WP 49_11

Andros Kourtellos
University of Cyprus, Cyprus
The Rimini Centre for Economic Analysis (RCEA)

Thanasis Stengos
University of Guelph, Canada
The Rimini Centre for Economic Analysis (RCEA)

Chih Ming Tan
Clark University, USA
The Rimini Centre for Economic Analysis (RCEA)

STRUCTURAL THRESHOLD REGRESSION

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§

November 3, 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 its asymptotic distribution and propose a method to construct bootstrap confidence intervals. We also provide inference for the slope parameters based on GMM. 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

§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. As we will show, 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 change-point models with restrictions across regimes. This finding also implies that existing methods as in Hansen (2000), Caner and Hansen (2004) that ignore in the endogeneity in threshold will still yield consistent estimates for the threshold parameter. However, the story is totally 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. In terms of inference the existing methods are problematic as they ignore the assumption of regime specific heteroskedasticity, which is inherent in our framework.

In particular, 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 change-point 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 regression (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, $\mathbf{g}_{\mathbf{x}i} \equiv \mathbf{g}_{\mathbf{x}}(\mathbf{z}_i; \boldsymbol{\pi}_{\mathbf{x}}) = E(\mathbf{x}_i|\mathbf{z}_i) = \mathbf{\Pi}'_{\mathbf{x}}\mathbf{z}_i$, is given by

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

where $E(\mathbf{v}_{\mathbf{x}i}|\mathbf{z}_i) = 0$.[2] For simplicity we assume that the error $\mathbf{v}_{\mathbf{x}i}$ is independent of the indicator function $I(q_i \leq \gamma)$.

Assuming joint normality of the errors conditionally on $\mathbf{z}_i$,

$$\begin{pmatrix} u_i \\ v_{qi} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & \kappa \\ \kappa & 1 \end{pmatrix} \right) \tag{2.7}$$

and using the properties of the truncated Normal distribution we can obtain the inverse Mills ratio terms

$$E(v_{qi}|\mathbf{z}_i, v_{qi} \leq \gamma - \mathbf{z}_i'\boldsymbol{\pi}_q) = \lambda_1(\gamma - z_i'\pi_q) = -\frac{\phi(\gamma - \mathbf{z}_i'\pi_q)}{\Phi(\gamma - \mathbf{z}_i'\pi_q)} \tag{2.8}$$

and

$$E(v_{qi}|\mathbf{z}_i, v_{qi} > \gamma - \mathbf{z}_i'\boldsymbol{\pi}_q) = \lambda_2(\gamma - z_i'\pi_q) = \frac{\phi(\gamma - \mathbf{z}_i'\pi_q)}{1 - \Phi(\gamma - \mathbf{z}_i'\pi_q)}, \tag{2.9}$$

[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] Our framework can easily be extended to allow nonlinear reduced form models, such as a threshold model; see for example Caner and Hansen (2004).

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the normal pdf and cdf, respectively. Using the assumption of joint Normality (2.7) we can also get that $u_i = \kappa v_{qi} + \epsilon_i$, where ϵ_i is independent of v_{qi}. Then, under Regime 1 the conditional expectation becomes

$$E(u_i|\mathbf{z}_i, q_i \leq \gamma) = E(u_i|\mathbf{z}_i, v_{qi} \leq \gamma - \mathbf{z}_i'\boldsymbol{\pi}_q) = \kappa E(v_{qi}|\mathbf{z}_i, v_{qi} \leq \gamma - \mathbf{z}_i'\boldsymbol{\pi}_q) = \kappa\lambda_1(\gamma - z_i'\pi_q) \quad (2.10)$$

since $E(\epsilon_i|\mathbf{z}_i, v_{qi} \leq \gamma - \mathbf{z}_i'\boldsymbol{\pi}_q) = 0$. Similarly, under Regime 2 we get

$$E(u_i|\mathbf{z}_i, q_i > \gamma) = E(u_i|\mathbf{z}_i, v_{qi} > \gamma - \mathbf{z}_i'\boldsymbol{\pi}_q) = \kappa E(v_{qi}|\mathbf{z}_i, v_{qi} > \gamma - \mathbf{z}_i'\boldsymbol{\pi}_q) = \kappa\lambda_2(\gamma - z_i'\pi_q) \quad (2.11)$$

Define $\lambda_{1i}(\gamma) = \lambda_1(\gamma - z_i'\pi_q)$ and $\lambda_{2i}(\gamma) = \lambda_2(\gamma - z_i'\pi_q)$ and note that the independence of $\mathbf{v}_{\mathbf{x}i}$ with $I(q_i \leq \gamma)$ implies that $E(\mathbf{x}_i|\mathbf{z}_i, I(q_i \leq \gamma)) = E(\mathbf{x}_i|\mathbf{z}_i)$. Next, using equations (2.1), (2.2), (2.10), and (2.11) we obtain the following conditional expectations

$$E(y_i|\mathbf{z}_i, q_i \leq \gamma) = \boldsymbol{\beta}_{\mathbf{x}1}'E(\mathbf{x}_i|\mathbf{z}_i) + E(u_i|\mathbf{z}_i, q_i \leq \gamma) = \boldsymbol{\beta}_{\mathbf{x}1}'\mathbf{g}_{\mathbf{x}i} + \kappa\lambda_{1i}(\gamma) \quad (2.12)$$

$$E(y_i|\mathbf{z}_i, q_i > \gamma) = \boldsymbol{\beta}_{\mathbf{x}2}'E(\mathbf{x}_i|\mathbf{z}_i) + E(u_i|\mathbf{z}_i, q_i > \gamma) = \boldsymbol{\beta}_{\mathbf{x}2}'\mathbf{g}_{\mathbf{x}i} + \kappa\lambda_{2i}(\gamma) \quad (2.13)$$

that define the STR model

$$y_i = \boldsymbol{\beta}_{\mathbf{x}1}'\mathbf{g}_{\mathbf{x}i} + \kappa\lambda_{1i}(\gamma) + \varepsilon_{1i}, \quad q_i \leq \gamma \quad (2.14)$$

$$y_i = \boldsymbol{\beta}_{\mathbf{x}2}'\mathbf{g}_{\mathbf{x}i} + \kappa\lambda_{2i}(\gamma) + \varepsilon_{2i}, \quad q_i > \gamma \quad (2.15)$$

where $\varepsilon_{1i} = \boldsymbol{\beta}_{\mathbf{x}1}'\mathbf{v}_{\mathbf{x}i} - \kappa\lambda_{1i}(\gamma) + u_i$ and $\varepsilon_{2i} = \boldsymbol{\beta}_{\mathbf{x}2}'\mathbf{v}_{\mathbf{x}i} - \kappa\lambda_{2i}(\gamma) + u_i$.[3]

Following Hansen (2000) and a suggestion from the change-point literature we assume a "small threshold" effect. In particular, we assume that $\boldsymbol{\delta}_{\mathbf{x}n} = \boldsymbol{\beta}_{\mathbf{x}1} - \boldsymbol{\beta}_{\mathbf{x}2}$ and $\kappa = \kappa_n$ will both tend to zero slowly as n diverges. The latter assumption implies that the endogeneity bias κ_n vanishes as $n \to \infty$ to ensure that the bias correction (i.e. the inverse Mills ratio terms) to the endogeneity of the threshold will not be present when the model is linear (i.e. there is only one regime). Under this framework, Hansen (2000) showed in the case without regime specific heteroskedasticity that the threshold estimate has an asymptotic distribution free of nuisance parameters. As we show below 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.

[3] Note that equations (2.12) and (2.13) 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).

Let $\lambda_i(\gamma) = \lambda_{1i}(\gamma) I(q_i \le \gamma) + \lambda_{2i}(\gamma) I(q_i > \gamma)$, $\boldsymbol{\beta}_{\mathbf{x}} = \boldsymbol{\beta}_{\mathbf{x}2}$, and $\varepsilon_i = \varepsilon_{1i} I(q_i \le \gamma) + \varepsilon_{2i} I(q_i > \gamma)$. We can then express (2.14) and (2.15) as

$$y_i = \mathbf{g}'_{\mathbf{x}i}\boldsymbol{\beta}_{\mathbf{x}} + \mathbf{g}'_{\mathbf{x}i} I(q_i \le \gamma)\boldsymbol{\delta}_{\mathbf{x}n} + \lambda_i(\gamma)\kappa_n + \varepsilon_i, \tag{2.16}$$

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

A few remarks are in order. First, note that when the error structure in the two regimes (2.1) and (2.2) 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 IVTR model and if additionally $\mathbf{x}_i$ is exogenous then it coincides with Hansen (2000)'s TR model. In general, there are two main differences between STR and TR/IVTR. First, the inverse Mills ratio bias correction term is omitted from either TR or IVTR and as we will be arguing below this yields inconsistent estimates of the slope parameters $\boldsymbol{\beta}_{\mathbf{x}1}$ and $\boldsymbol{\beta}_{\mathbf{x}2}$. Second, the presence of different inverse Mills ratio terms in each of the regimes implies that the error term of the STR model in equation (2.16) is regime-specific heteroskedastic.

