Econometric Theory, **32**, 2016, 827–860.
doi:10.1017/S0266466615000067

STRUCTURAL THRESHOLD REGRESSION

Andros Kourtellos
University of Cyprus

Thanasis Stengos
University of Guelph

Chih Ming Tan
University of North Dakota

This paper introduces the structural threshold regression (STR) 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, *Econometrica* 68, 575–603) and Caner and Hansen (2004, *Econometric Theory* 20, 813–843) to allow for the endogeneity of the threshold variable and regime-specific heteroskedasticity. Our estimation of the threshold parameter is based on a two-stage concentrated least squares method that involves an inverse Mills ratio bias correction term in each regime. We derive its asymptotic distribution and propose a method to construct confidence intervals. We also provide inference for the slope parameters based on a generalized method of moments. Finally, we investigate the performance of the asymptotic approximations using a Monte Carlo simulation, which shows 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

We thank the editor Peter C. B. Phillips, the co-editor Oliver Linton, and two anonymous referees whose comments greatly improved the paper. We also thank Bruce Hansen for helpful comments and seminar participants at the Athens University of Economics and Business, Hebrew University of Jerusalem, Ryerson University, Simon Fraser University, Universit libre de Bruxelles, University of Cambridge, University of Palermo, University of Waterloo, the University of Western Ontario, 10th World Congress of the Econometric Society in Shanghai, 27th Annual Meeting of the Canadian Econometrics Study Group in Vancouver, and 23rd $(EC)^2$ conference. Kourtellos thanks the University of Cyprus for funding. Tan thanks the Greg and Cindy Page Faculty Distribution Fund for financial support. Address correspondence to Andros Kourtellos, Department of Economics, University of Cyprus, P.O. Box 537, CY 1678 Nicosia, Cyprus; e-mail: andros@ucy.ac.cy.

flexibility in functional form and at the same time is not as susceptible to curse of dimensionality problems as nonparametric methods.

In this paper, we introduce the structural threshold regression (STR) model, which is a threshold regression that allows for endogeneity in the threshold variable as well as in the slope regressors, and develop estimation and inference for weakly dependent data. Our research is related to several recent papers in the literature; see for example Hansen (2000), Caner and Hansen (2004), Seo and Linton (2007), Gonzalo and Wolf (2005), and Yu (2012, 2013a). The main difference of all these papers with our work is that they maintain the assumption that the threshold variable is exogenous. This assumption severely limits the usefulness of threshold regression models in practice, since in economics many plausible threshold variables are endogenous.

For example, Papageorgiou (2002) organized countries into multiple growth regimes using the trade share, defined as the ratio of imports plus exports to real GDP in 1985, as a threshold variable. Similarly, Tan (2010) classified countries into development clubs using the average expropriation risk from 1984–1997 as the threshold variable. In each of these cases, there is strong evidence in the growth literature; see Frankel and Romer (1999) and Acemoglu, Johnson, and Robinson (2001), respectively, that the proposed threshold variable is endogenous. As Yu (2013b) argues, if the threshold variable is endogenous, the existing threshold regression estimation methods of Hansen (2000) and Caner and Hansen (2004) yield inconsistent estimates. One way to understand the reason for bias is to note that, just as in the limited dependent variable framework, a set of inverse Mills ratio bias correction terms is required to restore the conditional mean zero assumption of the errors.

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

Specifically, we propose to estimate the threshold parameter using a two-step concentrated least squares (CLS) method and the slope parameters using a two-stage-least squares (2SLS) or a generalized method of moments (GMM). Then, we show the consistency of our estimators and derive the corresponding asymptotic distributions. In particular, our estimation approach follows Hansen (2000) and Caner and Hansen (2004) with the difference that the concentrated criterion involves inverse Mills ratio terms which are different across the two regimes. This, in effect, means that there is a cross-regime restriction, which in turn implies that the estimates cannot be analyzed using results obtained regime-by-regime. To overcome the problem that the model cannot be analyzed regime-by-regime, we explore the relationship between the constrained and unconstrained sum of squared errors. It turns out that when the constraints are valid, the rate of convergence of the threshold estimator is not improved relative to the unconstrained problem. We also find that, in large samples, the asymptotic distribution of the threshold estimator in the unconstrained optimization problem is equivalent to the distribution of the threshold estimator in the constrained problem. Our finding is similar to the result of Perron and Qu (2006) who consider change-point models with restrictions across regimes.[1]

An additional implication of having different inverse Mills ratio terms across regimes is that the errors of the STR model are regime-specific heteroskedastic. Our framework for the asymptotic distribution of the threshold parameter estimator follows Hansen (2000) who assumes that the threshold effect diminishes as the sample size increases. This assumption is the key to overcoming a problem that was first pointed out by Chan (1993). Chan shows that while the threshold estimate is superconsistent, its asymptotic distribution 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.[2] Under the assumption of the diminishing threshold effect we reduce the rate of convergence and obtain a useful asymptotic distribution, which is characterized by parameters associated with regime-specific heteroskedasticity as in the case of change-point models; see Bai (1997). More precisely, it involves two independent Brownian motions with two different scales. These scale parameters are estimable and by numerically inverting the likelihood ratio, we obtain an asymptotically valid confidence interval. To examine the finite sample properties of our estimators, we provide a Monte Carlo analysis.

Our paper is closely related to Yu and Phillips (2014) who propose a nonparametric estimator of the threshold parameter, namely the integrated difference kernel estimator. Using the fixed threshold effect framework of Chan (1993) they show that the threshold parameter can be identified and estimated without the use of any instruments at the rate n. Interestingly, instrumental variables are also not necessary for the identification and estimation of the threshold effect parameters, which are estimated at a nonparametric rate, that is, slower than $\sqrt{n}$. However, regime-specific regression coefficients can only be identified and estimated at the usual semiparametric $\sqrt{n}$ rate when instrumental variables are available. The instruments can also provide efficiency improvements to the nonparametric estimator of the threshold parameter and allow the estimation of the threshold effect parameters at a $\sqrt{n}$ rate. One important difference between Yu and Phillips (2014) and the current paper is that the former is restricted to $i.i.d.$ data while we allow for stationary and ergodic time series data, which is useful in many applications in macroeconomics and finance. Furthermore, our framework also allows for regime-specific heteroskedasticity, which is a consequence of the control function approach we employ to remedy the endogeneity of the threshold variable. Another challenge of the nonparametric approach of Yu and Phillips (2014) is the choice of bandwidth; their analysis is limited to constraints on rates and does not offer specific criteria for bandwidth selection.

In terms of the broader literature, our paper is related to Seo and Linton (2007) who allow the threshold variable to be a linear index of observed variables. They avoid the assumption of the shrinking threshold by proposing a smoothed least squares estimation strategy based on smoothing the objective function in the sense of Horowitz's smoothed maximum score estimator. Although they show that their estimator exhibits asymptotic normality, their estimation method depends on the choice of bandwidth. Other recent works have proposed alternative approaches to constructing the asymptotic distribution of threshold estimators. For example, Gonzalo and Wolf (2005) proposed subsampling to conduct inference in the context of threshold autoregressive models. Yu (2012) proposes a semiparametric empirical Bayes estimator of the threshold parameter and shows that it is semiparametrically efficient. Finally, Yu (2014a) 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 a discontinuous threshold regression model. He also finds that the asymptotic nonparametric bootstrap distribution of the threshold estimate depends on the sampling path of the original data.

