}}_1)\widehat{\mathbf{W}}_1(\frac{1}{n}\widehat{\mathbf{Z}}_1'\widehat{\mathbf{X}}_1)\right)^{-1}(\frac{1}{n}\widehat{\mathbf{X}}_1'\widehat{\mathbf{Z}}_1)\widehat{\mathbf{W}}_1(\frac{1}{n}\widehat{\mathbf{Z}}_1'\mathbf{e})$

Given $\widehat{\mathbf{W}}_1 \longrightarrow \mathbf{W}_1 > 0$, the first term goes to zero by a Glivenko-Cantelli theorem and the second term goes to zero since $P(\widehat{\gamma} < \gamma_0) \to 0$. Similarly we can show that

$\left((\frac{1}{n}\widehat{\mathbf{X}}_1'\widehat{\mathbf{Z}}_1)\widehat{\mathbf{W}}_1(\frac{1}{n}\widehat{\mathbf{Z}}_1'\widehat{\mathbf{X}}_1)\right)^{-1}(\frac{1}{n}\widehat{\mathbf{X}}_1'\widehat{\mathbf{Z}}_1)\widehat{\mathbf{W}}_1(\frac{1}{n}\widehat{\mathbf{Z}}_1'\mathbf{X}_2) \xrightarrow{p} 0$ and

$\left((\frac{1}{n}\widehat{\mathbf{X}}_1'\widehat{\mathbf{Z}}_1)\widehat{\mathbf{W}}_1(\frac{1}{n}\widehat{\mathbf{Z}}_1'\widehat{\mathbf{X}}_1)\right)^{-1}(\frac{1}{n}\widehat{\mathbf{X}}_1'\widehat{\mathbf{Z}}_1)\widehat{\mathbf{W}}_1(\frac{1}{n}\widehat{\mathbf{Z}}_1'\mathbf{e}) \xrightarrow{p} 0.$

The proof is completed by showing that

$$
\begin{aligned}
&||(\tfrac{1}{n}\mathbf{X}_1(\widehat{\gamma})'\mathbf{Z}_1 I(q \le \widehat{\gamma}))\widehat{\mathbf{W}}_1(\widehat{\gamma})(\tfrac{1}{n}\mathbf{Z}_1' I(q \le \widehat{\gamma})\mathbf{X}_1(\widehat{\gamma}))- \\
&\qquad E(\mathbf{z}_{1i}\mathbf{x}_{1i}(\gamma_0)' I(q_i \le \gamma_0)\mathbf{W}_1(\gamma_0)E(\mathbf{x}_{1i}(\gamma_0)'\mathbf{z}_{1i} I(q_i \le \gamma_0)|| = \\
&||(\tfrac{1}{n}\widehat{\mathbf{X}}_1'\widehat{\mathbf{Z}}_1)\widehat{\mathbf{W}}_1(\tfrac{1}{n}\widehat{\mathbf{Z}}_1'\widehat{\mathbf{X}}_1)- \\
&\qquad E(\mathbf{z}_{1i}\mathbf{x}_{1i}(\gamma_0)' I(q_i \le \gamma_0)\mathbf{W}_1(\gamma_0)E(\mathbf{x}_{1i}(\gamma_0)'\mathbf{z}_{1i} I(q_i \le \gamma_0)|| \le \\
&\sup_{\gamma \in (\gamma_0 - \varepsilon, \gamma_0 + \varepsilon)} ||(\tfrac{1}{n}\mathbf{X}_1(\gamma)'\mathbf{Z}_1 I(q \le \gamma))\widehat{\mathbf{W}}_1(\gamma)(\tfrac{1}{n}\mathbf{Z}_1'\mathbf{X}_1(\gamma) I(q \le \gamma))- \\
&\qquad E(\mathbf{z}_i\mathbf{x}_i' I(q_i \le \gamma)\mathbf{W}_1(\gamma)E(\mathbf{x}_i'\mathbf{z}_i I(q_i \le \gamma)||+ \\
&\qquad ||E(\mathbf{z}_{1i}\mathbf{x}_{1i}(\widehat{\gamma}) I(q_i \le \widehat{\gamma})\mathbf{W}_1(\widehat{\gamma})E(\mathbf{x}_{1i}(\widehat{\gamma})'\mathbf{z}_{1i} I(q_i \le \widehat{\gamma})- \\
&\qquad E(\mathbf{z}_{1i}\mathbf{x}_{1i}(\gamma_0) I(q_i \le \gamma_0)\mathbf{W}_1(\gamma_0)E(\mathbf{x}_{1i}(\gamma_0)\mathbf{z}_{1i} I(q_i \le \gamma_0)||
\end{aligned}
$$