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

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

$$\widetilde{\gamma} = \arg\min_{\gamma} S_n(\gamma) \tag{2.17}$$

where

$$S_n(\gamma) = \sum_{i=1}^{n} (y_i - \widehat{\mathbf{g}}'_{\mathbf{x}i}\boldsymbol{\beta}_{\mathbf{x}} - \widehat{\mathbf{g}}'_{\mathbf{x}i} I(q_i \le \gamma)\boldsymbol{\delta}_{\mathbf{x}n} - \widehat{\lambda}_i(\gamma)\kappa_n)^2 \tag{2.18}$$

Finally, once we obtain $\widetilde{\gamma}$, we estimate the slope parameters by 2SLS or GMM. Notice that conditional on γ, estimation in each regime mirrors the Heckman (1979) sample selection bias correction model, the Heckit model. Let $\mathbf{X}$ be the matrix of stacked vectors $\mathbf{x}_i(\widetilde{\gamma}) = (\mathbf{x}'_i, \mathbf{x}'_i I(q_i \le$

$\widetilde{\gamma}), \lambda_i(\widetilde{\gamma}))'$. Similarly, let $\mathbf{Z}$ be the matrix of stacked vectors $\mathbf{z}_i(\widetilde{\gamma}) = (\mathbf{z}_i', \mathbf{z}_i' I(q_i \leq \widetilde{\gamma}), \lambda_i(\widetilde{\gamma}))$. Given a weight matrix $\mathbf{W}$ we can define the class of GMM estimators for $\mathbf{\Theta} = (\boldsymbol{\beta}_{\mathbf{x}}', \boldsymbol{\delta}_{\mathbf{x}n}', \kappa_n)'$

$$\widetilde{\mathbf{\Theta}} = \left(\mathbf{X}'\mathbf{Z}\mathbf{W}\mathbf{Z}'\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{Z}\mathbf{W}\mathbf{Z}'\mathbf{Y}. \tag{2.19}$$

When $\mathbf{W} = (\mathbf{Z}'\mathbf{Z})^{-1}$ we obtain the 2SLS estimator $\widetilde{\mathbf{\Theta}}_{2SLS}$. The 2SLS residual is given by $\widetilde{\varepsilon}_{i,2SLS} = y_i - \mathbf{x}_i(\widetilde{\gamma})'\widetilde{\mathbf{\Theta}}_{2SLS}$. Define $\widetilde{\mathbf{\Sigma}} = \sum_{i=1}^{n} \mathbf{z}_i(\widetilde{\gamma})\mathbf{z}_i(\widetilde{\gamma})'\widetilde{\varepsilon}_{i,2SLS},$. When $\widetilde{\mathbf{W}} = \widetilde{\mathbf{\Sigma}}^{-1}$ then we obtain the efficient GMM estimator, $\widetilde{\mathbf{\Theta}}_{GMM}$.

While from a computational standpoint our estimation strategy is similar to the one employed by Caner and Hansen (2004) there is one key difference. The STR model includes different inverse Mills ratio terms in each regime. To put it differently, STR imposes the exclusion restrictions across the regimes that require that only $\lambda_{1i}(\gamma)$ appears in Regime 1 and only $\lambda_{2i}(\gamma)$ appears in Regime 2. As a result we cannot analyze the estimation problem using results obtained regime by regime. In particular, we cannot decompose the sum of squared errors into two separable regime specific terms due to overlaps. To overcome this problem we next recast the STR model in equation (2.16) as a threshold regression subject to restrictions and exploit the relationship between restricted and unrestricted estimation problems.

3 Threshold Regression with Restrictions

In this section we rewrite the STR model in equation (2.16) as a threshold regression subject to restrictions. In particular, the unrestricted problem generalizes Caner and Hansen (2004) by including both inverse Mills ratio terms in both regimes. We denote with "~" the restricted estimators and with "^" the unrestricted estimators.

Define the vector of inverse Mills ratio terms $\boldsymbol{\lambda}_i(\gamma) = (\lambda_{1i}(\gamma), \lambda_{2i}(\gamma))'$ and the corresponding slope parameters $\boldsymbol{\beta}_{\boldsymbol{\lambda}1} = (\kappa_{11}, \kappa_{12})'$, $\boldsymbol{\beta}_{\boldsymbol{\lambda}2} = (\kappa_{21}, \kappa_{22})'$. Let $\mathbf{g}_i(\gamma) = (\mathbf{g}_{\mathbf{x}i}', \boldsymbol{\lambda}_i(\gamma)')'$, $\boldsymbol{\beta}_1 = (\boldsymbol{\beta}_{\mathbf{x}1}', \boldsymbol{\beta}_{\boldsymbol{\lambda}1}')'$, and $\boldsymbol{\beta}_2 = (\boldsymbol{\beta}_{\mathbf{x}2}', \boldsymbol{\beta}_{\boldsymbol{\lambda}2}')'$. Then the unrestricted STR model takes the form

$$y_i = \mathbf{g}_i(\gamma)' I(q_i \leq \gamma)\boldsymbol{\beta}_1 + \mathbf{g}_i(\gamma)' I(q_i > \gamma)\boldsymbol{\beta}_2 + e_i, \tag{3.20}$$

or more compactly in terms of Regime 1

$$y_i = \mathbf{g}_i'(\gamma)\boldsymbol{\beta} + \mathbf{g}_i'(\gamma) I(q_i \leq \gamma)\boldsymbol{\delta}_n + e_i, \tag{3.21}$$

where $\boldsymbol{\beta} = \boldsymbol{\beta}_2$ and $\boldsymbol{\delta}_n = \boldsymbol{\beta}_1 - \boldsymbol{\beta}_2$ and the (unrestricted) error term, e_i, is given by

$$e_i = \left(\mathbf{v}'_{\mathbf{x}i}\boldsymbol{\beta}_{\mathbf{x}1} - \boldsymbol{\lambda}_i(\gamma)'\boldsymbol{\beta}_{\boldsymbol{\lambda}1}\right) I(q_i \leq \gamma) + \left(\mathbf{v}'_{\mathbf{x}i}\boldsymbol{\beta}_{\mathbf{x}2} - \boldsymbol{\lambda}_i(\gamma)'\boldsymbol{\beta}_{\boldsymbol{\lambda}2}\right) I(q_i > \gamma) + u_i. \tag{3.22}$$

Using consistent first stage estimates as in Section 2.1 we define $\widehat{\mathbf{g}}_i(\gamma) = (\widehat{\mathbf{g}}'_{\mathbf{x}i}, \widehat{\boldsymbol{\lambda}}_i(\gamma)')'$. Then we can estimate the threshold parameter γ by minimizing the unconstrained CLS problem

$$\widehat{\gamma} = \arg\min_{\gamma} S_n^U(\gamma) \tag{3.23}$$

where

$$S_n^U(\gamma) = \sum_{i=1}^{n}(y_i - \widehat{\mathbf{g}}'_i(\gamma)\boldsymbol{\beta} - \widehat{\mathbf{g}}'_i(\gamma)I(q_i \leq \gamma)\boldsymbol{\delta}_n)^2 \tag{3.24}$$

It is easy to verify that the STR model in equation (2.16) is a special case of (3.20) under the following restrictions

$$\kappa_{12} = \kappa_{21} = 0 \tag{3.25}$$

and

$$\kappa_{11} = \kappa_{22} = \kappa. \tag{3.26}$$

In general, define $\boldsymbol{\beta}^* = \left(\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2\right)'$ and the restriction

$$\mathbf{R}'\boldsymbol{\beta}^* = \boldsymbol{\vartheta} \tag{3.27}$$

with $\mathbf{R}$ a $2q \times r$ matrix of rank r, $\boldsymbol{\vartheta}$ a r dimensional vector of constants. Note that the criterion, $S_n(\gamma)$, in equation (2.18) is in fact the restricted sum of squared errors, $S_n^R(\gamma) = S_n(\gamma)$. Then estimation of the STR model in equation (2.16) is equivalent to the estimation of the unrestricted model in equation (3.21) subject to (3.27). In terms of the slope parameters, we can exploit the relationship between the restricted and unrestricted GMM estimators. Consider the unrestricted GMM estimator $\widehat{\boldsymbol{\beta}}^*$ and a consistent weight matrix $\widehat{\mathbf{W}}$. Then, the restricted GMM estimator for $\boldsymbol{\beta}^*$ is given by

$$\widetilde{\boldsymbol{\beta}}^* = \widehat{\boldsymbol{\beta}}^* - \widehat{\mathbf{W}}\mathbf{R}\left(\mathbf{R}'\widehat{\mathbf{W}}\mathbf{R}\right)^{-1}\left(\mathbf{R}'\widehat{\boldsymbol{\beta}}^* - \boldsymbol{\vartheta}\right). \tag{3.28}$$

As we show in Lemma 4 of the Appendix 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 change-point models. Therefore, we proceed by presenting the assumptions for the unrestricted threshold regression.

4 Inference

Define the sigma field $\mathcal{F}_{i-1}$ generated by $\{\mathbf{z}_{i-j}, \mathbf{v}_{i-j}, u_{i-j} : j > 0\}$ with $\mathbf{v}_{i-j} = (\mathbf{v}'_{\mathbf{x}i-j}, v_{qi-j})'$ and $\overline{\mathbf{g}}_i = \sup_{\gamma\in\Gamma}|\mathbf{g}_i(\gamma)|$ and $\overline{\mathbf{g}}_i|e_i| = \sup_{\gamma\in\Gamma}|\mathbf{g}_i(\gamma)e_i|$. Then define the moment functional $\mathbf{M}(\gamma) = E(\mathbf{g}_i(\gamma)\mathbf{g}_i(\gamma)')$ and 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:

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

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

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

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

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

(1.2) $E(u_i|\mathcal{F}_{i-1}) = 0$,

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

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

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

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

(1.7) $\mathbf{D}_1(\gamma)$, $\mathbf{D}_1(\gamma)$, $\mathbf{\Omega}_2(\gamma)$, and $\mathbf{\Omega}_2(\gamma)$ 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) $f_q(\gamma) > 0$, $\mathbf{c}'\mathbf{D}_1(\gamma)\mathbf{c} > 0$, $\mathbf{c}'\mathbf{\Omega}_1(\gamma)\mathbf{c} > 0$, $\mathbf{c}'\mathbf{D}_2(\gamma)\mathbf{c} > 0$, $\mathbf{c}'\mathbf{\Omega}_2(\gamma)\mathbf{c} > 0$

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

(1.11) for all $\gamma \in \Gamma$,$\widetilde{\gamma} = \arg\min_{\gamma\in\Gamma} \sum_{i=1}^{n}(y_i - \mathbf{g}'_i(\gamma)\boldsymbol{\beta} - \mathbf{g}'_i(\gamma)I(q_i \leq \gamma)\boldsymbol{\delta}_n)^2$ 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). Assumption 1.1 excludes time trends and integrated processes. This assumption is trivially satisfied for *i.i.d.* data. Assumptions 1.2 and 1.3 imply that we assume the correct specification of the conditional mean in the structural equation and reduced form. Assumptions 1.4 and 1.5 are unconditional and conditional moment bounds. Assumptions 1.6 and 1.7 require the threshold variable to have a continuous distribution and the conditional variance $E(e_i^2|q_i = \gamma)$ to be semi-continuous at γ_0. This is different from Hansen (2000) and Caner and Hansen (2004) as we are dealing with an asymmetric two sided argmax distribution with different scales; see Stryhn (1996). 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 $n \to \infty$. Assumptions 1.9 and 1.10 are full rank conditions needed to have nondegenerate asymptotic distributions. Assumption 1.11 is an identification condition, which is trivially satisfied given the monotonicity of the inverse Mills ratio terms. The above assumptions are also sufficient to guarantee that the first stage regressions are consistent for the true conditional means i.e. $\widehat{\mathbf{r}} = (\widehat{\mathbf{r}}'_{\mathbf{x}i}, \widehat{\mathbf{r}}'_{\boldsymbol{\lambda}i})' = \mathbf{g}_i(\gamma) - \widehat{\mathbf{g}}_i(\gamma) = o_p(1)$.

