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
Tufts University§

November 12, 2009

Abstract

This paper extends the simple threshold regression framework of Hansen (2000) and Caner and Hansen (2004) to allow for endogeneity of the threshold variable. We develop a concentrated least squares estimator of the threshold parameter based on an inverse Mills ratio bias correction. We show that our estimator is consistent and investigate its performance 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 participants of ++++
†P.O. Box 537, CY 1678 Nicosia, Cyprus, e-mail: andros@ucy.ac.cy.
‡Guelph, Ontario N1G 2W1, Canada, email: tstengos@uoguelph.ca
§8 Upper Campus Road, Medford, MA 02155, email: chihming.tan@tufts.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.

Sample splitting and threshold regression models were studied by Hansen (2000) who proposed a concentrated least squares approach for estimating the sample split value. Caner and Hansen (2004) extended the Hansen (2000) framework to the case of endogeneity in the slope variables. Seo and Linton (2005) allow the threshold variable to be a linear index of observed variables and propose a smoothed least squares estimation strategy based on smoothing the objective function in the sense of Horowitz's smoothed maximum scored estimator.

In all these studies a crucial assumption 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, in the empirical growth context, one could posit that countries are organized into different growth processes depending on whether their quality of institutions is above a threshold value. But, as Acemoglu, Johnson, and Robinson (2001) have argued, quality of institutions is very likely an endogenous variable.

In this paper we introduce the Structural Threshold Regression (STR) model and propose an estimation strategy that extends Hansen (2000) and Caner and Hansen (2004) to the case where the threshold variable is endogenous. In particular, we propose a concentrated least squares estimator of the threshold parameter when the threshold variable is endogenous and based on the sample split implied by the threshold estimate, we estimate the slope parameters by 2SLS or GMM. Using a similar set of assumptions as in Hansen (2000) and Caner and Hansen (2004) we show that our estimators are consistent. To examine the finite sample properties of our estimators we provide a Monte Carlo analysis.

The main strategy in this paper is to exploit the intuition 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 (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.

Just as in the limited dependent variable framework, we show that consistent estimation of slope parameters under Normality requires the inclusion of a set of inverse Mills ratio bias correction terms. It also becomes clear that the slope parameter estimates of the threshold regression by Hansen (2000) and Caner and Hansen (2004) will be inconsistent in the endogenous threshold variable case because both strategies omit the inverse Mills ratio bias correction terms. Our Monte Carlo results confirm the above insight. While all three approaches perform similarly in terms of estimating the threshold variable, unlike STR, for both Hansen (2000) and Caner and Hansen (2004), the distribution of the slope estimate fails to center upon the true slope parameter when the threshold variable is endogenous.

In terms of inference, Chan (1993) showed that the asymptotic distribution of the threshold estimate is a functional of a compound Poisson process. This distribution is too complicated for inference as it depends on nuisance parameters. Hansen (2000) developed a more useful asymptotic distribution theory for both the threshold parameter estimate and the regression slope coefficients under the assumption that the threshold effect becomes smaller as the sample increases. Using a similar set of assumptions, Gonzalo and Wolf (2005) proposed subsampling to conduct inference in the context of threshold autoregressive models. Seo and Linton (2005) show that their estimator exhibits asymptotic normality but it depends on the choice of bandwidth. In the STR context, under the assumption of the diminishing threshold effect and non-regime specific heteroskedasticity, we derive an asymptotic distribution for the threshold estimate, which is similar to Caner and Hansen (2004) and propose bootstrap confidence intervals for the threshold estimator.

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

2 The Structural Threshold Regression (STR)

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 (STR),

$$y_i = \mathbf{x}_i'\beta_1 + u_i, \quad q_i \leq \gamma \tag{2.1}$$

$$y_i = \mathbf{x}_i'\beta_2 + u_i, \quad q_i > \gamma \tag{2.2}$$

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 selection equation that determines which regime applies is given by

$$q_i = \mathbf{z}_i'\pi_q + v_{q,i} \tag{2.3}$$

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

STR is similar in nature to the case of the error interdependence that exists in limited dependent variable models between the equation of interest and the sample selection equation, see Heckman (1979). However, in sample selection and endogenous dummy variable models, we observe the assignment of observations to regimes. However, the variable that is responsible for this assignment is latent. In the STR case, we have the opposite problem. Here, we do not know which observations belong to which regime, but we can observe the assignment (threshold) variable. To put it differently, while limited dependent variable models treat q_i as unobserved and the sample split as observed (e.g., via the known dummy variable), here we treat the sample split value as unknown and we estimate it.

Let us consider the following partition $\mathbf{x}_i = (\mathbf{x}_{1,i}, \mathbf{x}_{2,i})$ where $\mathbf{x}_{1,i}$ are endogenous and $\mathbf{x}_{2i}$ are exogenous and the $l \times 1$ vector of instrumental variables $\mathbf{z}_i = (\mathbf{z}_{1,i}, \mathbf{z}_{2,i})$ where $\mathbf{x}_{2,i} \in \mathbf{z}_i$. If both q_i and $\mathbf{x}_i$ are exogenous then we get the threshold model studied by Hansen (2000). If q_i and $\mathbf{x}_{2,i}$ are exogenous and $\mathbf{x}_{1i}$ is not a null set, then we get the threshold model studied by Caner and Hansen (2004). If $v_{q,i} = 0$ 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 $\mathbf{x}_{1,i}$ is not a null set[1].

By defining the indicator function

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

[1]Note that we exclude the special case of a continuous threshold model; see Hansen (2000) and Chan and Tsay (1998)

and $I(q_i > \gamma) = 1 - I(q_i \leq \gamma)$, we can rewrite the structural model (1)-(2) as

$$y_i = \boldsymbol{\beta}_1' \mathbf{x}_i I(q_i \leq \gamma) + \boldsymbol{\beta}_2' \mathbf{x}_i I(q_i > \gamma) + u_i \tag{2.5}$$

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

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

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

such that

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

Since the reduced form model (2.6) does not depend on the threshold q_i, we have the following conditional expectations

$$E(y_i|\mathbf{z}_i, q_i \leq \gamma) = \boldsymbol{\beta}_1' \mathbf{g}_i + E(u_i|\mathbf{z}_i, q_i \leq \gamma) \tag{2.9}$$

$$E(y_i|\mathbf{z}_i, q_i > \gamma) = \boldsymbol{\beta}_2' \mathbf{g}_i + E(u_i|\mathbf{z}_i, q_i > \gamma) \tag{2.10}$$

Define further $\kappa = \sigma_{uv} = \rho\sigma_u$[3]. Then by the properties of the joint normal distribution we obtain

$$E(u_i|\mathbf{z}_i, q_i \leq \gamma) = \kappa E(v_{q,i}|\mathbf{z}_i, q_i \leq \gamma) = \kappa\lambda_1\left(\gamma - \mathbf{z}_i'\boldsymbol{\pi}_q\right) \tag{2.11}$$

$$E(u_i|\mathbf{z}_i, q_i > \gamma) = \kappa E(v_{q,i}|\mathbf{z}_i, q_i > \gamma) = \kappa\lambda_2\left(\gamma - \mathbf{z}_i'\boldsymbol{\pi}_q\right) \tag{2.12}$$

where $\lambda_1(\gamma - z_i'\pi_q) = -\frac{\phi(\gamma - \mathbf{z}_i'\pi_q)}{\Phi(\gamma - \mathbf{z}_i'\pi_q)}$ and $\lambda_2(\gamma - z_i'\pi_q) = \frac{\phi(\gamma - \mathbf{z}_i'\pi_q)}{1 - \Phi(\gamma - \mathbf{z}_i'\pi_q)}$ are the inverse Mills ratio bias correction terms and $\phi(\cdot)$ and $\Phi(\cdot)$ are the normal pdf and cdf, respectively[4].

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

[3] For simplicity we assume that the covariance between the $v_{q,i}$ and u_i is the same across both regimes. Our model can easily be extended to the case of different degrees of endogeneity across regimes.

