Abstract

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

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

A crucial assumption in all the studies of the current literature is that the threshold variable is exogenous. This assumption severely limits the usefulness of threshold regression models in practice, since in economics many plausible threshold variables are endogenous. For example, our proposed threshold regression model can be used to understand the effect of the rapid rise in government debt in the wake of the recent financial crisis. Reinhart and Rogoff (2010) suggest that a country’s economic growth is negatively affected when gross public debt exceeds a threshold. One way to allow for nonlinearities in the regression without over-parameterization is to employ a threshold regression using public debt as a threshold. The process of public debt is not exogenous but rather depends on a range of factors including lack of commitment, political disagreement, and macroeconomic policy. Another example revisits the role of institutions in economic growth. 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. These are just two examples of endogenous thresholds, in fact there is a large number of potential applications in both economic and financial applications.

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.

In this paper we introduce the Structural Threshold Regression (STR) model that allows for the endogeneity of the threshold variable as well as the slope regressors. Just as in the limited
dependent variable framework, STR includes a set of inverse Mills ratio bias correction terms to restore the orthogonality of the errors. In particular, STR can be viewed as a threshold regression with restrictions across regimes. This problem is however nontrivial for two reasons. First, the estimates cannot be analyzed using results obtained regime by regime in the presence of restrictions across regimes and second the errors of STR are regime specific heteroskedastic. To overcome the first problem we explore the relationship between the restricted and unrestricted sum of squared errors. The second problem gives rise to an asymmetric distribution for the threshold estimate.

Interestingly, we show that the threshold estimate has the same properties with or without restrictions, which implies that ignoring the endogeneity in threshold will result in the same estimates and inference. To put it differently, if one ignores the endogeneity in threshold and employ existing methods as in Hansen (2000) and Caner and Hansen (2004) the result will be the same as far as the properties of the threshold are concerned. Our finding is similar to the result of Perron and Qu (2006) who consider structural change models with restrictions across regimes. The story is different in the estimates of the slope parameters, which suffer from bias and inefficiency when one ignores the endogeneity in the threshold.

We propose to estimate the threshold parameter using a concentrated least squares method and the slope estimates using 2SLS or GMM. We derive the consistency and asymptotic distribution of the threshold estimate under the assumption of the diminishing threshold effect. The asymptotic distribution of the threshold estimate is nonstandard because the threshold parameter is not identified under the null. More precisely, it involves two independent Brownian motions with two different scales and two different drifts. While these parameters are in principle estimable, inverting the likelihood ratio to obtain a confidence interval is not trivial as it involves a nonlinear algorithm. Instead, we employ a bootstrap inverted likelihood ratio approach. To examine the finite sample properties of our estimators we provide a Monte Carlo analysis.

Our research is related to several papers in the literature; see for example Hansen (2000) and Caner and Hansen (2004), Seo and Linton (2007), Gonzalo and Wolf (2005), and Yu (2010, 2011). The main difference of all these papers with our work is that they maintain the assumption that the threshold variable is exogenous. In particular our paper is related to Hansen (2000) and Caner and Hansen (2004) who develop inference for the threshold model under the assumption that the threshold effect becomes smaller as the sample increases. This assumption is the key to overcome a problem that was first realized by Chan (1993). While Chan proves that the threshold estimate is superconsistent, the asymptotic distribution of the threshold estimate turns out to be too complicated for inference as it depends on nuisance parameters, including the marginal distribution of the regressors and all the regression coefficients. The major difference between Hansen (2000) and Caner and Hansen (2004) and our paper is that if the threshold variable is endogenous, their approach will yield inconsistent slope coefficients for the two regimes because they omit the inverse Mills ratio bias.
correction terms. Another difference is while they exclude regime specific heteroskedasticity, we allow for it.

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

The paper is organized as follows. Section 2 describes the model and the setup. Section 3 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 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 (STR),

\[
y_i = \beta_1^\prime x_i + u_i, \quad q_i \leq \gamma
\]

\[
y_i = \beta_2^\prime x_i + u_i, \quad q_i > \gamma
\]

where \(E(u_i|z_i) = 0\). 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 \(\gamma\) being the sample split (threshold) value. The selection equation that determines which regime applies is given by

\[
q_i = \pi_i^\prime z_i + v_{qi}
\]

where \(E(v_{qi}|z_i) = 0\).

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 $x_i = (x_{1i}, x_{2i})$ where $x_{1i}$ are endogenous and $x_{2i}$ are exogenous and the $l \times 1$ vector of instrumental variables $z_i = (z_{1i}, z_{2i})$ where $x_{2i} \in z_i$. If both $q_i$ and $x_i$ are exogenous then we get the threshold model studied by Hansen (2000). If $q_i$ and $x_{2i}$ are exogenous and $x_{1i}$ is not a null set, then we get the threshold model studied by Caner and Hansen (2004). If $v_{qi} = 0$ then we get the smoothed exogenous threshold model as in Seo and Linton (2005), which allows the threshold variable to be a linear index of observed variables. In this paper we focus on the case where $q_i$ is endogenous and the general case where $x_{1i}$ is not a null set$^1$.

By defining the indicator function

$$I(q_i \leq \gamma) = \begin{cases} 
1 & \text{iff } q_i \leq \gamma \Leftrightarrow v_{qi} \leq \gamma - z_i' \pi_q : \text{Regime 1} \\
0 & \text{iff } q_i > \gamma \Leftrightarrow v_{qi} > \gamma - z_i' \pi_q : \text{Regime 2}
\end{cases}$$

(2.4)

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

$$y_i = \beta' x_{1i} x_i I(q_i \leq \gamma) + \beta' x_{2i} x_i I(q_i > \gamma) + u_i$$

(2.5)

The reduced form model$^2$, $g_{xi} \equiv g_x(z_i; \pi) = E(x_i | z_i) = \Pi' z_i$, is given by

$$x_i = \Pi' z_i + v_{xi}$$

(2.6)

$$E(v_{xi} | z_i) = 0$$

(2.7)

---

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

$^2$One may easily consider alternative reduced form models, such as a threshold model; see Caner and Hansen (2004).
We assume joint normality of the errors such that
\[
\begin{pmatrix}
u_i \\
v_{q,i} \\
v_i
\end{pmatrix}
\sim N
\left(
\begin{pmatrix}
0 \\
0 \\
0
\end{pmatrix},
\begin{pmatrix}
\sigma_u^2 & \sigma_{uv} & \sigma_u \\
\sigma_{uv} & \sigma_{vq} & \sigma_v \\
\sigma_u & \sigma_v & \Sigma_v
\end{pmatrix}
\right).
\tag{2.8}
\]

Then assuming that \(\sigma_{vq} = 0\) we can then obtain the following conditional expectations
\[
E(y_i|z_i, q_i \leq \gamma) = \beta'_{x1}g_{x_i} + E(u_i|z_i, q_i \leq \gamma) = \beta'_{x1}g_{x_i} + \kappa \lambda_{1i}(\gamma)
\tag{2.9}
\]
\[
E(y_i|z_i, q_i > \gamma) = \beta'_{x2}g_{x_i} + E(u_i|z_i, q_i > \gamma) = \beta'_{x2}g_{x_i} + \kappa \lambda_{2i}(\gamma)
\tag{2.10}
\]
where \(\lambda_{1i}(\gamma) = \lambda_1(\gamma - z_i'\pi_q) = -\frac{\phi(\gamma - z_i'\pi_q)}{\Phi(\gamma - z_i'\pi_q)}\) and \(\lambda_{1i}(\gamma) = \lambda_2(\gamma - z_i'\pi_q) = \frac{\phi(\gamma - z_i'\pi_q)}{1 - \Phi(\gamma - z_i'\pi_q)}\) are the inverse Mills ratio bias correction terms and \(\phi(\cdot)\) and \(\Phi(\cdot)\) are the normal pdf and cdf, respectively\(^3\).

Then we can write the STR model as follows
\[
y_i = \beta'_{x1}g_{x_i} + \kappa \lambda_{1i}(\gamma) + e_{1i}, \quad q_i \leq \gamma
\tag{2.11}
\]
\[
y_i = \beta'_{x2}g_{x_i} + \kappa \lambda_{2i}(\gamma) + e_{2i}, \quad q_i > \gamma
\tag{2.12}
\]
or
\[
y_i = (\beta'_{x1}g_{x_i} + \kappa \lambda_{1i}(\gamma)) I(q_i \leq \gamma) + (\beta'_{x2}g_{x_i} + \kappa \lambda_{2i}(\gamma)) I(q_i > \gamma) + e_i
\tag{2.13}
\]
where
\[
e_i = (\beta'_{x1}v_{x_i} - \kappa \lambda_{1i}(\gamma)) I(q_i \leq \gamma) + (\beta'_{x2}v_{x_i} - \kappa \lambda_{2i}(\gamma)) I(q_i > \gamma) + u_i
\tag{2.14}
\]

A few remarks are in order. First, note that when the error structure in the two regimes (2.2) and (2.1) is different \(u_1 \neq u_2\) then the slope coefficient of the inverse Mills ratio terms \(\kappa_1\) and \(\kappa_2\) can be different across the two regimes \(\kappa_1 \neq \kappa_2\). Here, for simplicity we assume \(\kappa_1 = \kappa_2\) but our results carry over to the more general case. Second, when \(\kappa = 0\), this model nests Caner and Hansen’s TR model and if additionally \(x_i\) is exogenous then it coincides with Hansen (2000)’s TR model. In general, the main difference with both of the above cases is that in the latter the inverse Mills ratio bias correction term is omitted and as we will be arguing below this yields inconsistent estimates of the slope parameters \(\beta_{x1}\) and \(\beta_{x2}\).

In the following section we propose a consistent profile estimation procedure for STR that takes

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

Let \( \lambda_i(\gamma) = \lambda_{1i}(\gamma) I(q_i \leq \gamma) + \lambda_{2i}(\gamma) I(q_i > \gamma) \) and define \( \delta_{x,n} = \beta_{x1} - \beta_{x2} \). For estimation purposes it is easier to write the regression model as

\[
y_i = g'_i x_i \beta_x + g'_i x_i I(q_i \leq \gamma) \delta_{x,n} + \lambda_i(\gamma) \kappa_n + e_i
\]

or

\[
y_i = g'_i x_i \beta_x + g'_i(\gamma) \delta_n + e_i
\]

with regression parameters \( g_i(\gamma) = (g'_i x_i I(q_i \leq \gamma), \lambda_i(\gamma))' \), \( \beta_x = \beta_{x2} \), and \( \delta_n = (\delta_{x,n}, \kappa_n)' \) and \( E(e_i|z_i) = 0 \).

First, we estimate by LS the reduced models (2.3) and (2.6) to obtain \( \hat{\Pi} \) and \( \hat{\pi}_q \). The fitted values are then given by \( \hat{q}_i = \pi'_i z_i \) and \( \hat{\xi}_i = \hat{\pi}_i x_i = \hat{\Pi} z_i \) along with first stage residuals as \( \hat{v}_{xi} = x_i - \hat{\xi}_i \) and \( \hat{v}_{qi} = q_i - \hat{q}_i \), respectively. We can also define the following functions of \( \gamma \), \( \hat{\lambda}_{1i}(\gamma) = \lambda_1(\gamma - z'_i \hat{\pi}_q) \) and \( \hat{\lambda}_{2i}(\gamma) = \lambda_2(\gamma - z'_i \hat{\pi}_q) \), and \( \hat{\lambda}_i(\gamma) = \hat{\lambda}_{1i}(\gamma) I(q_i \leq \gamma) + \hat{\lambda}_{2i}(\gamma) I(q_i > \gamma) \).

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

\[
\hat{\gamma} = \arg \min_{\gamma} S_n(\gamma)
\]

where

\[
S_n(\gamma) = \sum_{i=1}^{n} (y_i - \hat{g}'_{x_i} \beta_x - \hat{g}'_{x_i} I(q_i \leq \gamma) \delta_{x,n} - \hat{\lambda}_i(\gamma) \kappa_n)^2 = \sum_{i=1}^{n} (y_i - \hat{g}'_{x_i} \beta_x - \hat{g}'(\gamma) \delta_n)^2.
\]

Finally, once we obtain the split samples implied by \( \hat{\gamma} \), 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 \( \gamma \), estimation in each regime mirrors the Heckman (1979) sample selection bias correction model, the Heckit model.

Let \( \bar{X}_\gamma \) and \( \bar{X}_\perp \) the matrices of stacked vectors \( \bar{x}_{x_i(\gamma)} = (x_i' I(q_i \leq \hat{\gamma}), \lambda_{1,i}(\gamma) I(q_i \leq \hat{\gamma}))' \) and \( \bar{x}_{i(\hat{\gamma})} = (x_i' I(q_i > \hat{\gamma}), \lambda_{2,i}(\hat{\gamma}) I(q_i > \hat{\gamma}))' \). Similarly, \( \bar{Z}_\gamma \) and \( \bar{Z}_\perp \) denote the matrices of stacked vectors \( \bar{z}_\gamma = z_i I(q_i \leq \hat{\gamma}) \) and \( \bar{z}_\perp = z_i I(q_i > \hat{\gamma}) \).

Next, define the following matrices \( \bar{X} = (\bar{X}_\gamma, \bar{X}_\perp) \) and \( \bar{Z} = (\bar{Z}_\gamma, \bar{Z}_\perp) \) and the weight matrix \( \hat{W} \). Note that these matrices have a block diagonal form due to the indicator function. Then, we can
define the class of GMM estimators \( \tilde{\beta} = (\tilde{\beta}_1', \tilde{\beta}_2')' \) for \( \beta_1 = (\beta_{x1}', \kappa_1)' \) and \( \beta_2 = (\beta_{x2}', \kappa_2)' \)

\[
\tilde{\beta} = (\hat{X}'\hat{Z}\hat{W}\hat{Z}'\hat{X})^{-1}\hat{X}'\hat{Z}\hat{W}\hat{Z}'\hat{Y}. \tag{2.19}
\]

When \( \hat{W} = (\hat{Z}'\hat{Z})^{-1} \) we obtain the 2SLS estimator \( \tilde{\beta}_{2\text{SLS}} = (\beta_{1\text{SLS}}', \beta_{2\text{SLS}}')' \)

\[
\tilde{\beta}_{2\text{SLS}} = (\hat{X}'\hat{Z}(\hat{Z}'\hat{Z})^{-1}\hat{Z}'\hat{X})^{-1}\hat{X}'\hat{Z}(\hat{Z}'\hat{Z})^{-1}\hat{Z}'\hat{Y}. \tag{2.20}
\]

The 2SLS residual is given by \( \tilde{e}_{i,2\text{SLS}} = y_i - x_i(\hat{\gamma})'\beta_{1\text{SLS}} - x_i(\hat{\gamma})'\beta_{2\text{SLS}} \).