■

LEMMA 7 Consider the unrestricted threshold model in equation (3.22) and recall that $\mathbf{x}_i(\gamma) = (\mathbf{x}_i, \lambda_1(\gamma), \lambda_2(\gamma))'$. If $\widehat{\mathbf{W}}_j \xrightarrow{p} \mathbf{W}_j > \mathbf{0}$ for $j = 1, 2$ then the unconstrained minimum distance class estimators defined by equation (2.17) are asymptotically Normal:

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_j(\widehat{v}) - \boldsymbol{\beta}_j) \Longrightarrow N(0, \mathbf{V}_j) \tag{A.24}$$

where $\mathbf{V}_j = (\mathbf{S}_j'\mathbf{W}_j\mathbf{S}_j)^{-1}(\mathbf{S}_j'\mathbf{W}_j\mathbf{Q}_j\mathbf{W}_j\mathbf{S}_j)(\mathbf{S}_j'\mathbf{W}_j\mathbf{S}_j)^{-1}$.

Proof of Lemma 7

We show that the unconstrained estimators are asymptotically Normal. Let $\mathbf{X}_v(v), \mathbf{X}_\perp(v), \Delta\mathbf{X}_v(v), \mathbf{Z}_v$

denote the matrices obtained by stacking the following unrestricted vectors

$\mathbf{x}_i(\gamma_0 + n^{-(1-2\alpha)}\upsilon)' I(q_i \leq \gamma_0 + n^{-(1-2\alpha)}\upsilon)$,

$\mathbf{x}_i(\gamma_0 + n^{-(1-2\alpha)}\upsilon)' I(q_i > \gamma_0 + n^{-(1-2\alpha)}\upsilon)$,

$\mathbf{x}_i(\gamma_0 + n^{-(1-2\alpha)}\upsilon)' I(q_i \leq \gamma_0 + n^{-(1-2\alpha)}\upsilon) - \mathbf{x}_i(\gamma_0 + n^{-(1-2\alpha)}\upsilon) I(q_i > \gamma_0)$,

$\mathbf{z}_i{}' I(q_i \leq \gamma_0 + n^{-(1-2\alpha)}\upsilon)$.

From Theorem 2 of Hansen (1996), Lemma 1, and Lemma A.10 of Hansen (2000) we can deduce that uniformly on $\upsilon \in [-\overline{\upsilon}, \overline{\upsilon}]$

$$\frac{1}{n}\mathbf{Z}_\upsilon' \mathbf{X}_\upsilon(\upsilon) \xrightarrow{p} \mathbf{S}_1 \tag{A.25}$$

$$\frac{1}{\sqrt{n}}\mathbf{Z}_\upsilon' \mathbf{X}_\upsilon(\gamma) \xrightarrow{p} N(0, \mathbf{\Sigma}_1) \tag{A.26}$$

$$\frac{1}{n^{2\alpha}}\mathbf{Z}_\upsilon' \Delta\mathbf{X}_\upsilon \xrightarrow{p} O_p(1) \tag{A.27}$$

Following Hansen and Caner (2004) let

$$\widehat{\boldsymbol{\beta}}_1(\upsilon) = \left(\mathbf{X}_\upsilon' \widehat{\mathbf{Z}}_\upsilon \widehat{\mathbf{W}}_1 \widehat{\mathbf{Z}}_\upsilon' \mathbf{X}_\upsilon\right)^{-1} \widehat{\mathbf{X}}_\upsilon' \widehat{\mathbf{Z}}_\upsilon \widehat{\mathbf{W}}_1 \widehat{\mathbf{Z}}_\upsilon' \mathbf{Y}, \;\; j = 1, 2.$$

and write the unrestricted model as

$$\mathbf{Y} = \mathbf{X}_\upsilon(\upsilon)\boldsymbol{\beta}_1 + \mathbf{X}_\perp(\upsilon)\boldsymbol{\beta}_2 - \Delta\mathbf{X}_\upsilon(\upsilon)\delta_n + u$$