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

Therefore, using (2.9), (2.10), (2.11), (2.12) we can define the STR model as follows

$$y_i = \boldsymbol{\beta}_1' \mathbf{g}_i + \kappa\lambda_1 \left(\gamma - \mathbf{z}_i'\boldsymbol{\pi}_q\right) + \varepsilon_{1,i}, \quad q_i \leq \gamma \tag{2.13}$$

$$y_i = \boldsymbol{\beta}_2' \mathbf{g}_i + \kappa\lambda_2 \left(\gamma - \mathbf{z}_i'\boldsymbol{\pi}_q\right) + \varepsilon_{2,i}, \quad q_i > \gamma \tag{2.14}$$

where

$$E(\varepsilon_{1,i}|\mathbf{z}_i, q_i \leq \gamma) = 0 \tag{2.15}$$

$$E(\varepsilon_{2,i}|\mathbf{z}_i, q_i > \gamma) = 0 \tag{2.16}$$

We can also rewrite (2.13)-(2.14) as a single equation

$$y_i = \left(\boldsymbol{\beta}_1' \mathbf{g}_i + \kappa\lambda_1 \left(\gamma - \mathbf{z}_i'\boldsymbol{\pi}_q\right)\right) I(q_i \leq \gamma) + \left(\boldsymbol{\beta}_2' \mathbf{g}_i + \kappa\lambda_2 \left(\gamma - \mathbf{z}_i'\boldsymbol{\pi}_q\right)\right) I(q_i > \gamma) + \varepsilon_i \tag{2.17}$$

where the error ε_i is given by

$$\varepsilon_i = \left(\boldsymbol{\beta}_1' \mathbf{v}_i - \kappa\lambda_1 \left(\gamma - \mathbf{z}_i'\boldsymbol{\pi}_q\right)\right) I(q_i \leq \gamma) + \left(\boldsymbol{\beta}_2' \mathbf{v}_i - \kappa\lambda_2 \left(\gamma - \mathbf{z}_i'\boldsymbol{\pi}_q\right)\right) I(q_i > \gamma) + u_i \tag{2.18}$$

Notice that when the threshold variable q_i is exogenous, i.e. $\kappa = 0$, (2.17) becomes the threshold regression model of Caner and Hansen (2004)

$$y_i = \boldsymbol{\beta}_1' \mathbf{g}_i I(q_i \leq \gamma) + \boldsymbol{\beta}_2' \mathbf{g}_i I(q_i > \gamma) + \varepsilon_i \tag{2.19}$$

Additionally, when $\mathbf{x}_i$ is also exogenous then we get the threshold regression model of Hansen (2000). In both cases, the inverse Mills ratio bias correction terms are omitted so that naively estimating the STR model using Hansen (2000) or Caner and Hansen (2004) would result in inconsistent estimates of the slope parameters $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$.

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

2.1 Estimation

Define

$$\lambda_{1,i}(\gamma) \equiv \lambda_1 \left(\gamma - \mathbf{z}_i'\boldsymbol{\pi}_q\right) \tag{2.20}$$

$$\lambda_{2,i}(\gamma) \equiv \lambda_2 \left(\gamma - \mathbf{z}_i'\boldsymbol{\pi}_q\right) \tag{2.21}$$

and

$$\lambda_i(\gamma) = \lambda_{1,i}(\gamma)I(q_i \leq \gamma) + \lambda_{2,i}(\gamma)I(q_i > \gamma) \tag{2.22}$$

Define $\mathbf{g}_{i,\gamma} = \mathbf{g}_i I(q_i \leq \gamma)$ and $\boldsymbol{\beta}_1 = \boldsymbol{\delta}_n + \boldsymbol{\beta}_2$, then we can rewrite equation (2.17) in terms of the lower regime (Regime 1)

$$y_i = \mathbf{g}_i'\beta + \mathbf{g}_{i,\gamma}'\delta_n + \kappa_n\lambda_i(\gamma) + \varepsilon_i \tag{2.23}$$

We estimate the parameters of (2.23) in three steps. First, we estimate the reduced form parameter $\mathbf{\Pi}$ in (2.6) by LS. Given a LS estimator $\widehat{\mathbf{\Pi}}$, let us denote the fitted values for $\mathbf{x}_i$ as $\widehat{\mathbf{x}}_i = \widehat{\mathbf{g}}_i = \widehat{\mathbf{\Pi}}'\mathbf{z}_i$ and define the first stage residuals as $\widehat{\mathbf{v}}_i = \mathbf{g}_i - \widehat{\mathbf{g}}_i$.

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

$$\widehat{\gamma} = \arg\min_{\gamma} S_n(\gamma) \tag{2.24}$$

where

$$S_n(\gamma) = \sum_{i=1}^{n}(y_i - \widehat{\mathbf{g}}_i'\boldsymbol{\beta} - \widehat{\mathbf{g}}_{i,\gamma}'\boldsymbol{\delta}_n - \kappa_n\widehat{\lambda}_i(\gamma))^2 \tag{2.25}$$

where $\widehat{\mathbf{g}}_{i,\gamma} = \widehat{\mathbf{g}}_i I(q_i \leq \gamma)$, $\widehat{\lambda}_i(\gamma) = \widehat{\lambda}_{1,i}(\gamma) + \widehat{\lambda}_{2,i}(\gamma)$,with $\widehat{\lambda}_{1,i}(\gamma) = \lambda_1\left(\gamma - \mathbf{z}_i'\widehat{\boldsymbol{\pi}}_q\right)$ and $\widehat{\lambda}_{2,i}(\gamma) = \lambda_2\left(\gamma - \mathbf{z}_i'\widehat{\boldsymbol{\pi}}_q\right)$.

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

3 Inference

Let $\mathbf{g}_i(\gamma) = (\mathbf{g}_i, \lambda_i(\gamma))'$ and $\widehat{\mathbf{x}}_i(\gamma) = (\widehat{\mathbf{x}}_i', \widehat{\lambda}_i(\gamma))'$. Then define the moment functionals

$$\mathbf{M}(\gamma) = E\left(\mathbf{g}_i(\gamma)\mathbf{g}_i(\gamma)'I(q_i \leq \gamma)\right) \tag{3.26}$$

$$\mathbf{M}^{\perp}(\gamma) = E(\mathbf{g}_i(\gamma)\mathbf{g}_i(\gamma)'I(q_i > \gamma)) \tag{3.27}$$

$$\mathbf{D}_1(\gamma) = E\left(\mathbf{g}_i(\gamma)\mathbf{g}_i(\gamma)'|q_i = \gamma\right) \tag{3.28}$$

$$\mathbf{D}_2(\gamma) = E\left(\mathbf{g}_i(\gamma)\mathbf{g}_i(\gamma)'\varepsilon_i^2|q_i = \gamma\right) \tag{3.29}$$

Note that $f_q(q)$ denotes the density function of q, γ_0 denotes the true value of γ, $\mathbf{D}_1 = \mathbf{D}_1(\gamma_0)$, $\mathbf{D}_2 = \mathbf{D}_2(\gamma_0)$, $f_q = f_q(\gamma_0)$, and $\mathbf{M} = E(\mathbf{g}_i\mathbf{g}_i')$.

Assumption 1

1.1 $\{\mathbf{z}_i, \mathbf{g}_i, u_i, \mathbf{v}_i\}$ 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 $E(v_i|\mathcal{F}_{i-1}) = 0$,

1.4 $E|g_i|^4 < \infty$ and $E|\mathbf{g}_i\varepsilon_i|^4 < \infty$,

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

1.6 for all $\gamma \in \Gamma$, $0 < f_q(\gamma) \leq \overline{f} < \infty$

1.7 $\mathbf{D}_1(\gamma)$, $\mathbf{D}_2(\gamma)$, and $f_q(\gamma)$, is continuous at $\gamma = \gamma_0$

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

1.9 $f_q > 0$, $\mathbf{c}'\mathbf{D}_1\mathbf{c} > 0$, $\mathbf{c}'\mathbf{D}_2\mathbf{c} > 0$, where $\mathbf{c} = (\mathbf{c}_\delta, c_\kappa)$.

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

1.11 Let $\mathbf{H}_i = \{\mathbf{g}_i, \widehat{\lambda}_i(\gamma), \varepsilon_i, \widehat{\mathbf{v}}_i\}$ and $a_n = n^{1-2a}$, $\sup_{\gamma\in\Gamma} |\frac{1}{\sqrt{n}} \sum_i \mathbf{H}_i \widehat{\mathbf{v}}_i' I(q_i \leq \gamma)| = O_p(1)$

1.12 There exists a $0 < B < \infty$ such that for all $\epsilon > 0$ and $\zeta > 0$ there is a $\overline{\varepsilon} < \infty$ and $\overline{n} < \infty$ such that for all $n \geq \overline{n}$

$$P\Big(\sup_{\overline{\varepsilon}/a_n \leq |\gamma-\gamma_0| \leq B} \Big| \frac{\sum_i \mathbf{H}_i \widehat{\mathbf{v}}_i I_{\{q_i \leq \gamma\}} - I_{\{q_i \leq \gamma_0\}}}{n^{1-\alpha}|\gamma - \gamma_0|} \Big| > \zeta\Big) < \epsilon \tag{3.30}$$

and

$$\sup_{|\nu|\leq\overline{\varepsilon}} n^{-\alpha} \Big| \sum_i \mathbf{H}_i \widehat{\mathbf{v}}_i I_{\{q_i \leq \gamma+\nu/a_n\}} - I_{\{q_i \leq \gamma_0\}} \Big| \to 0 \tag{3.31}$$

This set of assumptions is similar to Hansen (2000) and Caner and Hansen (2004). While most assumptions are rather standard, Assumption 1.8 is not. Assumption 1.8 assumes a "small threshold" asymptotic framework in the sense that $\boldsymbol{\delta}_n$ will tend to zero rather slowly as $n \to \infty$. Under this assumption Hansen (2000) showed that the threshold estimate has an asymptotic distribution free of nuisance parameters. Assumption 1.7 excludes the case of regime-dependent heteroskedasticity and hence $E(\varepsilon_i^2|q_i = \gamma)$ is continuous at γ_0. [5] As in Caner and Hansen (2004), Assumption 1.11 is needed to ensure that the reduced form fitted values are consistent for the true reduced form conditional mean given in (2.6). Assumption 1.12 is necessary for inference.