Define \( \hat{\Sigma} = \sum_{i=1}^{n} z_i z_i' \tilde{e}_{i,2\text{SLS}}, j = 1, 2 \). When \( \hat{W} = \hat{\Sigma}^{-1} = \text{diag}(\hat{\Sigma}_1^{-1}, \hat{\Sigma}_2^{-1}) \) then we obtain the efficient GMM estimator

\[
\tilde{\beta}_{GMM} = (\hat{X}'\hat{Z}\hat{\Sigma}^{-1}\hat{Z}'\hat{X})^{-1}\hat{X}'\hat{Z}\hat{W}\hat{Z}'\hat{Y}. \tag{2.21}
\]

### 3 Threshold Regression with Restrictions

In this section we rewrite the STR model in equation (2.16) as a threshold regression subject to restrictions. We do so because the restricted problem above cannot be analyzed using results obtained regime by regime. Therefore we first analyze the unrestricted threshold regression and then relate to the restricted problem by using the relationship between restricted and unrestricted problem. In particular, the unrestricted problem generalizes Caner and Hansen (2004) by including both inverse mills ratio terms in both regimes.

Define \( \lambda_i(\gamma) = (\lambda_{i1}(\gamma), \lambda_{i2}(\gamma))' \), \( \beta_\lambda = \beta_{\lambda 2} = (\kappa_{21}, \kappa_{22})' \), \( \beta_{\lambda 1} = (\kappa_{11}, \kappa_{12})' \), and \( \delta_{\lambda n} = \beta_{\lambda 1} - \beta_{\lambda 2} \).

Then the unrestricted model takes the form

\[
y_i = g_{xi}'\beta_x + \lambda_i(\gamma)'\beta_\lambda + g_{xi}'I(q_i \leq \gamma)\delta_{xn} + \lambda_i(\gamma)'I(q_i \leq \gamma)\delta_{\lambda n} + e_i, \tag{3.22}
\]

where \( \beta = (\beta_x', \beta_\lambda')' \) and \( \delta_n = (\delta_{xn}', \delta_{\lambda n}')' \) are the slope coefficients and \( e_i \) is the error of the unrestricted threshold model

\[
e_i = (g_{xi}'\beta_{x1} - \lambda_i(\gamma)'\beta_{\lambda 1}) I(q_i \leq \gamma) + (g_{xi}'\beta_{x2} - \lambda_i(\gamma)'\beta_{\lambda 2}) I(q_i > \gamma) + u_i. \tag{3.23}
\]

It is easy to verify that the STR model in equation (2.16) is a special case of (3.22) under the following restrictions

\[
\kappa_{12} = \kappa_{21} = 0 \tag{3.24}
\]
and
\[ \kappa_{11} = \kappa_{22} = \kappa. \] (3.25)

In general, let \( g_i(\gamma) = (g'_x_i, \lambda_i(\gamma)', \theta = (\beta', \delta_n)' \). Then we can write the model as
\[ y_i = g'_i(\gamma)\beta + g'_i(\gamma)I(q_i \leq \gamma)\delta_n + e_i, \] (3.26)

subject to the restriction
\[ R'\theta = \theta \] (3.27)

with \( R \) a \( 2q \times r \) matrix of rank \( r \) and \( \theta \) a \( r \) dimensional vector of constants and \( e_i = e_1I(q_i \leq \gamma) + e_2I(q_i > \gamma) \).

Then, the estimate of the threshold parameter \( \gamma \) can be viewed as the minimizer of the unconstrained Concentrated Least Squares (CLS) problem subject to the constraint in equation (3.27).

Consider the vector \( x_i(\gamma) = (x_i, \lambda_1(\gamma), \lambda_2(\gamma))' \). Let \( \bar{X}_\gamma \) and \( \bar{X}_\perp \) the matrices of stacked vectors \( \bar{x}_i(\gamma) = (x'_iI(q_i \leq \gamma), \lambda_{1,i}(\gamma)I(q_i \leq \gamma))' \) and \( \bar{x}_i(\bar{\gamma}) = (x'_iI(q_i > \bar{\gamma}), \lambda_{2,i}(\bar{\gamma})I(q_i > \bar{\gamma}))' \). Similarly, \( \bar{z}_\gamma \) and \( \bar{z}_\perp \) denote the matrices of stacked vectors \( \bar{z}_\gamma = z_iI(q_i \leq \gamma) \) and \( \bar{z}_\perp = z_iI(q_i > \gamma) \). Define further \( \bar{X} = (\bar{X}_\gamma, \bar{X}_\perp) \) and \( \bar{Z} = (\bar{Z}_\gamma, \bar{Z}_\perp) \) and the weight matrix \( \bar{W} \). Then, the relationship between the unrestricted and restricted slope coefficients is given by
\[ \bar{\beta} = \hat{\beta} - \bar{W}R \left( R'\bar{W}R \right)^{-1} (R'\hat{\beta} - \theta) \] (3.28)

where \( \hat{\beta} \) is the unconstrained vector of slope estimators that correspond to equation (2.19).

Below we proceed with inference by presenting the assumptions for the general threshold regression.

4 Inference

Define \( \bar{g}_i = \sup_{\gamma \in F} |g_i(\gamma)| \) and \( \bar{g}_i |e_i| = \sup_{\gamma \in F} |g_i(\gamma)e_i| \). Then define the moment functionals
\[
\begin{align*}
M(\gamma) &= E(g_i(\gamma)g_i(\gamma)'), \\
D(\gamma) &= E(g_i(\gamma)g_i(\gamma)' | q_i = \gamma), \\
\Omega(\gamma) &= E(g_i(\gamma)g_i(\gamma)' e_i^2 | q_i = \gamma).
\end{align*}
\]

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

$$D_1 = \lim_{\gamma \to \gamma_0} D(\gamma) = \lim_{\gamma \to \gamma_0} E(g_i(\gamma)g_i(\gamma)'|q_i = \gamma)$$

$$D_2 = \lim_{\gamma \to \gamma_0} D(\gamma) = \lim_{\gamma \to \gamma_0} E(g_i(\gamma)g_i(\gamma)'|q_i = \gamma)$$

$$\Omega_1 = \lim_{\gamma \to \gamma_0} \Omega(\gamma) = \lim_{\gamma \to \gamma_0} E(g_i(\gamma)g_i(\gamma)'e_i^2|q_i = \gamma)$$

$$\Omega_2 = \lim_{\gamma \to \gamma_0} \Omega(\gamma) = \lim_{\gamma \to \gamma_0} E(g_i(\gamma)g_i(\gamma)'e_i^2|q_i = \gamma)$$

Then define the moment functionals

**Assumption 1**

(1.1) \{z_i, g_i(\gamma), u_i, v_i, v_{qi}\} is strictly stationary and ergodic with $\rho$ mixing coefficients $\sum_{m=1}^{\infty} \rho_m^{1/2} < \infty$.

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

(1.3) $E(v_i|F_{i-1}) = 0$,

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

(1.5) for all $\gamma \in \Gamma, E(|\mathbf{g}_i|^4|q_i = \gamma) \leq C, \lim_{\gamma \to \gamma_0} E(|\mathbf{g}_i|^4|q_i = \gamma) \leq C, \lim_{\gamma \to \gamma_0} E(|\mathbf{g}_i(\gamma)|^4e_i^2|q_i = \gamma) \leq C$, and for some $C < \infty$,

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

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

(1.8) $\delta_n = \beta_1 - \beta_2 = \mathbf{c}n^{-\alpha} \to 0, \mathbf{c} \neq 0, \alpha \in (0, 1/2)$,

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

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

(1.11) for all $\gamma \in \Gamma$, $\hat{\gamma} = \arg\min_{\gamma \in \Gamma} \sum_{i=1}^{n}(y_i - g_i'(\gamma)\beta - g_i'(\gamma)I(q_i \leq \gamma)\delta_n)^2$ exists and it is unique. Furthermore, $\hat{\gamma}$ lies in the interior of $\Gamma$, with $\Gamma$ compact and convex.

This set of assumptions is similar to Hansen (2000) and Caner and Hansen (2004). While most assumptions are rather standard, Assumption 1.8 is not. Assumption 1.8 assumes that a “small threshold” asymptotic framework in the sense that $\delta_n = (\delta_{\lambda_n}, \delta_{\lambda_n}')$ will tend to go 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. In the case of the STR model in equation (2.16), this assumption also implies that \( \kappa_n \) vanishes as \( n \to \infty \) at the same rate as \( \delta_{\lambda_n} \) to ensure that the bias correction (i.e. the inverse Mills ration terms) to the endogeneity of the threshold will not be present when the model is linear. Assumption 1.11 is satisfied given the monotonicity of the inverse Mills ratios.

The above assumptions are also sufficient to guarantee that the first stage regressions are consistent for the true conditional means i.e. \( \hat{r} = (\hat{r}_{x}, \hat{r}_{u1}, \hat{r}_{u2}) = g_i(\gamma) - \hat{g}_0(\gamma) = o_p(1) \). Note that we have \( p \) linear regressions, \( g_{xi} = \Pi z_i \), and two nonlinear \( g_{a1} = \lambda_{1}(\gamma) \), \( g_{u2} = \lambda_{2}(\gamma) \).

4.1 Threshold Estimate

**Proposition 4.1 Consistency of \( \hat{\gamma} \)**

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

\[
\hat{\gamma} \xrightarrow{p} \gamma_0
\]

The proof is given in the appendix.

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

**Remark 1** When we ignore the endogeneity in the threshold we would still get a consistent estimator for \( \gamma_0 \).

**Remark 2** When we ignore the endogeneity in slope the CLS will still yield a consistent estimator for \( \gamma_0 \).

**Remark 3** Although the endogeneity in the threshold does not generate bias in the threshold estimate, it does yield a bias for the estimation of the slope coefficients. As in the standard omitted variable case the bias will depend on the degree of correlation between the omitted inverse Mills ratio term.

To obtain the asymptotic distribution let us first define two independent standard Wiener processes \( W_1(s) \) and \( W_2(s) \) defined on \([0, \infty)\).
Let
\[ T(s) = \begin{cases} 
-\frac{1}{2}|s| + W_1(-s), & \text{if } v \leq 0 \\
-\frac{1}{2}\xi|s| + \sqrt{\phi}W_2(s), & \text{if } v > 0
\end{cases}, \]
where \( \xi = \frac{\mathbf{e}'D_1\mathbf{c}}{\mathbf{c}'D_1\mathbf{c}} \), and \( \varphi = \frac{\mathbf{e}'\Omega_1\mathbf{c}}{\mathbf{c}'\Omega_1\mathbf{c}} \). The case of the asymmetric two sided Brownian motion argmax distribution with unequal variances was first examined by Stryhn (1996).

**Theorem 4.1 Asymptotic Distribution of \( \hat{\gamma} \)**

Under Assumption 1
\[ n^{1-2\alpha}(\hat{\gamma} - \gamma_0) \xrightarrow{d} \omega T \]
where \( \omega = \frac{\mathbf{e}'\Omega_1\mathbf{c}}{(\mathbf{c}'D_1\mathbf{c})^{1/2}} \) and \( T = \arg\max_{-\infty<s<\infty} T(s) \). The proof is given in the appendix.

The distribution function of \( T \) is given by Bai (1997). For \( x < 0 \), the cdf of \( T \) is given by
\[
P(T \leq x) = -\sqrt{\frac{|x|}{2\pi}} \exp\left(-\frac{|x|}{8}\right) - c \exp(\mathbf{a}|x|) \Phi(-b\sqrt{|x|}) + (d - 2 - \frac{|x|}{2}) \Phi(-\frac{\sqrt{|x|}}{2}),
\]
where \( a = \frac{1}{2} \xi (1 + \frac{\xi}{\varphi}),\ b = \frac{1}{2} + \frac{\xi}{\varphi},\ c = \frac{\varphi(\varphi + 2\xi)}{\xi(\varphi + \xi)}, \) and \( d = \frac{(\varphi + 2\xi)^2}{\xi(\varphi + \xi)} \). For \( x > 0 \),
\[
P(T \leq x) = 1 + \xi \sqrt{\frac{x}{2\pi\varphi}} \exp\left(-\frac{\xi^2x}{8\varphi}\right) - c \exp(\mathbf{a}|x|) \Phi(-b\sqrt{x}) + (-d + 2 - \frac{\xi^2x}{2\varphi}) \Phi(-\frac{\xi\sqrt{x}}{2\varphi}),
\]
where \( a = \frac{\varphi + \xi}{2},\ b = \frac{\varphi + \xi}{2\sqrt{\varphi}},\ c = \frac{\xi(\xi + 2\varphi)}{\varphi(\varphi + \xi)}, \) and \( d = \frac{(\xi + 2\varphi)^2}{\varphi(\varphi + \xi)} \). The distribution is not symmetric when \( \varphi \neq 1 \) or \( \xi \neq 1 \). In the case of \( \varphi = \xi = 1 \), we get the symmetric case; see for example Hansen (2000).

Note that a simpler case occurs when we assume regime specific heteroskedasticity but homoskedasticity within each regime. In this case we get \( \Omega_1 = \sigma_{\varepsilon_1}^2 \mathbf{D}_1,\ \Omega_2 = \sigma_{\varepsilon_2}^2 \mathbf{D}_2 \), where \( \sigma_{\varepsilon_1}^2 = E(\varepsilon_{1i}^2|q = \gamma),\ \sigma_{\varepsilon_2}^2 = E(\varepsilon_{2i}^2|q = \gamma) \). This implies that \( \omega = \frac{\sigma_{\varepsilon_1}^2}{\mathbf{e}'\Omega_1\mathbf{c}},\ \varphi = \frac{\sigma_{\varepsilon_2}^2}{\mathbf{e}'\Omega_2\mathbf{c}} \). Furthermore, note that when \( \mathbf{D}_1 = \mathbf{D}_2 = \mathbf{D},\ \Omega_1 = \Omega_2 = \Omega \) we obtain the leading case that excludes regime specific heteroskedasticity. In this case we obtain \( \xi = 1,\ \varphi = 1,\ \omega = \frac{\mathbf{e}'\Omega_1\mathbf{c}}{(\mathbf{c}'\Omega_1\mathbf{c})^{1/2}} \).

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

Next using the distributional result in Theorem 4.1 we can construct confidence intervals for \( \gamma_0 \). We consider the pseudo Likelihood Ratio (LR) statistic
\[
\text{LR}_n(\gamma) = n \frac{S_n(\gamma) - S_n(\hat{\gamma})}{S_n(\hat{\gamma})},
\]
where
Define
\[ \eta^2 = \frac{c' \Omega c}{(c' D c) \sigma^2} \]
and
\[ \psi = \sup_{-\infty < s < \infty} \left( \left( -\frac{1}{2} |s| + W_1(-s) \right) I(q < \gamma_0) + \left( -\frac{1}{2} |s| + \sqrt{\phi} W_2(s) \right) I(q > \gamma_0) \right) \]
Then we have the following theorem.