The paper is organized as follows. Section 2 describes the model. Section 3 presents the estimation approach. Section 4 develops the asymptotic theory for our estimators. Section 5 presents our Monte Carlo experiments. Section 6 concludes. In the appendix, we collect the proofs of the main results. Supplementary proofs are given in Kourtellos, Stengos, and Tan (2014)-henceforth, we will refer to this as the Internet Appendix.

2. THE MODEL

Let $\{y_i, z_i, x_i, q_i\}_{i=1}^n$ be an i.i.d or a weakly dependent observed sample, where y_i is real valued, z_i is a $l \times 1$ vector, x_i is a $p \times 1$ vector such that $l \geq p$, and q_i is a scalar. Consider the following structural threshold regression model

$$y_i = \beta_{x1}' x_i + u_i, \quad q_i \leq \gamma, \tag{2.1a}$$

$$y_i = \beta_{x2}' x_i + u_i, \quad q_i > \gamma, \tag{2.1b}$$

where q_i is the threshold variable that splits the sample into two regimes each of which obeys a linear model. In each of the two linear models, y_i is a dependent variable, x_i is a vector of slope variables (regressors) including an intercept, and u_i is the equation error with $E(u_i|\mathcal{F}_{i-1}) = 0$, where the sigma field $\mathcal{F}_{i-1}$ is generated by $\{z_{i-j}, x_{i-1-j}, q_{i-1-j}, u_{i-1-j} : j \geq 0\}$. The parameters of interest, which are assumed to be unknown, include the scalar threshold parameter or sample split value, $\gamma \in \Gamma$, where Γ is a strict subset of the support of q_i and the slope (or regression) coefficients $\beta_x = (\beta_{x1}', \beta_{x2}')' \in R^{2p}$.[3]

https://doi.org/10.1017/S0266466615000067 Published online by Cambridge University Press

2.1. Endogeneity Only in the Threshold Variable

Consider the case where x_i is a vector of strictly exogenous regressors and a strict subset of z_i. Then the problem of endogeneity bias arises when conditional on $\mathcal{F}_{i-1}$, u_i is contemporaneously correlated with q_i. In this case, as Yu (2013b) shows, the standard CLS estimator of Hansen (2000) is biased and inconsistent. In particular, consider the reduced form model for the threshold variable q_i given by

$$q_i = \pi_q' z_i + v_{qi} \tag{2.2}$$

where $E(v_{qi}|\mathcal{F}_{i-1}) = 0$. Then, the endogeneity in the threshold variable amounts to $E(u_i|\mathcal{F}_{i-1}, v_{qi}) \neq 0$. Equation (2.2) is analogous to a selection equation that appears in the literature on limited dependent variable models; see Heckman (1979). The main difference is that while limited dependent variable models treat q_i as latent and the sample split as observed, here we treat the sample split value as an unknown parameter and we estimate it. In this paper, we allow for the equation error u_i to be correlated with both the threshold variable q_i and the regressors x_i.

We proceed to account for the "selection" bias by making the following assumptions.

Assumption 1.

1.1 $E(u_i|\mathcal{F}_{i-1}) = 0$
1.2 $E(v_{qi}|\mathcal{F}_{i-1}) = 0$
1.3 $E(u_i|\mathcal{F}_{i-1}, v_{qi}) = E(u_i|v_{qi})$
1.4 $E(u_i|v_{qi}) = \kappa v_{qi}$
1.5 $v_{qi} \sim N(0,1)$

Assumption 1.1 and 1.2 impose that the errors u_i and v_{qi} are martingale differences. Assumption 1.3 assumes conditional mean independence between u_i and $\mathcal{F}_{i-1}$. Assumption 1.4 assumes a linear conditional expectation between the errors of the structural and the reduced form equations. Assumption 1.5 assumes normality for the error of the reduced form equation of q_i. Although not trivial, Assumptions 1.4 and 1.5 can be relaxed and the bias correction terms can be estimated by semiparametric methods such as a series approximation; see Li and Wooldridge (2002).

Using Assumption 1 we get

$$\begin{aligned} E\left(u_i|\mathcal{F}_{i-1}, v_{qi} \leq \gamma - z_i'\pi_q\right) &= \kappa E\left(v_{qi}|v_{qi} \leq \gamma - z_i'\pi_q\right) \\ &= \kappa \int_{-\infty}^{\gamma - z_i'\pi_q} v_q f\left(v_q|v_q \leq \gamma - z_i'\pi_q\right) dv_q \\ &= \kappa\lambda_1\left(\gamma - z_i'\pi_q\right), \end{aligned} \tag{2.3a}$$

$$\begin{aligned} E\left(u_i|\mathcal{F}_{i-1}, v_{qi} > \gamma - z_i'\pi_q\right) &= \kappa E\left(v_{qi}|v_{qi} > \gamma - z_i'\pi_q\right) \\ &= \kappa \int_{\gamma - z_i'\pi_q}^{+\infty} v_q f\left(v_q|v_q > \gamma - z_i'\pi_q\right) dv_q \\ &= \kappa\lambda_2\left(\gamma - z_i'\pi_q\right), \end{aligned} \tag{2.3b}$$

where $\lambda_1(\gamma - z_i'\pi_q) = -\frac{\phi(\gamma - z_i'\pi_q)}{\Phi(\gamma - z_i'\pi_q)}$ and $\lambda_2(\gamma - z_i'\pi_q) = \frac{\phi(\gamma - z_i'\pi_q)}{1-\Phi(\gamma - z_i'\pi_q)}$ are the inverse Mills ratio terms. $\phi(\cdot)$ and $\Phi(\cdot)$ are the normal pdf and cdf, respectively. Note that while we do not make any specific distributional assumption about u_i, the normality of v_{qi} is key for the derivation of the inverse Mills ratio terms.

Denote the inverse Mills ratio terms at the true value π_{q0} as $\lambda_{1i}(\gamma) = \lambda_1(\gamma - z_i'\pi_{q0})$ and $\lambda_{2i}(\gamma) = \lambda_2(\gamma - z_i'\pi_{q0})$. Then taking conditional expectations in equations (2.1a)–(2.1b) yields

$$E\left(y_i|\mathcal{F}_{i-1}, v_{qi} \le \gamma - z_i'\pi_{q0}\right) = \beta_{x1}'x_i + E\left(u_i|\mathcal{F}_{i-1}, v_{qi} \le \gamma - z_i'\pi_{q0}\right)$$
$$= \beta_{x1}'x_i + \kappa\lambda_{1i}(\gamma) \quad \textbf{(2.4a)}$$
$$E\left(y_i|\mathcal{F}_{i-1}, v_{qi} > \gamma - z_i'\pi_{q0}\right) = \beta_{x2}'x_i + E\left(u_i|\mathcal{F}_{i-1}, v_{qi} > \gamma - z_i'\pi_{q0}\right),$$
$$= \beta_{x2}'x_i + \kappa\lambda_{2i}(\gamma) \quad \textbf{(2.4b)}$$

The STR model is then defined by

$$y_i = \beta_{x1}'x_i + \kappa\lambda_{1i}(\gamma) + \varepsilon_{1i}, \quad q_i \le \gamma \quad \textbf{(2.5a)}$$
$$y_i = \beta_{x2}'x_i + \kappa\lambda_{2i}(\gamma) + \varepsilon_{2i}, \quad q_i > \gamma \quad \textbf{(2.5b)}$$

where $\varepsilon_{1i} = -\kappa\lambda_{1i}(\gamma) + u_i$ and $\varepsilon_{2i} = -\kappa\lambda_{2i}(\gamma) + u_i$.