4.1 Threshold Estimate

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

Under Assumption 1, the estimator for γ *obtained by minimizing the CLS criterion (2.18),* $\widehat{\gamma}$*, is consistent. That is,*

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

The proof is given in the appendix.

Corollary 4.1 *Under Assumption 1, the estimator for* γ *obtained by minimizing the CLS based on a restricted projection,* $\widetilde{\gamma}$*, is also consistent for* γ_0*. 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 estimate for γ_0, regardless of whether there is endogeneity in the slope. This means that the estimators of Hansen's TR and Caner-Hansen's IVTR that ignore the endogeneity in the threshold will both yield consistent estimates for γ_0.

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

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

where $\xi = \frac{\mathbf{c'D_2c}}{\mathbf{c'D_1c}}$, and $\varphi = \frac{\mathbf{c'\Omega_2c}}{\mathbf{c'\Omega_1c}}$.[4]

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

Under Assumption 1

$$n^{1-2\alpha}(\widehat{\gamma} - \gamma_0) \xrightarrow{d} \omega T \tag{4.29}$$

where $\omega = \frac{\mathbf{c'\Omega_1c}}{(\mathbf{c'D_1c})^2 f}$ *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) in the context of change-point models.[5] For $x < 0$, the cdf of T is given by

$$P(T \leq x) = -\sqrt{\frac{|x|}{2\pi}}\exp(-\frac{|x|}{8}) - c\exp(a|x|)\Phi(-b\sqrt{|x|}) + (d - 2 + \frac{|x|}{2})\Phi(-\frac{\sqrt{|x|}}{2}), \tag{4.30}$$

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 $x > 0$,

$$P(T \leq x) = 1 + \xi\sqrt{\frac{x}{2\pi\varphi}}\exp(-\frac{\xi^2 x}{8\varphi}) - c\exp(ax)\Phi(-b\sqrt{x}) + (-d + 2 - \frac{\xi^2 x}{2\varphi})\Phi(-\frac{\xi}{2}\sqrt{\frac{x}{\varphi}}), \tag{4.31}$$

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 $\mathbf{\Omega}_1 = \sigma^2_{e_1}\mathbf{D}_1$, $\mathbf{\Omega}_2 = \sigma^2_{e_2}\mathbf{D}_2$, where $\sigma^2_{e_1} = E(e^2_{1i}|q = \gamma), \sigma^2_{e_2} = E(e^2_{2i}|q = \gamma)$. This implies that $\omega = \frac{\sigma^2_{e_1}}{(\mathbf{c'D_1c})f}$, and $\varphi = \frac{\sigma^2_{e_2}}{\sigma^2_{e_1}}\xi$. 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 $\xi = 1$, $\varphi = 1$, $\omega = \frac{\mathbf{c'\Omega c}}{(\mathbf{c'Dc})^2 f}$. Hence,

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

[5] However, change-point 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.

when we define $W(s) = W_1(s)$ for $s \leq 0$ and $W(s) = W_2(s)$ for $s > 0$, we can easily see that the distribution coincides with the two sided Wiener distribution established in Hansen (2000) and Caner and Hansen (2004).

Next we investigate the construction of confidence intervals for γ_0 using the distributional result in Theorem 4.1. Let us first consider the pseudo Likelihood Ratio (LR) statistic

$$LR_n(\gamma) = n\frac{S_n(\gamma) - S_n(\widehat{\gamma})}{S_n(\widehat{\gamma})}. \tag{4.32}$$

Define

$$\eta^2 = \frac{\mathbf{c}'\mathbf{\Omega}_1\mathbf{c}}{(\mathbf{c}'\mathbf{D}_1\mathbf{c})\sigma_e^2} \tag{4.33}$$

and

$$\psi = \sup_{-\infty<s<\infty}\left(\left(-\frac{1}{2}|s| + W_1(-s)\right)I(s<0) + \left(-\frac{1}{2}\xi|s| + \sqrt{\phi}W_2(s)\right)I(s>0)\right) \tag{4.34}$$

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

$$LR_n(\gamma_0) \xrightarrow{d} \eta^2\psi \tag{4.35}$$

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.

4.2 The Bootstrap

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

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

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

To construct bootstrap confidence intervals for γ we follow the test-inversion method of Hansen (2000). Using the bootstrap estimates $(\widetilde{\boldsymbol{\delta}}^b_{\mathbf{x}n}, \widetilde{\boldsymbol{\beta}}^b_{\mathbf{x}}, \widetilde{\kappa}^b_n, \widehat{\mathbf{g}}^b_{\mathbf{x}i}, \widehat{\lambda}^b_i(\widetilde{\gamma}))$ we propose to use the following non-pivotal bootstrap statisitic

$$LR_n^b(\gamma) = n\frac{S_n^b(\gamma) - S_n^b(\widehat{\gamma}^b)}{S_n^b(\widehat{\gamma}^b)((\widehat{\eta}_1^b)^2 + (\widehat{\eta}_2^b)^2)},$$

where $(\widetilde{\eta}_1^b)^2 = \frac{\mathbf{c}'\boldsymbol{\Omega}_1^b\mathbf{c}}{(\mathbf{c}'\mathbf{D}_1^b\mathbf{c})(\sigma_e^b)^2} = (\widetilde{\eta}^b)^2$ and $(\widetilde{\eta}_2^b)^2 = \frac{\mathbf{c}'\boldsymbol{\Omega}_2^b\mathbf{c}}{(\mathbf{c}'\mathbf{D}_2^b\mathbf{c})(\sigma_e^b)^2}$.[6] We store likelihood ratio values from bootstraps $\{LR_n^{b(1)}(\gamma), ..., LR_n^{b(B)}(\gamma)\}$ and sort them to determine the $a(B+1)^{th}$ LR value, $LR_n^b\left(c_a^b\right)$, as the critical value for the $1-a$ confidence level. Then we construct the bootstrapped inverted LR confidence region for γ_0, $\widetilde{\Gamma}^b = \left\{\gamma : LR_n(\gamma) \le LR_n^b(c_a^b)\right\}$, where $LR_n(\gamma)$ is computed from the data.

One difficulty with the above bootstrap procedure is that its validity relies heavily on the assumptions of the underlying model and in particular on the assumption of the diminishing threshold effect. Furthermore, it is not clear how one can distinguish whether a given dataset follows the STR model with the diminishing or fixed threshold effect as in Chan (1993). This is a problem because as Yu (2010) shows, the nonparametric bootstrap is invalid in the framework of Chan (1993) and while the parametric bootstrap is valid it is typically not feasible as one needs to specify a complete likelihood. Therefore, to overcome these problems we rely on the framework of an asymptotically diminishing threshold effect, which guarantees the validity of bootstrap at least under the assumption of regime specific homoskedasticity and Normal *i.i.d.* errors. The validity of the bootstrap under the assumptions of an asymptotically diminishing threshold and *i.i.d.* errors was established by Antoch et al (1995) in the context of change-point models. Using

[6] We have also investigated the alternative bootstrap statistic, $LR_n^b(\gamma) = n\frac{S_n^b(\gamma)-S_n^b(\widehat{\gamma}^b)}{S_n^b(\widehat{\gamma}^b)(\widehat{\eta}^b)^2}$. We have found similar patterns, albeit a bit weaker interval coverage.

similar arguments one can easily extend these results to threshold regression.[7]

4.3 Slope Parameters

Consider the unrestricted vector of covariates $\mathbf{x}_i(\gamma_0) = (\mathbf{x}_i', \boldsymbol{\lambda}_i(\gamma_0)')'$. Then, the inference on the slope parameters of the STR model can be viewed as the restricted problem of Caner and Hansen (2004). Let us define the following matrices

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

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

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

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

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

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

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

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

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

Then the following theorem establishes the asymptotic distributions of the (restricted) 2SLS and GMM slope estimators of the STR model in equation (2.16)

Theorem 4.3 *Under Assumption 1 and restrictions given in equation (3.27)*

(a)

$$\sqrt{n}(\widetilde{\boldsymbol{\beta}}^*_{2SLS} - \boldsymbol{\beta}^*) \xrightarrow{d} N(0, \widetilde{\mathbf{V}}_{2SLS}) \tag{4.36}$$

where

$$\begin{aligned}\widetilde{\mathbf{V}}_{2SLS} &= \mathbf{V} - \mathbf{Q}^{-1}\mathbf{R}\left(\mathbf{R}'\mathbf{Q}^{-1}\mathbf{R}\right)^{-1}\mathbf{R}'\mathbf{V} - \mathbf{V}\mathbf{R}\left(\mathbf{R}'\mathbf{Q}^{-1}\mathbf{R}\right)^{-1}\mathbf{R}'\mathbf{Q}^{-1} \\ &\quad + \mathbf{Q}^{-1}\mathbf{R}\left(\mathbf{R}'\mathbf{Q}^{-1}\mathbf{R}\right)^{-1}\mathbf{R}'\mathbf{V}\mathbf{R}\left(\mathbf{R}'\mathbf{Q}^{-1}\mathbf{R}\right)^{-1}\mathbf{R}'\mathbf{Q}^{-1}.\end{aligned} \tag{4.37}$$

(b)

$$\sqrt{n}(\widetilde{\boldsymbol{\beta}}^*_{GMM} - \boldsymbol{\beta}^*) \xrightarrow{d} N(0, \widetilde{\mathbf{V}}_{GMM}) \tag{4.38}$$

[7] However, the problem of proving the validity of bootstrap in the general case of regime specific heteroskedasticity is left for future research.

where

$$\widetilde{\mathbf{V}}_{GMM} = \overline{\mathbf{V}} - \overline{\mathbf{V}}\mathbf{R}\left(\mathbf{R}'\overline{\mathbf{V}}\mathbf{R}\right)^{-1}\mathbf{R}'\overline{\mathbf{V}} \tag{4.39}$$

The proof is given in the appendix.