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

Under Assumption 1, the asymptotic distribution of the likelihood ratio test under the \( H_0 \) is given by
\[ LR_n(\gamma_0) \xrightarrow{d} \eta^2 \psi \]
(4.32)
where the distribution of \( \psi \) is \( P(\psi \leq x) = (1 - e^{-x/2})(1 - e^{-\xi x/2})\sqrt{\phi} \)

The proof is given in the appendix.

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

Define \( \hat{\Gamma} = \{ \gamma : LR_n(\gamma) \leq c \} \) and let \( 1 - a \) denote the desired asymptotic confidence level and let \( c = c_{\Psi}(1 - a) \) be the critical value for \( \psi \). Assuming \( \alpha = 1, \xi = \varphi = 1, \eta^2 = 1 \) and Gaussian errors we can invoke Theorem 3 of Hansen (2000) to show that the likelihood ratio test is asymptotically conservative. This implies at least in this special case inferences based on the confidence region \( \hat{\Gamma} \) are asymptotically valid. In general, although in principle, \( \eta^2, \xi, \) and \( \varphi \) can be estimated it is quite difficult to apply the test-inversion method of Hansen (2000) to construct an asymptotic confidence interval for \( \gamma_0 \) because there is no closed form solution for \( 1 - a = (1 - e^{-x/2})(1 - e^{-\xi x/2})\sqrt{\phi} \).
Therefore we propose to use a bootstrap inverted likelihood ratio approach that we describe next.

The bootstrap for the threshold regression model has been studied by Yu (2010) who shows that the parametric bootstrap is consistent while the nonparametric bootstrap is inconsistent for inference on the threshold estimate. The problem is that typically the parametric bootstrap is not feasible as one needs to specify a complete likelihood. To overcome this problem we follow Hansen (2000) who under the framework of an asymptotically diminishing threshold effect he shows that the confidence interval constructed by inverting the likelihood ratio statistic is asymptotically valid.
In this framework the bootstrap is valid. 4

Given consistent estimates for \((\delta_{X}, \tilde{\beta}^{b}_{X}, \tilde{\alpha}_{X}, \tilde{\beta}^{b}_{x}, \tilde{\lambda}_{i}(\gamma))\) we define the residuals of the STR model

\[ \tilde{e}_{i} = y_{i} - \tilde{g}_{X}^{b} \tilde{\beta} - \tilde{g}_{X}^{b} I(q_{i} \leq \gamma) \tilde{\delta}_{X} - \tilde{\lambda}_{i}(\gamma) \tilde{\alpha}_{X} \]

Then following Hansen (1996) we fix the regressors and define the bootstrap dependent variable \(y^{*}_{i} = \tilde{e}_{i}(\gamma) \zeta_{i} \), where \(\zeta_{i}\) is Normal \(i.i.d.\) and \(\tilde{e}_{i}\) is the recentered residual \(\tilde{e}_{i}\).

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

\[ LR^{*}_{n}(\gamma) = n \frac{S^{*}_{n}(\gamma) - S^{*}_{n}(\tilde{\gamma}^{*})}{S^{*}_{n}(\tilde{\gamma}^{*})} \]

We store likelihood ratio values from bootstraps \(\{LR_{n}^{*(1)}(\gamma), \ldots, LR_{n}^{*(B)}(\gamma)\}\) and sort them to determine the \(a(B+1)^{th}\) LR value, \(LR_{n}^{*}(e^{*}_{n})\) as the critical value for \(1 - a\) confidence level. Then we construct the bootstrapped inverted LR confidence region for \(\gamma_{0} \), \(\bar{\Gamma}^{*} = \{\gamma : LR_{n}(\gamma) \leq LR_{n}(e^{*}_{n})\}\), where \(LR_{n}(\gamma)\) is computed from the data.

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

4.2 Slope Parameters

Consider the unrestricted vector of covariates \(x_{i}(\gamma_{0}) = (x_{i}, \lambda_{1i}(\gamma_{0}), \lambda_{2i}(\gamma_{0}))^{t}\). Then , the inference on the slope parameters follows a restricted version of Caner and Hansen (2004). Let us define the following matrices:

\[ Q_{1} = E(z_{i}z_{i}^{t}I(q_{i} \leq \gamma_{0}), Q_{2} = E(z_{i}z_{i}^{t}I(q_{i} > \gamma_{0}) \]

\[ S_{1} = E(z_{i}x_{i}(\gamma_{0})^{t}I(q_{i} \leq \gamma_{0}), S_{2} = E(z_{i}x_{i}(\gamma_{0})^{t}I(q_{i} > \gamma_{0}) \]

\[ \Sigma_{1} = E(z_{i}z_{i}^{t}u_{i}^{2}I(q_{i} \leq \gamma_{0}), \Sigma_{2} = E(z_{i}z_{i}^{t}u_{i}^{2}I(q_{i} > \gamma_{0}) \]

\[ V_{1} = (S_{1}^{-1}Q_{1}^{-1})^{-1}S_{1}^{-1}Q_{1}^{-1}\Sigma_{1}Q_{1}^{-1}S_{1}(S_{1}^{-1}Q_{1}^{-1})^{-1} \]

4 Antoch et al (1995) established the validity of the nonparametric bootstrap under the assumptions of an asymptotically diminishing threshold \(i.i.d\) errors in the context of structural change models.
Theorem 4.3 Under Assumption 1, 

(a) 
\[ \sqrt{n}(\hat{\beta}_{2SLS} - \beta) \rightarrow N(0, \tilde{V}_{2SLS}) \]  

where 
\[ \tilde{V}_{2SLS} = V - Q^{-1}R(R'Q^{-1}R)^{-1}R'V - VR (R'Q^{-1}R)^{-1}R'Q^{-1} \\
+ Q^{-1}R(R'Q^{-1}R)^{-1}R'VR (R'Q^{-1}R)^{-1}R'Q^{-1}. \]  

(b) 
\[ \sqrt{n}(\hat{\beta}_{GMM} - \beta) \rightarrow N(0, \tilde{V}_{GMM}) \]  

where 
\[ \tilde{V}_{GMM} = V - VR (R'VR)^{-1}R'V \]

We should note that the asymptotic variances are functions of \( \gamma_0 \) since they are functions of the inverse Mills ratio terms.

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

\[ y_i = \begin{cases} 
\beta_{1,1} + \beta_{1,2} x_i + u_i, & q_i \leq 2 \\
\beta_{2,1} + \beta_{2,2} x_i + u_i, & q_i > 2 
\end{cases} \]  

(5.37)

where

\[ q_i = 2 + z_{1,i} + v_{q,i} \]  

(5.38)

with \( z_{1,i}, v_{q,i}, \varepsilon_i \sim NIID(0,1) \) and \( u_i = v_x + \kappa v_{q,i} + (0.1) N(0,1) \). The degree of endogeneity of the threshold variable is controlled by \( \kappa \), where \( \kappa = 0.01 \sqrt{\delta^2/(1-\delta^2)} \). We sampled \( v_x \) and \( v_q \) independently. We fix \( \bar{\kappa} = 0.95 \) and set \( \beta_{2,1} = \beta_{2,2} = \beta_2 = 1 \) and \( \beta_{1,1} = \beta_{1,2} = \beta_1 \), and vary \( \beta_1 \) 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. 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 \( v_i = 0.5 u_i \). 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 \( \gamma \). 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 \( \delta \) 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 \( \delta \) 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 have conducted a large number of experiments and the results are similar. Specifically, our experiments investigated a broader range of values of \( \delta \), different degrees of threshold endogeneity \( (\sigma_{uq_i}) \), 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 see that the performance of the threshold estimator of STR improves as $\delta$ 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 $\delta$ 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 $\beta_2$. As in the case of the threshold estimates we find that the performance of the slope coefficient estimate of STR improves as $\delta$ 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 $\hat{\beta}_2$ 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 converges 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 $\hat{\Pi}$ using 300 bootstrap replications. We report results where $\delta$ 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 $\delta$ 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 $\delta$ increases. Our bootstrap results are consistent with the simulation findings of Caner and Hansen (2004), which are based on the distribution theory.

6 Conclusion

In this paper we introduce the Structural Threshold Regression (STR) model that allows for the endogeneity of the threshold variable as well as the slope regressors. We develop a concentrated least squares estimator that deals with the problem of endogeneity in the threshold variable by including a correction term based on the inverse Mills ratios in each regime to produce. Using the asymptotic framework of Hansen (2000) and Caner and Hansen (2004), which relies on the assumption of the diminishing threshold effect we derive the asymptotic distribution of our estimators. Our proposed estimator performs well for a variety of sample sizes and parameter
combinations.
References


A Appendix

The model in matrix notation

Recall that $g_i(\gamma) = ((g_{x_i}, \lambda_1(\gamma), \lambda_2(\gamma)))'$ and use the following convention to denote a regime specific matrix $A_{\gamma} = \{a_iI(q_i \leq \gamma)\}$. Then let us write (2.15) in a matrix form by stacking $y_i, g_{x_i}, g_{x_i}'I(q_i \leq \gamma), \lambda_i(\gamma)$ and $e_i$ to define $Y, G_x, G_{x,\gamma}, \Lambda(\gamma)$, and $e$, respectively.

Let us write (2.15) in a matrix form by stacking $y_i, g_{i}(\gamma), g_{i}(\gamma)'I(q_i \leq \gamma), g_{i}(\gamma)'I(q_i > \gamma)$ and $e_i$ to define $Y, G(\gamma), G_{\gamma(\gamma)}, G_{\perp(\gamma)}$ and $e$, respectively.

\[
Y = G(\gamma)\beta + G_{\gamma(\gamma)}\delta_n + e 
\]  
(A.1)

or

\[
Y = G^*(\gamma)\beta^* + e
\]  
(A.2)

where $G^*(\gamma) = (G_{\gamma(\gamma)}, G_{\perp(\gamma)})$ and $\beta^* = (\beta_1', \beta_2')'$. 

Note that $\hat{x}_i = \hat{g}_{x_i}$ and hence we can define $\tilde{G}_x = \tilde{X}$. In a similar manner we can define the $\tilde{G}^*(\gamma) = (\tilde{G}_{\gamma(\gamma)}, \tilde{G}_{\perp(\gamma)})$, where $G_{\gamma}(\gamma) = (\tilde{G}_{x,\gamma}, \tilde{A}_{1,\gamma}(\gamma), \tilde{A}_{2,\gamma}(\gamma))$, $G_{\perp(\gamma)} = (\tilde{G}_{x,\perp}, \tilde{A}_{1,\perp}(\gamma), \tilde{A}_{2,\perp}(\gamma))$. Define for any $\gamma$ the following orthogonal matrices $\tilde{X}_{\gamma(\gamma)} = (\tilde{X}_{\gamma}, \tilde{A}_{1,\gamma}(\gamma), \tilde{A}_{2,\gamma}(\gamma))$ and $\tilde{X}_{\perp(\gamma)} = (\tilde{X}_{\perp}, \tilde{A}_{1,\perp}(\gamma), \tilde{A}_{2,\perp}(\gamma))$ that correspond to the orthogonal stacked vectors $\tilde{x}_{\gamma(\gamma)} = (\tilde{x}(\gamma), \tilde{A}_{1,\gamma}(\gamma), \tilde{A}_{2,\gamma}(\gamma))(\tilde{x}(\gamma), \tilde{A}_{1,\gamma}(\gamma), \tilde{A}_{2,\gamma}(\gamma))'$ and 

$\tilde{x}_{\perp(\gamma)} = (\tilde{x}(\gamma), \tilde{A}_{1,\gamma}(\gamma), \tilde{A}_{2,\gamma}(\gamma))(\tilde{x}(\gamma), \tilde{A}_{1,\gamma}(\gamma), \tilde{A}_{2,\gamma}(\gamma))'$. 

Note that in the absence of any restrictions $P_{\gamma(\gamma)}$ and $P_{\perp(\gamma)}$ are orthogonal such that $P^*(\gamma) = P_{\gamma(\gamma)} + P_{\perp(\gamma)}$, where $P^*(\gamma) = \tilde{x}^*(\gamma)(\tilde{x}^*(\gamma)'\tilde{x}^*(\gamma))^{-1}\tilde{x}^*(\gamma)'$ and $\tilde{x}^*(\gamma) = (\tilde{x}(\gamma), \tilde{X}_{\perp(\gamma)}, \tilde{X}_{\perp(\gamma)})$. The problem, however, is that this orthogonality collapses once we impose exclusions restrictions across the two regimes. Therefore, we cannot analyze the two regimes separately.

Let $\tilde{X}_{\gamma}$ and $\tilde{X}_{\perp}$ be the orthogonal stacked vectors of $\tilde{x}(\gamma)$ and $\tilde{x}(\gamma)'$, respectively. Let us also denote $G(\gamma_0) = (G_{x,0}, \Lambda(\gamma_0))$ the matrix at the true value $\gamma = \gamma_0$.

