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‡

June 13, 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

*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 to 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 Threshold Regression with Endogenous Threshold variable (THRET) 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 THRET, 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.

While there are several econometric studies on the statistical inference of threshold regression models, there is as yet no available inference when the threshold variable itself is endogenous. 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 THRET context, the introduction of the inverse Mills ratio bias correction terms results in regime-dependent heteroskedasticity. We propose an asymptotic distribution for the threshold estimator for this case.

The paper is organized as follows. Section 2 describes the model and the setup. Section 3 describes the properties of the estimators. Section 4 presents our Monte Carlo experiments. Section 5 concludes.

2 The Threshold Regression with Endogenous Threshold (THRET) 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 (THRET),

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

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

where 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 what regime applies is given by

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

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

THRET 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 THRET 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

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

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

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

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

such that

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

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.

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

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

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

[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.

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 THRET model using Hansen (2000) or Caner and Hansen (2004) would generally result in inconsistent estimation.

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

2.1 Estimation

For estimation purposes it is convenient to rewrite the model in terms of the lower regime (Regime 1). 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 express (2.17) as

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

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

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

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 $\widehat{\lambda}_{1,i}(\gamma) = \lambda_1(\gamma - \mathbf{z}_i'\widehat{\boldsymbol{\pi}}_q)$ and $\widehat{\lambda}_{2,i}(\gamma) = \lambda_2(\gamma - \mathbf{z}_i'\widehat{\boldsymbol{\pi}}_q)$.

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).

3 Asymptotic Properties

3.1 Consistency

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

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

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

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

$$\mathbf{D}_2(\gamma) = E\left(\mathbf{g}_i(\gamma)\mathbf{g}_i(\gamma)'\varepsilon_{1,i}^2 | q_i = \gamma\right) \tag{3.29}$$

$$\mathbf{D}_2^{\perp}(\gamma) = E\left(\mathbf{g}_i(\gamma)\mathbf{g}_i(\gamma)'\varepsilon_{2,i}^2 | q_i = \gamma\right) \tag{3.30}$$

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 $\mathbf{D}_1(\gamma)$ 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 $f_q > 0$, $\mathbf{c}'\mathbf{D}_1\mathbf{c} > 0$, $\mathbf{c}'\mathbf{D}_2\mathbf{c} > 0, \mathbf{c}'\mathbf{D}_2^{\perp}\mathbf{c}$, 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 $a_n = n^{1-2a}$, $\underset{\gamma\in\Gamma}{sup}|\frac{1}{\sqrt{n}}\sum_i \mathbf{H}_i\widehat{\mathbf{v}}_i' I(q_i \leq \gamma)| = O_p(1)$

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. One important difference is that in our case we have regime-dependent heteroskedasticity and hence $E(\varepsilon_i^2|q_i = \gamma)$ is not continuous at γ_0. This arises because the inverse Mills ratio terms are different across the two regimes. 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).

Theorem 1: Consistency

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

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

We provide a proof in the appendix.

3.2 Asympotic distribution

The difference between the framework of THRET vis-a-vis the framework of Hansen (2000) and Caner and Hansen (2004) is not limited to the inverse Mills ratio terms in the conditional mean but, as can be seen from (2.18), extends to the conditional variance. This implies that there is a discontinuity in the argmax probability density of $\widehat{\gamma}$ at γ_0 and the distribution is asymmetric. The case of the argmax probability density with unequal drifts and scalings has been investigated by Stryhn (1996).

One of the complications of the derivation of such a two-sided argmax distribution within Hansen's (2000) framework is the asymmetric treatment of the conditional variance and the cross moment matrices. For consistency within Hansen's framework we need continuity of the cross moment matrix D_1 whereas for the asympototic distribution one also needs continuity of the conditional variance matrix.

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 $T_2 = \arg\max_\nu \left[\xi_2(-\frac{|\nu|}{2} + W(\nu))\right]$. Seo and Linton (2007) explore the case of regime dependent heteroskedasticity with a discontinuity at γ_0, which is however smoothened out in their framework.

In the THRET framework, the argmax distribution of $\widehat{\gamma}$ that is analogous to the Seo and Linton case would take the form

$$n^{1-2\alpha}(\widehat{\gamma} - \gamma_0) \xrightarrow{d} T = \arg\max_\nu \left(\xi_1(-\frac{|\nu|}{2} + W(\nu))I(q < \gamma_0) + \xi_2(-\frac{|\nu|}{2} + W(\nu))I(q > \gamma_0)\right) \tag{3.31}$$

where $W(\nu)$ denotes a two-sided Brownian motion on the real line, and where $\xi_1 = \frac{cD_2c}{(cD_1c)^2 f}$ and $\xi_2 = \frac{cD_2^{\perp}c}{(cD_1c)^2 f}$.