It is useful to write the model in a single equation by making the following definitions

$$I(\cdot) = \begin{cases} 1 & \text{iff } q_i \le \gamma \\ 0 & \text{iff } q_i > \gamma \end{cases}$$
$$\Lambda_i(\gamma) = \lambda_{1i}(\gamma)I(q_i \le \gamma) + \lambda_{2i}(\gamma)I(q_i > \gamma) \quad \textbf{(2.6)}$$
$$\varepsilon_i = \varepsilon_{1i}I(q_i \le \gamma) + \varepsilon_{2i}I(q_i > \gamma) \quad \textbf{(2.7)}$$

We can then express equations (2.5a) and (2.5b) as

$$y_i = \beta_{x1}'x_iI(q_i \le \gamma) + \beta_{x2}'x_iI(q_i > \gamma) + \kappa\Lambda_i(\gamma) + \varepsilon_i, \quad \textbf{(2.8)}$$

where $E(\varepsilon_i|\mathcal{F}_{i-1}) = 0$.

Note that equation (2.8) shows that the STR model nests the threshold regression model of Hansen (2000); henceforth TR model, when $\kappa = 0$. However, when u_i is correlated with q_i, we get $\kappa \ne 0$. This implies that estimating equations (2.1a)–(2.1b) using the estimators of the TR model results in the omission of the inverse Mills ratio bias correction terms. This, in turn, yields inconsistent estimates of the slope parameters β_{x1} and β_{x2}. Another difference between STR and TR is that the presence of different inverse Mills ratio terms in each of the regimes in STR necessarily implies the presence of regime-specific heteroskedasticity as can be seen in equation (2.7).

Our asymptotic framework is based on the mathematical device of the "small threshold" effect. In particular, we assume that the threshold effect, $\beta_{x1} - \beta_{x2} = \delta_{xn}$, and the degree of endogeneity bias, $\kappa = \kappa_n$, will both tend to

https://doi.org/10.1017/S0266466615000067 Published online by Cambridge University Press

zero slowly as n diverges. The latter assumption implies that the endogeneity bias vanishes as $n \to \infty$ to ensure that the bias correction (i.e. the inverse Mills ratio terms) to the endogeneity of the threshold will not be present when the model is linear (i.e., when there is only one regime). Using the assumption of a diminishing threshold effect and allowing for nonregime specific heteroskedasticity Hansen (2000) showed that the threshold estimate has an asymptotic distribution that only depends on a scale parameter. Similarly, in our case, using this assumption but allowing for regime-specific heteroskedasticity we will derive below an asymptotic distribution of the threshold estimate that depends on two scale parameters.

2.2. Endogeneity in Both the Threshold and Slope Variables

When the slope variables are also endogenous and x_i is not a subset of z_i the reduced form model for x_i takes the form

$$x_i = \Pi_x' z_i + v_{xi}, \tag{2.9}$$

where $E(v_{xi}|\mathcal{F}_{i-1}) = 0$ and Π_x is a $l \times p$ matrix of unknown parameters. Denote the conditional expectation at the true value Π_{x0} as $g_{xi} = E(x_i|\mathcal{F}_{i-1}) = \Pi_{x0}' z_i$. It is important to note that the assumption of the correct specification of the conditional mean for x_i is crucial for our theory. The assumptions that are needed to restore the conditional mean zero property of the error u_i, in this case, are Assumptions 1.1–1.5 augmented with

1.6 $E(v_{xi}|\mathcal{F}_{i-1}) = 0$.
1.7 $v_{xi} \perp I(v_{qi} \leq \gamma - z_i'\pi_{q0})|\mathcal{F}_{i-1}$

Assumptions 1.6 and 1.7 allow us to write $E(x_i|\mathcal{F}_{i-1}, v_{qi} \leq \gamma - z_i'\pi_{q0}) = E(x_i|\mathcal{F}_{i-1}) = \Pi_{x0}' z_i$ and $E(x_i|\mathcal{F}_{i-1}, v_{qi} > \gamma - z_i'\pi_{q0}) = E(x_i|\mathcal{F}_{i-1}) = \Pi_{x0}' z_i$.[4] Then under Assumption 1 the corresponding equations to (2.4a) and (2.4b) become

$$E(y_i|\mathcal{F}_{i-1}, v_{qi} \leq \gamma - z_i'\pi_{q0}) = \beta_{x1}' g_{xi} + \kappa \lambda_{1i}(\gamma) \tag{2.10a}$$

$$E(y_i|\mathcal{F}_{i-1}, v_{qi} > \gamma - z_i'\pi_{q0}) = \beta_{x2}' g_{xi} + \kappa \lambda_{2i}(\gamma). \tag{2.10b}$$

and, using analogous definitions as in Section 2.1 as well as equation (2.9) evaluated at the true value, the STR model that allows for endogeneity in both the threshold and slope variables can be written as follows

$$y_i = \beta_{x1}' g_{xi} I(q_i \leq \gamma) + \beta_{x2}' g_{xi} I(q_i > \gamma) + \kappa \Lambda_i(\gamma) + e_i^*, \tag{2.11}$$

where $e_i^* = \beta_{x1}' v_{xi} I(q_i \leq \gamma) + \beta_{x2}' v_{xi} I(q_i > \gamma) + \varepsilon_i$ with $E(e_i^*|\mathcal{F}_{i-1}) = 0$. Notice that the instrumental variable threshold regression model of Caner and Hansen (2004); henceforth IVTR model, arises as a special case of the STR model in equation (2.11) when $\kappa = 0$.

One possible concern in applied work is the assumption of linearity in the reduced form of x_i. This assumption can be relaxed to allow for nonlinearities such

https://doi.org/10.1017/S0266466615000067 Published online by Cambridge University Press

as a threshold regression in the first stage as in the IVTR model. However, this extension is not trivial. For example, Boldea, Hall, and Han (2012) and Hall, Han, and Boldea (2012) studied the problem of having an unstable reduced form in the context of change-point models with endogenous regressors and found that inference is not invariant to the nature of the reduced form. In particular, Boldea et al. (2012) derived a limiting distribution theory and constructed approximate large sample confidence intervals for the break points under the following three assumptions: (i) the reduced form is unstable; (ii) the magnitudes of the parameter change in both the equation of interest and the reduced form shrink with the sample size; and (iii) the break shifts are nearly weakly identified at different rates and locations for the structural equation and reduced form. We expect that similar difficulties and solutions may apply in the context of threshold regression.