Then, $\sqrt{n}(\widehat{\boldsymbol{\beta}}_1(\upsilon) - \boldsymbol{\beta}_1) =$

$\left((\frac{1}{n}\mathbf{X}_\upsilon(\upsilon)' \mathbf{Z}_\upsilon)\widehat{\mathbf{W}}_1(\frac{1}{n}\mathbf{Z}_\upsilon' \mathbf{X}_\upsilon(\upsilon))\right)^{-1} \left(\frac{1}{n}\mathbf{X}_\upsilon(\upsilon)' \mathbf{Z}_\upsilon \widehat{\mathbf{W}}_1 \left(\frac{1}{\sqrt{n}}\mathbf{Z}_\upsilon{}' u - \frac{1}{\sqrt{n}}\mathbf{Z}_\upsilon' \Delta\mathbf{X}_\upsilon(\upsilon)\delta_n\right)\right)$

$\Longrightarrow (\mathbf{S}_1' \mathbf{W}_1 \mathbf{S}_1)^{-1} \mathbf{S}_1' \mathbf{W}_1 N(0, \mathbf{\Sigma}_1)$.

Since $\widehat{\upsilon} = n^{1-2\alpha}(\widehat{\gamma} - \gamma_0) = O_p(1)$,

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_1(\widehat{\upsilon}) - \boldsymbol{\beta}_1) \Longrightarrow N(0, \mathbf{V}_1)$$

where $\mathbf{V}_1 = (\mathbf{S}_1' \mathbf{W}_1 \mathbf{S}_1)^{-1}(\mathbf{S}_1' \mathbf{W}_1 \mathbf{Q}_1 \mathbf{W}_1 \mathbf{S}_1)(\mathbf{S}_1' \mathbf{W}_1 \mathbf{S}_1)^{-1}$.

Similarly we can get $\sqrt{n}(\widehat{\boldsymbol{\beta}}_1(\upsilon) - \boldsymbol{\beta}_2) \Longrightarrow N(0, \mathbf{V}_2)$ as stated.

LEMMA 8 The restricted estimators defined in equation (2.17) are asymptotically Normal.

$$\sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) \Longrightarrow N(0, \widetilde{\mathbf{V}})$$

where

$$\begin{aligned} \widetilde{\mathbf{V}} &= \mathbf{V} - \widehat{\mathbf{W}}\mathbf{R}\left(\mathbf{R}'\widehat{\mathbf{W}}\mathbf{R}\right)^{-1}\mathbf{R}'\boldsymbol{\upsilon} - \boldsymbol{\upsilon}\mathbf{R}\left(\mathbf{R}'\widehat{\mathbf{W}}\mathbf{R}\right)^{-1}\mathbf{R}'\widehat{\mathbf{W}} \\ &\quad + \widehat{\mathbf{W}}\mathbf{R}\left(\mathbf{R}'\widehat{\mathbf{W}}\mathbf{R}\right)^{-1}\mathbf{R}'\boldsymbol{\upsilon}\mathbf{R}\left(\mathbf{R}'\widehat{\mathbf{W}}\mathbf{R}\right)^{-1}\mathbf{R}'\widehat{\mathbf{W}}. \end{aligned} \tag{A.28}$$

Proof of Lemma 8

Let $\widetilde{\boldsymbol{\beta}}^* = (\widetilde{\boldsymbol{\beta}}_1, \widetilde{\boldsymbol{\beta}}_2)'$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)'$, $\widehat{\mathbf{W}} = diag(\widehat{\mathbf{W}}_1, \widehat{\mathbf{W}}_2)$, $\mathbf{V} = diag(\mathbf{V}_1, \mathbf{V}_2)$

Recalling that $\mathbf{R}'\widehat{\boldsymbol{\beta}} = \boldsymbol{\vartheta}$ the restricted estimator of the STR model can be written as

$$\widetilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}} - \widehat{\mathbf{W}}\mathbf{R}\left(\mathbf{R}'\widehat{\mathbf{W}}\mathbf{R}\right)^{-1}\left(\mathbf{R}'\widehat{\boldsymbol{\beta}} - \boldsymbol{\vartheta}\right) \tag{A.29}$$

then using Lemma 7 we get

$$\sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) \Rightarrow \left(\mathbf{I} - \widehat{\mathbf{W}}\mathbf{R}\left(\mathbf{R}'\mathbf{V}\mathbf{R}\right)^{-1}\mathbf{R}'\right)\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \mathbf{N}(\mathbf{0}, \widetilde{\mathbf{V}}) \tag{A.30}$$

as stated.