Finally, let us also define the second stage residuals $\tilde{e}_i = \tilde{x}_i\beta + e_i$ and its vector form $\tilde{e} = \tilde{x}\beta + e$.
**LEMMA 1.** For some $B < \infty$ and $\gamma \leq \gamma' \leq \gamma \leq \gamma$ and $r \leq 4$, uniformly in $\gamma$

\[
Eh_i^r(\gamma, \gamma') \leq B|\gamma - \gamma'| \quad \text{(A.3)}
\]
\[
Ek_i^r(\gamma, \gamma') \leq B|\gamma - \gamma'| \quad \text{(A.4)}
\]

**Proof of Lemma 1.**

Define $d_i(\gamma) = I_{\{\gamma_i \leq \gamma\}}$ and $d_i^+(\gamma) = I_{\{\gamma_i > \gamma\}}$. Define $h_i(\gamma, \gamma') = |(h_i(\gamma) - h_i(\gamma')) \varepsilon_i|$ and $k_i(\gamma, \gamma') = |(h_i(\gamma) - h_i(\gamma'))|$. In the case of the STR model in equation (2.13) $h_i(\gamma) = (g_i d_i(\gamma), \lambda_i(\gamma))$ and thus $h_i(\gamma, \gamma')$ takes the form

\[
h_i(\gamma, \gamma') = \left( \frac{|g_i \varepsilon_i| |d_i(\gamma) - d_i(\gamma')|}{|\lambda_i(\gamma) \varepsilon_i - \lambda_i(\gamma') \varepsilon_i|} \right)
\]

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

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

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

\[
= E|((\lambda_{2i}(\gamma) - \lambda_{2i}(\gamma') + (\lambda_{1i}(\gamma)d_i(\gamma) - \lambda_{1i}(\gamma')d_i(\gamma')) - (\lambda_{2i}(\gamma)d_i(\gamma) - \lambda_{2i}(\gamma')d_i(\gamma'))) \varepsilon_i|^r
\]

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

\[
\leq (E|\lambda_{2i} - \lambda_{2i} \varepsilon_i|^r)^{1/r} + (E|\lambda_{1i} - \lambda_{1i} \varepsilon_i|^r)^{1/r} + (E|\lambda_{2i} - \lambda_{2i} \varepsilon_i|^r)^{1/r}
\]

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

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

■

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

\[
\frac{1}{n} \hat{X}^*(\gamma) \hat{X}^*(\gamma) = \frac{1}{n} \sum_{i=1}^{n} \hat{X}_i^*(\gamma) \hat{X}_i^*(\gamma)^r \xrightarrow{p} M(\gamma) \quad \text{(A.5)}
\]

22
\[
\frac{1}{n} \hat{\mathbf{X}}^*(\gamma_0) G^*(\gamma_0) = \frac{1}{n} \sum_{i=1}^{n} \hat{x}_i^*(\gamma) \hat{x}_i^*(\gamma)' \xrightarrow{p} \mathbf{M}(\gamma_0) \quad (A.6)
\]
\[
\frac{1}{\sqrt{n}} \hat{\mathbf{X}}^*(\gamma)' \hat{\mathbf{e}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{x}_i^*(\gamma) \hat{e}_i = O_p(1) \quad (A.7)
\]

**Proof of Lemma 2.**

To show (A.5) note that
\[
\frac{1}{n} \hat{\mathbf{X}}_\gamma^*(\gamma)' \hat{\mathbf{X}}_\gamma^*(\gamma)
\]
\[
= \begin{pmatrix}
\frac{1}{n} \hat{\mathbf{X}}_\gamma^*(\gamma)' \hat{\mathbf{X}}_\gamma^*(\gamma) & \frac{1}{n} \hat{\mathbf{X}}_\gamma^*(\gamma)' \hat{\mathbf{A}}_{1\gamma}(\gamma) & \frac{1}{n} \hat{\mathbf{X}}_\gamma^*(\gamma)' \hat{\mathbf{A}}_{2\gamma}(\gamma) \\
\frac{1}{n} \hat{\mathbf{A}}_{1\gamma}(\gamma)' \hat{\mathbf{X}}_\gamma^*(\gamma) & \frac{1}{n} \hat{\mathbf{A}}_{1\gamma}(\gamma)' \hat{\mathbf{A}}_{1\gamma}(\gamma) & \frac{1}{n} \hat{\mathbf{A}}_{1\gamma}(\gamma)' \hat{\mathbf{A}}_{2\gamma}(\gamma) \\
\frac{1}{n} \hat{\mathbf{A}}_{2\gamma}(\gamma)' \hat{\mathbf{X}}_\gamma^*(\gamma) & \frac{1}{n} \hat{\mathbf{A}}_{2\gamma}(\gamma)' \hat{\mathbf{A}}_{1\gamma}(\gamma) & \frac{1}{n} \hat{\mathbf{A}}_{2\gamma}(\gamma)' \hat{\mathbf{A}}_{2\gamma}(\gamma)
\end{pmatrix}
\]
\[
= \begin{pmatrix}
\frac{1}{n} \sum_i (\hat{x}_i \hat{x}_i'(q_i \leq \gamma)) & \frac{1}{n} \sum_i \hat{\lambda}_{1i}(\gamma) \hat{x}_i I(q_i \leq \gamma) & \frac{1}{n} \sum_i \hat{\lambda}_{2i}(\gamma) \hat{x}_i I(q_i \leq \gamma) \\
\frac{1}{n} \sum_i \hat{\lambda}_{1i}(\gamma) \hat{x}_i I(q_i \leq \gamma) & \frac{1}{n} \sum_i (\hat{\lambda}_{1i}(\gamma))^2 I(q_i \leq \gamma) & \frac{1}{n} \sum_i \hat{\lambda}_{2i}(\gamma) \hat{\lambda}_{2i}(\gamma) I(q_i \leq \gamma) \\
\frac{1}{n} \sum_i \hat{\lambda}_{2i}(\gamma) \hat{x}_i I(q_i \leq \gamma) & \frac{1}{n} \sum_i \hat{\lambda}_{2i}(\gamma) \hat{\lambda}_{2i}(\gamma) I(q_i \leq \gamma) & \frac{1}{n} \sum_i (\hat{\lambda}_{2i}(\gamma))^2 I(q_i \leq \gamma)
\end{pmatrix}
\]
and recall that \( \hat{x}_i = g_{xi} = g_{xi} - \hat{r}_{xi}, \hat{\lambda}_{1i}(\gamma) = \lambda_{1i}(\gamma) - \hat{r}_{u1i}, \) and \( \hat{\lambda}_{2i}(\gamma) = \lambda_{2i}(\gamma) - \hat{r}_{u2i}. \)

First note that \( \frac{1}{n} \sum_i (\hat{x}_i \hat{x}_i'(q_i \leq \gamma)) \xrightarrow{p} E(g_{xi}g_{xi}'(q_i \leq \gamma)) \) follows from Caner and Hansen (2004) and Lemma 1 of Hansen (1996). Since the first stage regressions are consistently estimated and Lemma 1 of Hansen (1996) we get for \( j = 1, 2 \)
\[
\frac{1}{n} \sum_i \hat{\lambda}_{1i}(\gamma) \hat{x}_i I(q_i \leq \gamma) = \frac{1}{n} \sum_i \hat{\lambda}_{1i}(\gamma) g_{xi} I(q_i \leq \gamma) - \frac{1}{n} \sum_i \hat{r}_{xi} \hat{\lambda}_{1i}(\gamma) I(q_i \leq \gamma)
\]
\[
= \frac{1}{n} \sum_i \lambda_{1i}(\gamma) g_{xi} I(q_i \leq \gamma) - \frac{1}{n} \sum_i \hat{r}_{u1i} g_{xi} I(q_i \leq \gamma) - \frac{1}{n} \sum_i \hat{r}_{xi} \lambda_{1i}(\gamma) I(q_i \leq \gamma) + \frac{1}{n} \sum_i \hat{r}_{u1i} \hat{r}_{u2i} I(q_i \leq \gamma)
\]
\[
\frac{1}{n} \sum_i (\hat{\lambda}_{1i}(\gamma))^2 I(q_i \leq \gamma) = \frac{1}{n} \sum_i (\lambda_{1i}(\gamma))^2 I(q_i \leq \gamma) - 2 \frac{1}{n} \sum_i \lambda_{1i}(\gamma) \hat{r}_{u1i} I(q_i \leq \gamma) + \frac{1}{n} \sum_i \hat{r}_{u1i}^2 I(q_i \leq \gamma)
\]
Similarly, we can show that
\[
\frac{1}{n} \sum_i \hat{\lambda}_{2i}(\gamma) \hat{\lambda}_{2i}(\gamma) I(q_i \leq \gamma) = \frac{1}{n} \sum_i \lambda_{2i}(\gamma) \lambda_{2i}(\gamma) I(q_i \leq \gamma) - \frac{1}{n} \sum_i \lambda_{1i}(\gamma) \hat{r}_{u1i} I(q_i \leq \gamma) - \frac{1}{n} \sum_i \hat{r}_{u1i} \hat{r}_{u2i} I(q_i \leq \gamma)
\]
Therefore, uniformly in \( \gamma \in \Gamma, \frac{1}{n} \hat{\mathbf{X}}_\gamma^*(\gamma)' \hat{\mathbf{X}}_\gamma^*(\gamma) \xrightarrow{p} E(g_{\gamma i}(\gamma)g_{\gamma i}(\gamma)') = \mathbf{M}_\gamma(\gamma), \) where
\[
\mathbf{M}_\gamma(\gamma) = \begin{pmatrix}
E(g_{\gamma i}g_{\gamma i}'(q_i \leq \gamma)) & E(\lambda_{1i}(\gamma) g_{\gamma i}(q_i \leq \gamma)) & E(\lambda_{2i}(\gamma) g_{\gamma i}(q_i \leq \gamma)) \\
E(\lambda_{1i}(\gamma) g_{\gamma i}'(q_i \leq \gamma)) & E((\lambda_{1i}(\gamma))^2 I(q_i \leq \gamma)) & E(\lambda_{1i}(\gamma) \lambda_{2i}(\gamma) I(q_i \leq \gamma)) \\
E(\lambda_{2i}(\gamma) g_{\gamma i}'(q_i \leq \gamma)) & E(\lambda_{2i}(\gamma) \lambda_{1i}(\gamma) I(q_i \leq \gamma)) & E((\lambda_{2i}(\gamma))^2 I(q_i \leq \gamma))
\end{pmatrix}
\]
Similarly we can show that, \[ \frac{1}{n} \hat{X}_\perp(\gamma)' \hat{X}_\perp(\gamma) \xrightarrow{p} E(g_{\perp i}(\gamma)g_{\perp i}(\gamma)' = M(\gamma). \] Then, we get (A.5)
\[ \frac{1}{n} \hat{X}^*(\gamma)' \hat{X}^*(\gamma) \xrightarrow{p} M(\gamma) = \begin{pmatrix} M_{\gamma}(\gamma) & 0 \\ 0 & M(\gamma) \end{pmatrix} \]

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

First note that \( \frac{1}{n} \sum_i (x_i \hat{e}' I(q_i \leq \gamma)) = O_p(1) \) follows from Caner and Hansen (2004). Second, from Lemma A.4 of Hansen (2000) and Theorem 1 of Hansen (1996) we can obtain for \( j = 1, 2 \)
\[ \frac{1}{n} \sum_i \hat{\lambda}_{ji} (\gamma) \hat{e}' I(q_i \leq \gamma) = \frac{1}{n} \sum_i \lambda_{ji} (\gamma) \hat{e}' I(q_i \leq \gamma) - \frac{1}{n} \sum_i \hat{\tau}_{uij} \hat{e}' I(q_i \leq \gamma)
\]
\[ = \frac{1}{n} \sum_i \lambda_{ji} (\gamma) \beta_{ri} \hat{e} I(q_i \leq \gamma) + \frac{1}{n} \sum_i \lambda_{ji} (\gamma) \hat{e}' I(q_i \leq \gamma)
\]
\[ - \frac{1}{n} \sum_i \hat{\tau}_{uij} \beta_{\hat{e}} \hat{e} I(q_i \leq \gamma) - \frac{1}{n} \sum_i \hat{\tau}_{uij} e I(q_i \leq \gamma)
\]
\[ = O_p(1) \]

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

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

\[ \blacksquare \]

**Proof of Proposition 1.**

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

Define \( \hat{e} = \hat{r} \beta + \epsilon \). Given that \( G(\gamma) = \hat{G}(\gamma) + \hat{V} \) and \( \hat{G}(\gamma) = \hat{X}(\gamma) \) is in the span of \( \hat{X}^*(\gamma) \) then \( (I - P^*(\gamma))G(\gamma) = (I - P^*(\gamma))\hat{r} \) and
\[ (I - P^*(\gamma))Y = (I - P^*(\gamma))(G(\gamma_0)\beta + G_0(\gamma_0)\delta_n + \hat{\epsilon}) \]

Then
\[ S_n^U(\gamma) = Y'(I - P^*(\gamma))Y \]  
\[ = (n^{-\alpha}c'G_0(\gamma)' + \tilde{e}')(I - P^*(\gamma))(G_0(\gamma)n^{-\alpha}c + \tilde{e}) \]  
\[ = (n^{-\alpha}c'G_0(\gamma)' + \tilde{e}')(G_0(\gamma)n^{-\alpha}c + \tilde{e}) \]  
\[ -(n^{-\alpha}c'G_0(\gamma)' + \tilde{e})P^*(\gamma)(G_0(\gamma)n^{-\alpha}c + \tilde{e}) \]

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

\[ S_n^{*U}(\gamma) = n^{-1+2\alpha}(n^{-\alpha}c'G_0(\gamma)' + \tilde{e})P^*(\gamma)G_0(\gamma)n^{-\alpha}c + \tilde{e} \]  
\[ = n^{-1+2\alpha}\tilde{e}P^*(\gamma)\tilde{e} + 2n^{-1+\alpha}c'G_0(\gamma)'P^*(\gamma)\tilde{e} + n^{-1}c'G_0(\gamma)'P^*(\gamma)G_0(\gamma)c \]

Let us now examine \( S_n^{*U}(\gamma) \) for \( \gamma \in (\gamma_0, \overline{\gamma}] \). Note that \( G_0(\gamma_0)'P_\perp(\gamma) = 0 \)

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

\[ n^{-1+2\alpha}\tilde{e}P_\gamma(\gamma)\tilde{e} = n^{-1+2\alpha}(\frac{1}{\sqrt{n}}\tilde{e}'\tilde{X}_\gamma(\gamma))(\frac{1}{\sqrt{n}}\tilde{X}_\gamma(\gamma)'\tilde{X}_\gamma(\gamma))^{-1}(\frac{1}{\sqrt{n}}\tilde{X}_\gamma(\gamma)'\tilde{e}) \xrightarrow{P} 0 \]
\[ n^{-1+2\alpha}\tilde{e}P_\perp(\gamma)\tilde{e} = n\alpha^{-1}(\frac{1}{\sqrt{n}}\tilde{e}'\tilde{X}_\perp(\gamma))(\frac{1}{\sqrt{n}}\tilde{X}_\perp(\gamma)'\tilde{X}_\perp(\gamma))^{-1}(\frac{1}{\sqrt{n}}\tilde{X}_\perp(\gamma)'\tilde{e}) \xrightarrow{P} 0 \]

and

\[ n^{-1+\alpha}c_0'G_0(\gamma_0)'P_\gamma(\gamma)\tilde{e} = n^{\alpha-1/2}(\frac{1}{\sqrt{n}}G_0(\gamma_0)'\tilde{X}_0(\gamma))(\frac{1}{\sqrt{n}}\tilde{X}_\gamma(\gamma)'\tilde{X}_\gamma(\gamma))^{-1}(\frac{1}{\sqrt{n}}\tilde{X}_\gamma(\gamma)'\tilde{e}) \xrightarrow{P} 0 \]

So

\[ S_n^{*U}(\gamma) = n^{-1+2\alpha}\tilde{e}P_\gamma(\gamma)\tilde{e} + n^{-1+2\alpha}\tilde{e}P_\perp(\gamma)\tilde{e} + 2n^{-1+\alpha}c'G_0(\gamma_0)'P_\gamma(\gamma)\tilde{e} + n^{-1}c'G_0(\gamma_0)'P_\gamma(\gamma)G_0(\gamma_0)c \]

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

\[ \frac{1}{n}\tilde{X}_1(\gamma)'G_0(\gamma_0) = \begin{pmatrix} \frac{1}{n}\tilde{X}_1(\gamma)'G_{x,0} \\ \frac{1}{n}\tilde{X}_1(\gamma)'A_{1,0}(\gamma_0) \\ \frac{1}{n}\tilde{X}_1(\gamma)'A_{1,0}(\gamma_0) \\ \frac{1}{n}\tilde{X}_1(\gamma)'A_{2,0}(\gamma_0) \end{pmatrix} = \begin{pmatrix} \frac{1}{n}\sum_i \tilde{e}_x i I(q_i \leq \gamma_0) \\ \frac{1}{n}\sum_i \lambda_1(i(\gamma_0) \tilde{e}_x I(q_i \leq \gamma_0) \\ \frac{1}{n}\sum_i \lambda_2(i(\gamma_0) \tilde{e}_x I(q_i \leq \gamma_0) \\ \frac{1}{n}\sum_i \lambda_3(i(\gamma_0) \tilde{e}_x I(q_i \leq \gamma_0) \end{pmatrix} \]
Thus, the conditions of Theorem 5.7 (p.45) by Van der Vaart (1998) are satisfied. First, given the uniform convergence of $S^*_n(\gamma)$, i.e.

$$\sup_{\gamma \in \Gamma} |S^*_{n}(\gamma) - S^*_n(\gamma_0)| \overset{p}{\to} 0 \quad \text{as} \quad n \to \infty.$$  

Second, by the compactness of $\Gamma$ and the fact that $S^*_n(\gamma)$ is uniquely maximized at $\gamma_0$, we can have

$$S^*_n(\gamma) \to c'M_0(\gamma_0, \gamma)'M_\gamma(\gamma)^{-1}M_0(\gamma_0, \gamma)c \quad (A.11)$$

by a Glivenko-Cantelli theorem for stationary ergodic processes.