Theorem 1: Consistency

Under Assumption 1, the estimator for γ obtained by minimizing the CLS criterion (2.25), $\widehat{\gamma}$, is consistent. That is,

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

The proof is given in the appendix.

To obtain the asymptotic distribution let us first define $\omega = \frac{\mathbf{c'D_2c}}{(\mathbf{c'D_1c})^2 f}, \eta^2 = \frac{\mathbf{c'D_2c}}{(\mathbf{c'D_1c})^2 \sigma_\varepsilon^2}$, where $\sigma_\varepsilon^2 = E(\varepsilon_i^2)$, and $W(\nu)$ denote a two-sided Brownian motion on the real line. Let

$$T = \underset{-\infty<v<\infty}{\arg\max} \left(-\frac{1}{2}|\nu| + W(\nu) \right)$$

and

$$\xi = \sup_{-\infty<v<\infty} \left(-|\nu| + 2W(\nu) \right). \tag{3.32}$$

Furthermore, we define the likelihood ratio statistic for $H_0 : \gamma = \gamma_0$ as follows

$$LR_n(\gamma) = n\frac{S_n(\gamma) - S_n(\widehat{\gamma})}{S_n(\widehat{\gamma})} \tag{3.33}$$

Theorem 2: Asymptotic Distribution of $\widehat{\gamma}$

[5] An alternative assumption is to specify regime dependent heteroskedasticity. However, this potentially involves the argmax probability density with unequal drifts and scalings, which has been investigated by Stryhn (1996). A route for deriving such a distribution within Hansen's (2000) framework would be as follows. Under the assumption of equal scaling in both regimes, where the scale of the lower regime is also assumed to hold true for the upper regime, one can derive the argmax distribution T_1 as $T_1 = \arg\max_\nu \left[\xi_1(-\frac{|\nu|}{2} + W(\nu)) \right]$. Of course the above distribution would only be correct for the lower regime $I(q \leq \gamma_0)$. Similarly, reversing the definition of regimes one can obtain the argmax distribution that applies to the upper regime T_2 as $T_2 = \arg\max_\nu \left[\xi_2(-\frac{|\nu|}{2} + W(\nu)) \right]$. $T = \arg\max_\nu \left(\xi_1(-\frac{|\nu|}{2} + W(\nu))I(q < \gamma_0) + \xi_2(-\frac{|\nu|}{2} + W(\nu))I(q > \gamma_0) \right)$,where $\xi_1 = \frac{c' D_2 c}{(c' D_1 c)^2 f}$ and $\xi_2 = \frac{c' D_2^\perp c}{(c' D_1^\perp c)^2 f}$, and $\xi = \sup_v (\xi_1(-\mid \nu \mid + 2W(\nu))I(q < \gamma_0) + \xi_2(-\mid \nu \mid + 2W(\nu))I(q > \gamma_0))$. D_j and $D_j^\perp, j = 1, 2$, are the moment functionals corresponding to the upper and lower regimes, respectively. A similar suggestion was made by Seo and Linton (2007).

Under Assumption 1, the argmax distribution of $\widehat{\gamma}$ for the STR model would take the form

$$n^{1-2\alpha}(\widehat{\gamma}-\gamma_0) \xrightarrow{d} T \tag{3.34}$$

$$LR_n(\gamma) \xrightarrow{d} \eta^2\xi \tag{3.35}$$

We can then employ the test-inversion method of Hansen (2000) to construct an asymptotic confidence interval for γ_0. To do so, first, let μ be the 95th percentile of the distribution of ξ. Then, $\widehat{\Gamma}$ is an asymptotically valid 95% confidence region for γ_0, and is given by

$$\widehat{\Gamma} = \left\{\gamma : LR_n(\gamma) \leq \widehat{\eta}^2\mu\right\} \tag{3.36}$$

where $\widehat{\eta}^2$ is an estimate of η^2 based on a second-order polynomial expansion of the threshold variable q_i or a kernel regression; see Hansen (2000). The proof is given in the appendix.

4 The Bootstrap

The bootstrap for the threshold regression model has been studied by Antoch et al (1995) and Yu (2009). In particular, Antoch et al established the validity of the nonparametric bootstrap under the assumptions of an asymptotically diminishing threshold $i.i.d$ errors. Note that once the inverse Mills bias correction terms are included the conditional mean zero assumption of the error is restored and therefore the STR model can be reduced to the model studied in Antoch et al (1995). Therefore, appealing to Theorem 3 of Antoch et al (1995) shows that the bootstrap approximation will converge to the asymptotic distribution given in (3.34).

Given consistent estimates for $(\widehat{\beta}_1, \widehat{\beta}_2, \widehat{\kappa}, \widehat{\boldsymbol{\pi}}_q, \widehat{\boldsymbol{\Pi}}, \widehat{\gamma})$, and fitted values $\widehat{\mathbf{g}}_i$, we define the residuals

$$\widehat{\varepsilon}_i = y_i - \left(\widehat{\boldsymbol{\beta}}_1'\widehat{\mathbf{g}}_i + \widehat{\kappa}\lambda_1\left(\widehat{\gamma}-\mathbf{z}_i'\widehat{\boldsymbol{\pi}}_q\right)\right) I(q_i \leq \widehat{\gamma}) + \left(\widehat{\boldsymbol{\beta}}_2'\widehat{\mathbf{g}}_i + \widehat{\kappa}\lambda_2\left(\widehat{\gamma}-\mathbf{z}_i'\widehat{\boldsymbol{\pi}}_q\right)\right) I(q_i > \widehat{\gamma})$$

$$\widehat{v}_{q,i} = q_i - \mathbf{z}_i'\widehat{\pi}_q \tag{4.37}$$

$$\widehat{\mathbf{v}}_i = \mathbf{x}_i - \widehat{\boldsymbol{\Pi}}'\mathbf{z}_i \tag{4.38}$$

for $i = 1, 2.., n$. These residuals are recentered to get $(\widetilde{\varepsilon}_i, \widetilde{v}_{q,i}, \widetilde{\mathbf{v}}_i)$. Then using the EDF of $(\widetilde{\varepsilon}_i, \widetilde{v}_{q,i}, \widetilde{\mathbf{v}}_i)$ and keeping $\mathbf{z}_i$ is fixed, $\mathbf{z}_i^* = \mathbf{z}_i$, we draw the bootstrap sample of $(\mathbf{z}_j^*, \varepsilon_j^*, v_{q,j}^*, \mathbf{v}_j^*)$ and obtain

$(y_j^*, \mathbf{x}_j^*, q_j^*)$ using equations (2.3), (2.6), and (2.5).

To construct bootstrap confidence intervals for γ we follow the test-inversion method of Hansen (2000) and then obtain the bootstrap distribution of the likelihood ratio statistic using the bootstrap estimates $(\widehat{\beta}_1^*, \widehat{\beta}_2^*, \widehat{\kappa}^*, \widehat{\boldsymbol{\pi}}_q^*, \widehat{\boldsymbol{\Pi}}^*, \widehat{\gamma}^*)$

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

to construct the bootstrap confidence region for γ_0 , $\widehat{\Gamma} = \{\gamma : LR_n^*(\gamma) \leq LR_n(\gamma)\}$.

5 Monte Carlo

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

The Monte Carlo design is based on the following threshold regression

$$y_i = \begin{cases} \beta_{1,1} + \beta_{1,2} x_i + u_i, & q_i \leq 2 \\ \beta_{2,1} + \beta_{2,2} x_i + u_i, & q_i > 2 \end{cases} \tag{5.39}$$

where

$$q_i = 2 + z_{1,i} + v_{q,i} \tag{5.40}$$

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

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

the case of Heckman's estimator, our estimator becomes more efficient as this correlation decreases and the degree of multicollinearity between $\Pi' z$ and x is small.