Proof of Theorem 4.3

The 2SLS estimators $\widetilde{\boldsymbol{\beta}}_{2SLS}$ fall in the class of estimators (2.17) with $\widehat{\mathbf{W}} = diag(\widehat{\mathbf{W}}_1, \widehat{\mathbf{W}}_2)$

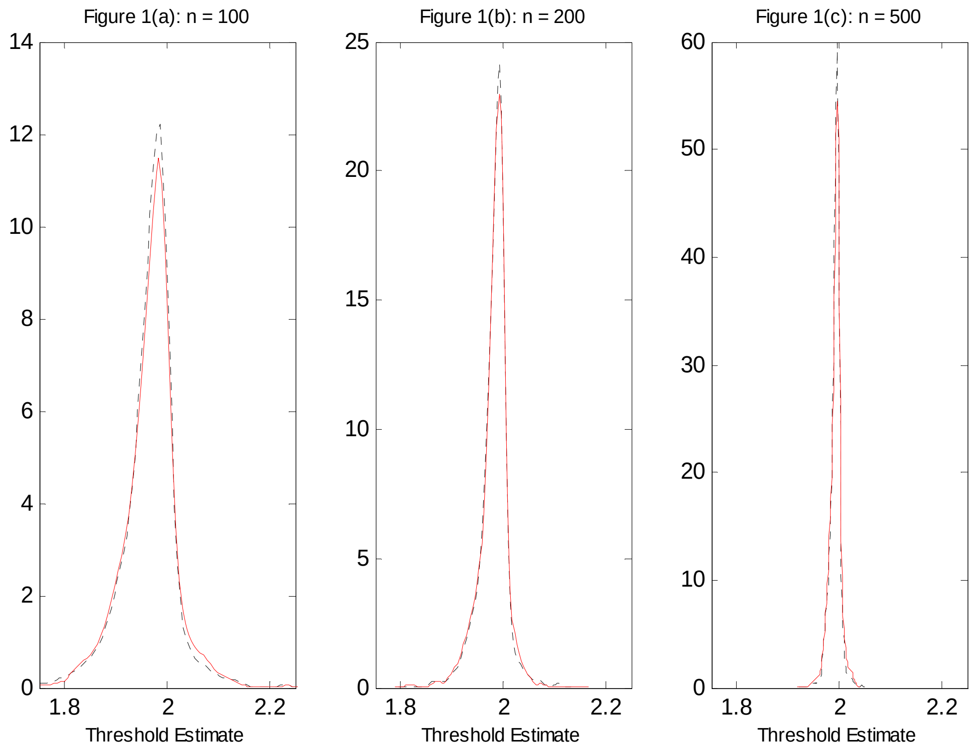
$$\begin{aligned} \widehat{\mathbf{W}}_1 &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i' I(q_i \leq \widehat{\gamma}) \\ \widehat{\mathbf{W}}_2 &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i' I(q_i > \widehat{\gamma}) \end{aligned}$$

The proof for (a) follows Theorem 2 of Caner and Hansen (2004). For the 2SLS estimator, we appeal to Lemma 1 of Hansen (1996), the consistency of $\widehat{\gamma}$, $\widehat{\mathbf{W}}_1 \xrightarrow{p} \mathbf{Q}_1$ and $\widehat{\mathbf{W}}_2 \xrightarrow{p} \mathbf{Q}_2$. Therefore, $\widetilde{\boldsymbol{\beta}}_{2SLS}$ is asymptotically Normal with covariance matrix as stated in (A.28) with $\mathbf{Q} = diag(\mathbf{Q}_1, \mathbf{Q}_2)$ replacing $\widehat{\mathbf{W}} = diag(\widehat{\mathbf{W}}_1, \widehat{\mathbf{W}}_2)$.