Given the monotonicity of the inverse Mills ratio, $M_0(\gamma_0, \gamma_0 + \epsilon) \geq M_0(\gamma_0)$ for any $\epsilon > 0$ with equality at $\gamma = \gamma_0$. To see this note that for $\epsilon > 0$, $\lambda_{ii}(\gamma_0 + \epsilon) > \lambda_{ii}(\gamma_0)$ and $\lambda_{2i}(\gamma_0 + \epsilon) > \lambda_{2i}(\gamma_0)$.

Therefore, we need to show that $S^*_n(\gamma) < M_0(\gamma_0)$ for any $\gamma \in (\gamma_0, \overline{\gamma}]$. It is sufficient to show that $M_0(\gamma_0)'M_\gamma(\gamma)^{-1}M_0(\gamma_0) < M_0(\gamma_0)$, which reduces to $M_\gamma(\gamma) > M_0(\gamma_0)$ for any $\gamma \in (\gamma_0, \overline{\gamma}]$.

To see recall that $M_\gamma(\gamma) = E\left( g_{\gamma i}(\gamma)g'_{\gamma i}(\gamma) \right)$.

$$M_\epsilon(\gamma_0 + \epsilon) - M_0(\gamma_0) = \int_{\gamma_0}^{\gamma_0 + \epsilon} E(g_i(t)g_i(t)'|q = t)f_q(t)dt$$

$$> \inf_{\gamma_0 < \gamma \leq \gamma_0 + \epsilon} E g_i(\gamma)g_i'(\gamma)|q = \gamma \left( \int_{\gamma_0}^{\gamma_0 + \epsilon} f(\nu)d\nu \right)$$

$$= \inf_{\gamma_0 < \gamma \leq \gamma_0 + \epsilon} D_1(\gamma) \left( \int_{\gamma_0}^{\gamma_0 + \epsilon} f(\nu)d\nu \right) > 0$$

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

Thus, the conditions of Theorem 5.7 (p.45) by Van der Vaart (1998) are satisfied. First, given the uniform convergence of $S^*_n(\gamma)$, i.e.
Suppose $S^*_n(\gamma) < S^*_n(\gamma_0)$ for every $\epsilon > 0$. Therefore, it follows that $\hat{\gamma} \xrightarrow{p} \gamma_0$ for the unrestricted problem.

Assuming the restrictions hold,

$$S^*_n(\gamma) = S^*_n(\gamma_0) \leq S^*_n(\gamma)$$ \hspace{1cm} (A.12)

When $\hat{\gamma}$ is not consistent it must be the case that $S^*_n(\hat{\gamma}) \geq S^*_n(\gamma) + C||\beta_{10} - \beta_1||^2 + ||\beta_{20} - \beta_2||^2 + o_p(1)$, where $\beta_{10}$ and $\beta_{20}$ are the true slope coefficients for the two regimes. But since $S^*_n(\hat{\gamma}) \leq S^*_n(\hat{\gamma})$ we also have $S^*_n(\hat{\gamma}) \geq S^*_n(\gamma) + C||\beta_{10} - \beta_1||^2 + ||\beta_{20} - \beta_2||^2 + o_p(1)$, which yields a contradiction with (A.12). This completes the proof.

\[\blacksquare\]

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

**Proof of Lemma 3.**

Note that $S^*_n(\gamma) = S^*_n(\gamma) + (\vartheta - \tilde{\vartheta})(\vartheta' - \tilde{\vartheta}')(\tilde{\vartheta}^*(\gamma)'\tilde{\vartheta}^*(\gamma))^{-1}\vartheta$. The proof proceeds in steps. First we establish that the unrestricted and the restricted problems share the same rate of convergence.

$$S^*_n(\gamma) - S^*_n(\gamma_0) \geq S^*_n(\gamma) - S^*_n(\gamma_0) + o_p(1) \hspace{1cm} (A.13)$$

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

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

$$\tilde{\vartheta}^*(\gamma)'\tilde{\vartheta}^*(\gamma))^{-1} = (\tilde{\vartheta}^*(\gamma_0)'\tilde{\vartheta}^*(\gamma_0))^{-1} + O_p(\frac{|\gamma - \gamma_0|}{n^2})$$ \hspace{1cm} (A.14)

and

$$(\vartheta'(\tilde{\vartheta}^*(\gamma)'\tilde{\vartheta}^*(\gamma))^{-1}\vartheta)^{-1} = (\vartheta'(\tilde{\vartheta}^*(\gamma_0)'\tilde{\vartheta}^*(\gamma_0))^{-1}\vartheta)^{-1} + O_p(|\gamma - \gamma_0|).$$ \hspace{1cm} (A.15)

Consider $\Delta \vartheta = \vartheta - \vartheta_0 = (\tilde{\vartheta}^*(\gamma)'\tilde{\vartheta}^*(\gamma))^{-1}\tilde{\vartheta}^*(\gamma)(\tilde{\vartheta}^*(\gamma_0)'\theta_0 + \epsilon) - \theta_0 - (\tilde{\vartheta}^*(\gamma_0)'\tilde{\vartheta}^*(\gamma_0))^{-1}\tilde{\vartheta}^*(\gamma_0)'\epsilon$.

$$= (\tilde{\vartheta}^*(\gamma_0)'\tilde{\vartheta}^*(\gamma_0))^{-1}(\tilde{\vartheta}^*(\gamma - \tilde{\vartheta}^*(\gamma_0))'+\theta_0 + (\tilde{\vartheta}^*(\gamma - \tilde{\vartheta}^*(\gamma_0)))'\epsilon) + |\gamma - \gamma_0|O_p(\frac{1}{n})$$

$= (\tilde{\vartheta}^*(\gamma_0)'\tilde{\vartheta}^*(\gamma_0))^{-1/2}A_n$

with

27
\[ A_n = \hat{X}^*(\gamma_0)'\hat{X}^*(\gamma_0) - 1/2(\tilde{X}^*(\gamma) - \hat{X}^*(\gamma_0))'\hat{X}^*(\gamma_0)\theta_0 \]
\[ + (\tilde{X}^*(\gamma_0)\tilde{X}^*(\gamma_0))^{-1/2}(\tilde{X}^*(\gamma) - \tilde{X}^*(\gamma_0))'e + |\gamma - \gamma_0|O_p(\frac{1}{\sqrt{n}}) \]
\[ = |\gamma - \gamma_0|O_p(n^{-1/2}) + |\gamma - \gamma_0|O_p(1) + |\gamma - \gamma_0|O_p(\frac{1}{\sqrt{n}}) \]
\[ = |\gamma - \gamma_0|O_p(n^{-1/2}) \] and hence \( \Delta \tilde{\theta} = |\gamma - \gamma_0|O_p(n^{-1}) \) using (A.14) for the first equality and the fact that \( \tilde{X}^*(\gamma) - \hat{X}^*(\gamma_0) \) and \( (\tilde{X}^*(\gamma) - \tilde{X}^*(\gamma_0))'e \) are \( |\gamma - \gamma_0|O_p(1) \) terms.

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

\[ S_n^R(\gamma) - S_n^R(\gamma_0) = \]
\[ [S_n^U(\gamma) - S_n^U(\gamma_0)] + \]
\[ [(\theta - R\tilde{\theta})'(R'(\tilde{X}^*(\gamma)\tilde{X}^*(\gamma))^{-1}R)^{-1}(\theta - R\tilde{\theta}) - \]
\[ (\theta - R\tilde{\theta}_0)',(R'(\tilde{X}^*(\gamma_0)\tilde{X}^*(\gamma_0))^{-1}R)^{-1}(\theta - R\tilde{\theta}_0)] = \]
\[ [S_n^U(\gamma) - S_n^U(\gamma_0)] + \]
\[ [(\theta - R\tilde{\theta})'(R'(\tilde{X}^*(\gamma_0)\tilde{X}^*(\gamma_0))^{-1}R)^{-1}(\theta - R\tilde{\theta}) - \]
\[ (\theta - R\tilde{\theta}_0)',(R'(\tilde{X}^*(\gamma_0)\tilde{X}^*(\gamma_0))^{-1}R)^{-1}(\theta - R\tilde{\theta}_0)] + (\gamma - \gamma_0)^2O_p(n^{-1}) = \]
\[ [S_n^U(\gamma) - S_n^U(\gamma_0)] + \]
\[ (\theta_0 + \Delta \tilde{\theta})',R'(\tilde{X}^*(\gamma_0)\tilde{X}^*(\gamma_0))^{-1}R)^{-1}R'(\theta_0 + \Delta \tilde{\theta}) - \]
\[ \tilde{\theta}_0',R'(\tilde{X}^*(\gamma_0)\tilde{X}^*(\gamma_0))^{-1}R)^{-1}R'\tilde{\theta}_0 - \]
\[ 2\theta'_R\tilde{\theta}'R'(\tilde{X}^*(\gamma_0)\tilde{X}^*(\gamma_0))^{-1}R)^{-1}R'(\hat{\theta}_1 - \tilde{\theta}_0) + \]
\[ (\gamma - \gamma_0)^2O_p(n^{-1}) \]
\[ = [S_n^U(\gamma) - S_n^U(\gamma_0)] + \]
\[ 2\Delta \tilde{\theta}'R'(\tilde{X}^*(\gamma_0)\tilde{X}^*(\gamma_0))^{-1}R)^{-1}R'(\tilde{X}^*(\gamma_0)\tilde{X}^*(\gamma_0))^{-1}\tilde{X}^*(\gamma_0)e + \]
\[ \Delta \tilde{\theta}'R'(\tilde{X}^*(\gamma_0)\tilde{X}^*(\gamma_0))^{-1}R)^{-1}R'\Delta \tilde{\theta} + \]
\[ 2\Delta \tilde{\theta}'R'(\tilde{X}^*(\gamma_0)\tilde{X}^*(\gamma_0))^{-1}R)^{-1}(R\theta_0 - \theta) + \]
\[ (\gamma - \gamma_0)^2O_p(n^{-1}) \]

Note that the second and fourth term are \( o_p(1) \) when divided by \( (\gamma - \gamma_0) \) since \( \Delta \tilde{\theta} = |\gamma - \gamma_0|O_p(n^{-1}) \). The third term is always positive. Therefore, we can now focus on the first term, which is the rate of convergence for the unrestricted problem.
Our proof follows in spirit Yu (2010). In this lemma we use the notation for empirical processes in van der Vaart and Wellner (1996). Define \( M_n(\theta) = P_n m(\theta) \), where \( P_n \) denotes the empirical measure \( P_n = \frac{1}{n} \sum_{i=1}^{n} x_i \), such that for any class of measurable function \( f : x \to \mathbb{R} \), we denote \( P_n f = \frac{1}{n} \sum_{i=1}^{n} f(x_i) \). We also define \( M(\theta) = P m(\theta) \), where \( P m(\theta) = \int f(x) P(dx) \). Finally, define the empirical process \( G_n = \sqrt{n} (P_n - P) \) so that \( G_n m(\theta) = \sqrt{n} (M_n(\theta) - M) \).

Given that the theorem is for the maximization problem we will consider \( P \) measure \( n \mathbb{1}_0 = n ) = 1 \) and let \( \theta = (\beta_1', \beta_2', \gamma')' \). Recall that \( \gamma \in \Gamma = [\gamma, \bar{\gamma}] \), then we have \( I(q_i \leq \gamma) \leq I(q \leq q \wedge \gamma_0) \) and \( I(q_i > \gamma) \geq I(q_0 < q \leq \gamma \lor \gamma_0) \), where “\( \wedge \)” and “\( \lor \)” denote the minimum and maximum, respectively.