5 Monte Carlo

We proceed below with an exhaustive simulation that investigates the finite sample performance of our estimators. We explore two sets of simulation experiments. The first set of simulations assume an endogenous threshold variable but retain the assumption of an exogenous slope variable. In this case we compare our results with TR of Hansen (2000). In the second set of simulations we allow for endogeneity in both the threshold and the slope variable and compare our results with IVTR of Caner and Hansen (2004).

Specifically, we assume that the threshold is determined by

$$q_i = 2 + z_{qi} + v_{qi},$$

where v_{qi} is *i.i.d.* $N(0,1)$. The first set of simulations are based on the following threshold regression

$$\text{Model 1}: \; y_i = \beta_1 + \beta_2 x_i + (\delta_1 + \delta_2 x_i) I\{q_i \leq 2\} + u_i, \tag{5.40}$$

where

$$z_i = (w x_i + (1-w)\varsigma_{zi}) / \sqrt{w^2 + (1-w)^2} \tag{5.41}$$

and

$$u_i = 0.1\varsigma_{ui} + \kappa v_{qi}, \tag{5.42}$$

where ς_{zi} and ς_{ui} are independent *i.i.d.* $N(0,1)$ random variables. The degree of endogeneity of the threshold is controlled by κ. The degree of correlation between the instrumental variable z_i and the included exogenous slope variable x_i is controlled by w. We fix $\kappa = 0.95$, $w = 0.5$, $\beta_1 = \beta_2 = 1$, and $\delta_1 = 0$ and vary δ_2 over the values of $1, 2, 3, 4, 5$, which correspond to a range of small to large threshold effects.

The second set of simulations are based on a model that includes both an endogenous, x_{1i}, and an exogenous slope variable, x_{2i},

$$\text{Model 2: } y_i = \beta_1 + \beta_2 x_{1i} + \beta_2 x_{2i} + (\delta_1 + \delta_2 x_{1i} + \delta_3 x_{1i}) I\{q_i \leq 2\} + u_i, \tag{5.43}$$

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

where

$$z_i = (wx_i + (1-w)\varsigma_{zi}) \,/ \sqrt{w^2 + (1-w)^2}, \tag{5.44}$$

and

$$u_i = (c_{xu}v_{xi} + c_{qu}v_{qi} + (1 - c_{xu} - c_{qu})\varsigma_{ui}) \,/ \sqrt{c_{xu}^2 + c_{qu}^2 + (1 - c_{xu} - c_{qu})^2}, \tag{5.45}$$

where ς_{zi} and ς_{ui} are independent *i.i.d.* $N(0,1)$ random variables. The degree of endogeneity of the threshold, κ, is controlled by the correlation coefficient between u_i and v_{qi} given by $c_{qu}/\sqrt{c_{xu}^2 + c_{qu}^2 + (1 - c_{xu} - c_{qu})^2}$. Similarly, the degree of endogeneity of x_1 between u_i and v_{xi} given by $c_{xu}/\sqrt{c_{xu}^2 + c_{qu}^2 + (1 - c_{xu} - c_{qu})^2}$. We fix $c_{xu} = c_{qu} = 0.45$, $w = 0.5$, $\beta_1 = \beta_2 = 1$, and $\delta_1 = \delta_2 = 0$ and vary δ_3 over the values of $1, 2, 3, 4, 5$.

In both cases we consider sample sizes of 100, 250, 500, and 1000 using 1000 monte carlo replications simulations. We also investigated different degrees of endogeneity and correlation between x_i and z_i and our results are qualitatively similar. We then examined what happened when we allowed for a threshold effect in all slope variables (including the intercept) as well as when we varied the degree of endogeneity. We also considered various degrees of correlation between the instrumental variables z's and the exogenous slope variables x_2's. All the results are qualitatively similar and are available upon request.

First, we discuss the monte carlo findings on the estimation of the threshold value, γ, based on the STR model as described in Section 3.1. Table 1 presents the 5th, 50th, and 95th quantiles for the distribution of the threshold estimate of $\hat{\gamma}$ for Model 1 and Model 2 in equations (5.40) and (5.43), respectively. We also compare our STR results with the results obtained if we ignore endogeneity in the threshold and simply employ the TR of Hansen (2000) in the case of Model 1 and the IVTR of Caner and Hansen (2004) in the case of Model 2. We see that the performance of the STR estimator improves as the parameter of the threshold effect, δ_2 or δ_3, and/or the sample size, n, increases. Specifically, the 50th quantile approaches the true threshold parameter, $\gamma_0 = 2$, as the sample size increases and the width of the distribution becomes smaller as δ increases. We also find that both TR and IVTR, which both ignore the endogeneity in the threshold variable estimate the threshold parameter accurately and exhibit similar behavior to STR. This finding verifies Corollary 4.1.

The results of Table 1 are also verified by Figures 1 and 2 that present the Gaussian kernel density estimates, using Silverman's bandwidth, for $\hat{\gamma}$, over different sample sizes and different threshold effects, respectively. Specifically, Figures 1(a)-(d) and Figures 1(e)-(h) present the density estimates for Model 1 (using $\delta_2 = 2$) and Model 2 (using $\delta_3 = 2$), respectively, for $n = 100$, 250, 500, and 1000. Similarly, Figures 2(a)-(e) present the density estimates for Model 1 using $n = 1000$ and $\delta_2 = 1, 2, 3, 4, 5$ and Figures 2(e)-(h) present the density estimates for Model 2 using $n = 1000$ and $\delta_3 = 1, 2, 3, 4, 5$. The solid red line shows the STR estimates while the black dashed line shows the

TR or IVTR estimates, which ignore endogeneity in the threshold variable. The similar behavior of all three estimators is evident for all threshold effects and sample sizes. Furthermore, all estimators exhibit efficiency gains the larger the threshold effect and/or the larger the sample size.

Table 2 presents bootstrap coverage probabilities of a nominal 90% interval $\widehat{\Gamma}^*$ using 300 bootstrap replications.[8] We constructed $\widehat{\Gamma}^*$ using the parametric correction of heteroskedasticity within each regime as explained in Section 3.4 of Hansen (2000). We find that the coverage probability increases with either the size of the threshold effect or the sample size and becomes conservative for larger values. In particular, while for a small threshold effect $\delta_2 = 1$ or $\delta_3 = 1$ the bootstrap coverage is far from the nominal coverage, for a large threshold effect $\delta_2 = 5$ or $\delta_3 = 5$ the coverage is conservative even for a small sample size of 100. Interestingly, our bootstrap findings are similar, albeit less conservative, to the simulation findings of Hansen (2000) and Caner and Hansen (2004), which are based on an asymptotic distribution, under the assumption of regime specific homoskedasticity. Furthermore, our results are consistent with Theorem 3 of Hansen (2000), which suggests that under the assumption of *i.i.d.* Gaussian errors and regime specific homoskedastacity the confidence interval is asymptotically conservative for fixed parameter value as n becomes large.

Next, we discuss the monte carlo evidence on the estimation of the slope parameters, β_2, δ_2 (or δ_3) and κ. Table 3 presents the quantiles of the distributions of the slope coefficients β_2 and δ_2. In Panel A we present the LS estimates for Model 1 and in Panel B we present the GMM estimates for Model 2. As in the case of the threshold estimates we find that STR accurately estimates the parameters for both models, for different sample sizes, and for different threshold effects. The performance of both slope coefficient estimates improves as the threshold effect or the sample size increases. In sharp contrast to the results for the threshold estimate, we find that TR in the case of Model 1 and IVTR in the case of Model 2 yield substantial bias in the estimation of β_2. More precisely, while the true value of $\beta_2 = 1$, in the case of Model 1, TR converges about the value 0.81 and in the case of Model 2, IVTR converges about the value of 0.74. Nevertheless the slope estimates for δ_2 and δ_3 appear to be accurate implying that the bias in the estimation of β_1 is of an equal magnitude. 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.

Table 4 presents the quantiles of the coefficient of the inverse Mills ratio term and verifies that STR accurately estimates, κ, for both models, for different sample sizes, and for different threshold effects. The true value for κ is 0.95 for Model 1 and 0.70 for Model 2 as implied by equations (5.44) and (5.45), respectively. In both cases, the 50th quantile approaches the true value of κ, as the sample size increases and the width of the distribution becomes smaller as δ increases.

[8] More accurate results will need a much larger number of replications. Unfortunately, computational power restricted our monte carlo experiments to 300 bootstrap replications.

Finally, we discuss the inference of the slope parameters. The fact that the threshold estimator enjoys a faster rate of convergence than the slope estimators implies that we can estimate the slope coefficients without error by simply treating the threshold estimate as known as described in Section 2.1. In the case of Model 2, Theorem 4.3 shows that the GMM slope estimates are asymptotically normal and asymptotic standard errors can be computed by consistently estimating the asymptotic covariance matrix. It is also easy to show that in the case of Model 1, the LS estimates are also asymptotically normal. This implies that we can construct conventional asymptotic confidence intervals using the normal approximation. As in the case of Caner and Hansen (2004) we focus on the threshold effect parameter and report the nominal 95% confidence interval coverage for Models 1 and 2 in Table 5. Generally, coverage improves as the sample size increases and especially as the threshold effect becomes larger. However, coverage is rather poor for a small threshold effect $\delta_3 = 1$ in the case of Model 2. In principle, one can employ a bootstrap version of the Bonferroni-type approach, which is employed in Caner and Hansen (2004), in order to account for the uncertainty concerning γ. One difficulty is that the asymptotic distribution of the threshold estimator in the case of STR is not practical (as explained in Section 4.1) and therefore a Bonferroni-type approach will have to rely on bootstrap approximation. However, such an approach would be extremely computationally intensive, and it is not clear how practical it would be to implement in applied settings. We plan to follow up on this issue in future research.