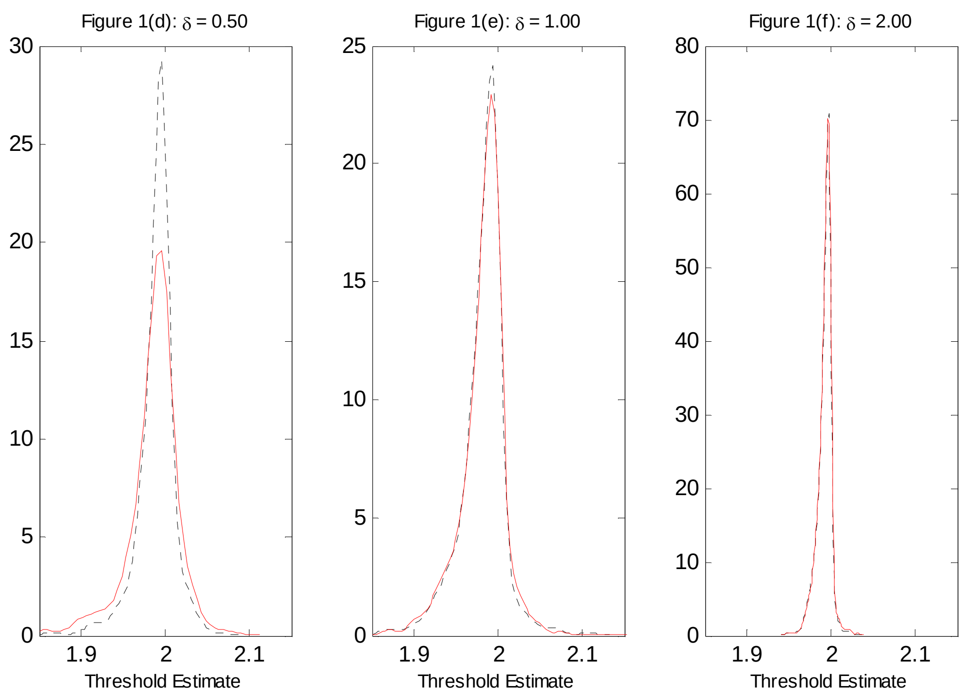
The proof for (b) follows Theorem 3 of Caner and Hansen (2004), which is used to establish that $\widehat{\boldsymbol{\Sigma}}_1(\gamma) \xrightarrow{p} E\left(\mathbf{z}_i\mathbf{z}_i' u_i I(q_i \leq \gamma\right)$ uniformly in $\gamma \in \Gamma$. Then, by the consistency of $\widehat{\gamma}$, the fact that $n^{-1}\widehat{\boldsymbol{\Sigma}}_1 = n^{-1}\widehat{\boldsymbol{\Sigma}}_1(\gamma) \xrightarrow{p} \boldsymbol{\Sigma}_1$, and Lemmas 7 and 8 we obtain Theorem 4.3 (b).

Figures[⊥] 1(a) – (f) : MC Kernel Densities of the Threshold Estimate (Exogenous Slope Variable)

Estimates based on STR and TR for δ = 1 and various sample sizes



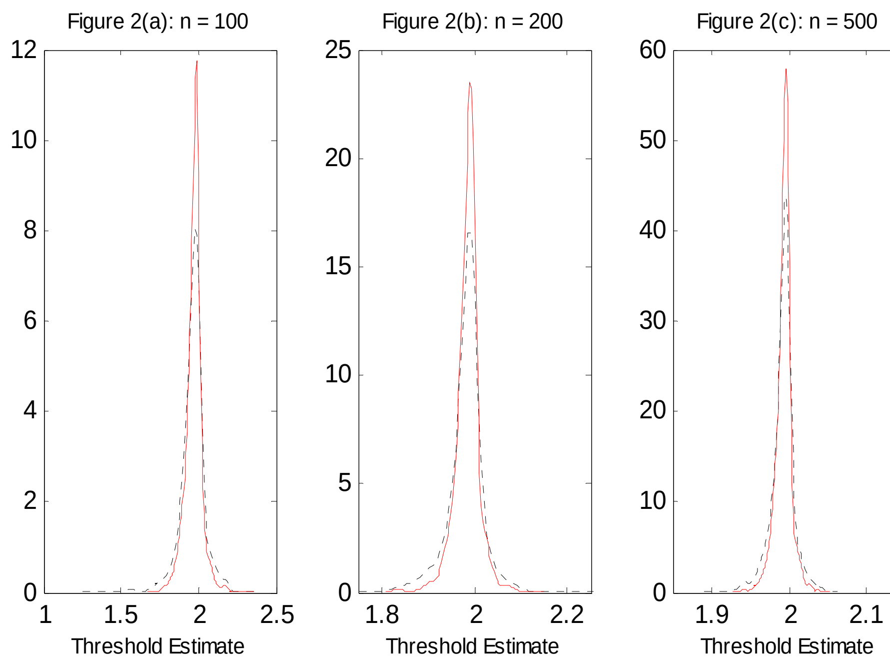
Estimates based on STR and TR for n =500 and various values of δ