We can derive the following formula.

\[
m(\theta) = -(y_i - g_i(\gamma)' \beta_i I(q_i \leq \gamma) - g_i(\gamma)' \beta_2 I(q_i > \gamma))^2
\]

\[
= -g_i(\gamma)' \beta_{10} - g_i(\gamma)' \beta_1 + e_{i1}^2 I(q_i \leq \gamma \land \gamma_0)
\]

\[
- [g_i(\gamma)' \beta_{20} - g_i(\gamma)' \beta_2 + e_{i2}]^2 I(q_i > \gamma \lor \gamma_0)
\]

\[
- g_i(\gamma)' \beta_{10} - g_i(\gamma)' \beta_2 + e_{i1}^2 I(\gamma \land \gamma_0 < q_i \leq \gamma_0)
\]

\[
- g_i(\gamma)' \beta_{20} - g_i(\gamma)' \beta_1 + e_{i2}^2 I(\gamma_0 < q_i \leq \gamma \lor \gamma_0)
\]

\[
- [g_i(\gamma)'(\beta_{x10} - \beta_{x1}) + \lambda_i(\gamma)'(\beta_{\lambda10} - \beta_{\lambda1}) + (\lambda_i(\gamma) - \lambda_i(\gamma_0))' \beta_{\lambda1} + e_{i1}]^2 I(q \leq \gamma \land \gamma_0)
\]

\[
- [g_i(\gamma)'(\beta_{x20} - \beta_{x2}) + \lambda_i(\gamma)'(\beta_{\lambda20} - \beta_{\lambda2}) + (\lambda_i(\gamma) - \lambda_i(\gamma_0))' \beta_{\lambda2} + e_{i2}]^2 I(q > \gamma \lor \gamma_0)
\]

\[
- [g_i(\gamma)'(\beta_{x10} - \beta_{x2}) + \lambda_i(\gamma)'(\beta_{\lambda10} - \beta_{\lambda2}) + (\lambda_i(\gamma) - \lambda_i(\gamma_0))' \beta_{\lambda2} + e_{i1}]^2 I(\gamma \land \gamma_0 < q \leq \gamma_0)
\]

\[
- [g_i(\gamma)'(\beta_{x20} - \beta_{x1}) + \lambda_i(\gamma)'(\beta_{\lambda20} - \beta_{\lambda1}) + (\lambda_i(\gamma) - \lambda_i(\gamma_0))' \beta_{\lambda1} + e_{i2}]^2 I(\gamma_0 < q \leq \gamma \lor \gamma_0)
\]

Define

\[
T(\theta_{1,0}, \theta_{1}) = (g_i(\gamma)'(\beta_{x10} - \beta_{x1}) + (\lambda_i(\gamma) - \lambda_i(\gamma_0))' \beta_{\lambda1} + e_{i1})^2 - e_{i1}^2,
\]
To show this let us first define the class of functions

\[ T(\theta_{2,0}, \theta_1) = (g_i(\gamma_0) (\beta_{x20} - \beta_{x2}) + (\lambda_i(\gamma) - \lambda_i(\gamma_0)) \beta_{x2} + e_{2i})^2 - e_{2i}^2, \]

\[ T(\theta_{1,0}, \theta_2) = (g_i(\gamma) (\beta_{x10} - \beta_{x2}) + (\lambda_i(\gamma) - \lambda_i(\gamma_0)) \beta_{x2} + e_{1i})^2 - e_{1i}^2, \]

\[ T(\theta_{2,0}, \theta_1) = (g_i(\gamma_0) (\beta_{x20} - \beta_{x1}) + (\lambda_i(\gamma) - \lambda_i(\gamma_0)) \beta_{x1} + e_{2i})^2 - e_{2i}^2. \]

Define the discrepancy function \( d(\theta, \theta_0) = ||\beta - \beta_0|| + |\gamma_0 - \gamma| + \sqrt{F_q(\gamma) - F_q(\gamma_0)} \) for \( \theta \) in the neighborhood of \( \theta_0 \). Note that \( d(\theta, \theta_0) \to 0 \) if and only if \( ||\beta - \beta_0|| \to 0 \) and \( |\gamma - \gamma_0| \to 0 \).

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

Consider

\[ M(\theta) - M(\theta_0) = \]

\[ -E[T(\theta_{1,0}, \theta_1) I(q_i \leq \gamma \land \gamma_0)] \]

\[ -E[T(\theta_{2,0}, \theta_2) I(q_i > \gamma \lor \gamma_0)] \]

\[ -E[T(\theta_{1,0}, \theta_2) I(\gamma \land \gamma_0 < q_i \leq \gamma_0)] \]

\[ -E[T(\theta_{2,0}, \theta_1) I(\gamma_0 < q_i \leq \gamma \lor \gamma_0)] \]

\[ \leq \]

\[ -(\beta_{10} - \beta_1)' E(g_i(\gamma_0) g_i(\gamma_0)' I(q_i \leq \gamma \land \gamma_0))(\beta_{10} - \beta_1) \]

\[ -(\beta_{20} - \beta_2)' (Eg_i(\gamma_0) g_i(\gamma_0)' I(q_i > \gamma \lor \gamma_0))(\beta_{20} - \beta_2) \]

\[ -(\beta_{10} - \beta_2)' E(g_i(\gamma_0) g_i(\gamma_0)' I(\gamma \land \gamma_0 < q_i \leq \gamma_0))(\beta_{20} - \beta_1) \]

\[ -(\beta_{20} - \beta_1)' E(g_i(\gamma_0) g_i(\gamma_0)' I(\gamma_0 < q_i \leq \gamma \lor \gamma_0))(\beta_{10} - \beta_2) - C_\lambda |\gamma_0 - \gamma|^2 \]

\[ \leq -C \left( ||\beta_{10} - \beta_1||^2 + ||\beta_{20} - \beta_2||^2 + |\gamma_0 - \gamma|^2 + |F_q(\gamma) - F_q(\gamma_0)| \right) \]

\[ = -C d^2(\theta, \theta_0), \]

where the first inequality is due to the monotonicity of \( \lambda_1(\cdot) \) and \( \lambda_2(\cdot) \), Assumption 1, and Lemma 1.

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

\[ E^* \left( \sup_{d(\theta, \theta_0)} |G_n(m(w|\theta) - m(w|\theta_0))| \right) \leq C\epsilon, \]

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

To show this let us first define the class of functions

\[ \mathcal{M}_\epsilon = \{m(\theta) - m(\theta_0) : d(\theta, \theta_0) < \epsilon\} \]
Let us also write $m(\theta) - m(\theta_0)$ as follows

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

$$-T(\theta_{1,0}, \theta_1)I(q_i \leq \gamma \wedge \gamma_0) - T(\theta_{2,0}, \theta_2)I(q_i > \gamma \vee \gamma_0)$$

$$-T(\theta_{1,0}, \theta_2)I(\gamma \wedge \gamma_0 < q_i \leq \gamma_0) - T(\theta_{2,0}, \theta_1)I(\gamma_0 < q_i \leq \gamma \vee \gamma_0)$$

$$= A + B + C + D,$$

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

Note that $\{T(\theta_{1,0}, \theta_1) : d(\theta, \theta_0) < \tilde{\delta}\}$ is a finite-dimensional vector space of real valued functions. Then Lemma 2.4 of Pakes and Pollard (1989) implies that $\{I(q \leq \gamma \wedge \gamma_0) : d(\theta, \theta_0) < \tilde{\delta}\}$ is a VC subgraph class of functions. Then it follows that $\{A_n : d(\theta, \theta_0) < \tilde{\delta}\}$ is also a VC subgraph by Lemma 2.14 (ii) of Pakes and Pollard (1989). Similarly, we can show that $\{B_n : d(\theta, \theta_0) < \tilde{\delta}\}, \{C_n : d(\theta, \theta_0) < \tilde{\delta}\}, \{D_n : d(\theta, \theta_0) < \tilde{\delta}\}$ are VC-classes.

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

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

and $F$ is the envelope function of the class of functions defined by $\{m(w|\theta) - m(w|\theta_0) : d(\theta, \theta_0) < \tilde{\delta}\}$. Given the functional form of $m(w|\theta) - m(w|\theta_0),$ $\sqrt{PF^2} \leq C\tilde{\delta}$ by Assumption 1.4 and 1.5. This shows the second condition.

Corollary 3.2.6 of van der Vaart and Wellner (1996) implies that $\phi(\tilde{\delta}) = \tilde{\delta}$ and thus $\phi(\tilde{\delta})/\tilde{\delta}^\alpha = \delta^{1-\alpha}$ is decreasing for any $\alpha \in (1, 2)$, hence Theorem 14.4 in Kosorok (2008) is satisfied. Since $r_n^2 \phi(r_n^{-1}) = r_n$ and hence $\sqrt{n}d(\hat{\theta}, \theta_0) = O_p(1)$. By the definition of $d$, we get that $||\hat{\beta} - \beta_0|| = O_p(n^{-1/2})$ and $|\hat{\gamma} - \gamma_0| = |F(\hat{\gamma}) - F(\gamma_0)| = O_p(n^{-1/2}) + O_p(n^{-1}) = O_p(n^{-1})$.

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

**Lemma 4.** $\arg\min_{\gamma/\gamma_0 \leq B} S_n^R(\gamma) - S_n^R(\gamma_0) = \arg\min_{\gamma/\gamma_0 \leq B} S_n^U(\gamma) - S_n^U(\gamma_0) + o_p(1)$

**Proof of Lemma 4.**
Recall that $S_n^R(\gamma) = S_n^U(\gamma) + (\theta - \bar{R}\hat{\theta})'(R'(\hat{X}^*(\gamma)\hat{X}^*(\gamma))^{-1}R)^{-1}(\theta - \bar{R}\hat{\theta})$.

Then

$$S_n^R(\gamma) - S_n^R(\gamma_0) = [S_n^U(\gamma) - S_n^U(\gamma_0)] + [\theta - \bar{R}\hat{\theta}]'(R'(\hat{X}^*(\gamma)\hat{X}^*(\gamma))^{-1}R)^{-1}(\theta - \bar{R}\hat{\theta}) - (\theta - \bar{R}\hat{\theta}_0)'(R'(\hat{X}^*(\gamma_0)\hat{X}^*(\gamma_0))^{-1}R)^{-1}(\theta - \bar{R}\hat{\theta}_0)$$

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

Define $\Delta(\gamma) = I(q \leq \gamma) - I(q \leq \gamma_0)$ and $\bar{I} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. Let us consider the case of $\gamma > \gamma_0$.

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

$$\frac{1}{n}||\sum_i g_i(\gamma)g_i(\gamma)'\Delta(\gamma) - \sum_i g_i(\gamma)\bar{f}'(\gamma)\Delta(\gamma) - \sum_i g_i(\gamma)\bar{f}''(\gamma) + \sum_i \bar{f}'(\gamma)\bar{f}'(\gamma) || \leq$$

$$\frac{1}{n}||\sum_i g_i(\gamma)g_i(\gamma)'\Delta(\gamma) \otimes \bar{I}|| + 2\frac{1}{n}||\sum_i g_i(\gamma)\bar{f}'(\gamma)\Delta(\gamma) \otimes \bar{I}|| + ||\sum_i \bar{f}'(\gamma)\bar{f}'(\gamma) \otimes \bar{I}|| \leq$$

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

$$\sqrt{2\frac{1}{n}(tr(\sum_i g_i(\gamma_0 + \epsilon)\bar{f}'(\gamma)))^{2/2}} +$$

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

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

$$\left(\frac{1}{n}\hat{X}^*(\gamma)'\hat{X}^*(\gamma)\right)^{-1} = \left(\frac{1}{n}\hat{X}^*(\gamma_0)'\hat{X}^*(\gamma_0)\right)^{-1} + o_p(1).$$

and

$$\frac{1}{n}(R'(\hat{X}^*(\gamma)'\hat{X}^*(\gamma))^{-1}R)^{-1} = \frac{1}{n}(R'(\hat{X}^*(\gamma_0)'\hat{X}^*(\gamma_0))^{-1}R)^{-1} + o_p(1).$$

To show this note first that $S_n^U(\gamma) - S_n^U(\gamma_0) = o_p(1)$. Then,

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

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

$$[(\theta - \bar{R}\hat{\theta})'(R'(\hat{X}^*(\gamma)'\hat{X}^*(\gamma))^{-1}R)^{-1}(\theta - \bar{R}\hat{\theta}) -$$

$$(\theta - \bar{R}\hat{\theta}_0)'(R'(\hat{X}^*(\gamma_0)'\hat{X}^*(\gamma_0))^{-1}R)^{-1}(\theta - \bar{R}\hat{\theta}_0)] =$$

$$[(\theta - \bar{R}\hat{\theta})'(R'(\hat{X}^*(\gamma_0)'\hat{X}^*(\gamma_0))^{-1}R)^{-1}(\theta - \bar{R}\hat{\theta}) -$$

$$(\theta - \bar{R}\hat{\theta}_0)'(R'(\hat{X}^*(\gamma_0)'\hat{X}^*(\gamma_0))^{-1}R)^{-1}(\theta - \bar{R}\hat{\theta}_0)] + o_p(1) =$$
We proceed by studying the behavior of each term: (i) $\left(n^{-1/2}(\theta_0 - \tilde{\theta})' R (R'(\hat{X}^*(\gamma_0)/\tilde{X}^*(\gamma_0))^{-1} R)^{-1} R' n^{-1/2}(\theta_0 - \tilde{\theta}) - n^{-1/2}(\theta_0 - \tilde{\theta})' R (R'(\hat{X}^*(\gamma_0)/\tilde{X}^*(\gamma_0))^{-1} R)^{-1} R' n^{-1/2}(\theta_0 - \tilde{\theta}) + o_p(1) = o_p(1)$ since $n^{-1/2}(\hat{\theta} - \theta_0) = n^{-1/2}(\tilde{\theta} - \theta_0) + o_p(1)$.

This completes the proof.

**Lemma 5.** On $[-\bar{v}, \bar{v}]$,

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

where $\mu_i = c'D_i c_f$ and $\zeta_i = c'\Omega_i c_f$, for $i = 1, 2$.

**Proof of Lemma 5.**

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

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

$$Q_n(v) = S_n^U(\gamma_0) - S_n^U(\gamma_0 + v/a_n) = (n^{-\alpha} c' G(\gamma_0)' + \tilde{e}') P^*(v) (G(\gamma_0) c n^{-\alpha} + \tilde{e}) - (n^{-\alpha} c' G(\gamma_0)' + \tilde{e}') P^*(\gamma_0) (G(\gamma_0) c n^{-\alpha} + \tilde{e}) = n^{-2\alpha} c' G(\gamma_0)' (P^*(v) - P^*(\gamma_0)) G(\gamma_0) c + 2n^{-\alpha} c' G(\gamma_0)' (P^*(v) - P^*(\gamma_0)) \tilde{e} + \tilde{e}' (P^*(v) - P^*(\gamma_0)) \tilde{e}$$

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

(i)

Define $\tilde{X}_\gamma(\gamma, \gamma_0) = (\tilde{X}_\gamma, \tilde{A}_{1,\gamma}(\gamma_0), \tilde{A}_{2,\gamma}(\gamma_0))$ and $\tilde{X}_{\gamma}(\gamma_0) = \tilde{X}_\gamma(\gamma_0, \gamma_0)$. Then

Recall that

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

$$n^{-2\alpha} \left| \frac{1}{n} \tilde{X}_\gamma(\gamma)' \tilde{X}_\gamma(\gamma) - \frac{1}{n} \tilde{X}_0(\gamma)' \tilde{X}_0(\gamma) \right|$$
This establishes that uniformly on 

\[ \gamma_0 + \frac{\pi}{a_n} \leq \gamma \leq \gamma_0 + B \]

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

\[ n^{-2\alpha} \sum_{i=1}^{n} |g_i(v)|^2 \Delta_i(v) + 2n^{-2\alpha} \left| \sum_{i=1}^{n} g_i(v) \hat{e}_i' \Delta_i(v) \right| + n^{-2\alpha} \left| \sum_{i=1}^{n} \hat{e}_i \hat{e}_i' \Delta_i(v) \right| \implies \]

\( \{ \begin{array}{l}
|D_1 f| |v|, \quad v \in [-\pi, 0] \\
|D_2 f| |v|, \quad v \in [0, \pi]
\end{array} \)

Therefore,\( n^{-2\alpha} \sup_{|v| \leq \pi} |\hat{X}_v(v)'\hat{X}_v(v) - \hat{X}_0(\gamma_0)'\hat{X}_0(\gamma_0)| = O_p(1) \)

We also know from Lemma 2 that

\[ \frac{1}{n} \hat{X}_v(v)'\hat{X}_v(v) \implies M(\gamma_0) \quad (A.18) \]

Our analysis below will be restricted to the region \( [\gamma_0 + \pi/a_n \leq \gamma \leq \gamma_0 + B] \) for some constant \( B > 0 \), which follows from Lemma 1. Note that this restriction implies that \( \hat{X}_v'G_{X,0} = \hat{X}_0'G_{X,0}, \hat{X}_v'\hat{X}_0 = \hat{X}_0'\hat{X}_0 \)

The analysis for the case \( [\gamma_0 - \pi/a_n \geq \gamma \geq \gamma_0 - B] \) is similar.