3. ESTIMATION

We proceed in three steps to estimate equation (2.11): a two-step concentrated LS method to estimate the threshold parameter and an additional step to produce estimates of the slope coefficients.

3.1. Threshold Estimation

First, we estimate the reduced form parameters π_q and Π_x by LS in equations (2.2) and (2.9) to obtain $\widehat{\pi}_q$ and $\widehat{\Pi}_x$, respectively. The fitted values are then given by $\widehat{q}_i = \widehat{\pi}_q' z_i$ and $\widehat{x}_i = \widehat{g}_{xi} = \widehat{\Pi}_x' z_i$ along with first stage residuals, $\widehat{v}_{xi} = x_i - \widehat{x}_i$ and $\widehat{v}_{qi} = q_i - \widehat{q}_i$, respectively.

For any γ, define the following predicted objects. Define the predicted inverse Mills ratio term $\widehat{\Lambda}_i(\gamma) = \widehat{\lambda}_{1i}(\gamma) I(q_i \leq \gamma) + \widehat{\lambda}_{2i}(\gamma) I(q_i > \gamma)$, where $\widehat{\lambda}_{1i}(\gamma) = \lambda_1(\gamma - z_i'\widehat{\pi}_q)$ and $\widehat{\lambda}_{2i}(\gamma) = \lambda_2(\gamma - z_i'\widehat{\pi}_q)$. Let $\widehat{\mathfrak{x}}_i(\gamma) = (x_i' I(q_i \leq \gamma), x_i' I(q_i > \gamma), \widehat{\Lambda}_i(\gamma))'$ and $\widehat{\mathfrak{z}}_i(\gamma) = (z_i' I(q_i \leq \gamma), z_i' I(q_i > \gamma), \widehat{\Lambda}_i(\gamma))'$.

Second, we estimate the threshold parameter γ using the predicted values of the endogenous regressors $\widehat{x}_i$ and predicted inverse Mills ratio term $\widehat{\Lambda}_i(\gamma)$ by concentration. Conditional on γ, the estimation problem is linear in the slope parameters $\theta = (\beta_{x1}', \beta_{x2}', \kappa)'$, yielding conditional 2SLS or GMM estimator $\widehat{\theta}(\gamma) = (\widehat{\beta}_{x1}(\gamma)', \widehat{\beta}_{x2}(\gamma)', \widehat{\kappa}(\gamma))'$ by regressing y_i on $\widehat{\mathfrak{x}}_i(\gamma)$ and instruments $\widehat{\mathfrak{z}}_i(\gamma)$.[5] Define the CLS criterion

$$\begin{aligned} S_n(\gamma) &= S_n(\gamma, \widehat{\theta}(\gamma)) \\ &= \sum_{i=1}^{n} (y_i - \widehat{\beta}_{x1}(\gamma)' \widehat{g}_{xi} I(q_i \leq \gamma) - \widehat{\beta}_{x2}(\gamma)' \widehat{g}_{xi} I(q_i > \gamma) - \widehat{\kappa}(\gamma) \widehat{\Lambda}_i(\gamma))^2 \end{aligned} \quad \textbf{(3.12)}$$

Then, we can estimate γ by minimizing the CLS criterion

$$\widehat{\gamma} = \underset{\gamma}{\operatorname{argmin}}\ S_n(\gamma) \quad \textbf{(3.13)}$$

3.2. Slope Estimation

Once we obtain the threshold estimate $\widehat{\gamma}$, we proceed with estimation of the slope parameters θ by 2SLS or GMM. Denote $\widehat{\mathcal{X}}(\gamma)$ and $\widehat{\mathcal{Z}}(\gamma)$ the matrices of stacked vectors, $\widehat{\chi}_i(\gamma)$ and $\widehat{z}_i(\gamma)$, respectively. Let also Y be the stacked vector of y_i. By suppressing their dependence on $\widehat{\gamma}$, let $\widehat{\mathcal{X}} = \widehat{\mathcal{X}}(\widehat{\gamma})$ and $\widehat{\mathcal{Z}} = \widehat{\mathcal{Z}}(\widehat{\gamma})$ denote the matrices $\widehat{\mathcal{X}}(\gamma)$ and $\widehat{\mathcal{Z}}(\gamma)$ evaluated at $\widehat{\gamma}$. Then, the 2SLS estimator of $\theta = (\beta_{x1}', \beta_{x2}', \kappa)'$ is given by

$$\widehat{\theta}_{2SLS} = (\widehat{\mathcal{X}}'\widehat{\mathcal{Z}}(\widehat{\mathcal{Z}}'\widehat{\mathcal{Z}})^{-1}\widehat{\mathcal{Z}}'\widehat{\mathcal{X}})^{-1}\widehat{\mathcal{X}}'\widehat{\mathcal{Z}}(\widehat{\mathcal{Z}}'\widehat{\mathcal{Z}})^{-1}\widehat{\mathcal{Z}}'Y. \quad \textbf{(3.14)}$$

Using the 2SLS residual, $\widehat{e}^*_{i,2SLS} = y_i - \widehat{\chi}_i(\widehat{\gamma})'\widehat{\theta}_{2SLS}$, construct the weight matrix, $\widehat{\Sigma}^* = \sum_{i=1}^{n} \widehat{z}_i(\widehat{\gamma})\widehat{z}_i(\widehat{\gamma})'\widehat{e}^{*2}_{i,2SLS}$. Then we can also define the GMM estimator

$$\widehat{\theta}_{GMM} = (\widehat{\mathcal{X}}'\widehat{\mathcal{Z}}\widehat{\Sigma}^{*-1}\widehat{\mathcal{Z}}'\widehat{\mathcal{X}})^{-1}\widehat{\mathcal{X}}'\widehat{\mathcal{Z}}\widehat{\Sigma}^{*-1}\widehat{\mathcal{Z}}'Y. \quad \textbf{(3.15)}$$

with estimated covariance matrix, $\widehat{V}_{GMM} = (\widehat{\mathcal{X}}'\widehat{\mathcal{Z}}\widehat{\Sigma}^{*-1}\widehat{\mathcal{Z}}'\widehat{\mathcal{X}})^{-1}$.

While from a computational standpoint our estimation strategy is similar to the one employed by Caner and Hansen (2004), there is one key difference. The STR model includes different inverse Mills ratio terms in each regime. To put it differently, STR imposes the exclusion restrictions across regimes that require that only $\lambda_{1i}(\gamma)$ appears in Regime 1 and only $\lambda_{2i}(\gamma)$ in Regime 2. As a result, we cannot analyze the estimation problem using results obtained regime-by-regime. In particular, we cannot decompose the sum of squared errors into two separable regime-specific terms due to overlaps. To overcome this problem we next recast the STR model in equation (2.11) as a threshold regression subject to restrictions and exploit the relationship between constrained and unconstrained estimation problems. This allows us to decompose the sum of squared errors into two separable regime-specific terms and derive the asymptotic theory of the above estimators.