We can also define the likelihood ratio statistic

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

and

$$\xi = \sup_v (\xi_1(-|\nu| + 2W(\nu))I(q < \gamma_0) + \xi_2(-|\nu| + 2W(\nu))I(q > \gamma_0)) \tag{3.32}$$

and

$$\eta^2 = \frac{c' D_2 c + c' D_2^{\perp} c}{\sigma_{\varepsilon}^2 c' D_1 c} \tag{3.33}$$

Then following Theorem 1 of Caner and Hansen (2004), under Assumption 1 and the further assumption that 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 $n \geq \overline{n}$

$$P(\sup_{\overline{\varepsilon}/a_n \leq |\gamma - \gamma_0| \leq B} | \frac{\sum_i \mathbf{H}_i \widehat{\mathbf{v}}_i I_{\{q_i \leq \gamma\}} - I_{\{q_i \leq \gamma_0\}}}{n^{1-\alpha} |\gamma - \gamma_0|} | > \zeta) < \epsilon \tag{3.34}$$

$$\sup_{|\nu| \leq \overline{\varepsilon}} n^{-\alpha} | \sum_i \mathbf{H}_i \widehat{\mathbf{v}}_i I_{\{q_i \leq \gamma + \nu/a_n\}} - I_{\{q_i \leq \gamma_0\}} | \rightarrow 0 \tag{3.35}$$

we obtain that,

$$LR_n(\gamma) \xrightarrow{d} \eta^2 \xi \tag{3.36}$$

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 90th percentile of the distribution of ξ. Then, $\widehat{\Gamma}$ is an asymptotically valid 90% confidence region for γ_0, and is given by

$$\widehat{\Gamma} = \left\{ \gamma : LR_n(\gamma) \leq \widehat{\eta}^2 \mu \right\} \tag{3.37}$$

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).

4 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 threshold 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

$$y_i = \begin{cases} \beta_{1,1} + \beta_{1,2} x_i + u_i, & q_i \leq 2 \\ \beta_{2,1} + \beta_{2,2} x_i + u_i, & q_i > 2 \end{cases} \tag{4.38}$$

where

$$q_i = 2 + z_{1,i} + v_{q,i} \tag{4.39}$$

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[4]. In the case of endogenous threshold and endogenous slope variable we assume that $x_i = z_{2,i} + v_i$, where $z_{2,i} \sim NIID(0,1)$ and $v_i = 0.5u_i$. Finally we consider sample sizes of 100, 200, and 500 using 1000 Monte Carlo simulations. We also investigated what happened when we varied the degree of the correlation between the instrumental variables z and the exogenous slope variables x_2. As in the case of Heckman's estimator, our estimator becomes more efficient as this correlation decreases and the degree of multicollinearity between $\Pi' z$ and x is small.

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

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 THRET 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 THRET improves as δ and/or n increases. We also find that the threshold estimates of THRET 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 $\gamma = 2$. THRET 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

[4]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.

we find that the performance of the slope coefficient estimate of THRET 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 THRET. Table 2 suggests that the distribution of $\widehat{\beta}_2$ for THRET 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.

5 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. By means of an extensive simulation study we examine the performance of our estimator when compared with its competitors. Our proposed estimator performs well for a variety of sample sizes and parameter combinations.

References

[1] 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.

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

[3] 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.

[4] 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.

[5] Easterly, W. and R. Levine (2003), "Tropics, Germs, and Crops: How Endowments Influence Economic Development," *Journal of Monetary Economics*, 50(1), p. 3-39.

[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] Sachs, J. (2003), "Institutions Don't Rule: Direct Effects of Geography on Per Capita Income," National Bureau of Economic Research Working Paper No. 9490.

[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.