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

\[ n^{-2\alpha} \left| c'G(\gamma_0)'(P^*(v) - P^*(\gamma_0))G(\gamma_0)c \right| = n^{-2\alpha} \left| c'G(\gamma_0)'(P_v(v) - P_0(\gamma_0))G(\gamma_0)c \right| \]

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

\[ n^{-2\alpha} \left| c'G(\gamma_0)'(P^*(v) - P^*(\gamma_0))G(\gamma_0)c \right| = n^{-2\alpha} \left| c' \left( \hat{X}_v(v)'\hat{X}_v(v) - \hat{X}_0(\gamma_0)'\hat{X}_0(\gamma_0) \right) \right| \]

\[ = -c' \left( \hat{X}_v(v)'\hat{X}_v(v) - \hat{X}_0(\gamma_0)'\hat{X}_0(\gamma_0) \right) \left( I - (\hat{X}_v(v)'\hat{X}_v(v))^{-1}\hat{X}_0(\gamma_0)'\hat{X}_0(\gamma_0) \right) c \]

\[ - c \left( I - G_0(\gamma_0)'\hat{X}_0(\gamma_0)'\hat{X}_0(\gamma_0) \right) \left( \hat{X}_v(v)'\hat{X}_v(v) - \hat{X}_0(\gamma_0)'\hat{X}_0(\gamma_0) \right) \left( \hat{X}_v(v)'\hat{X}_v(v) \right)^{-1}\hat{X}_0(\gamma_0)'G_0(\gamma_0)c + o_p(1) \]

\[ = n^{-2\alpha} \sum_{i=1}^{n} |g_i(v)|^2 \Delta_i(v) + o_p(1) \implies \mu_2 |v|. \]

This establishes that uniformly on \( [\gamma_0 + \pi/a_n \leq \gamma \leq \gamma_0 + B] \),

\[ n^{-2\alpha} \left| c'G(\gamma_0)'(P^*(\gamma_0) - P^*(v))G(\gamma_0)c \right| \implies \mu_2 |v| \quad (A.19) \]

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

34
\[ n^{-a}c' G_0(\gamma_0)' (P^*(\gamma_0) - P^*(v)) \tilde{e} = n^{-a}c' G_0(\gamma_0)' (P_0(\gamma_0) - P_0(v)) \tilde{e} \]

\[ = \left[ G_0(\gamma_0)' \hat{X}_0(\gamma_0) (\hat{X}_0(\gamma_0)' \hat{X}_0(\gamma_0))^{-1} \right] \left[ n^{-2a} (\hat{X}_v(v)' \hat{X}_v(v) - \hat{X}_0(\gamma_0)' \hat{X}_0(\gamma_0)) \right] \left[ n^a (\hat{X}_v(v)' \hat{X}_v(v))^{-1} \hat{X}_0(\gamma_0)' \tilde{e} \right] \]

Note that by Lemma 2 and (A.18) we can get uniform in \( v \in [0, \nu] \),

\[ n^a (\hat{X}_v(v)' \hat{X}_v(v))^{-1} \hat{X}_0(\gamma_0)' \tilde{e} = \frac{1}{n} \hat{X}_v(v)' \hat{X}_v(v)^{-1} \hat{X}_0(\gamma_0)' \tilde{e} = o_p(1) \quad (A.20) \]

and

\[ n^{-a} (\hat{X}_v(v)' - \hat{X}_0(\gamma_0)') \tilde{e} = n^{-a} \sum_{i=1}^{n} \hat{g}_i(v) \hat{e}_i \Delta_i(v) \]

\[ = n^{-a} \sum_{i=1}^{n} \hat{g}_i(v) \hat{r}_i \beta \Delta_i(v) + n^{-a} \sum_{i=1}^{n} \hat{g}_i(v) \hat{e}_i \beta \Delta_i(v) - n^{-a} \sum_{i=1}^{n} \hat{r}_i \hat{e}_i \Delta_i(v) \]

\[ \Rightarrow n^{-a} \sum_{i=1}^{n} \hat{g}_i(v) \hat{e}_i \Delta_i(v) + o_p(1) = B_1(v). \quad (A.21) \]

Then, it follows that

\[ n^{-a}c' G_0(\gamma_0)' (P^*(\gamma_0) - P^*(v)) \tilde{e} \Rightarrow B_1(v). \]

where \( B_1(v) \) a vector Brownian motion with covariance matrix \( \Omega_1 f \) and hence

\[ n^{-a}c' G(\gamma_0)' (P^*(\gamma_0) - P^*(v)) \tilde{e} \Rightarrow \zeta_1^{1/2} W_1(v) \quad (A.22) \]

(iii)

\[ \hat{e}'(P^*(\gamma_0) - P^*(v)) = \left[ n^a \hat{e}' \hat{X}_0(\gamma_0) (\hat{X}_0(\gamma_0)' \hat{X}_0(\gamma_0))^{-1} \right] \left[ n^{-2a} (\hat{X}_v(v)' \hat{X}_v(v) - \hat{X}_0(\gamma_0)' \hat{X}_0(\gamma_0)) \right] \left[ n^a (\hat{X}_v(v)' \hat{X}_v(v))^{-1} \hat{X}_0(\gamma_0)' \tilde{e} \right] \]

\[ = o_p(1). \quad \text{Hence,} \]

\[ \hat{e}'(P^*(\gamma_0) - P^*(v)) \tilde{e} \Rightarrow 0. \quad (A.23) \]

Using equation (A.10) and (A.19)-(A.23) we get
\[ Q_n(v) = S_n(\gamma_0) - S_n(\gamma_0 + v/a_n) \]
\[ = (n^{-\alpha}c'G(\gamma_0)' + \bar{c}')P^*(v)(G(\gamma_0)cn^{-\alpha} + \bar{e}) - (n^{-\alpha}c'G(\gamma_0)' + \bar{c}')P^*(\gamma_0)(G(\gamma_0)cn^{-\alpha} + \bar{r}) \]
\[ = n^{-2\alpha}c'G(\gamma_0)'(P^*(\gamma_0) - P^*(v))G(\gamma_0)c + 2n^{-\alpha}c'G(\gamma_0)'(P^*(\gamma_0) - P^*(v))\bar{e} + \bar{c}'(P^*(\gamma_0) - P^*(v))\bar{e} \]
\[ \Longrightarrow -\mu_1|v| + 2\zeta_1^{1/2}W_1(v), \text{ uniformly on } v \in [-\varepsilon, 0] \]

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

\[ \blacksquare \]

**Proof of Theorem 4.1**

By Lemma 3, \( a_n(\hat{\gamma} - \gamma_0) = \arg \max_v Q_n(v) = O_p(1) \) and by Lemma 4,

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

. Then, by Theorem 2.7 of Kim and Pollard (1990) and Theorem 1 of Hansen (2000) we can get \( n^{1-2\alpha}(\hat{\gamma} - \gamma_0) \overset{d}{\rightarrow} \arg \max Q_n(v) \).

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

\[ \arg \max_{-\infty < v < \infty} Q_n(v) \]
\[ = \begin{cases} 
\arg \max_{-\infty < v < \infty} \left( -\frac{\zeta_1}{\mu_1^2} |s| + 2\zeta_1^{1/2}W_1((\zeta_1/\mu_1^2)s) \right) = \omega \arg \max_{-\infty < v < \infty} \left( \frac{1}{2} |s| + W_1(s) \right), & \text{if } s \in [-\varepsilon, 0] \\
\arg \max_{-\infty < v < \infty} \left( -\frac{\zeta_1}{\mu_1^2} |s| + 2\zeta_2^{1/2}W_1((\zeta_1/\mu_1^2)s) \right) = \omega \arg \max_{-\infty < v < \infty} \left( \frac{1}{2} |s| + \varphi W_2(s) \right), & \text{if } s \in [0, \varepsilon] 
\end{cases} \]

where \( \xi = \mu_2/\mu_1 \) and \( \varphi = \zeta_2/\zeta_1 \). Hence, \( n^{1-2\alpha}(\hat{\gamma} - \gamma_0) \overset{d}{\rightarrow} \arg \max_{-\infty < v < \infty} \omega T(s) \), where

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

.    

\[ \blacksquare \]
Proof of Theorem 4.2

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

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

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

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

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

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

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

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

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

\[ \blacksquare \]

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

**Proof of Lemma 6.**

$$\hat{\beta}_1 = \left( \frac{1}{n} \tilde{X}_1^T \tilde{Z}_1 \tilde{W}_1 \tilde{Z}_1^T \tilde{X}_1 \right)^{-1} \tilde{X}_1^T \tilde{Z}_1 \tilde{W}_1 \tilde{Z}_1 \left( X_1/\beta_{10} + X_2/\beta_{20} + e \right) =$$

$$\left( \frac{1}{n} \tilde{X}_1^T \tilde{Z}_1 \right) \left( \frac{1}{n} \tilde{Z}_1^T \tilde{X}_1 \right)^{-1} \left( \frac{1}{n} \tilde{X}_1^T \tilde{Z}_1 \right) \left( \frac{1}{n} \tilde{Z}_1^T \tilde{X}_1 \right) \beta_{10} +$$

$$\left( \frac{1}{n} \tilde{X}_1^T \tilde{Z}_1 \right) \left( \frac{1}{n} \tilde{Z}_1^T \tilde{X}_1 \right)^{-1} \left( \frac{1}{n} \tilde{X}_1^T \tilde{Z}_1 \right) \left( \frac{1}{n} \tilde{Z}_1^T \tilde{X}_2 \right) \beta_{20} +$$

$$\left( \frac{1}{n} \tilde{X}_1^T \tilde{Z}_1 \right) \left( \frac{1}{n} \tilde{Z}_1^T \tilde{X}_1 \right)^{-1} \left( \frac{1}{n} \tilde{X}_1^T \tilde{Z}_1 \right) \left( \frac{1}{n} \tilde{Z}_1^T \tilde{e} \right)$$

Given $\tilde{W}_1 \xrightarrow{p} W_1 > 0$, the first term goes to zero by a Glivenko-Cantelli theorem and the second term goes to zero since $P(\hat{\gamma} < \gamma_0) \rightarrow 0$. Similarly we can show that

$$\left( \frac{1}{n} \tilde{X}_1^T \tilde{Z}_1 \right) \left( \frac{1}{n} \tilde{Z}_1^T \tilde{X}_1 \right)^{-1} \left( \frac{1}{n} \tilde{X}_1^T \tilde{Z}_1 \right) \left( \frac{1}{n} \tilde{Z}_1^T \tilde{X}_2 \right) \xrightarrow{p} 0$$

and

37
\[
\left(\frac{1}{n}\hat{X}_1'\hat{Z}_1\right)\hat{W}_1\left(\frac{1}{n}\hat{Z}_1'\hat{X}_1\right)^{-1}\left(\frac{1}{n}\hat{X}_1'\hat{Z}_1\right)\hat{W}_1\left(\frac{1}{n}\hat{Z}_1'e\right) \overset{p}{\to} 0.
\]

The proof is completed by showing that

\[
\|\left(\frac{1}{n}X_1(\hat{\gamma})'Z_1I(q \leq \hat{\gamma})\right)\hat{W}_1(\hat{\gamma})\left(\frac{1}{n}Z_1'X_1(\hat{\gamma})\right) - E(z_{1i}x_{1i}(\gamma_0)'I(q_i \leq \gamma_0)W_1(\gamma_0)E(x_{1i}(\gamma_0)'z_{1i}I(q_i \leq \gamma_0))|| =
\]

\[
\|\left(\frac{1}{n}\hat{X}_1'\hat{Z}_1\right)\hat{W}_1\left(\frac{1}{n}\hat{Z}_1'\hat{X}_1\right) -
\]

\[
\sup_{\gamma \in (\gamma_0 - \varepsilon, \gamma_0 + \varepsilon)}\|\left(\frac{1}{n}X_1(\gamma)'Z_1I(q \leq \gamma)\right)\hat{W}_1(\gamma)\left(\frac{1}{n}Z_1'X_1(\gamma)\right)I(q \leq \gamma) - E(z_{1i}x_{1i}(\gamma_0)'I(q_i \leq \gamma_0)W_1(\gamma_0)E(x_{1i}(\gamma_0)'z_{1i}I(q_i \leq \gamma_0))|| +
\]

\[
E(z_{1i}x_{1i}(\gamma_0)'I(q_i \leq \gamma_0)W_1(\gamma_0)E(x_{1i}(\gamma_0)'z_{1i}I(q_i \leq \gamma)) -
\]

\[
E(z_{1i}x_{1i}(\gamma_0)I(q_i \leq \gamma_0)W_1(\gamma_0)E(x_{1i}(\gamma_0)z_{1i}I(q_i \leq \gamma_0))||
\]

\[
\leq
\]

**LEMMA 7** Consider the unrestricted threshold model in equation (3.26) and recall that \(x_i(\gamma) = (x_i, \lambda_1(\gamma), \lambda_2(\gamma))'\). If \(\hat{W}_j \overset{p}{\to} W_j > 0\) for \(j = 1, 2\) then the unconstrained minimum distance class estimators defined by equation (2.19) are asymptotically Normal:

\[
\sqrt{n}(\hat{\beta}_j (\hat{\nu}) - \beta_j) \xrightarrow{d} N(0, V_j) \quad \text{(A.24)}
\]

where \(V_j = (S_j'W_jS_j)^{-1}(S_j'W_jQ_jW_jS_j)(S_j'W_jS_j)^{-1}\).