3.3. An Alternative Representation

Consider an auxiliary (unconstrained) STR model that generalizes Caner and Hansen (2004) by including both inverse Mills ratio terms in both regimes. Define $g_i(\gamma) = (g_{xi}', \lambda_{1i}(\gamma), \lambda_{2i}(\gamma))'$ and slope parameters $\beta = \left(\beta_1', \beta_2'\right)'$ with $\beta_1 = (\beta_{x1}', \kappa_{11}, \kappa_{12})'$, $\beta_2 = (\beta_{x2}', \kappa_{21}, \kappa_{22})'$. Then we can specify

$$y_i = \beta_1' g_i(\gamma) I(q_i \le \gamma) + \beta_2' g_i(\gamma) I(q_i > \gamma) + e_i, \quad \textbf{(3.16)}$$

where $e_i = (\beta_{x1}' v_{xi} - \kappa_{11}\lambda_{1i}(\gamma) - \kappa_{12}\lambda_{2i}(\gamma)) I(q_i \le \gamma) + (\beta_{x2}' v_{xi} - \kappa_{21}\lambda_{1i}(\gamma) - \kappa_{22}\lambda_{2i}(\gamma)) I(q_i > \gamma) + u_i$. The error e_i will play an important role because the asymptotic theory for the estimate of γ will behave as if $g_i(\gamma)$ were observable.

The STR model in equation (2.11) is equivalent to the (unconstrained) threshold regression in equation (3.16) subject to the constraints $\kappa_{12} = \kappa_{21} = 0$ and $\kappa_{11} = \kappa_{22} = \kappa$, which can be generally written as

$$R'\beta = \vartheta \tag{3.17}$$

where R is a $2(p+2) \times 3$ matrix of rank 3, and ϑ is a 3-dimensional vector of constants.

3.4. Minimum Distance Estimation

In this subsection we estimate the slope parameters β under the restriction in equation (3.17) using a minimum distance estimation method.

3.4.1. Unconstrained estimation. First, we consider the estimation of the unconstrained problem. The parameters of the unconstrained STR model in equation (3.16), β and γ, are estimated analogously to the constrained parameters in Section 3 using a three-step procedure. The first step is the same as in the case of the constrained problem, which yields consistent first stage estimates for Π_x and π_q. For any γ, we can then define $\widehat{x}_i(\gamma) = (x_i', \widehat{\lambda}_{1i}(\gamma), \widehat{\lambda}_{2i}(\gamma))'$ and $\widehat{z}_i(\gamma) = (z_i', \widehat{\lambda}_{1i}(\gamma), \widehat{\lambda}_{2i}(\gamma))'$. Let $\widehat{X}_1(\gamma)$, $\widehat{X}_2(\gamma)$, $\widehat{Z}_1(\gamma)$, and $\widehat{Z}_2(\gamma)$ denote the matrices of stacked vectors $\widehat{x}_i(\gamma)I(q_i \le \gamma)$, $\widehat{x}_i(\gamma)I(q_i > \gamma)$, $\widehat{z}_i(\gamma)I(q_i \le \gamma)$, and $\widehat{z}_i(\gamma)I(q_i > \gamma)$, respectively. Then, conditional on γ, we obtain the 2SLS or GMM estimator $\widehat{\beta}(\gamma) = (\widehat{\beta}_1(\gamma), \widehat{\beta}_2(\gamma))'$ by regressing Y on $\widehat{X}(\gamma) = (\widehat{X}_1'(\gamma), \widehat{X}_2'(\gamma))'$ and instruments $\widehat{Z}(\gamma) = (\widehat{Z}_1'(\gamma), \widehat{Z}_2'(\gamma))'$.

Second, for any γ, we define the (unconstrained) concentrated least squares criterion,

$$\begin{aligned} S_n^U(\gamma) &= S_n^U(\gamma, \widehat{\beta}(\gamma)) \\ &= \sum_{i=1}^{n} \left(y_i - \widehat{\beta}_1'\widehat{g}_i(\gamma)I(q_i \le \gamma) - \widehat{\beta}_2'\widehat{g}_i(\gamma)I(q_i > \gamma)\right)^2, \end{aligned} \tag{3.18}$$

where $\widehat{g}_i(\gamma) = (\widehat{g}_{xi}', \widehat{\lambda}_{1i}(\gamma), \widehat{\lambda}_{2i}(\gamma))'$. Then, the unconstrained estimator for γ is given by $\widetilde{\gamma} = \underset{\gamma}{\operatorname{argmin}}\, S_n^U(\gamma)$. Note that the criterion, $S_n(\gamma)$, in equation (3.12) is in fact the constrained sum of squared errors, $S_n(\gamma) = S_n^R(\gamma)$ so that $\widehat{\gamma} = \underset{\gamma}{\operatorname{argmin}}\, S_n^R(\gamma)$. The key difference between $S_n^R(\gamma)$ and $S_n^U(\gamma)$ is that the latter criterion can be decomposed into two separable regime-specific terms.

Third, we proceed with the estimation of the slope parameters β_1 and β_2 by splitting the sample into two sub-samples, based on $I(q_i \le \widehat{\gamma})$ and $I(q_i > \widehat{\gamma})$ using the constrained estimator $\widehat{\gamma}$.[6]

Let $\widehat{X}_1 = \widehat{X}_1(\widehat{\gamma})$, $\widehat{X}_2 = \widehat{X}_2(\widehat{\gamma})$, $\widehat{Z}_1 = \widehat{Z}_1(\widehat{\gamma})$, and $\widehat{Z}_2 = \widehat{Z}_2(\widehat{\gamma})$, then the unconstrained 2SLS estimators for the slope parameters β_1 and β_2 are given by

$$\widetilde{\beta}_{1,2SLS} = (\widehat{X}_1'\widehat{Z}_1(\widehat{Z}_1'\widehat{Z}_1)^{-1}\widehat{Z}_1'\widehat{X}_1)^{-1}\widehat{X}_1'\widehat{Z}_1(\widehat{Z}_1'\widehat{Z}_1)^{-1}\widehat{Z}_1'Y, \tag{3.19a}$$

$$\widetilde{\beta}_{2,2SLS} = (\widehat{X}_2'\widehat{Z}_2(\widehat{Z}_2'\widehat{Z}_2)^{-1}\widehat{Z}_2'\widehat{X}_2)^{-1}\widehat{X}_2'\widehat{Z}_2(\widehat{Z}_2'\widehat{Z}_2)^{-1}\widehat{Z}_2'Y, \tag{3.19b}$$

and the 2SLS residual is

$$\widetilde{e}_{i,2SLS} = y_i - \widehat{x}_i(\widehat{\gamma})'\widetilde{\beta}_{1,2SLS}I(q_i \leq \widehat{\gamma}) - \widehat{x}_i(\widehat{\gamma})'\widetilde{\beta}_{2,2SLS}I(q_i > \widehat{\gamma}). \tag{3.20}$$