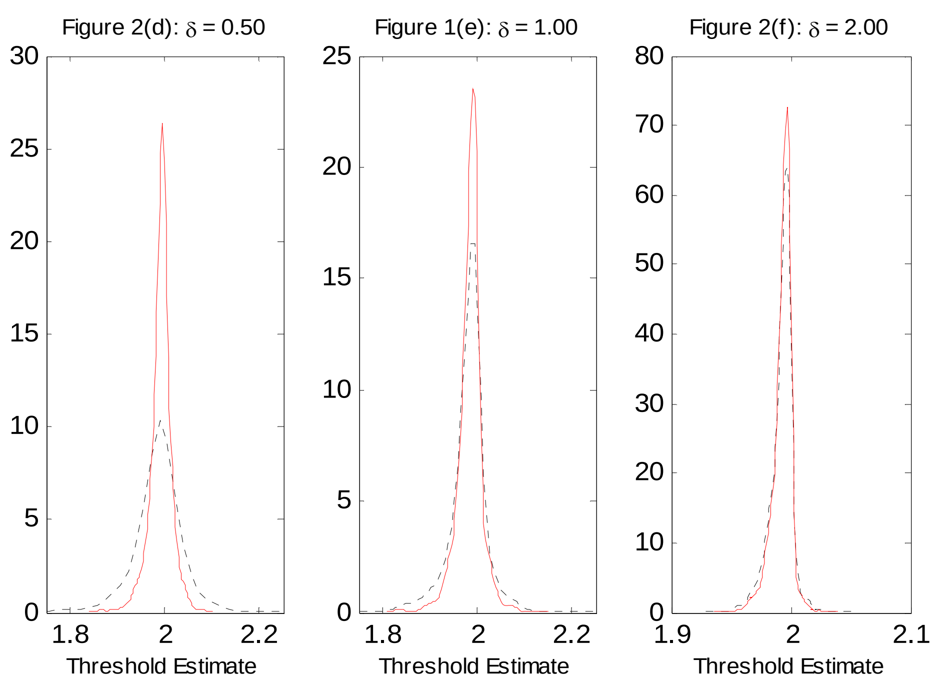
[⊥] The solid line represents the MC kernel density of the STR threshold estimate while the dotted line represents the corresponding density for the TR (Hansen, 2000) threshold estimate.

Figures[⊥] 2(a) – (f) : MC Kernel Densities of the Threshold Estimate (Endogenous Slope Variable)

Estimates based on STR and IVTR for δ = 1 and various sample sizes



Estimates based on STR and IVTR for n =500 and various values of δ



[⊥] The solid line represents the MC kernel density of the STR threshold estimate while the dotted line represents the corresponding density for the IVTR (Caner and Hansen, 2004) threshold estimate.

Table 1: Quantiles of Threshold Estimator, $\gamma = 2$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Exogenous Slope Variable						Endogenous Slope Variable					
	TR			STR			IVTR			STR		
Quantiles	5^{th}	50^{th}	95^{th}	5^{th}	50^{th}	95^{th}	5^{th}	50^{th}	95^{th}	5^{th}	50^{th}	95^{th}
δ = 0.50												
n = 100	1.645	1.964	2.090	1.580	1.958	2.189	0.613	1.953	2.517	1.692	1.971	2.195
n = 200	1.855	1.983	2.045	1.773	1.977	2.073	1.498	1.979	2.187	1.888	1.987	2.077
n = 500	1.950	1.994	2.019	1.922	1.992	2.027	1.887	1.991	2.060	1.952	1.994	2.026
δ = 1.00												
n = 100	1.874	1.975	2.032	1.874	1.974	2.041	1.829	1.974	2.082	1.878	1.978	2.044
n = 200	1.932	1.988	2.013	1.929	1.987	2.014	1.908	1.987	2.034	1.940	1.989	2.023
n = 500	1.975	1.994	2.005	1.973	1.995	2.008	1.964	1.994	2.015	1.975	1.995	2.009
δ = 2.00												
n = 100	1.888	1.975	2.001	1.889	1.976	2.010	1.882	1.976	2.023	1.893	1.978	2.012
n = 200	1.943	1.988	2.000	1.942	1.988	2.000	1.939	1.987	2.012	1.947	1.988	2.005
n = 500	1.976	1.995	2.000	1.976	1.995	2.001	1.974	1.994	2.003	1.978	1.995	2.001