**Proof of Lemma 7**

First we show that the unconstrained estimators are asymptotically Normal. Let \(X_v (v), X_{\perp} (v), \Delta X_v (v), Z_v\) denote the matrices obtained by stacking the following unrestricted vectors

\[
x_i(\gamma_0 + n^{-1-2\alpha}v)'I(q_i \leq \gamma_0 + n^{-1-2\alpha}v),
\]

\[
x_i(\gamma_0 + n^{-1-2\alpha}v)'I(q_i > \gamma_0 + n^{-1-2\alpha}v),
\]

\[
x_i(\gamma_0 + n^{-1-2\alpha}v)'I(q_i \leq \gamma_0 + n^{-1-2\alpha}v) - x_i(\gamma_0 + n^{-1-2\alpha}v)'I(q_i > \gamma_0),
\]

\[
z_i'\varepsilon_i \leq \gamma_0 + n^{-1-2\alpha}v).
\]

From Theorem 2 of Hansen (1996), Lemma 1 and Lemma A.10 of Hansen (2000) we can deduce uniformly on \(v \in [-\bar{v}, \bar{v}]\)
\[
\frac{1}{n} Z'_v X_v (v) \xrightarrow{p} S_1 \tag{A.25}
\]
\[
\frac{1}{\sqrt{n}} Z'_v X_v (\gamma) \xrightarrow{p} N(0, \Sigma_1) \tag{A.26}
\]
\[
\frac{1}{n^{2\alpha}} Z'_v \Delta X_v \xrightarrow{p} O_p(1) \tag{A.27}
\]

Following Hansen and Caner (2004) let
\[
\tilde{\beta}_1 (v) = \left( X'_v \hat{Z}_v \hat{W}_1 \hat{Z}_v' X_v \right)^{-1} \hat{X}_v \hat{Z}_v \hat{W}_1 \hat{Z}_v' Y, \ j = 1, 2.
\]

and write the unrestricted model as
\[
Y = X_v (v) \beta_1 + X_\perp (v) \beta_2 - \Delta X_v (v) \delta_n + u
\]

\[
\sqrt{n}(\tilde{\beta}_1 (v) - \beta_1) =
\left( \left( \frac{1}{n} X_v (v)' Z_v \right) \hat{W}_1 \left( \frac{1}{n} Z'_v X_v (v) \right) \right)^{-1} \left( \frac{1}{n} X_v (v)' Z_v \hat{W}_1 \left( \frac{1}{\sqrt{n}} Z_v' u - \frac{1}{\sqrt{n}} Z'_v \Delta X_v (v) \delta_n \right) \right)
\Rightarrow (S'_1 W_1 S_1)^{-1} S'_1 W_1 N(0, \Sigma_1).
\]

Since \( \tilde{\nu} = n^{1-2\alpha} (\tilde{\gamma} - \gamma_0) = O_p(1) \),
\[
\sqrt{n}(\tilde{\beta}_1 (\tilde{\nu}) - \beta_1) \Rightarrow N(0, V_1)
\]

where \( V_1 = (S'_1 W_1 S_1)^{-1} (S'_1 W_1 Q_1 W_1 S_1) (S'_1 W_1 S_1)^{-1} \). Similarly we can get \( \sqrt{n}(\tilde{\beta}_1 (v) - \beta_2) \Rightarrow N(0, V_2) \) as stated.

**Lemma 8** The restricted estimators defined in equation (2.19) are asymptotically Normal.

\[
\sqrt{n}(\tilde{\beta} - \beta) \Rightarrow N(0, \tilde{V})
\]

where
\[
\tilde{V} = V - WR \left( R' \hat{WR} \right)^{-1} R' v - v R \left( R' \hat{WR} \right)^{-1} R' \hat{W}
+ \hat{WR} \left( R' \hat{WR} \right)^{-1} R' v R \left( R' \hat{WR} \right)^{-1} R' \hat{W}. \tag{A.28}
\]

Proof of Lemma 8
Let \( \tilde{\beta}^* = (\tilde{\beta}_1, \tilde{\beta}_2)' \) and \( \beta = (\beta_1, \beta_2)' \), \( \tilde{W} = diag(\tilde{W}_1, \tilde{W}_2) \), \( V = diag(V_1, V_2) \)

Recalling that \( R'\tilde{\beta} = \vartheta \) the restricted estimator of the STR model can be written as

\[
\tilde{\beta} = \beta - \tilde{W} R (R' \tilde{W} R)^{-1} (R' \tilde{\beta} - \vartheta) \tag{A.29}
\]

then using Lemma 7 we get

\[
\sqrt{n}(\tilde{\beta} - \beta) \Rightarrow \left( I - \tilde{W} R (R' V R)^{-1} R' \right) \sqrt{n}(\tilde{\beta} - \beta) = N(0, \tilde{V}) \tag{A.30}
\]

as stated.

**Proof of Theorem 4.3**

The 2SLS estimators \( \tilde{\beta}_{2SLS} \) fall in the class of estimators (2.19) with \( \tilde{W} = diag(\tilde{W}_1, \tilde{W}_2) \)

\[
\tilde{W}_1 = \frac{1}{n} \sum_{i=1}^{n} z_i z_i' I(q_i \leq \tilde{\gamma})
\]

\[
\tilde{W}_2 = \frac{1}{n} \sum_{i=1}^{n} z_i z_i' I(q_i > \tilde{\gamma})
\]

The proof for (a) follows Theorem 2 of Caner and Hansen (2004). For the 2SLS estimator, we appeal to Lemma 1 of Hansen (1996), the consistency of \( \tilde{\gamma}, \tilde{W}_1 \xrightarrow{p} Q_1 \) and \( \tilde{W}_2 \xrightarrow{p} Q_2 \). Therefore, \( \tilde{\beta}_{2SLS} \) is asymptotically Normal with covariance matrix as stated in (A.28) with \( Q = diag(Q_1, Q_2) \) replacing \( \tilde{W} = diag(\tilde{W}_1, \tilde{W}_2) \).

The proof for (b) follows Theorem 3 of Caner and Hansen (2004), which is used to establish that \( \tilde{\Sigma}_1(\gamma) \xrightarrow{p} E \left( z_i z_i' I(q_i \leq \gamma) \right) \) uniformly in \( \gamma \in \Gamma \). Then, by the consistency of \( \tilde{\gamma} \), \( n^{-1} \tilde{\Sigma}_1 = n^{-1} \tilde{\Sigma}_1(\gamma) \xrightarrow{p} \Sigma_1 \), and Lemmas 7 and 8 we obtain Theorem 4.3 (b).
Figures 1(a) – (f): MC Kernel Densities of the Threshold Estimate (Exogenous Slope Variable)
Estimates based on STR and TR for $\delta = 1$ and various sample sizes

 inclined The solid line represents the MC kernel density of the STR threshold estimate while the dotted line represents the corresponding density for the TR (Hansen, 2000) threshold estimate.
Figures 2(a) – (f): MC Kernel Densities of the Threshold Estimate (Endogenous Slope Variable)

Estimates based on STR and IVTR for $\delta = 1$ and various sample sizes

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

† The solid line represents the MC kernel density of the STR threshold estimate while the dotted line represents the corresponding density for the IVTR (Caner and Hansen, 2004) threshold estimate.
Table 1: Quantiles of Threshold Estimator, $\gamma = 2$

<table>
<thead>
<tr>
<th>Quantiles</th>
<th>$5^{th}$</th>
<th>$50^{th}$</th>
<th>$95^{th}$</th>
<th>$5^{th}$</th>
<th>$50^{th}$</th>
<th>$95^{th}$</th>
<th>$5^{th}$</th>
<th>$50^{th}$</th>
<th>$95^{th}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta = 0.50$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 100</td>
<td>1.645</td>
<td>1.964</td>
<td>2.090</td>
<td>1.580</td>
<td>1.958</td>
<td>2.189</td>
<td>0.613</td>
<td>1.953</td>
<td>2.517</td>
</tr>
<tr>
<td>n = 200</td>
<td>1.855</td>
<td>1.983</td>
<td>2.045</td>
<td>1.773</td>
<td>1.977</td>
<td>2.073</td>
<td>1.498</td>
<td>1.979</td>
<td>2.187</td>
</tr>
<tr>
<td>n = 500</td>
<td>1.950</td>
<td>1.994</td>
<td>2.019</td>
<td>1.922</td>
<td>1.992</td>
<td>2.027</td>
<td>1.887</td>
<td>1.991</td>
<td>2.060</td>
</tr>
<tr>
<td>$\delta = 1.00$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 100</td>
<td>1.874</td>
<td>1.975</td>
<td>2.032</td>
<td>1.874</td>
<td>1.974</td>
<td>2.041</td>
<td>1.829</td>
<td>1.974</td>
<td>2.082</td>
</tr>
<tr>
<td>n = 200</td>
<td>1.932</td>
<td>1.988</td>
<td>2.013</td>
<td>1.929</td>
<td>1.987</td>
<td>2.014</td>
<td>1.908</td>
<td>1.987</td>
<td>2.034</td>
</tr>
<tr>
<td>n = 500</td>
<td>1.975</td>
<td>1.994</td>
<td>2.005</td>
<td>1.973</td>
<td>1.995</td>
<td>2.008</td>
<td>1.964</td>
<td>1.994</td>
<td>2.015</td>
</tr>
<tr>
<td>$\delta = 2.00$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 100</td>
<td>1.888</td>
<td>1.975</td>
<td>2.001</td>
<td>1.889</td>
<td>1.976</td>
<td>2.010</td>
<td>1.882</td>
<td>1.976</td>
<td>2.023</td>
</tr>
<tr>
<td>n = 200</td>
<td>1.943</td>
<td>1.988</td>
<td>2.000</td>
<td>1.942</td>
<td>1.988</td>
<td>2.000</td>
<td>1.939</td>
<td>1.987</td>
<td>2.012</td>
</tr>
<tr>
<td>n = 500</td>
<td>1.976</td>
<td>1.995</td>
<td>2.000</td>
<td>1.976</td>
<td>1.995</td>
<td>2.001</td>
<td>1.974</td>
<td>1.994</td>
<td>2.003</td>
</tr>
</tbody>
</table>

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 $\delta$. We consider two designs: (i) columns (1)-(6) consider the case where threshold variable is endogenous but the slope variable is exogenous and compare the results of Hansen’s (2000) TR model (equation (2.19) in the text, under $\sigma_{uv} = 0$) vis-à-vis STR (equation (2.17) in the text, under $\sigma_{uv} = 0$); (ii) columns (7)-(12) consider the case where both the threshold variable and slope variable are endogenous and compare the results of Caner and Hansen’s (2004) IVTR model (equation (2.19) in the text, under $\sigma_{uv} \neq 0$) vis-à-vis STR (equation (2.17) in the text under $\sigma_{uv} \neq 0$).
<table>
<thead>
<tr>
<th>Quantiles</th>
<th>$5^{th}$</th>
<th>$50^{th}$</th>
<th>$95^{th}$</th>
<th>$5^{th}$</th>
<th>$50^{th}$</th>
<th>$95^{th}$</th>
<th>$5^{th}$</th>
<th>$50^{th}$</th>
<th>$95^{th}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exogenous Slope Variable</td>
<td>TR</td>
<td>STR</td>
<td>TR</td>
<td>STR</td>
<td>TR</td>
<td>STR</td>
<td>TR</td>
<td>STR</td>
<td></td>
</tr>
<tr>
<td>$\delta = 0.50$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 100</td>
<td>0.843</td>
<td>0.917</td>
<td>0.99</td>
<td>0.903</td>
<td>0.999</td>
<td>1.115</td>
<td>1.121</td>
<td>1.194</td>
<td>1.322</td>
</tr>
<tr>
<td>n = 200</td>
<td>0.869</td>
<td>0.917</td>
<td>0.97</td>
<td>0.934</td>
<td>1.001</td>
<td>1.081</td>
<td>1.133</td>
<td>1.184</td>
<td>1.250</td>
</tr>
<tr>
<td>n = 500</td>
<td>0.888</td>
<td>0.917</td>
<td>0.946</td>
<td>0.959</td>
<td>1.000</td>
<td>1.045</td>
<td>1.144</td>
<td>1.175</td>
<td>1.211</td>
</tr>
<tr>
<td>$\delta = 1.00$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 100</td>
<td>0.844</td>
<td>0.918</td>
<td>0.987</td>
<td>0.902</td>
<td>0.996</td>
<td>1.110</td>
<td>1.111</td>
<td>1.178</td>
<td>1.244</td>
</tr>
<tr>
<td>n = 200</td>
<td>0.870</td>
<td>0.918</td>
<td>0.972</td>
<td>0.935</td>
<td>1.000</td>
<td>1.076</td>
<td>1.129</td>
<td>1.175</td>
<td>1.218</td>
</tr>
<tr>
<td>n = 500</td>
<td>0.888</td>
<td>0.918</td>
<td>0.946</td>
<td>0.959</td>
<td>1.000</td>
<td>1.044</td>
<td>1.142</td>
<td>1.172</td>
<td>1.203</td>
</tr>
<tr>
<td>$\delta = 2.00$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 100</td>
<td>0.845</td>
<td>0.918</td>
<td>0.988</td>
<td>0.904</td>
<td>0.997</td>
<td>1.112</td>
<td>1.108</td>
<td>1.175</td>
<td>1.240</td>
</tr>
<tr>
<td>n = 200</td>
<td>0.870</td>
<td>0.918</td>
<td>0.972</td>
<td>0.935</td>
<td>1.000</td>
<td>1.078</td>
<td>1.127</td>
<td>1.173</td>
<td>1.217</td>
</tr>
<tr>
<td>n = 500</td>
<td>0.888</td>
<td>0.918</td>
<td>0.946</td>
<td>0.959</td>
<td>1.000</td>
<td>1.044</td>
<td>1.142</td>
<td>1.172</td>
<td>1.203</td>
</tr>
</tbody>
</table>

This Table presents Monte Carlo results for the 5th, 50th, and 95th quantiles for the slope coefficient of the second regime $\beta = \beta_2 = 1$ when the threshold variable is endogenous for $\gamma = 2$ and various values of $\delta$. 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 $\sigma_{w} = 0$) vis-à-vis STR (equation (2.17) in the text under $\sigma_{w} = 0$); (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 $\sigma_{w} \neq 0$) vis-à-vis STR (equation (2.17) in the text, under $\sigma_{w} \neq 0$).
<table>
<thead>
<tr>
<th>Method</th>
<th>Exogenous Slope</th>
<th>Endogenous Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \delta_2 = 0.5 )</td>
<td>( n = 50 )</td>
<td>0.729</td>
</tr>
<tr>
<td></td>
<td>( n = 100 )</td>
<td>0.887</td>
</tr>
<tr>
<td></td>
<td>( n = 200 )</td>
<td>0.930</td>
</tr>
<tr>
<td></td>
<td>( n = 500 )</td>
<td>0.997</td>
</tr>
<tr>
<td>( \delta_2 = 1.00 )</td>
<td>( n = 50 )</td>
<td>0.794</td>
</tr>
<tr>
<td></td>
<td>( n = 100 )</td>
<td>0.969</td>
</tr>
<tr>
<td></td>
<td>( n = 200 )</td>
<td>0.995</td>
</tr>
<tr>
<td></td>
<td>( n = 500 )</td>
<td>1.000</td>
</tr>
<tr>
<td>( \delta_2 = 2.00 )</td>
<td>( n = 50 )</td>
<td>0.808</td>
</tr>
<tr>
<td></td>
<td>( n = 100 )</td>
<td>0.981</td>
</tr>
<tr>
<td></td>
<td>( n = 200 )</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td>( n = 500 )</td>
<td>1.000</td>
</tr>
</tbody>
</table>