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 $\mathbf{P}_\gamma^* = \widehat{\mathbf{X}}_\gamma^*(\widehat{\mathbf{X}}_\gamma^{*\prime}\widehat{\mathbf{X}}_\gamma^*)^{-1}\widehat{\mathbf{X}}_\gamma^{*\prime}$. 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 $\widehat{\mathbf{\Lambda}}_{1,\gamma}, \widehat{\mathbf{\Lambda}}_{2,\gamma}, \mathbf{G}_\gamma$ 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

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

and recall that $\widehat{\mathbf{x}}_i = \widehat{\mathbf{g}}_i = \mathbf{g}_i - \widehat{\mathbf{v}}_i$. Note that $\frac{1}{n}\sum_i(\widehat{\mathbf{x}}_i\widehat{\mathbf{x}}_i' I(q_i \leq \gamma)) \xrightarrow{p} E\left(\mathbf{g}_i\mathbf{g}_i' I(q_i \leq \gamma)\right)$ 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

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

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

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

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

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 $\gamma \in [\gamma_0, \overline{\gamma}]$

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

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

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

Proof of Lemma 2

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

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

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

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

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

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

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

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

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

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

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

B Proof of Consistency

We can express (2.23) in matrix notation

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

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

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

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

$$\begin{aligned} S^*(\gamma,\gamma_0) &= \mathbf{c}_\delta'\mathbf{M}_{\mathbf{XG}}(\gamma_0,\gamma)'\mathbf{M}(\gamma)^{-1}\mathbf{M}_{\mathbf{XG}}(\gamma_0,\gamma)\mathbf{c}_\delta \\ &\quad + c_\kappa\mathbf{M}_{\mathbf{X\Lambda}}(\gamma_0,\gamma)'\mathbf{M}(\gamma)^{-1}\mathbf{M}_{\mathbf{X\Lambda}}(\gamma_0,\gamma)c_\kappa \\ &\quad + c_\kappa\mathbf{M}_{\mathbf{X\Lambda}}^\perp(\gamma_0,\gamma)'\mathbf{M}^\perp(\gamma)^{-1}\mathbf{M}_{\mathbf{X\Lambda}}^\perp(\gamma_0,\gamma)c_\kappa \\ &\quad + 2\mathbf{c}_\delta'\mathbf{M}_{\mathbf{XG}}(\gamma_0,\gamma)'\mathbf{M}(\gamma)^{-1}\mathbf{M}_{\mathbf{X\Lambda}}(\gamma_0,\gamma)c_\kappa \end{aligned} \tag{B.9}$$

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

$$S^*(\gamma,\gamma_0) = \mathbf{c}'\mathbf{M}(\gamma_0,\gamma)'\mathbf{M}(\gamma)^{-1}\mathbf{M}(\gamma_0,\gamma)\mathbf{c} + \mathbf{c}'\widetilde{\mathbf{M}}^\perp(\gamma_0,\gamma)'\mathbf{M}^\perp(\gamma)^{-1}\widetilde{\mathbf{M}}^\perp(\gamma_0,\gamma)\mathbf{c} \tag{B.10}$$

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, $S^*(\gamma,\gamma_0) \le S^{**}(\gamma,\gamma_0)$ 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 $S^{**}(\cdot)$.

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

$$\frac{d\mathbf{M}(\gamma)}{d\gamma} = E(\mathbf{x}_i(\gamma)\mathbf{x}_i(\gamma)')|q = \gamma) f_q(\gamma) = D_1(\gamma) f_q(\gamma) \tag{B.11}$$

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

$$\frac{d\mathbf{M}^{\perp}(\gamma)}{d\gamma} = -E((\mathbf{x}_i(\gamma)\mathbf{x}_i(\gamma)')|q = \gamma) f_q(\gamma) = -D_1(\gamma) f_q(\gamma) \tag{B.12}$$

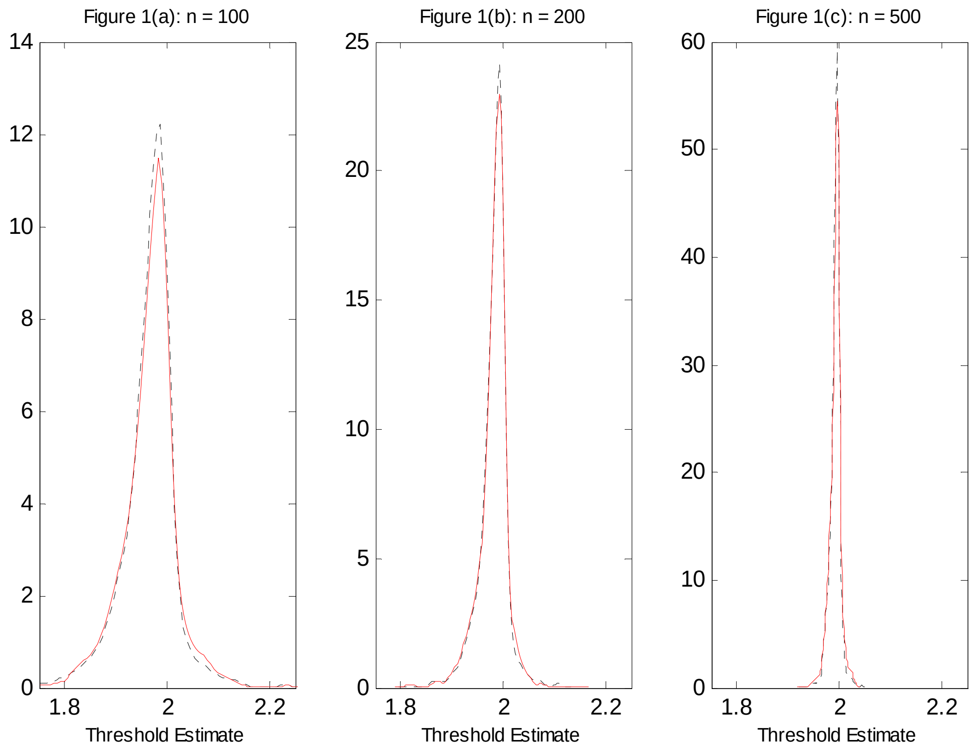
Then, using (B.11) and (B.12)