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 study 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 as well as a GMM estimator for the slope parameters. 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] Frankel, J. and D. Romer, (1999), "Does Trade Cause Growth?," *American Economic Review*, 89, 3, 379-399.

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

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

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

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

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

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

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

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

[16] 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.

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

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

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

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

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

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

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 (3.21) 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

$$Eh_i^r(\gamma, \gamma') \leq B|\gamma - \gamma'| \tag{A.3}$$

$$Ek_i^r(\gamma, \gamma') \leq B|\gamma - \gamma'| \tag{A.4}$$

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 (2.16) $h_i(\gamma) = (\mathbf{g}_i d_i(\gamma), \lambda_i(\gamma))$ 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 $n \longrightarrow \infty$

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

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

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

Proof of Lemma 2.

To show (A.5) note that

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

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

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

recall that $\widehat{\mathbf{x}}_i = \widehat{\mathbf{g}}_{\mathbf{x}i} = \mathbf{g}_{\mathbf{x}i} - \widehat{\mathbf{r}}_{\mathbf{x}i}$, $\widehat{\lambda}_i(\gamma) = (\widehat{\lambda}_{1i}(\gamma), \widehat{\lambda}_{2i}(\gamma))$, $\widehat{\lambda}_{1i}(\gamma) = \lambda_{1i}(\gamma) - \widehat{r}_{\lambda_1 i}$, and $\widehat{\lambda}_{2i}(\gamma) = \lambda_{2i}(\gamma) - \widehat{r}_{\lambda_2 i}$.

First note that $\frac{1}{n}\sum_i(\widehat{\mathbf{x}}_i\widehat{\mathbf{x}}_i'I(q_i \le \gamma)) \overset{p}{\to} E(\mathbf{g}_i\mathbf{g}_i'I(q_i \le \gamma))$ 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 $j = 1, 2$

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

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

Similarly, we can show that

$$\begin{aligned} \frac{1}{n}\sum_i\widehat{\lambda}_{1i}(\gamma)\widehat{\lambda}_{2i}(\gamma)I(q_i \le \gamma) &= \frac{1}{n}\sum_i\lambda_{1i}(\gamma)\lambda_{2i}(\gamma)I(q_i \le \gamma) - \frac{1}{n}\sum_i\lambda_{1i}(\gamma)\widehat{r}_{\lambda_1 i}I(q_i \le \gamma) \\ &\quad -\frac{1}{n}\sum_i\lambda_{2i}(\gamma)\widehat{r}_{\lambda_2 i}I(q_i \le \gamma) + \frac{1}{n}\sum_i\widehat{r}_{\lambda_1 i}\widehat{r}_{\lambda_2 i}I(q_i \le \gamma) \end{aligned}$$

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

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

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 $j = 1, 2$

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

Then,

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

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 $\gamma\in\Gamma$,

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

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

and

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

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,

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

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. $\sup_{\gamma\in\Gamma} |S_n^{*U}(\gamma) - S_n^{*U}(\gamma_0)| \xrightarrow{p} 0$ 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 $\sup_{|\gamma-\gamma_0|\geq\epsilon} S_n^{*U}(\gamma) < S_n^{*U}(\gamma_0)$ for every $\epsilon > 0$. Therefore, it follows that $\widehat{\gamma} \xrightarrow{p} \gamma_0$ for the unrestricted problem.

Assuming the restrictions in equation (3.27) 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 $S_n^R(\gamma) = S_n^U(\gamma) + (\boldsymbol{\vartheta} - \mathbf{R}'\boldsymbol{\beta}^*)'(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma)'\widehat{\mathbf{X}}^*(\gamma))^{-1}\mathbf{R})^{-1}(\boldsymbol{\vartheta} - \mathbf{R}'\boldsymbol{\beta}^*)$. The proof proceeds in steps. First we establish that the unrestricted and the restricted problems share the same rate of convergence.

Let $\widehat{\mathbf{X}}^*(\gamma)$ denote the partitioned regressor matrix associated with the unrestricted sum of squared residuals $S_n^U(\gamma)$, threshold value γ, and estimated coefficients $\widehat{\boldsymbol{\beta}}^*_{\gamma}$. Similarly, let $\widehat{\mathbf{X}}^*(\gamma_0)$ denote the partitioned regressor matrix associated with the unrestricted sum of squared residuals $S_n^U(\gamma_0)$, threshold value γ_0 and estimated coefficients $\widehat{\boldsymbol{\beta}}^*_{\gamma_0}$. We also use the subscript 0 to denote the parameter at the true value.

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

$$(\widehat{\mathbf{X}}^*(\gamma)'\widehat{\mathbf{X}}^*(\gamma))^{-1} = (\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1} + O_p(\frac{|\gamma - \gamma_0|}{n^2}) \tag{A.13}$$

and

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

Consider

$$\widehat{\boldsymbol{\beta}}^*_\Delta = \widehat{\boldsymbol{\beta}}^*_\gamma - \widehat{\boldsymbol{\beta}}^*_{\gamma_0}$$

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

$$= (\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}((\widehat{\mathbf{X}}^*(\gamma) - \widehat{\mathbf{X}}^*(\gamma_0))'\mathbf{G}^*(\gamma_0)\boldsymbol{\beta}^*_0$$

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

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

with

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

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

$= |\gamma - \gamma_0|O_p(n^{-1/2})$, where the first equality uses (A.13). To get the second equality note that $(\widehat{\mathbf{X}}^*(\gamma) - \widehat{\mathbf{X}}^*(\gamma_0))'\mathbf{G}^*(\gamma_0) = |\gamma - \gamma_0|O_p(1)$,

$\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1/2}(\widehat{\mathbf{X}}^*(\gamma) - \widehat{\mathbf{X}}^*(\gamma_0))'\mathbf{G}^*(\gamma_0)\boldsymbol{\beta}^*_0 = |\gamma - \gamma_0|O_p(\frac{1}{\sqrt{n}})$, and

$(\widehat{\mathbf{X}}^*(\gamma) - \widehat{\mathbf{X}}^*(\gamma_0))'\mathbf{e} = |\gamma - \gamma_0|O_p(1)$.

Therefore, $\widehat{\boldsymbol{\beta}}^*_\Delta = |\gamma - \gamma_0|O_p(n^{-1})$.

Furthermore, note that $\widehat{\boldsymbol{\beta}}^*_\Delta\mathbf{R} = |\gamma - \gamma_0|O_p(n^{-1})$ and $(\boldsymbol{\vartheta} - \mathbf{R}'\boldsymbol{\beta}^*)' = |\gamma - \gamma_0|O_p(n^{-1})$. Then,

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

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

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

$$-(\boldsymbol{\vartheta} - \mathbf{R}'\widehat{\boldsymbol{\beta}}^*_{\gamma_0})'(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\mathbf{R})^{-1}(\boldsymbol{\vartheta} - \mathbf{R}'\boldsymbol{\beta}^*_{\gamma_0})]$$

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

$$+[(\boldsymbol{\vartheta} - \mathbf{R}'\widehat{\boldsymbol{\beta}}^*_\gamma)'(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\mathbf{R})^{-1}(\boldsymbol{\vartheta} - \mathbf{R}'\widehat{\boldsymbol{\beta}}^*_\gamma)$$

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

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

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

$$-\widehat{\boldsymbol{\beta}}^{*\prime}_{\gamma_0}\mathbf{R}(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\mathbf{R})^{-1}\mathbf{R}'\widehat{\boldsymbol{\beta}}^*_{\gamma_0}$$

$$-2\boldsymbol{\vartheta}'\mathbf{R}(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\mathbf{R})^{-1}\mathbf{R}'(\widehat{\boldsymbol{\beta}}^*_\gamma - \widehat{\boldsymbol{\beta}}^*_{\gamma_0})$$

$$
\begin{aligned}
&+|\gamma-\gamma_0|^2 O_p(n^{-1})\\
&= [S_n^U(\gamma) - S_n^U(\gamma_0)]\\
&\quad +2\widehat{\boldsymbol{\beta}}_{\Delta}^{*\prime}\mathbf{R}(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\mathbf{R})^{-1}\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\widehat{\mathbf{X}}^*(\gamma_0)'\mathbf{e}\\
&\quad +\widehat{\boldsymbol{\beta}}_{\Delta}^{*\prime}\mathbf{R}(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\mathbf{R})^{-1}\mathbf{R}'\widehat{\boldsymbol{\beta}}_{\Delta}^{*}\\
&\quad +2\widehat{\boldsymbol{\beta}}_{\Delta}^{*\prime}\mathbf{R}(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\mathbf{R})^{-1}(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\widehat{\mathbf{X}}^*(\gamma_0)'\mathbf{G}^*(\gamma_0)\boldsymbol{\beta}_0^* - \boldsymbol{\vartheta})\\
&\quad +|\gamma-\gamma_0|^2 O_p(n^{-1}).\\
&= [S_n^U(\gamma) - S_n^U(\gamma_0)]\\
&\quad +2\widehat{\boldsymbol{\beta}}_{\Delta}^{*\prime}\mathbf{R}(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\mathbf{R})^{-1}\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\widehat{\mathbf{X}}^*(\gamma_0)'\mathbf{e}\\
&\quad +\widehat{\boldsymbol{\beta}}_{\Delta}^{*\prime}\mathbf{R}(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\mathbf{R})^{-1}\mathbf{R}'\widehat{\boldsymbol{\beta}}_{\Delta}^{*}\\
&\quad +2\widehat{\boldsymbol{\beta}}_{\Delta}^{*\prime}\mathbf{R}(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\mathbf{R})^{-1}(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\widehat{\mathbf{X}}^*(\gamma_0)'(\mathbf{G}^*(\gamma') - \widehat{\mathbf{X}}^*(\gamma_0))\boldsymbol{\beta}_0^*\\
&\quad +|\gamma-\gamma_0|^2 O_p(n^{-1}).
\end{aligned}
$$