First, we consider the estimation of the threshold value γ. Table 1 presents the 5th, 50th, and 95th quantiles for the distribution of the threshold estimate $\hat{\gamma}$ under STR, TR, and IVTR. Specifically, columns (1)-(6) of Table 1 consider the case where the threshold variable is endogenous but the slope variable is exogenous and compare the distribution of the TR estimates with those of STR. Columns (7)-(12) of Table 1 consider the case where both the threshold variable and slope variable are endogenous and compare the distribution of the IVTR estimates with those of STR.

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

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

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

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

6 Conclusion

In this paper we propose an extension of Hansen (2000) and Caner and Hansen (2004) that deals with the endogeneity of the threshold variable. We developed a concentrated least squares estimator that deals with the problem of endogeneity in the threshold variable by generating a correction term based on the inverse Mills ratios to produce consistent estimates for the threshold parameter and the slope coefficients. Our proposed estimator performs well for a variety of sample sizes and parameter combinations.

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, p. 1369-1401.

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

[4] Chan, K, . S. (1993), "Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model," *The Annals of Statistics*, 21, p. 520-533.

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

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

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

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

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

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

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

[12] Yu P. (2009), "Bootstrap in Threshold Regression," University of Auckland, mimeo.

A Preliminaries

Define for any γ the following $(p+1)\times 1$ vectors, $\widehat{\mathbf{x}}_i(\gamma) = (\widehat{\mathbf{x}}_i', \widehat{\lambda}_i(\gamma))'$, where $\widehat{\lambda}_i(\gamma) = \widehat{\lambda}_{1,i}(\gamma)\, I(q_i \leq \gamma) + \widehat{\lambda}_{2,i}(\gamma)\, I(q_i > \gamma)$. Let $\widehat{\mathbf{X}}_\gamma$ and $\widehat{\mathbf{X}}_\perp$ be the orthogonal stacked vectors of $\widehat{\mathbf{x}}_i(\gamma)I(q_i \leq \gamma)$ and $\widehat{\mathbf{x}}_i(\gamma)I(q_i > \gamma)$, respectively.

Consider the following projections spanned by the columns of $\widehat{\mathbf{X}}_\gamma$ and $\widehat{\mathbf{X}}_\perp$, respectively.

$$\mathbf{P}_\gamma = \widehat{\mathbf{X}}_\gamma(\widehat{\mathbf{X}}_\gamma'\widehat{\mathbf{X}}_\gamma)^{-1}\widehat{\mathbf{X}}_\gamma' \tag{A.1}$$

$$\mathbf{P}_\perp = \widehat{\mathbf{X}}_\perp(\widehat{\mathbf{X}}_\perp'\widehat{\mathbf{X}}_\perp)^{-1}\widehat{\mathbf{X}}_\perp' \tag{A.2}$$

Define further $\widehat{\mathbf{X}}^*_\gamma = (\widehat{\mathbf{X}}_\gamma, \widehat{\mathbf{X}}_\perp)$ and $\mathbf{P}^*_\gamma = \widehat{\mathbf{X}}^*_\gamma(\widehat{\mathbf{X}}^{*\prime}_\gamma\widehat{\mathbf{X}}^*_\gamma)^{-1}\widehat{\mathbf{X}}^{*\prime}_\gamma$. Note that by construction $\widehat{\mathbf{X}}_\gamma'\widehat{\mathbf{X}}_\perp = 0$ and hence

$$\mathbf{P}^*_\gamma = \mathbf{P}_\gamma + \mathbf{P}_\perp \tag{A.3}$$

Define $\mathbf{Y}$, $\widehat{\mathbf{G}}, \mathbf{G}, \widehat{\mathbf{V}}$, and $\boldsymbol{\varepsilon}$ by stacking the y_i, $\widehat{\mathbf{g}}_i$, $\mathbf{g}_i$, $\widehat{\mathbf{v}}_i$, and ε_i, respectively. Recall that $\widehat{\mathbf{x}}_i = \widehat{\mathbf{g}}_i = \mathbf{g}_i - \widehat{\mathbf{v}}_i$ then we can also write $\widehat{\mathbf{G}} = \widehat{\mathbf{X}}$. Similarly, define $\widehat{\boldsymbol{\Lambda}}_{1,\gamma}, \widehat{\boldsymbol{\Lambda}}_{2,\gamma}, \mathbf{G}_\gamma$ by stacking $\widehat{\lambda}_{1,i}(\gamma)\, I(q_i \leq \gamma)$, $\widehat{\lambda}_{2,i}(\gamma)\, I(q_i > \gamma)$, and $\mathbf{g}_i I(q_i \leq \gamma)$. Similarly, we can define $\Lambda(\gamma)$ and $\widehat{\Lambda}(\gamma)$ by stacking $\lambda_i(\gamma)$ and $\widehat{\lambda}_i(\gamma)$. Let us denote $\mathbf{G}_0$, and $\Lambda(0)$ the matrices at the true value $\gamma = \gamma_0$.

Lemma 1 Uniformly in $\gamma \in \Gamma$ as $n \longrightarrow \infty$

$$\frac{1}{n}\widehat{\mathbf{X}}_\gamma'\widehat{\mathbf{X}}_\gamma = \frac{1}{n}\sum_{i=1}^n \widehat{\mathbf{x}}_i(\gamma)\widehat{\mathbf{x}}_i(\gamma)' I(q_i \leq \gamma) \xrightarrow{p} \mathbf{M}(\gamma,\gamma) = \mathbf{M}(\gamma) \tag{A.4}$$

$$\frac{1}{n}\widehat{\mathbf{X}}_\perp'\widehat{\mathbf{X}}_\perp = \frac{1}{n}\sum_{i=1}^n \widehat{\mathbf{x}}_i(\gamma)\widehat{\mathbf{x}}_i(\gamma)' I(q_i > \gamma) \xrightarrow{p} \mathbf{M}^\perp(\gamma,\gamma) = \mathbf{M}^\perp(\gamma) \tag{A.5}$$

$$\frac{1}{\sqrt{n}}\widehat{\mathbf{X}}_\gamma'\widehat{\mathbf{r}} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \widehat{\mathbf{x}}_i(\gamma)\widehat{r}_i I(q_i \leq \gamma) = O_p(1) \tag{A.6}$$

$$\frac{1}{\sqrt{n}}\widehat{\mathbf{X}}_\perp'\widehat{\mathbf{r}} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \widehat{\mathbf{x}}_i(\gamma)\widehat{r}_i I(q_i > \gamma) = O_p(1) \tag{A.7}$$

Proof of Lemma 1

To show (A.4) note that

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

and recall that $\widehat{\mathbf{x}}_i = \widehat{\mathbf{g}}_i = \mathbf{g}_i - \widehat{\mathbf{v}}_i$. Note that $\frac{1}{n}\sum_i(\widehat{\mathbf{x}}_i\widehat{\mathbf{x}}_i' I(q_i \le \gamma)) \xrightarrow{p} E\left(\mathbf{g}_i\mathbf{g}_i' I(q_i \le \gamma)\right)$ follows from Caner and Hansen (2004) and (Assumption 1.11) and Lemma 1 of Hansen (1996). Based on (Assumption 1.11) and Lemma 1 of Hansen (1996) we also have

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

$$\begin{aligned} \frac{1}{n}\sum_i(\widehat{\lambda}_i(\gamma))^2 I(q_i \le \gamma) &= \frac{1}{n}\sum_i \left(\widehat{\lambda}_{1,i}(\gamma)\right)^2 I(q_i \le \gamma) + \frac{1}{n}\sum_i \left(\widehat{\lambda}_{2,i}(\gamma)\right)^2 I(q_i \le \gamma)I(q_i > \gamma) \\ &+ 2\frac{1}{n}\sum_i \left(\widehat{\lambda}_{1,i}(\gamma)\,\widehat{\lambda}_{2,i}(\gamma)\right) I(q_i \le \gamma)I(q_i > \gamma) = \frac{1}{n}\sum_i \left(\widehat{\lambda}_{1,i}(\gamma)\right)^2 I(q_i \le \gamma) \end{aligned}$$

Therefore, $\frac{1}{n}\widehat{\mathbf{X}}_\gamma'\widehat{\mathbf{X}}_\gamma \xrightarrow{p} E\left(\mathbf{x}_i(\gamma)\mathbf{x}_i(\gamma)' I(q_i \le \gamma)\right) = \mathbf{M}(\gamma)$, where

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

We should note that this moment does not depend on $\lambda_{2,i}(\gamma)$. Equation (A.5) follows similarly. In particular,

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

Finally, (A.6) and (A.7) follow directly from Assumption (1.11) and Lemma A.4 of Hansen (2000).