This Table presents Monte Carlo results for the 5th, 50th, and 95th quantiles of the threshold estimator when the threshold variable is endogenous for $\gamma = 2$ and various values of δ. We consider two designs: (i) columns (1)-(6) consider the case where threshold variable is endogenous but the slope variable is exogenous and compare the results of Hansen's (2000) TR model (equation (2.19) in the text, under σ_{uv} = 0) vis-à-vis STR (equation (2.17) in the text, under σ_{uv} = 0); (ii) columns (7)-(12) consider the case where both the threshold variable and slope variable are endogenous and compare the results of Caner and Hansen's (2004) IVTR model (equation (2.19) in the text, under $\sigma_{uv} \neq 0$) vis-à-vis STR (equation (2.17) in the text under $\sigma_{uv} \neq 0$).

Table 2: Quantiles of Slope Coefficient of the second regime $\beta = \beta_2 = 1$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Exogenous Slope Variable						Endogenous Slope Variable					
	TR			STR			IVTR			STR		
Quantiles	5^{th}	50^{th}	95^{th}	5^{th}	50^{th}	95^{th}	5^{th}	50^{th}	95^{th}	5^{th}	50^{th}	95^{th}
δ = 0.50												
n = 100	0.843	0.917	0.99	0.903	0.999	1.115	1.121	1.194	1.322	0.921	1.001	1.085
n = 200	0.869	0.917	0.97	0.934	1.001	1.081	1.133	1.184	1.250	0.949	1.002	1.049
n = 500	0.888	0.917	0.946	0.959	1.000	1.045	1.144	1.175	1.211	0.968	0.999	1.031
δ = 1.00												
n = 100	0.844	0.918	0.987	0.902	0.996	1.110	1.111	1.178	1.244	0.921	0.998	1.075
n = 200	0.870	0.918	0.972	0.935	1.000	1.076	1.129	1.175	1.218	0.949	1.002	1.048
n = 500	0.888	0.918	0.946	0.959	1.000	1.044	1.142	1.172	1.203	0.968	0.999	1.030
δ = 2.00												
n = 100	0.845	0.918	0.988	0.904	0.997	1.112	1.108	1.175	1.240	0.922	0.999	1.075
n = 200	0.870	0.918	0.972	0.935	1.000	1.078	1.127	1.173	1.217	0.949	1.002	1.049
n = 500	0.888	0.918	0.946	0.959	1.000	1.044	1.142	1.172	1.203	0.968	0.999	1.030

This Table presents Monte Carlo results for the 5th, 50th, and 95th quantiles for the slope coefficient of the second regime $\beta = \beta_2$ when the threshold variable is endogenous for $\gamma = 2$ and various values of δ. We consider two designs: (i) columns (1)-(6) consider the case where threshold variable is endogenous but the slope variable is exogenous and compare the results of Hansen's (2000) TR model (equation (2.19) in the text, under $\sigma_{uv} = 0$) vis-à-vis STR (equation (2.17) in the text under $\sigma_{uv} = 0$); (ii) columns (7)-(12) consider the case where both the threshold variable and slope variable are endogenous and compare the results of Caner and Hansen's (2004) IVTR model (equation (2.19) in the text, under $\sigma_{uv} \neq 0$) vis-à-vis STR (equation (2.17) in the text, under $\sigma_{uv} \neq 0$).

Table 3: 95% Confidence Interval Coverage for $\hat{\gamma}$

	Exogenous Slope	Endogenous Slope
Method	LS	GMM
$\delta_2 = 0.5$		
n = 50	0.729	0.770
n = 100	0.887	0.933
n = 200	0.930	0.941
n = 500	0.997	0.995
$\delta_2 = 1.00$		
n = 50	0.794	0.808
n = 100	0.969	0.970
n = 200	0.995	0.992
n = 500	1.000	1.000
$\delta_2 = 2.00$		
n = 50	0.808	0.818
n = 100	0.981	0.960
n = 200	0.999	0.998
n = 500	1.000	1.000