Now consider the second term divided by $|\gamma-\gamma_0|$

$$
\begin{aligned}
&||2\widehat{\boldsymbol{\beta}}_{\Delta}^{*\prime}\mathbf{R}(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\mathbf{R})^{-1}\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\widehat{\mathbf{X}}^*(\gamma_0)'\mathbf{e}||/n^{2\alpha-1}(\gamma-\gamma_0)\\
&= ||\mathbf{A}_n'((\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1/2}\mathbf{R}\left(\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1}\mathbf{R}\right)^{-1}\mathbf{R}'(\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1/2})\\
&\quad \cdot((\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1/2}\mathbf{e})||/n^{2\alpha-1}(\gamma-\gamma_0)\\
&\leq ||\mathbf{A}_n'||||((\widehat{\mathbf{X}}^*(\gamma_0)'\widehat{\mathbf{X}}^*(\gamma_0))^{-1/2}\mathbf{e})||/n^{2\alpha-1}(\gamma-\gamma_0) = o_p(1)
\end{aligned}
$$

Note that the third term is nonnegative and divided by $n^{2\alpha-1}(\gamma-\gamma_0)$ is also $o_p(1)$. The key in the fourth term is $(\mathbf{G}^*(\gamma') - \widehat{\mathbf{X}}^*(\gamma_0))\boldsymbol{\beta}_0^*$ which is also $o_p(1)$ when it is divided by $n^{2\alpha-1}(\gamma-\gamma_0)$.

Therefore,

$$
\frac{S_n^R(\gamma) - S_n^R(\gamma_0)}{n^{2\alpha-1}(\gamma-\gamma_0)} \geq \frac{S_n^U(\gamma) - S_n^U(\gamma_0)}{n^{2\alpha-1}(\gamma-\gamma_0)} + o_p(1) \tag{A.15}
$$

We can now focus on the unrestricted problem since the rates of convergence for the restricted and unrestricted problems ar the same. 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 $m(\boldsymbol{\theta}) = -(y_i - \mathbf{g}_i(\gamma)' \boldsymbol{\beta}_1 I(q_i \leq \gamma) - \mathbf{g}_i(\gamma)' \boldsymbol{\beta}_2 I(q_i > \gamma))^2$ and let $\boldsymbol{\theta} = \left(\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \gamma\right)'$. Recall that $\gamma \in \Gamma = [\underline{\gamma}, \overline{\gamma}]$, then we have $I(q_i \leq \gamma) \leq I(q \leq \gamma \wedge \gamma_0)$ and $I(q_i > \gamma) \geq I(\gamma_0 < q \leq \gamma \vee \gamma_0)$, where "$\wedge$" and "$\vee$" denote the minimum and maximum, respectively.

We can derive the following formula.

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

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

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

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

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

$$=$$

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

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

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

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

$$=$$

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

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

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

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

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

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

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

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

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

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

$\leq$

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

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

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

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

$\leq -C\left(||\boldsymbol{\beta}_{10}-\boldsymbol{\beta}_1||^2+||\boldsymbol{\beta}_{20}-\boldsymbol{\beta}_2||^2+|\gamma_0-\gamma|^2+|F_q(\gamma)-F_q(\gamma_0)|\right)=-Cd^2(\boldsymbol{\theta},\boldsymbol{\theta}_0),$ 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

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

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

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

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 $|\widehat{\gamma} - \gamma_0| + |F(\widehat{\gamma}) - F(\gamma_0)| = O_p(n^{-1/2}) + O_p(n^{-1}) = O_p(n^{-1})$.

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

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

We 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 $\gamma \leq \gamma_0$,

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

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

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

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

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

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

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

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

and

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

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

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

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

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

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

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

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

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

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

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

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 $\zeta_i=\mathbf{c}'\mathbf{\Omega}_i\mathbf{c}f$, 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

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

Our analysis below will be restricted to the region $[\gamma_0+\overline{v}/a_n \leq \gamma \leq \gamma_0+B]$ 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 $[\gamma_0 - \overline{v}/a_n \geq \gamma \geq \gamma_0 - B]$ is similar.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$=$

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

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

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

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

and

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

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 $\boldsymbol{\Omega}_1 f$ and hence

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

(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,

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

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,

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

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 $s \in [-\overline{\upsilon}, 0]$,

$$\underset{-\infty<\upsilon<\infty}{\arg\max}\, Q_n(\upsilon)$$

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

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

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

■

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,

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

$$= \frac{1}{\sigma^2} \sup_{-\infty<\upsilon<\infty} \left(\left(-\mu_1|\upsilon| + 2\zeta_1^{1/2}W_1(\upsilon)\right) I(\upsilon<0) + \left(-\mu_2|\upsilon| + 2\zeta_2^{1/2}W_2(\upsilon)\right) I(\upsilon>0)\right)$$

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

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

$$= \frac{1}{\sigma^2} \sup_{-\infty<\upsilon<\infty} \left(\left(-\mu_1|\tfrac{\zeta_1}{\mu_1^2}s| + 2\zeta_1^{1/2}W_1(\tfrac{\zeta_1}{\mu_1^2}s)\right) I(\upsilon<0) + \left(-\mu_2|\tfrac{\zeta_1}{\mu_1^2}s| + 2\zeta_2^{1/2}W_2(\tfrac{\zeta_1}{\mu_1^2}s)\right) I(\upsilon>0)\right)$$

$=\frac{\zeta_1}{\sigma^2\mu_1} \sup_{-\infty<\upsilon<\infty} \left((-|s| + 2W_1(s))\, I(\upsilon < 0) + \left(-\xi|s| + 2\sqrt{\varphi}W_2(s)\right) I(\upsilon > 0) \right)$

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

Note that $\psi = 2\max(\psi_1, \psi_2)$, where $\psi_1 = \sup_{s\leq 0} (-|s| + 2W_1(s))$ 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.

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

■

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.

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

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

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

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

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

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

$$||E(\mathbf{z}_{1i}\mathbf{x}_{1i}(\widehat{\gamma})I(q_i \leq \widehat{\gamma})\mathbf{W}_1(\widehat{\gamma})E(\mathbf{x}_{1i}(\widehat{\gamma})'\mathbf{z}_{1i}I(q_i \leq \widehat{\gamma})-$$

$$E(\mathbf{z}_{1i}\mathbf{x}_{1i}(\gamma_0)I(q_i \leq \gamma_0)\mathbf{W}_1(\gamma_0)E(\mathbf{x}_{1i}(\gamma_0)\mathbf{z}_{1i}I(q_i \leq \gamma_0)||$$

■

LEMMA 7 Consider the unrestricted threshold model in equation (3.21) 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 (2.19) are asymptotically Normal:

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

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 $\mathbf{X}_\upsilon\,(\upsilon)$,$\mathbf{X}_\perp\,(\upsilon)$,$\Delta\mathbf{X}_\upsilon\,(\upsilon)$,$\mathbf{Z}_\upsilon$ 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 $\upsilon \in [-\overline{\upsilon}, \overline{\upsilon}]$

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

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

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

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

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

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

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

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

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 (2.19) are asymptotically Normal.

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

where

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

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

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

as stated.■

Proof of Theorem 4.3

The 2SLS estimators $\widetilde{\boldsymbol{\beta}}_{2SLS}$ fall in the class of estimators (2.19) 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 $\widehat{\mathbf{W}} = diag(\widehat{\mathbf{W}}_1, \widehat{\mathbf{W}}_2)$.

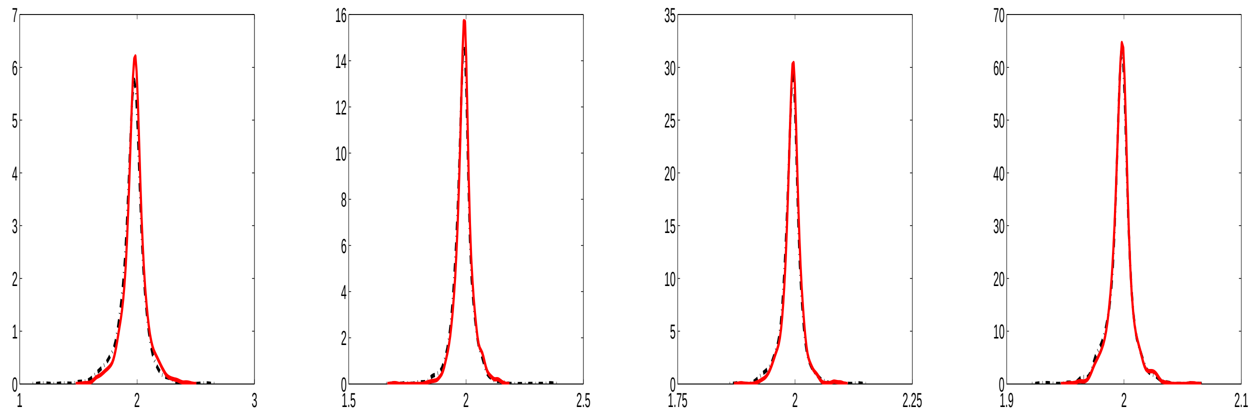
The proof for (b) follows Theorem 3 of Caner and Hansen (2004), which is used to establish that $\widehat{\boldsymbol{\Sigma}}_1(\gamma) \xrightarrow{p} E\left(\mathbf{z}_i\mathbf{z}_i' u_i I(q_i \leq \gamma\right)$ uniformly in $\gamma \in \Gamma$. Then, by the consistency of $\widehat{\gamma}$, the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

■

Figure 1: MC kernel densities of the threshold estimate for different sample sizes