To obtain the unconstrained GMM estimators define the matrices

$$\widetilde{\Sigma}_1 = \sum_{i=1}^{n} \widehat{z}_i(\widehat{\gamma})\widehat{z}_i(\widehat{\gamma})'\widetilde{e}_{i,2SLS}^2 I(q_i \leq \widehat{\gamma}) \tag{3.21a}$$

$$\widetilde{\Sigma}_2 = \sum_{i=1}^{n} \widehat{z}_i(\widehat{\gamma})\widehat{z}_i(\widehat{\gamma})'\widetilde{e}_{i,2SLS}^2 I(q_i > \widehat{\gamma}) \tag{3.21b}$$

The GMM estimators are then given by

$$\widetilde{\beta}_{1,GMM} = (\widehat{X}_1'\widehat{Z}_1\widetilde{\Sigma}_1^{-1}\widehat{Z}_1'\widehat{X}_1)^{-1}\widehat{X}_1'\widehat{Z}_1\widetilde{\Sigma}_1^{-1}\widehat{Z}_1'Y, \tag{3.22a}$$

$$\widetilde{\beta}_{2,GMM} = (\widehat{X}_2'\widehat{Z}_2\widetilde{\Sigma}_2^{-1}\widehat{Z}_2'\widehat{X}_2)^{-1}\widehat{X}_2'\widehat{Z}_2\widetilde{\Sigma}_2^{-1}\widehat{Z}_2'Y, \tag{3.22b}$$

with estimated covariances

$$\widetilde{V}_{1,GMM} = (\widehat{X}_1'\widehat{Z}_1\widetilde{\Sigma}_1^{-1}\widehat{Z}_1'\widehat{X}_1)^{-1} \tag{3.23a}$$

$$\widetilde{V}_{2,GMM} = (\widehat{X}_2'\widehat{Z}_2\widetilde{\Sigma}_2^{-1}\widehat{Z}_2'\widehat{X}_2)^{-1} \tag{3.23b}$$

3.4.2. Constrained estimation. We proceed to obtain the estimators of the constrained problem using a minimum distance estimation method.

Let $\widetilde{\beta}_{2SLS} = (\widetilde{\beta}_{1,2SLS}', \widetilde{\beta}_{2,2SLS}')'$ and $\widehat{W}_{2SLS} = diag((\widehat{Z}_1'\widehat{Z}_1)^{-1}, (\widehat{Z}_2'\widehat{Z}_2)^{-1})$. Define $J_n(\beta, \widehat{W}_{2SLS}) = n(\widetilde{\beta}_{2SLS} - \beta)'\widehat{W}_{2SLS}(\widetilde{\beta}_{2SLS} - \beta)$. The constrained 2SLS slope estimator of β is obtained by solving a minimum distance problem, which yields the constrained estimator, $\widehat{\beta}_{C2SLS} = \underset{R'\beta=\vartheta}{\operatorname{argmin}} J_n(\beta, \widehat{W}_{2SLS})$. This constrained estimator is related to the unconstrained estimator via

$$\widehat{\beta}_{C2SLS} = \widetilde{\beta}_{2SLS} - \widehat{W}_{2SLS}R(R'\widehat{W}_{2SLS}R)^{-1}(R'\widetilde{\beta}_{2SLS} - \vartheta). \tag{3.24}$$

Similarly, let $\widetilde{\beta}_{GMM} = \left(\widetilde{\beta}_{1,GMM}', \widetilde{\beta}_{2,GMM}'\right)'$, $\widetilde{V}_{GMM} = diag\left(\widetilde{V}_{1,GMM}, \widetilde{V}_{2,GMM}\right)$, and $\widetilde{W}_{GMM} = \widetilde{V}_{GMM}^{-1}$. Define $J_n\left(\beta, \widetilde{W}_{GMM}\right) = n\left(\widetilde{\beta}_{GMM} - \beta\right)'\widetilde{W}_{GMM}\left(\widetilde{\beta}_{GMM} - \beta\right)$. Then, we obtain the constrained GMM estimator by the minimum distance estimator $\widehat{\beta}_{CGMM} = \underset{R'\beta=\vartheta}{\operatorname{argmin}} J_n(\beta, \widetilde{W}_{GMM})$, which is related to the unconstrained estimator via

$$\widehat{\beta}_{CGMM} = \widetilde{\beta}_{GMM} - \widetilde{W}_{GMM}R\left(R'\widetilde{W}_{GMM}R\right)^{-1}\left(R'\widetilde{\beta}_{GMM} - \vartheta\right) \tag{3.25}$$

and estimated covariance

$$\widehat{V}_{CGMM} = \widetilde{V}_{GMM} - \widetilde{V}_{GMM}R\left(R'\widetilde{V}_{GMM}R\right)^{-1}R'\widetilde{V}_{GMM}. \tag{3.26}$$

Having derived the connection between the constrained and unconstrained problem, we proceed below with inference.[7]

https://doi.org/10.1017/S0266466615000067 Published online by Cambridge University Press

4. ASYMPTOTIC THEORY

4.1. Assumptions

Define $v_i = (v_{xi}', v_{qi})'$ and the following moment functionals

$$M(\gamma) = E(g_i(\gamma)g_i(\gamma)')$$
$$D = D(\gamma_0) = E(g_i(\gamma_0)g_i(\gamma_0)'|q_i = \gamma_0)$$
$$\Omega_1 = \lim_{\gamma \nearrow \gamma_0} \Omega(\gamma) = \lim_{\gamma \nearrow \gamma_0} E\left(g_i(\gamma)g_i(\gamma)'e_i^2|q_i = \gamma\right)$$
$$\Omega_2 = \lim_{\gamma \searrow \gamma_0} \Omega(\gamma) = \lim_{\gamma \searrow \gamma_0} E\left(g_i(\gamma)g_i(\gamma)'e_i^2|q_i = \gamma\right)$$

where $\lim_{\gamma \nearrow \gamma_0}$ and $\lim_{\gamma \searrow \gamma_0}$ denote the limits from below and above the threshold γ_0, respectively. Further, define $\bar{g}_i = \sup_{\gamma\in\Gamma}|g_i(\gamma)|$ and $\bar{M} = E(\bar{g}_i\bar{g}_i')$ and $f_q(q)$ be the density function of q_i and let γ_0 denote the true value of γ so that $f = f_q(\gamma_0)$.

Assumption 2.

2.1 $\{z_i, g_{xi}, u_i, v_i\}$ is strictly stationary and ergodic with ρ-mixing coefficients $\sum_{m=1}^{\infty} \rho_m^{1/2} < \infty$,

2.2 $E|\bar{g}_i|^4 < \infty$ and $E|\bar{g}_i e_i|^4 < \infty$,

2.3 for all $\gamma \in \Gamma$, $E\left(|\bar{g}_i|^4|q_i = \gamma\right) \le C$, $E\left(|\bar{g}_i|^4 e_i^4|q_i = \gamma\right) \le C$, a.s., for some $C < \infty$,

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

2.5 $D(\gamma)$ is continuous at $\gamma = \gamma_0$; $\Omega_1(\gamma)$, and $\Omega_2(\gamma)$ are semi-continuous at $\gamma = \gamma_0$.