$$\begin{aligned} \tfrac{dS^{**}(\gamma,\gamma_0)}{d\gamma} = -\mathbf{c}' \left(\mathbf{M}(\gamma_0) + \widetilde{\mathbf{M}}^{\perp}(\gamma_0)\right) (\mathbf{M}(\gamma)^{-1} D_1(\gamma) f_q(\gamma) \mathbf{M}(\gamma)^{-1} \\ -\mathbf{M}^{\perp}(\gamma)^{-1} D_1(\gamma) f_q(\gamma) \mathbf{M}^{\perp}(\gamma)^{-1}) \left(\mathbf{M}(\gamma_0) + \widetilde{\mathbf{M}}^{\perp}(\gamma_0)\right) \mathbf{c} < 0 \end{aligned}$$

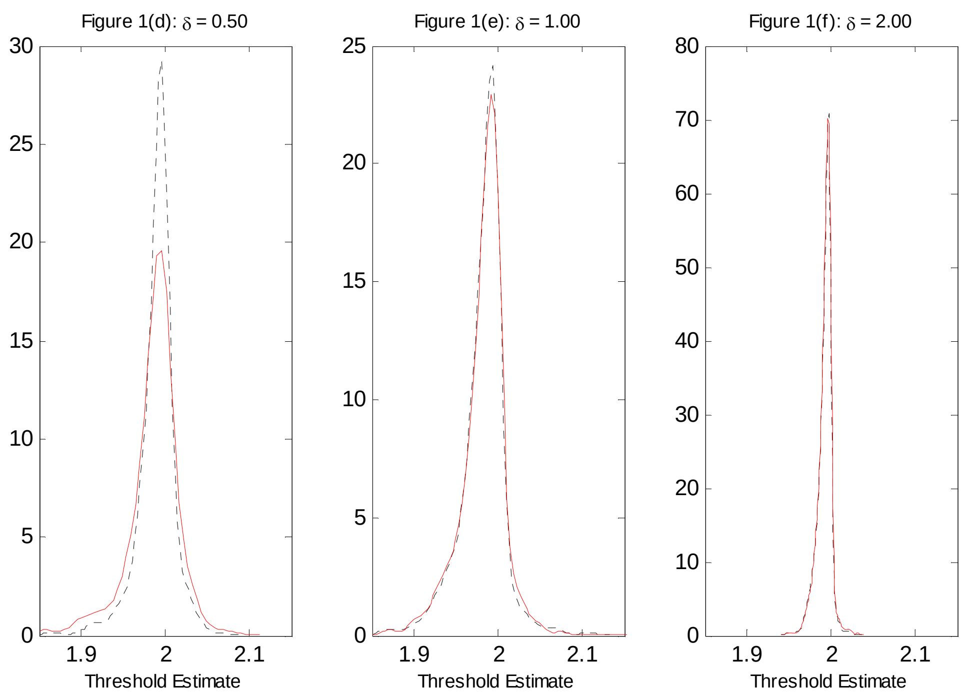
is continuous and weakly decreasing on $\gamma \in [\gamma_0, \overline{\gamma}]$ since $\mathbf{c}' D_1(\gamma) f_q(\gamma) \mathbf{c} > 0$ by Assumption (1.7), and $\alpha' \left(\mathbf{M}^{\perp}(\gamma) - \mathbf{M}(\gamma)\right) \alpha > 0$ for any α since $\lambda_{1,i}(\gamma) \leq \lambda_{2,i}(\gamma)$ 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 $\widehat{\gamma} \xrightarrow{p} \gamma_0$.