Lemma 2 The following sample moment functionals defined uniformly in $\gamma \in [\gamma_0, \overline{\gamma}]$

$$\frac{1}{n}\widehat{\mathbf{X}}_\gamma'\mathbf{G}_0 = \begin{pmatrix} \frac{1}{n}\sum_i(\widehat{\mathbf{x}}_i\mathbf{g}_i' I(q_i \le \gamma_0)) \\ \frac{1}{n}\sum_i \widehat{\lambda}_i(\gamma)\,\mathbf{g}_i' I(q_i \le \gamma_0)) \end{pmatrix} \xrightarrow{p} \mathbf{M}_{\mathbf{XG}}(\gamma_0, \gamma)$$

$$\frac{1}{n}\widehat{\mathbf{X}}_\gamma'\Lambda(0) = \begin{pmatrix} \frac{1}{n}\sum_i \widehat{\mathbf{x}}_i\lambda_i(\gamma_0)\,I(q_i \le \gamma) \\ \frac{1}{n}\sum_i \widehat{\lambda}_i(\gamma)\,\lambda_i(\gamma_0)\,I(q_i \le \gamma) \end{pmatrix} \xrightarrow{p} \mathbf{M}_{\mathbf{X\Lambda}}(\gamma_0, \gamma)$$

$$\frac{1}{n}\widehat{\mathbf{X}}'_{\perp}\Lambda(0)=\left(\begin{array}{c}\frac{1}{n}\sum_i \widehat{\mathbf{x}}_i\lambda_i\left(\gamma_0\right)I(q_i>\gamma)\\ \frac{1}{n}\sum_i \widehat{\lambda}_i\left(\gamma\right)\lambda_i\left(\gamma_0\right)I(q_i>\gamma)\end{array}\right)\xrightarrow{p}\mathbf{M}^{\perp}_{\mathbf{X\Lambda}}(\gamma_0,\gamma)$$

Proof of Lemma 2

Using (Assumption 1.11) and Lemma 1 of Hansen (1996),

$$\frac{1}{n}\sum_i(\widehat{\mathbf{x}}_i\mathbf{g}'_iI(q_i\leq\gamma_0))=\frac{1}{n}\sum_i\mathbf{g}_i\mathbf{g}'_iI(q_i\leq\gamma)-\frac{1}{n}\sum_i\widehat{\mathbf{v}}_i\mathbf{g}_iI(q_i\leq\gamma)\xrightarrow{p}E\left(\mathbf{g}_i\mathbf{g}'_iI(q_i\leq\gamma)\right)$$

$$\frac{1}{n}\sum_i\widehat{\lambda}_i\left(\gamma\right)\mathbf{g}'_iI(q_i\leq\gamma_0))\xrightarrow{p}E(\mathbf{g}'_i\lambda_i\left(\gamma\right)I(q_i\leq\gamma_0))$$

$$\begin{aligned}\frac{1}{n}\sum_i\widehat{\mathbf{x}}_i\lambda_i\left(\gamma_0\right)I(q_i\leq\gamma) &= \frac{1}{n}\sum_i\mathbf{g}_i\lambda_i\left(\gamma_0\right)I(q_i\leq\gamma)-\frac{1}{n}\sum_i\widehat{\mathbf{v}}_i\lambda_i\left(\gamma_0\right)I(q_i\leq\gamma)\\ &\xrightarrow{p} E(\mathbf{g}_i\lambda_i\left(\gamma_0\right)I(q_i\leq\gamma))\end{aligned}$$

$$\frac{1}{n}\sum_i\widehat{\lambda}_i\left(\gamma\right)\lambda_i\left(\gamma_0\right)I(q_i\leq\gamma)\xrightarrow{p}E(\lambda_i\left(\gamma\right)\lambda_i\left(\gamma_0\right)I(q_i\leq\gamma))$$

$$\begin{aligned}\frac{1}{n}\sum_i\widehat{\mathbf{x}}_i\lambda_i\left(\gamma_0\right)I(q_i>\gamma) &= \frac{1}{n}\sum_i\mathbf{g}_i\lambda_i\left(\gamma_0\right)I(q_i>\gamma)-\frac{1}{n}\sum_i\widehat{\mathbf{v}}_i\lambda_i\left(\gamma_0\right)I(q_i>\gamma)\\ &\xrightarrow{p} E(\mathbf{g}_i\lambda_i\left(\gamma_0\right)I(q_i>\gamma))\end{aligned}$$

$$\frac{1}{n}\sum_i\widehat{\lambda}_i\left(\gamma\right)\lambda_i\left(\gamma_0\right)I(q_i>\gamma)\xrightarrow{p}E(\lambda_i\left(\gamma\right)\lambda_i\left(\gamma_0\right)I(q_i>\gamma))$$

Note that $I(q_i\leq\gamma)I(q_i\leq\gamma_0)=I(q_i\leq\gamma_0)$, $I(q_i<\gamma)I(q_i>\gamma_0)=I(\gamma_0\leq q_i<\gamma)$, $I(q_i>\gamma)I(q_i\leq\gamma_0)=0$, $I(q_i\leq\gamma_0)I(q_i>\gamma_0)=0$, and $I(q_i\leq\gamma)I(q_i>\gamma)=0$. Then using (2.22) we can further express these functionals as follows

$$\mathbf{M}_{\mathbf{XG}}(\gamma_0,\gamma)=\left(\begin{array}{c}E\left(\mathbf{g}_i\mathbf{g}'_iI(q_i\leq\gamma)\right)\\ E(\mathbf{g}'_i\lambda_{1,i}\left(\gamma\right)I(q_i\leq\gamma_0))\end{array}\right)$$

$$\mathbf{M}_{\mathbf{X\Lambda}}(\gamma_0,\gamma)=\left(\begin{array}{c}E(\mathbf{g}_i\lambda_{1,i}\left(\gamma_0\right)I(q_i\leq\gamma_0))+E(\mathbf{g}_i\lambda_{2,i}\left(\gamma_0\right)I(\gamma_0\leq q_i<\gamma))\\ E(\lambda_{1,i}\left(\gamma_0\right)\lambda_{1,i}\left(\gamma\right)I(q_i\leq\gamma_0)+E(\lambda_{2,i}\left(\gamma_0\right)\lambda_{1,i}\left(\gamma\right)I(\gamma_0<q_i\leq\gamma)\end{array}\right)$$

$$\mathbf{M}^{\perp}_{\mathbf{X\Lambda}}(\gamma_0,\gamma)=\left(\begin{array}{c}E(\mathbf{g}_i\lambda_{2,i}\left(\gamma_0\right)I(q_i>\gamma))\\ E(\lambda_{2,i}\left(\gamma_0\right)\lambda_{2,i}\left(\gamma\right)I(q_i>\gamma_0)\end{array}\right)$$

Proof of Theorem 1

We can express (2.23) in matrix notation

$$\mathbf{Y} = \mathbf{G}\boldsymbol{\beta} + \mathbf{G}_0\boldsymbol{\delta}_n + \kappa_n \Lambda(0) + \boldsymbol{\varepsilon} \tag{A.8}$$

Let $\boldsymbol{\delta}_n = \mathbf{c}_\delta n^{-\alpha}$ and $\kappa_n = c_\kappa n^{-\alpha}$. Given that $\mathbf{G} = \widehat{\mathbf{G}} + \widehat{\mathbf{V}}$ and $\widehat{\mathbf{X}} = \widehat{\mathbf{G}}$ is in the span of $\widehat{\mathbf{X}}^*_\gamma$ then $(\mathbf{I} - \mathbf{P}^*_\gamma)\mathbf{G} = (\mathbf{I} - \mathbf{P}^*_\gamma)\widehat{\mathbf{V}}$ and

$$(\mathbf{I} - \mathbf{P}^*_\gamma)\mathbf{Y} = (\mathbf{I} - \mathbf{P}^*_\gamma)(n^{-\alpha}\mathbf{c}'_{\boldsymbol{\delta}}\mathbf{G}'_0 + n^{-\alpha}c_\kappa\Lambda(0)' + \widehat{\mathbf{r}})$$

where

$$\widehat{\mathbf{r}} = \widehat{\mathbf{V}}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