2.6 $\delta_n = \left(\delta_{xn}', \delta_{\lambda_1 n}, \delta_{\lambda_2 n}\right)' = cn^{-\alpha} \to 0$, with $c \neq 0$ and $\alpha \in (0, 1/2)$, where $c = \left(c_\delta', c_{\kappa_1}, c_{\kappa_2}\right)'$, $\delta_{xn} = \beta_{x1} - \beta_{x2} = c_\delta n^{-\alpha}$, $\delta_{\lambda_1 n} = \kappa_{11} - \kappa_{21} = c_{\lambda_1} n^{-\alpha}$, and $\delta_{\lambda_2 n} = \kappa_{12} - \kappa_{22} = c_{\lambda_2} n^{-\alpha}$.

2.7 $f > 0$, $c'Dc > 0$, $c'\Omega_1 c > 0$, $c'\Omega_2 c > 0$.

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

This set of assumptions is similar to Hansen (2000) and Caner and Hansen (2004). Assumption 2.1 excludes time trends and integrated processes. This assumption is trivially satisfied for i.i.d. data. Assumptions 2.2 and 2.3 are unconditional and conditional moment bounds. Assumptions 2.4 and 2.5 require the threshold variable to have a continuous distribution and the conditional variance $E(e_i^2|q_i = \gamma)$ to be semi-continuous at γ_0. This assumption allows for regime-specific heteroskedasticity. Assumption 2.6 assumes that a "small threshold" asymptotic framework applies to the threshold effect of x_i, $\delta_{xn} \to 0$ as well as to the threshold effects of $\lambda_{1i}(\gamma)$ and $\lambda_{2i}(\gamma)$, $\delta_{\lambda_1 n} \to 0$ and $\delta_{\lambda_2 n} \to 0$, respectively.[8] Assumptions 2.7 and 2.8 are full rank conditions needed to have nondegenerate asymptotic distributions.[9]

Assumption 3. The constraint in equation (3.17) is valid.

Given Assumptions 1–3 we proceed to derive the consistency and asymptotic distribution of the threshold and slope parameters of equation (3.16) subject to the constraint in (3.17).

4.2. Threshold Estimate

4.2.1. Consistency.

PROPOSITION 1. *Consistency of* $\widehat{\gamma}$
Under Assumptions 1–3, the estimator $\widehat{\gamma}$ *of* γ, *obtained by minimizing the CLS criterion in equation (3.12) (or, equivalently,* $S_n^U(\gamma)$ *subject to the constraints in (3.17)) is consistent. That is,*

$$\widehat{\gamma} \xrightarrow{p} \gamma_0$$

COROLLARY 1. *The estimator* $\widetilde{\gamma}$ *of* γ *obtained by minimizing the unconstrained CLS criterion* $S_n^U(\gamma)$ *is also consistent for* γ_0.

This corollary suggests that when the constraints are valid, the estimated threshold parameter for both the constrained and unconstrained problem will converge to the same true value. Therefore, in large samples, splitting the sample into two subsamples using the indicators $I(q_i \leq \widetilde{\gamma})$ and $I(q_i > \widetilde{\gamma})$ is equivalent to using $I(q_i \leq \widehat{\gamma})$ and $I(q_i > \widehat{\gamma})$ assuming that the constraints are valid.

Next, we proceed with the derivation of the asymptotic distribution by first showing that the rate of convergence of the constrained estimator for the threshold parameter is not improved, which implies that the threshold estimate may not be sensitive to additional information given by the valid constraints. We then proceed to show that the asymptotic distribution for the unconstrained threshold estimator $\widetilde{\gamma}$ is the same as that for the constrained estimator $\widehat{\gamma}$.

4.2.2. Asymptotic Distribution. Define $a_n = n^{1-2\alpha}$ and let the constants $B > 0$ and $\bar{\upsilon} > 0$. Then, we have the following lemma.

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

Lemma 1 says that the effect of the restrictions on the threshold estimates becomes negligible asymptotically. Thus, the constrained minimization problem reduces to the unconstrained minimization problem and the limit distribution of the threshold estimators are the same in both cases. This allows us to focus on the distribution of the unconstrained problem. We note that Perron and Qu (2006) obtained a similar finding in the context of change-point models.

Define

$$\varphi = \frac{c'\Omega_2 c}{c'\Omega_1 c}, \qquad \omega = \frac{c'\Omega_1 c}{(c'Dc)^2 f}$$

https://doi.org/10.1017/S0266466615000067 Published online by Cambridge University Press

and $\sigma_e^2 = E(e_i^2)$. Let $\mathcal{W}_1(s)$ and $\mathcal{W}_2(s)$ be two independent standard Wiener processes defined on $[0, \infty)$ and let $\mathcal{T}(s)$ denote the asymmetric two-sided Brownian motion on the real line.[10]

$$\mathcal{T}(s) = \begin{cases} -\frac{1}{2}|s| + \mathcal{W}_1(-s), & \text{if } s \leq 0 \\ -\frac{1}{2}|s| + \sqrt{\varphi}\mathcal{W}_2(s), & \text{if } s > 0 \end{cases} \tag{4.27}$$

THEOREM 1. *Asymptotic Distribution of* $\widehat{\gamma}$ *Under Assumptions 1–3*

$$n^{1-2\alpha}(\widehat{\gamma} - \gamma_0) \xrightarrow{d} \omega \mathcal{T} \tag{4.28}$$

where $\mathcal{T} = \underset{-\infty < s < \infty}{\arg\max}\ \mathcal{T}(s)$.

For $x < 0$, the cdf of $\mathcal{T}$ is given by

$$P(\mathcal{T} \leq x) = -\sqrt{\frac{|x|}{2\pi}} \exp\left(-\frac{|x|}{8}\right) - c\exp(a|x|)\Phi\left(-b\sqrt{|x|}\right) + \left(d - 2 + \frac{|x|}{2}\right)\Phi\left(-\frac{\sqrt{|x|}}{2}\right), \tag{4.29}$$

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

For $x > 0$,

$$P(\mathcal{T} \leq x) = 1 + \sqrt{\frac{x}{2\pi\varphi}} \exp\left(-\frac{x}{8\varphi}\right) - c\exp(ax)\Phi(-b\sqrt{x}) + (-d + 2 - \frac{x}{2\varphi})\Phi\left(-\frac{1}{2}\sqrt{\frac{x}{\varphi}}\right), \tag{4.30}$$

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

Theorem 1 shows that the asymptotic distribution of the threshold estimate, under the assumption of the diminishing threshold effect, features unequal scales for each regime and takes a similar form to the one found in Bai (1997) in the context of change-point models that assume stationarity within each regime and not for the whole sample.[11] While the asymptotic distribution is generally asymmetric, it becomes symmetric in the special case that excludes regime-specific heteroskedasticity. To see this note that when $\Omega_1 = \Omega_2 = \Omega$, then $\varphi = 1$ and scaling ratio $\omega = \frac{c'\Omega c}{(c'Dc)^2 f}$. In this case defining $\mathcal{W}(s) = \mathcal{W}_1(s) = \mathcal{W}_2(s)$ in equation (4.27), we get the two sided Wiener distribution scaled by ω derived in Hansen (2000). Moreover, under conditional homoskedasticity, $\sigma_e^2 = E\left(e_i^2 | q_i = \gamma_0\right)$, we get that $\Omega = \sigma_e^2 D$, and the scaling ratio simplifies to $\omega = \frac{\sigma_e^2}{(c'Dc)^2 f}$.