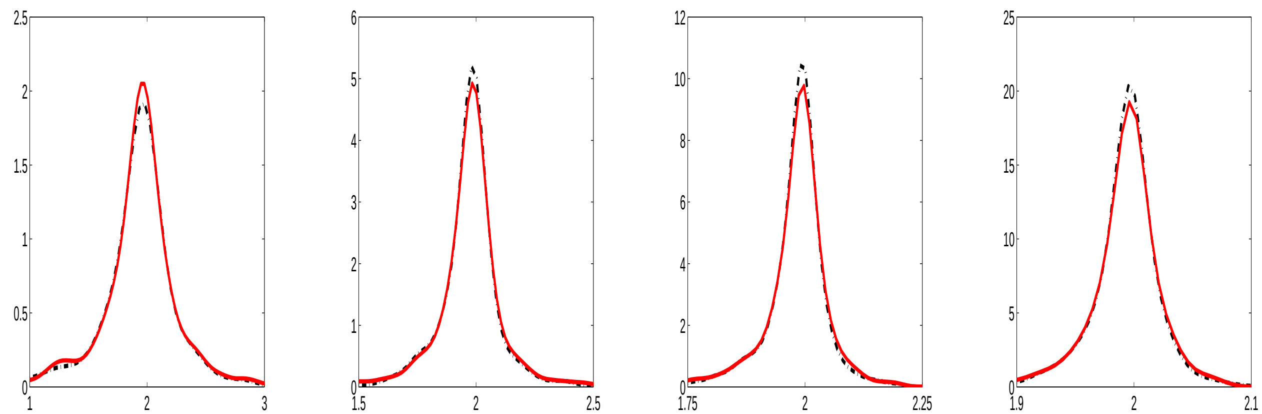
DGP: Model 1 - endogeneity only in the threshold variable, $\delta_2 = 2$

(a) $n = 100$ **(b)** $n = 250$ **(c)** $n = 500$ **(d)** $n = 1000$



DGP: Model 2 - endogeneity in both the threshold and slope variables, $\delta_3 = 2$

(e) $n = 100$ **(f)** $n = 250$ **(g)** $n = 500$ **(h)** $n = 1000$

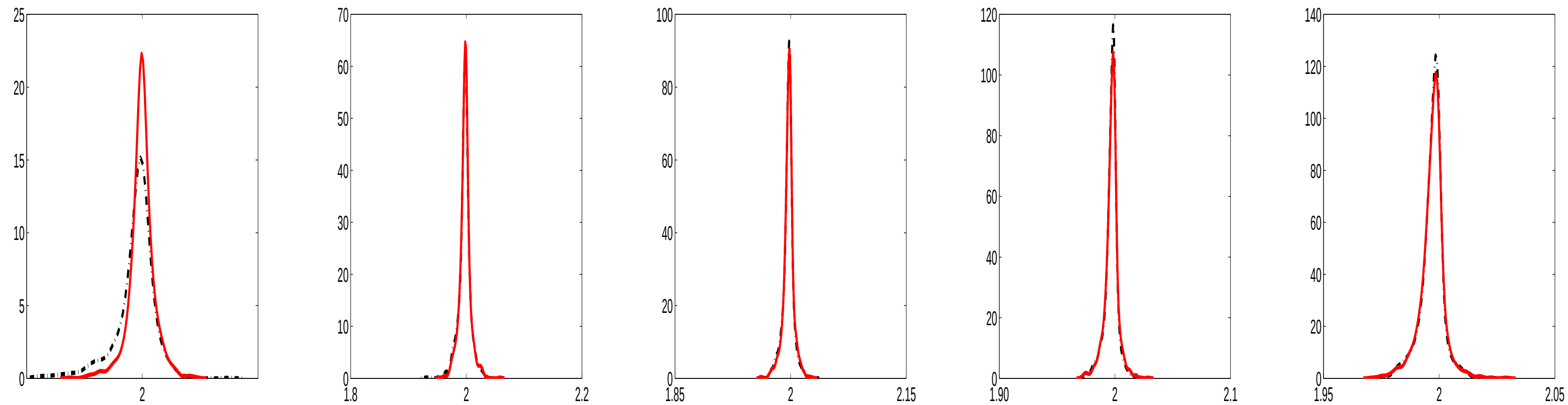


Note: The solid red line represents the MC kernel density of the STR threshold estimate while the black dashed line represents the corresponding densities for the TR of Hansen (2000) and IVTR of Caner and Hansen(2004).

Figure 2: MC kernel densities of the threshold estimate for different threshold effects

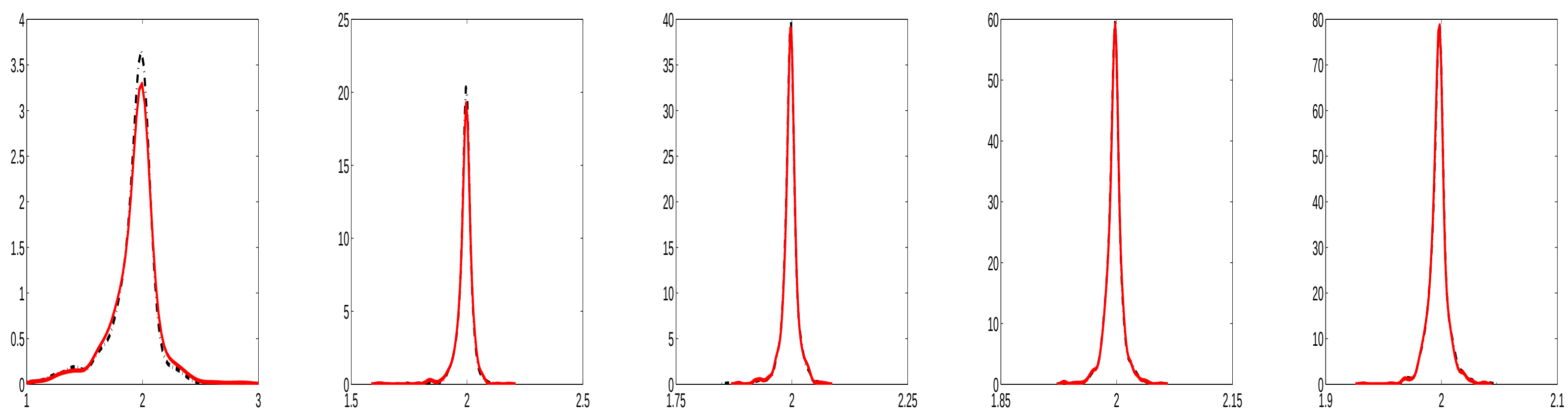
DGP: Model 1 - endogeneity only in the threshold variable, $n = 1000$

(a) $\delta_2 = 1$ **(b)** $\delta_2 = 2$ **(c)** $\delta_2 = 3$ **(d)** $\delta_2 = 4$ **(e)** $\delta_2 = 5$



DGP: Model 2 - endogeneity in both the threshold and slope variables, $n = 1000$

(f) $\delta_3 = 1$ **(g)** $\delta_3 = 2$ **(h)** $\delta_3 = 3$ **(i)** $\delta_3 = 4$ **(j)** $\delta_3 = 5$



Note: The solid red line represents the MC kernel density of the STR threshold estimate while the black dashed line represents the corresponding densities for the TR of Hansen (2000) and IVTR of Caner and Hansen(2004).

Table 1: Quantiles of the distribution of $\hat{\gamma}$

	DGP: Model 1 *endogeneity only in the threshold variable*						DGP: Model 2 *endogeneity in both the threshold and slope variables*					
	TR			STR			IVTR			STR		
Quantile / Sample size	5th	50th	95th	5th	50th	95th	5th	50th	95th	5th	50th	95th
	$\delta_2 = 1.00$						$\delta_3 = 1.00$					
100	1.321	1.936	2.328	1.607	1.983	2.340	0.373	1.886	3.146	1.081	1.914	2.770
250	1.657	1.973	2.162	1.800	1.993	2.148	0.862	1.933	2.603	1.192	1.960	2.600
500	1.810	1.986	2.084	1.906	1.997	2.092	1.217	1.953	2.311	1.408	1.962	2.413
1000	1.893	1.994	2.039	1.954	1.999	2.042	1.515	1.966	2.132	1.534	1.964	2.192
	$\delta_2 = 2.00$						$\delta_3 = 2.00$					
100	1.761	1.973	2.116	1.820	1.982	2.153	1.246	1.958	2.428	1.345	1.959	2.437
250	1.918	1.990	2.053	1.933	1.993	2.061	1.744	1.984	2.168	1.735	1.984	2.183
500	1.955	1.995	2.023	1.960	1.996	2.023	1.861	1.992	2.068	1.855	1.992	2.076
1000	1.978	1.998	2.013	1.982	1.998	2.014	1.938	1.997	2.033	1.933	1.996	2.038
	$\delta_2 = 3.00$						$\delta_3 = 3.00$					
100	1.851	1.976	2.065	1.861	1.978	2.073	1.682	1.974	2.17	1.686	1.974	2.194
250	1.944	1.992	2.031	1.947	1.992	2.032	1.88	1.988	2.078	1.874	1.988	2.078
500	1.972	1.995	2.013	1.972	1.996	2.016	1.936	1.994	2.034	1.935	1.994	2.041
1000	1.984	1.998	2.007	1.985	1.998	2.008	1.967	1.998	2.02	1.967	1.997	2.021
	$\delta_2 = 4.00$						$\delta_3 = 4.00$					
100	1.873	1.976	2.045	1.877	1.978	2.056	1.789	1.976	2.12	1.793	1.977	2.137
250	1.951	1.992	2.026	1.95	1.992	2.024	1.915	1.989	2.049	1.919	1.99	2.052
500	1.976	1.995	2.009	1.976	1.995	2.012	1.956	1.995	2.023	1.955	1.995	2.024
1000	1.986	1.998	2.004	1.987	1.998	2.006	1.979	1.998	2.011	1.98	1.998	2.013
	$\delta_2 = 5.00$						$\delta_3 = 5.00$					
100	1.879	1.976	2.036	1.887	1.977	2.039	1.822	1.977	2.092	1.823	1.977	2.105
250	1.955	1.992	2.018	1.955	1.992	2.017	1.934	1.99	2.039	1.934	1.99	2.041
500	1.977	1.995	2.007	1.977	1.995	2.008	1.965	1.996	2.017	1.964	1.996	2.017
1000	1.987	1.998	2.004	1.988	1.998	2.004	1.984	1.998	2.01	1.984	1.998	2.01