The sum of squared errors is given by

$$\begin{aligned} S_n(\gamma,\gamma_0) &= \mathbf{Y}'(\mathbf{I} - \mathbf{P}^*_\gamma)\mathbf{Y} \qquad \text{(A.9)} \\ &= (n^{-\alpha}\mathbf{c}'_{\boldsymbol{\delta}}\mathbf{G}'_0 + n^{-\alpha}c_\kappa\Lambda(0)' + \widehat{\mathbf{r}}')(\mathbf{I} - \mathbf{P}^*_\gamma)(\mathbf{G}_0\mathbf{c}_{\boldsymbol{\delta}}n^{-\alpha} + \Lambda(0)c_\kappa n^{-\alpha} + \widehat{\mathbf{r}}) \\ &= (n^{-\alpha}\mathbf{c}'_{\boldsymbol{\delta}}\mathbf{G}'_0 + n^{-\alpha}c_\kappa\Lambda(0)' + \widehat{\mathbf{r}}')\left(\mathbf{G}_0\mathbf{c}_{\boldsymbol{\delta}}n^{-\alpha} + \Lambda(0)c_\kappa n^{-\alpha} + \widehat{\mathbf{r}}\right) \\ &\quad -(n^{-\alpha}\mathbf{c}'_{\boldsymbol{\delta}}\mathbf{G}'_0 + n^{-\alpha}c_\kappa\Lambda(0)' + \widehat{\mathbf{r}}')\mathbf{P}^*_\gamma\left(\mathbf{G}_0\mathbf{c}_{\boldsymbol{\delta}}n^{-\alpha} + \Lambda(0)c_\kappa n^{-\alpha} + \widehat{\mathbf{r}}\right) \end{aligned}$$

Notice that to minimize $S_n(\gamma)$ it is sufficient to maximize

$$\begin{aligned} S^*_n(\gamma,\gamma_0) &= n^{2\alpha-1}(n^{-\alpha}\mathbf{c}'_{\boldsymbol{\delta}}\mathbf{G}'_0 + n^{-\alpha}c_\kappa\Lambda(0)' + \widehat{\mathbf{r}}')\mathbf{P}^*_\gamma\left(\mathbf{G}_0\mathbf{c}_{\boldsymbol{\delta}}n^{-\alpha} + \Lambda(0)c_\kappa n^{-\alpha} + \widehat{\mathbf{r}}\right) \\ &= n^{-1}\mathbf{c}'_\delta\mathbf{G}'_0\mathbf{P}^*_\gamma\mathbf{G}_0\mathbf{c}_\delta + n^{-1}c_\kappa\Lambda(0)'\mathbf{P}^*_\gamma\Lambda(0)c_\kappa + 2n^{-1}\mathbf{c}'_\delta\mathbf{G}'_0\mathbf{P}^*_\gamma\Lambda(0)c_\kappa \\ &\quad +2n^{\alpha-1}\mathbf{c}'_{\boldsymbol{\delta}}\mathbf{G}'_0\mathbf{P}^*_\gamma\widehat{\mathbf{r}} + 2n^{\alpha-1}c_\kappa\Lambda(0)'\mathbf{P}^*_\gamma\widehat{\mathbf{r}} + n^{2\alpha-1}\widehat{\mathbf{r}}'\mathbf{P}^*_\gamma\widehat{\mathbf{r}} \end{aligned}$$

Let us first consider the problem when $\gamma \in [\gamma_0, \overline{\gamma}]$.

Recall that $\mathbf{P}^*_\gamma = \mathbf{P}_\gamma + \mathbf{P}_\perp$ so that $\mathbf{P}_\perp\mathbf{G}_0 = 0$ and so $\mathbf{P}^*_\gamma\mathbf{G}_0 = \mathbf{P}_\gamma\mathbf{G}_0$. Let us examine each of the six terms in $S^*_n(\gamma,\gamma_0)$. Using Lemmas 1 and 2 we have

(i)

$$\begin{aligned} \tfrac{1}{n}\mathbf{G}'_0\mathbf{P}^*_\gamma\mathbf{G}_0 &= \tfrac{1}{n}\mathbf{G}'_0\mathbf{P}_\gamma\mathbf{G}_0 \\ &= (\tfrac{1}{n}\mathbf{G}'_0\widehat{\mathbf{X}}_\gamma)(\tfrac{1}{n}\widehat{\mathbf{X}}'_\gamma\widehat{\mathbf{X}}_\gamma)^{-1}(\tfrac{1}{n}\widehat{\mathbf{X}}'_\gamma\mathbf{G}_0) \\ &\xrightarrow{p} \mathbf{M}_{\mathbf{XG}}(\gamma_0,\gamma)'\mathbf{M}(\gamma)^{-1}\mathbf{M}_{\mathbf{XG}}(\gamma_0,\gamma) \end{aligned}$$

(ii)

$$\begin{aligned}\tfrac{1}{n}\Lambda(0)'\mathbf{P}^*_\gamma\Lambda(0) &= (\tfrac{1}{n}\Lambda(0)'\widehat{\mathbf{X}}_\gamma)(\tfrac{1}{n}\widehat{\mathbf{X}}'_\gamma\widehat{\mathbf{X}}_\gamma)^{-1}(\tfrac{1}{n}\widehat{\mathbf{X}}'_\gamma\Lambda(0)) \\ &\quad +(\tfrac{1}{n}\Lambda(0)'\widehat{\mathbf{X}}_\perp)(\tfrac{1}{n}\widehat{\mathbf{X}}'_\perp\widehat{\mathbf{X}}_\perp)^{-1}(\tfrac{1}{n}\widehat{\mathbf{X}}'_\perp\Lambda(0)) \\ &\xrightarrow{p} \mathbf{M}_{\mathbf{X\Lambda}}(\gamma_0,\gamma)'\mathbf{M}(\gamma)^{-1}\mathbf{M}_{\mathbf{X\Lambda}}(\gamma_0,\gamma) \\ &\quad +\mathbf{M}^\perp_{\mathbf{X\Lambda}}(\gamma_0,\gamma)'\mathbf{M}^\perp(\gamma)^{-1}\mathbf{M}^\perp_{\mathbf{X\Lambda}}(\gamma_0,\gamma)\end{aligned}$$

(iii)

$$\begin{aligned}\tfrac{1}{n}\mathbf{G}'_0\mathbf{P}^*_\gamma\Lambda(0) &= (\tfrac{1}{n}\mathbf{G}'_0\widehat{\mathbf{X}}_\gamma)(\tfrac{1}{n}\widehat{\mathbf{X}}'_\gamma\widehat{\mathbf{X}}_\gamma)^{-1}(\tfrac{1}{n}\widehat{\mathbf{X}}'_\gamma\Lambda(0)) \\ &\xrightarrow{p} \mathbf{M}_{\mathbf{XG}}(\gamma_0,\gamma)'\mathbf{M}(\gamma)^{-1}\mathbf{M}_{\mathbf{XG}}(\gamma_0,\gamma)\end{aligned}$$

(iv)

$$n^{\alpha-1}\mathbf{G}'_0\mathbf{P}^*_\gamma\widehat{\mathbf{r}} = n^{\alpha-1/2}(\frac{1}{n}\mathbf{G}'_0\widehat{\mathbf{X}}_\gamma)(\frac{1}{n}\widehat{\mathbf{X}}'_\gamma\widehat{\mathbf{X}}_\gamma)^{-1}(\frac{1}{\sqrt{n}}\widehat{\mathbf{X}}'_\gamma\widehat{\mathbf{r}}) \xrightarrow{p} 0$$

(v) Recall that from Lemma 1, $\frac{1}{\sqrt{n}}\widehat{\mathbf{X}}'_\gamma\widehat{\mathbf{r}} \xrightarrow{p} 0$ and $\frac{1}{\sqrt{n}}\widehat{\mathbf{X}}'_\perp\widehat{\mathbf{r}} \xrightarrow{p} 0$, then

$$\begin{aligned}n^{\alpha-1}\Lambda(0)'\mathbf{P}^*_\gamma\widehat{\mathbf{r}} &= n^{\alpha-1/2}\Lambda(0)'\left(\mathbf{P}_\gamma+\mathbf{P}_\perp\right)\widehat{\mathbf{r}} \\ &= (\tfrac{1}{n}\Lambda(0)'\widehat{\mathbf{X}}_\gamma)(\tfrac{1}{n}\widehat{\mathbf{X}}'_\gamma\widehat{\mathbf{X}}_\gamma)^{-1}(\tfrac{1}{\sqrt{n}}\widehat{\mathbf{X}}'_\gamma\widehat{\mathbf{r}}) \\ &\quad + (\tfrac{1}{n}\Lambda(0)\widehat{\mathbf{X}}_\perp)(\tfrac{1}{n}\widehat{\mathbf{X}}'_\perp\widehat{\mathbf{X}}_\perp)^{-1}(\tfrac{1}{\sqrt{n}}\widehat{\mathbf{X}}'_\perp\widehat{\mathbf{r}}) \\ &\xrightarrow{p} 0\end{aligned}$$