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

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



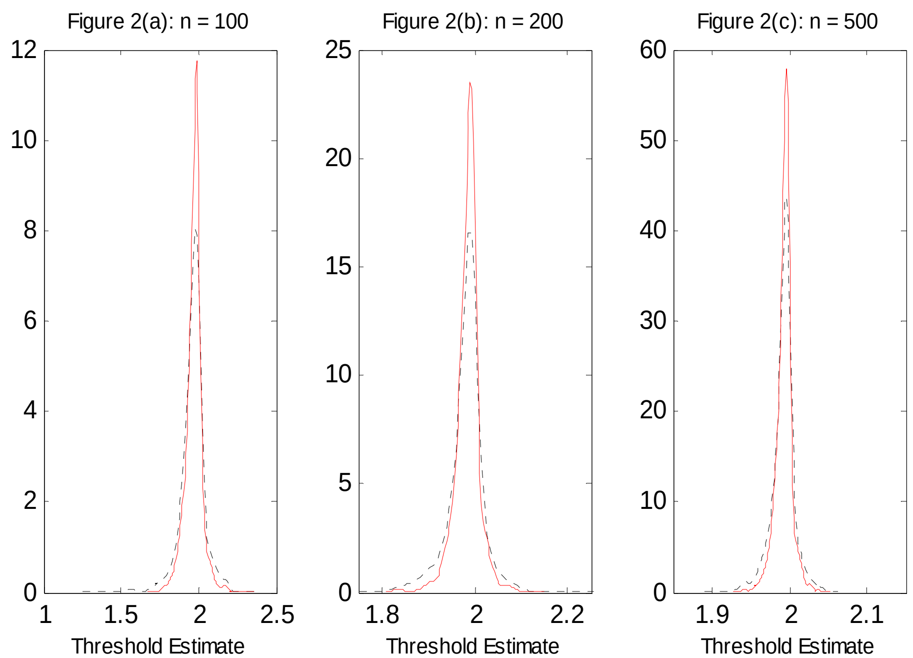
Estimates based on THRET and TR for n =500 and various values of δ



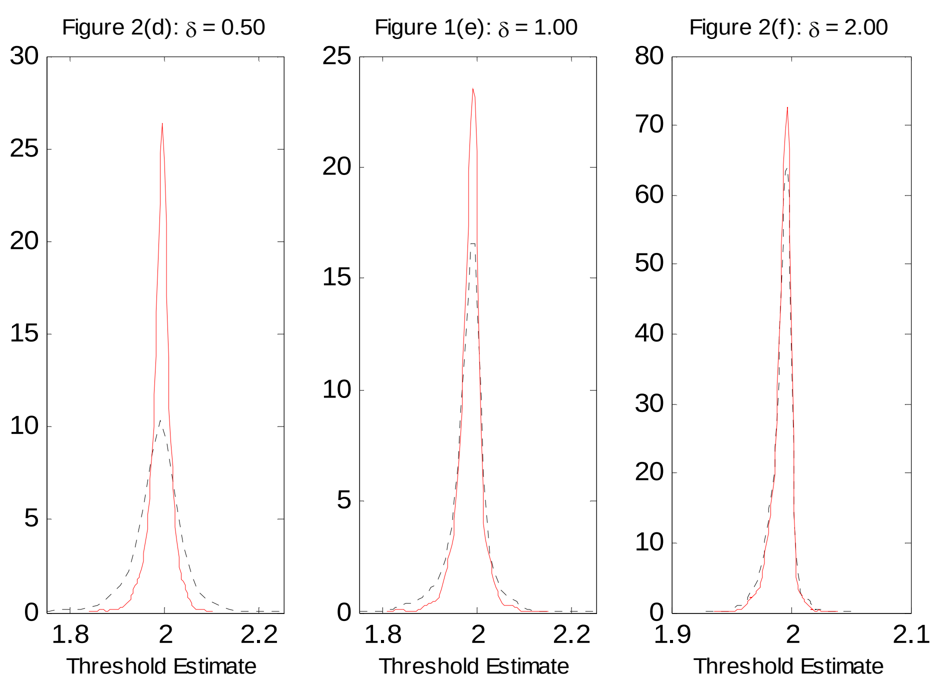
[⊥] The solid line represents the MC kernel density of the THRET threshold estimate while the dotted line represents the corresponding density for the TR (Hansen, 2000) threshold estimate.

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

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



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



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

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

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

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

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

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

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