Table 2: Bootstrap confidence interval for γ for 90% nominal coverage

DGP: Model 1 - *endogeneity only in the threshold variable*

sample size	$\delta_2 = 1$	$\delta_2 = 2$	$\delta_2 = 3$	$\delta_2 = 4$	$\delta_2 = 5$
100	68	84	89	90	91
250	68	89	94	96	97
500	74	91	95	96	97
1000	72	89	94	96	98

DGP: Model 2 - *endogeneity in both the threshold and slope variables*

sample size	$\delta_3 = 1$	$\delta_3 = 2$	$\delta_3 = 3$	$\delta_3 = 4$	$\delta_3 = 5$
100	70	86	90	92	93
250	71	90	95	96	97
500	71	93	97	99	99
1000	71	95	98	99	99

Table 3: Quantiles of the distributions of slope coefficients

Panel A, DGP: Model 1 - *endogeneity only in the threshold variable*

	Quantiles of Slope Coefficient of the slope β_2						Quantiles of Slope Coefficient of the slope δ_3					
	TR			STR			TR			STR		
Quantile	5th	50th	95th	5th	50th	95th	5th	50th	95th	5th	50th	95th
Sample size												
						$\delta_3 = 1.00$						
100	0.539	0.765	1.074	0.752	1.009	1.342	0.573	0.936	1.282	0.642	0.953	1.225
250	0.623	0.760	0.935	0.847	1.006	1.189	0.753	0.973	1.159	0.803	0.980	1.146
500	0.659	0.751	0.866	0.887	1.000	1.129	0.836	0.978	1.114	0.877	0.983	1.100
1000	0.682	0.748	0.824	0.913	1.000	1.084	0.895	0.995	1.082	0.920	0.994	1.076
						$\delta_3 = 2.00$						
100	0.546	0.740	0.975	0.739	0.994	1.321	1.669	1.985	2.284	1.719	1.985	2.269
250	0.622	0.744	0.868	0.843	0.999	1.173	1.820	1.995	2.175	1.840	1.997	2.159
500	0.658	0.744	0.831	0.884	0.997	1.130	1.877	1.992	2.122	1.886	1.994	2.108
1000	0.681	0.744	0.809	0.913	0.998	1.083	1.918	2.002	2.090	1.927	1.999	2.080
						$\delta_3 = 3.00$						
100	0.546	0.734	0.962	0.739	0.989	1.316	2.697	2.992	3.289	2.734	3.000	3.271
250	0.622	0.744	0.865	0.843	0.999	1.173	2.828	3.000	3.175	2.846	3.001	3.160
500	0.658	0.744	0.829	0.886	0.997	1.128	2.880	2.993	3.123	2.890	2.995	3.109
1000	0.681	0.744	0.808	0.914	0.998	1.082	2.918	3.003	3.089	2.930	3.000	3.081
						$\delta_3 = 4.00$						
100	0.544	0.734	0.961	0.736	0.989	1.315	3.701	3.994	4.289	3.742	4.002	4.273
250	0.622	0.744	0.865	0.843	0.999	1.176	3.828	4.001	4.175	3.850	4.004	4.160
500	0.658	0.744	0.829	0.886	0.997	1.123	3.880	3.993	4.123	3.889	3.996	4.109
1000	0.681	0.744	0.809	0.914	0.998	1.082	3.918	4.003	4.090	3.931	4.000	4.081
						$\delta_3 = 5.00$						
100	0.543	0.734	0.958	0.737	0.988	1.315	4.707	4.991	5.289	4.743	5.004	5.273
250	0.623	0.745	0.866	0.842	0.999	1.176	4.828	5.001	5.175	4.850	5.004	5.160
500	0.658	0.743	0.829	0.886	0.997	1.123	4.880	4.994	5.123	4.892	4.997	5.109
1000	0.681	0.743	0.809	0.914	0.998	1.082	4.918	5.003	5.090	4.931	5.001	5.081

Table continued on next page ...

Table 3 continued

Panel B, DGP: Model 2 - *endogeneity in both the threshold and slope variables*

	Quantiles of Slope Coefficient of the slope β_2						Quantiles of Slope Coefficient of the slope δ_2					
	TR			STR			TR			STR		
Quantile	5th	50th	95th	5th	50th	95th	5th	50th	95th	5th	50th	95th
Sample size												
						$\delta_3 = 1.00$						
100	0.452	0.860	1.350	0.702	1.067	1.449	0.161	0.912	1.393	0.428	0.912	1.318
250	0.630	0.840	1.204	0.827	1.053	1.294	0.495	0.941	1.204	0.609	0.935	1.190
500	0.689	0.840	1.112	0.875	1.036	1.235	0.667	0.955	1.131	0.737	0.955	1.119
1000	0.741	0.829	1.016	0.920	1.025	1.183	0.751	0.967	1.094	0.792	0.966	1.086
						$\delta_3 = 2.00$						
100	0.535	0.833	1.297	0.707	1.033	1.453	1.440	1.958	2.354	1.469	1.955	2.335
250	0.655	0.822	1.036	0.822	1.013	1.242	1.698	1.988	2.211	1.723	1.983	2.198
500	0.697	0.820	0.948	0.875	1.014	1.159	1.835	1.988	2.144	1.835	1.987	2.134
1000	0.740	0.813	0.897	0.914	1.004	1.101	1.879	1.995	2.098	1.884	1.993	2.093
						$\delta_3 = 3.00$						
100	0.561	0.831	1.142	0.706	1.011	1.382	2.583	2.972	3.363	2.585	2.982	3.340
250	0.662	0.818	0.992	0.823	1.007	1.218	2.745	3.001	3.211	2.751	2.996	3.205
500	0.702	0.816	0.924	0.867	1.007	1.143	2.849	2.996	3.146	2.855	2.991	3.142
1000	0.740	0.811	0.887	0.914	1.000	1.095	2.887	2.996	3.099	2.894	2.998	3.098
						$\delta_3 = 4.00$						
100	0.566	0.823	1.078	0.710	1.000	1.347	3.628	3.987	4.363	3.634	3.994	4.339
250	0.662	0.817	0.974	0.824	1.004	1.198	3.763	4.001	4.211	3.774	4.001	4.205
500	0.699	0.815	0.919	0.868	1.004	1.140	3.852	3.999	4.149	3.857	3.996	4.145
1000	0.740	0.810	0.882	0.914	0.999	1.093	3.891	3.999	4.100	3.895	3.999	4.098
						$\delta_3 = 5.00$						
100	0.570	0.821	1.068	0.712	0.995	1.322	4.639	4.991	5.362	4.651	4.998	5.336
250	0.663	0.817	0.964	0.819	1.002	1.188	4.772	5.004	5.215	4.776	5.003	5.209
500	0.699	0.814	0.916	0.866	1.004	1.137	4.854	5.000	5.148	4.863	4.998	5.146
1000	0.741	0.810	0.882	0.911	1.001	1.088	4.898	5.002	5.101	4.901	5.002	5.098

Table 4: Quantiles of the coefficient of the inverse Mills ratio

DGP: Model 1 - *endogeneity only in the threshold variable*

Quantile	5th	50th	95th	5th	50th	95th	5th	50th	95th	5th	50th	95th	5th	50th	95th
Sample size															
	$\delta_2 = 1.00$			$\delta_2 = 2.00$			$\delta_2 = 3.00$			$\delta_2 = 4.00$			$\delta_2 = 5.00$		
100	0.457	0.936	1.515	0.514	0.960	1.53	0.516	0.965	1.53	0.522	0.966	1.524	0.532	0.967	1.528
250	0.635	0.941	1.276	0.665	0.955	1.295	0.665	0.955	1.297	0.672	0.955	1.297	0.672	0.958	1.311
500	0.736	0.940	1.183	0.744	0.949	1.193	0.743	0.951	1.192	0.743	0.951	1.191	0.743	0.95	1.19
1000	0.799	0.950	1.11	0.804	0.952	1.112	0.801	0.953	1.113	0.801	0.953	1.113	0.803	0.953	1.112

DGP: Model 2 - *endogeneity in both the threshold and slope variables*

Quantile	5th	50th	95th	5th	50th	95th	5th	50th	95th	5th	50th	95th	5th	50th	95th
Sample size															
	$\delta_3 = 1.00$			$\delta_2 = 3.00$			$\delta_2 = 3.00$			$\delta_3 = 4.00$			$\delta_3 = 5.00$		
	5th	50th	95th	5th	50th	95th	5th	50th	95th	5th	50th	95th	5th	50th	95th
100	0.201	0.784	1.442	0.133	0.751	1.415	0.164	0.730	1.364	0.173	0.723	1.333	0.186	0.718	1.315
250	0.42	0.739	1.126	0.392	0.713	1.064	0.377	0.700	1.051	0.388	0.695	1.038	0.394	0.696	1.053
500	0.504	0.734	0.999	0.473	0.714	0.968	0.469	0.710	0.951	0.468	0.705	0.952	0.469	0.704	0.945
1000	0.557	0.716	0.897	0.549	0.704	0.883	0.549	0.702	0.884	0.547	0.702	0.879	0.541	0.703	0.862

Table 5: Nominal 95% confidence interval coverage for the threshold effect parameter

DGP: Model 1 - *endogeneity only in the threshold variable*
Nominal 95% confidence interval coverage for δ_2

Sample Size	$\delta_2 = 1$	$\delta_2 = 2$	$\delta_2 = 3$	$\delta_2 = 4$	$\delta_2 = 5$
100	88	92	93	93	93
250	92	94	94	94	95
500	95	96	96	96	96
1000	95	95	95	95	95

DGP: Model 2 - *endogeneity in both the threshold and slope variables*
Nominal 95% confidence interval coverage for δ_3

Sample Size	$\delta_3 = 1$	$\delta_3 = 2$	$\delta_3 = 3$	$\delta_3 = 4$	$\delta_3 = 5$
100	81	86	90	92	92
250	82	90	92	93	93
500	82	93	94	94	94
1000	82	93	94	94	95