(vi)

$$\begin{aligned}n^{2\alpha-1}\widehat{\mathbf{r}}'\mathbf{P}^*_\gamma\widehat{\mathbf{r}} &= n^{2\alpha-1}\left(\tfrac{1}{\sqrt{n}}\widehat{\mathbf{r}}'\widehat{\mathbf{X}}_\gamma\right)\left(\tfrac{1}{n}\widehat{\mathbf{X}}'_\gamma\widehat{\mathbf{X}}_\gamma\right)^{-1}\left(\tfrac{1}{\sqrt{n}}\widehat{\mathbf{X}}'_\gamma\widehat{\mathbf{r}}\right) \\ &\quad + n^{2\alpha-1}\left(\tfrac{1}{\sqrt{n}}\widehat{\mathbf{r}}'\widehat{\mathbf{X}}_\perp\right)\left(\tfrac{1}{n}\widehat{\mathbf{X}}'_\perp\widehat{\mathbf{X}}_\perp\right)^{-1}\left(\tfrac{1}{\sqrt{n}}\widehat{\mathbf{X}}'_\perp\widehat{\mathbf{r}}\right) \\ &\xrightarrow{p} 0\end{aligned}$$

Therefore, uniformly on $\gamma \in [\gamma_0, \overline{\gamma}]$

$$S^*_n(\gamma,\gamma_0) \xrightarrow{p} S^*(\gamma,\gamma_0)$$

where

$$\begin{aligned}S^*(\gamma,\gamma_0) &= \mathbf{c}'_\delta\mathbf{M}_{\mathbf{XG}}(\gamma_0,\gamma)'\mathbf{M}(\gamma)^{-1}\mathbf{M}_{\mathbf{XG}}(\gamma_0,\gamma)\mathbf{c}_\delta \\ &\quad + c_\kappa\mathbf{M}_{\mathbf{X\Lambda}}(\gamma_0,\gamma)'\mathbf{M}(\gamma)^{-1}\mathbf{M}_{\mathbf{X\Lambda}}(\gamma_0,\gamma)c_\kappa \\ &\quad + c_\kappa\mathbf{M}^\perp_{\mathbf{X\Lambda}}(\gamma_0,\gamma)'\mathbf{M}^\perp(\gamma)^{-1}\mathbf{M}^\perp_{\mathbf{X\Lambda}}(\gamma_0,\gamma)c_\kappa \\ &\quad + 2\mathbf{c}'_\delta\mathbf{M}_{\mathbf{XG}}(\gamma_0,\gamma)'\mathbf{M}(\gamma)^{-1}\mathbf{M}_{\mathbf{X\Lambda}}(\gamma_0,\gamma)c_\kappa\end{aligned} \tag{A.10}$$

Define $\mathbf{c} = (\mathbf{c}_\delta, c_k)'$ and note that $\mathbf{M}(\gamma_0,\gamma) = \begin{pmatrix}\mathbf{M}'_{\mathbf{XG}}(\gamma_0,\gamma)\\ \mathbf{M}'_{\mathbf{X\Lambda}}(\gamma_0,\gamma)\end{pmatrix}$ and $\widetilde{\mathbf{M}}^\perp(\gamma_0,\gamma) =$

$\begin{pmatrix} \mathbf{0} \\ \mathbf{M}_{\mathbf{X\Lambda}}^{\perp}(\gamma_0,\gamma) \end{pmatrix}$. Then by Lemma 2 we get

$$S^*(\gamma,\gamma_0) = \mathbf{c}'\mathbf{M}(\gamma_0,\gamma)'\mathbf{M}(\gamma)^{-1}\mathbf{M}(\gamma_0,\gamma)\mathbf{c} + \mathbf{c}'\widetilde{\mathbf{M}}^{\perp}(\gamma_0,\gamma)'\mathbf{M}^{\perp}(\gamma)^{-1}\widetilde{\mathbf{M}}^{\perp}(\gamma_0,\gamma)\mathbf{c} \tag{A.11}$$

We restrict $\gamma \in [\gamma_0, \overline{\gamma}]$ to the region where $\lambda_{2,i}(\gamma)$ is non-decreasing. Notice that, in this case, both $\lambda_{1,i}(\gamma)$ and $\lambda_{2,i}(\gamma)$ are monotonically increasing in the range $\gamma \in [\gamma_0, \overline{\gamma}]$ and $\lambda_{1,i}(\gamma) < \lambda_{2,i}(\gamma)$, and hence, for any α, $\alpha^{/}\left(\mathbf{M}(\gamma_0) - \mathbf{M}(\gamma_0,\gamma)\right)\alpha > 0$ and $\alpha^{/}\left(\mathbf{M}^{\perp}(\gamma_0) - \mathbf{M}^{\perp}(\gamma_0,\gamma)\right)\alpha > 0$, so that, $S^*(\gamma,\gamma_0) \leq S^{**}(\gamma,\gamma_0)$ with equality at $\gamma = \gamma_0$, where

$$\begin{aligned} S^{**}(\gamma,\gamma_0) &= \mathbf{c}'\mathbf{M}(\gamma_0)\mathbf{M}(\gamma)^{-1}\mathbf{M}(\gamma_0)\mathbf{c} + \mathbf{c}'\widetilde{\mathbf{M}}^{\perp}(\gamma_0)'\mathbf{M}^{\perp}(\gamma)^{-1}\widetilde{\mathbf{M}}^{\perp}(\gamma_0)\mathbf{c} \\ &= \mathbf{c}'\left(\mathbf{M}(\gamma_0) + \widetilde{\mathbf{M}}^{\perp}(\gamma_0)\right)\left(\mathbf{M}(\gamma)^{-1} + \mathbf{M}^{\perp}(\gamma)^{-1}\right)\left(\mathbf{M}(\gamma_0) + \widetilde{\mathbf{M}}^{\perp}(\gamma_0)\right)\mathbf{c} \end{aligned}$$

Hence, maximizing $S^{**}(\cdot)$ is equivalent to maximizing $S^*(\cdot)$, as $S^{**}(\cdot)$ will be shown to be a decreasing function in γ.

Given that $\mathbf{M}(\gamma) = \int_{-\infty}^{\gamma} E(\mathbf{x}_i(t)\mathbf{x}_i(t)'|q=t)f_q(t)dt$, the derivative of $\mathbf{M}(\gamma)$ is

$$\frac{d\mathbf{M}(\gamma)}{d\gamma} = E(\mathbf{x}_i(\gamma)\mathbf{x}_i(\gamma)')|q=\gamma)f_q(\gamma) = \mathbf{D}_1(\gamma)f_q(\gamma) \tag{A.12}$$

Similarly, since $\mathbf{M}^{\perp}(\gamma) = \int_{\gamma}^{+\infty} E(\mathbf{x}_i(t)\mathbf{x}_i(t)'|q=t)f_q(t)dt$,

$$\frac{d\mathbf{M}^{\perp}(\gamma)}{d\gamma} = -E((\mathbf{x}_i(\gamma)\mathbf{x}_i(\gamma)')|q=\gamma)f_q(\gamma) = -\mathbf{D}_1(\gamma)f_q(\gamma) \tag{A.13}$$

Then, using (A.12) and (A.13)

$$\begin{aligned} \tfrac{dS^{**}(\gamma,\gamma_0)}{d\gamma} = -\mathbf{c}'\left(\mathbf{M}(\gamma_0) + \widetilde{\mathbf{M}}^{\perp}(\gamma_0)\right)&(\mathbf{M}(\gamma)^{-1}\mathbf{D}_1(\gamma)f_q(\gamma)\mathbf{M}(\gamma)^{-1} \\ &-\mathbf{M}^{\perp}(\gamma)^{-1}\mathbf{D}_1(\gamma)f_q(\gamma)\mathbf{M}^{\perp}(\gamma)^{-1})\left(\mathbf{M}(\gamma_0) + \widetilde{\mathbf{M}}^{\perp}(\gamma_0)\right)\mathbf{c} < 0 \end{aligned}$$

is continuous and weakly decreasing on $\gamma \in [\gamma_0, \overline{\gamma}]$ since $\mathbf{c}'\mathbf{D}_1(\gamma)f_q(\gamma)\mathbf{c} > 0$ by Assumption (1.7), and $\alpha^{/}\left(\mathbf{M}^{\perp}(\gamma) - \mathbf{M}(\gamma)\right)\alpha > 0$ for any α since $\lambda_{1,i}(\gamma) \leq \lambda_{2,i}(\gamma)$ for all $\gamma \in [\gamma_0, \overline{\gamma}]$, so that $S_n^{**}(\gamma,\gamma_0)$ is uniquely maximized at γ_0. A symmetric argument can be made to show that $S_n^{**}(\gamma,\gamma_0)$ is uniquely maximized at γ_0 when $\gamma \in [\underline{\gamma}, \gamma_0]$. Since, $\widehat{\gamma}$ maximizes $S_n^{**}(\gamma,\gamma_0)$ for $\gamma \in \Gamma$, therefore $\widehat{\gamma} \xrightarrow{p} \gamma_0$.

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

Proof: Lemma 4 of Caner and Hansen (2004) carries over to our framework as follows. Recall that $\mathbf{g}_i(\gamma) = (\mathbf{g}_i, \lambda_i(\gamma))'$ and let the stacked version of $\mathbf{g}_i(\gamma)$ evaluated at γ_0 be $\mathbf{G}(\gamma_0)$. Let the constants B, d, k be defined as $B > 0$, $0 < d < \infty$, $0 < k < \infty$.

Then,

$$\inf_{\bar{\varepsilon}/a_n \le |\gamma-\gamma_0| \le B} c'(\frac{\mathbf{G}(\gamma_0)'(\mathbf{P}_0^* - \mathbf{P}_\gamma^*)\mathbf{G}(\gamma_0)}{n(\gamma-\gamma_0)})c \ge 5d/6 \tag{A.14}$$

$$\sup_{\bar{\varepsilon}/a_n \le |\gamma-\gamma_0| \le B} |\frac{c'\mathbf{G}(\gamma_0)'(\mathbf{P}_0^* - \mathbf{P}_\gamma^*)\widehat{\mathbf{r}}}{n^{1-\alpha}(\gamma-\gamma_0)}| \le d/12 \tag{A.15}$$

$$\sup_{\bar{\varepsilon}/a_n \le |\gamma-\gamma_0| \le B} |\frac{\widehat{\mathbf{r}}'(\mathbf{P}_0^* - \mathbf{P}_\gamma^*)\widehat{\mathbf{r}}}{n^{1-2\alpha}(\gamma-\gamma_0)}| \le d/6 \tag{A.16}$$

where $d \in (0, \infty)$.

Hence we can write $S_n(\gamma) - S_n(\gamma_0)$ for $\bar{\varepsilon}/a_n \le |\gamma - \gamma_0| \le B$ as

$$\frac{S_n(\gamma) - S_n(\gamma_0)}{n^{1-2\alpha}(\gamma-\gamma_0)} = \frac{\widehat{\mathbf{r}}'(\mathbf{P}_0^* - \mathbf{P}_\gamma^*)\widehat{\mathbf{r}}}{n^{1-2\alpha}(\gamma-\gamma_0)} + 2\frac{\widehat{\mathbf{r}}'(\mathbf{P}_0^* - \mathbf{P}_\gamma^*)\mathbf{F}_0 c}{n^{1-\alpha}(\gamma-\gamma_0)} + \frac{c'\mathbf{F}_0'(\mathbf{P}_0^* - \mathbf{P}_\gamma^*)\mathbf{F}_0 c}{n(\gamma-\gamma_0)} \ge d/2 \tag{A.17}$$

Using equations (A.14) to (A.16) above, and similar bounding conditions as in Caner and Hansen (2004) and since $S_n(\widehat{\gamma}) \le S_n(\gamma_0)$, equation (A.17) implies that $|\widehat{\gamma} - \gamma_0| \le \bar{\varepsilon}/a_n$.

Now as all the conditions used to derive (A.17) hold jointly with probability more than $1 - \epsilon$ we have that $P(n^{1-2\alpha}|\widehat{\gamma} - \gamma_0| > \bar{\varepsilon}) = \epsilon$ for $n \ge \bar{n}$. Hence, $a_n(\widehat{\gamma} - \gamma_0) = O_p(1)$.

Lemma 4 On $\nu \in [-\bar{\varepsilon}, \bar{\varepsilon}]$

$$n^{-2a}c'\mathbf{G}(\gamma_0)'(\mathbf{P}_0^* - \mathbf{P}_\nu^*)\mathbf{G}(\gamma_0)c \Longrightarrow \mu|\nu| \tag{A.18}$$

$$n^{-a}c'\mathbf{G}(\gamma_0)'(\mathbf{P}_0^* - \mathbf{P}_\nu^*)\widehat{\mathbf{r}} \Longrightarrow \xi^{1/2}W(\nu) \tag{A.19}$$

$$\widehat{\mathbf{r}}'(\mathbf{P}_0^* - \mathbf{P}_\nu^*)\widehat{\mathbf{r}} \Longrightarrow 0. \tag{A.20}$$

Proof:

Let us reparameterize all functions of γ as functions of ν. For example $\widehat{\mathbf{X}}_\nu = \widehat{\mathbf{X}}_{\gamma_0+\nu/a_n}$, $\mathbf{P}_\nu = \mathbf{P}_{\gamma_0+\nu/a_n}$, and for $\Delta_i(\gamma) = I(q_i \leq \gamma) - I(q_i \leq \gamma_0)$ we have $\Delta_i(\nu) = \Delta_i(\gamma_0 + \nu/a_n)$.

Suppose that $\nu \in [-\bar{\varepsilon}, \bar{\varepsilon}]$, then using equation (A.9) and Lemma 5 of Caner and Hansen we get

$$\begin{aligned} Q_n(\nu) &= S_n(\gamma_0) - S_n(\gamma_0 + \nu/a_n) \\ &= (n^{-\alpha}\mathbf{c}'\mathbf{G}(\gamma_0)' + \widehat{\mathbf{r}}')\mathbf{P}_\nu^*(\mathbf{G}(\gamma_0)\mathbf{c}n^{-\alpha} + \widehat{\mathbf{r}}) - (n^{-\alpha}\mathbf{c}'\mathbf{G}(\gamma_0)' + \widehat{\mathbf{r}}')\mathbf{P}_0(\mathbf{G}(\gamma_0)\mathbf{c}n^{-\alpha} + \widehat{\mathbf{r}}) \\ &= n^{-2a}c'\mathbf{G}(\gamma_0)'(\mathbf{P}_\nu^* - \mathbf{P}_0^*)\mathbf{G}(\gamma_0)c + 2n^{-a}c'\mathbf{G}(\gamma_0)'(\mathbf{P}_\nu^* - \mathbf{P}_0^*)\widehat{\mathbf{r}} + \widehat{\mathbf{r}}'(\mathbf{P}_0^* - \mathbf{P}_\nu^*)\widehat{\mathbf{r}} \\ &\Longrightarrow \mu|\nu| + \xi^{1/2}W(\nu), \text{ uniformly on } \nu \in [-\bar{\varepsilon}, \bar{\varepsilon}] \end{aligned}$$

where $\mu = \mathbf{c}'\mathbf{D}_1\mathbf{c}f$ and $\xi = \mathbf{c}'\mathbf{D}_2\mathbf{c}f$.

Proof of Theorem 2

Using Assumption 1 and Lemma 4 we establish that

$n^{-2a}c'\mathbf{G}(\gamma_0)'(\mathbf{P}_0^* - \mathbf{P}_\nu^*)\mathbf{F}_0c = n^{-2a}\sum_{i=1}^{n}(c'\mathbf{g}_{0i}^*)^2\Delta_i(\nu) + o_p(1) \Longrightarrow \mu|\nu|$, $n^{-a}c'\mathbf{G}(\gamma_0)'(\mathbf{P}_0^* - \mathbf{P}_\nu^*)\widehat{\mathbf{r}} = n^{-a}\sum_{i=1}^{n}\mathbf{g}_{0i}^*\boldsymbol{\varepsilon}_i\Delta_i(\nu) + o_p(1) \Longrightarrow \xi^{1/2}W(\nu)$, and $\widehat{\mathbf{r}}'(\mathbf{P}_0^* - \mathbf{P}_\nu^*)\widehat{\mathbf{r}} = o_p(1)$.

Therefore, using the results that $a_n(\widehat{\gamma} - \gamma_0) = \arg\max_\nu Q_n(\nu) = O_p(1)$ and that $Q_n(\nu) \Longrightarrow -\mu|\nu| + 2\xi^{1/2}W(\nu)$ where the limit functional is continuous with a unique maximum almost surely. Then, equations (3.34) and (3.35) are established by following the argument in the proofs Caner and Hansen (2004).