4.2.3. Likelihood Ratio Test. Consider the likelihood ratio statistic under the auxiliary assumption that e_i is *i.i.d.* $N(0, \sigma_e^2)$ for the hypothesis $H_0 : \gamma = \gamma_0$. Let

$$LR_n(\gamma) = n \frac{S_n(\gamma) - S_n(\widehat{\gamma})}{S_n(\widehat{\gamma})}. \tag{4.31}$$

Define

$$\eta^2 = \frac{c'\Omega_1 c}{(c'Dc)\sigma_e^2} \tag{4.32}$$

and

$$\psi = \sup_{-\infty < s < \infty} \left(\left(-\frac{1}{2}|s| + \mathcal{W}_1(-s) \right) I(s < 0) + \left(-\frac{1}{2}|s| + \sqrt{\varphi}\mathcal{W}_2(s) \right) I(s > 0) \right) \tag{4.33}$$

Then we have the following theorem.

THEOREM 2. *Asymptotic Distribution of* $LR(\gamma_0)$

Under Assumptions 1–3, the asymptotic distribution of the likelihood ratio test under H_0 *is given by*

$$LR_n(\gamma_0) \xrightarrow{d} \eta^2 \psi \tag{4.34}$$

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

Theorem 2 says that the asymptotic distribution of $LR_n(\gamma_0)$ is nonstandard and depends on two nuisance parameters, η^2 and φ. Note that the distribution does not have a closed form solution but we can compute the critical value $c_\psi(1-\alpha, \varphi)$ by numerically solving the equation $(1 - e^{-x/2})(1 - e^{-\sqrt{\varphi}x/2}) = 1 - \alpha$ for known values of φ. Hence, we reject the hypothesis $H_0 : \gamma = \gamma_0$ with asymptotic size of the test, α, when $LR_n(\gamma_0) > \eta^2 c_\psi(1-\alpha, \varphi)$. Under the special case that excludes regime-specific heteroskedasticity we obtain $\varphi = 1$ and the distribution is identical to the distribution of Hansen (2000). Moreover, under homoskedasticity, the $LR_n(\gamma_0)$ statistic is free of nuisance parameters and simplifies further to $LR_n(\gamma_0) = \psi$ since $\eta^2 = 1$.

4.2.4. Nuisance Parameters. The nuisance parameters, η^2 and φ, can be estimated by adapting the estimation method proposed by Hansen (2000). Let us first define the following random variables $r_{1i}^L = (\beta_1' \widehat{x}_i(\gamma) I(q_i \leq \gamma))^2 e_i^2/\sigma_e^2$, $r_{1i}^U = (\beta_2' \widehat{x}_i(\gamma) I(q_i > \gamma))^2 e_i^2/\sigma_e^2$, and $r_{2i} = (\delta_n' \widehat{x}_i(\gamma))^2$ as well as their sample analogues using the constrained 2SLS or GMM estimators $\widehat{\beta}_1$, $\widehat{\beta}_2$, and $\widehat{\delta}_n = \widehat{\beta}_1 - \widehat{\beta}_2$ defined in Section 3.4.2, $\widehat{r}_{1i}^L = (\widehat{\beta}_1' \widehat{x}_i(\widehat{\gamma}) I(q_i \leq \widehat{\gamma}))^2 \widehat{e_i}^2/\widehat{\sigma}_e^2$, $\widehat{r}_{1i}^U = (\widehat{\beta}_2' \widehat{x}_i(\widehat{\gamma}) I(q_i > \widehat{\gamma}))^2 \widehat{e_i}^2/\widehat{\sigma}_e^2$, and $\widehat{r}_{2i} = (\widehat{\delta}_n' \widehat{x}_i(\widehat{\gamma}))^2$, with

https://doi.org/10.1017/S0266466615000067 Published online by Cambridge University Press

$\widehat{e}_i = y_i - \widehat{\beta}_1' \widehat{x}_i(\widehat{\gamma}) I(q_i \leq \widehat{\gamma}) - \widehat{\beta}_2' \widehat{x}_i(\widehat{\gamma}) I(q_i > \widehat{\gamma}))$ and $\widehat{\sigma}_e^2 = \widehat{e}_i{}' \widehat{e}_i / n$. Further define the following ratios of conditional expectations

$$\eta^2 = \frac{\lim_{\gamma \nearrow \gamma_0} E(r_{1i}^L | q_i = \gamma)}{E(r_{2i} | q_i = \gamma_0)} \tag{4.35}$$

$$\varphi = \frac{\lim_{\gamma \searrow \gamma_0} E(r_{1i}^U | q_i = \gamma)}{\lim_{\gamma \nearrow \gamma_0} E(r_{1i}^L | q_i = \gamma)} \tag{4.36}$$

The estimation of these ratios of conditional expectations can be based on a quadratic polynomial in q_i regression or kernel regression as in Hansen (2000). For brevity we only present the former method. For $j = 1, 2$, consider the estimated LS regressions

$$\widehat{r}_{1i}^L = \widehat{\mu}_{10}^L + \widehat{\mu}_{11}^L q_i + \widehat{\mu}_{12}^L q_i^2 + \widehat{\epsilon}_{1i}^L$$
$$\widehat{r}_{1i}^U = \widehat{\mu}_{10}^U + \widehat{\mu}_{11}^U q_i + \widehat{\mu}_{12}^U q_i^2 + \widehat{\epsilon}_{1i}^U$$
$$\widehat{r}_{2i} = \widehat{\mu}_{20} + \widehat{\mu}_{21} q_i + \widehat{\mu}_{22} q_i^2 + \widehat{\epsilon}_{2i}$$

and then set

$$\widehat{\eta}^2 = \frac{\widehat{\mu}_{10}^L + \widehat{\mu}_{11}^L \widehat{\gamma} + \widehat{\mu}_{12}^L \widehat{\gamma}^2}{\widehat{\mu}_{20} + \widehat{\mu}_{21} \widehat{\gamma} + \widehat{\mu}_{22} \widehat{\gamma}^2}$$
$$\widehat{\varphi} = \frac{\widehat{\mu}_{10}^U + \widehat{\mu}_{11}^U \widehat{\gamma} + \widehat{\mu}_{12}^U \widehat{\gamma}^2}{\widehat{\mu}_{10}^L + \widehat{\mu}_{11}^L \widehat{\gamma} + \widehat{\mu}_{12}^L \widehat{\gamma}^2}$$

4.2.5. Confidence Intervals. We construct confidence intervals for the threshold parameter by inverting the likelihood ratio test statistic, LR_n. This approach follows Hansen (2000) who argues that under certain conditions this approach yields an asymptotically valid confidence region. In particular, assuming a constant threshold effect, conditional homoskedasticity, and Gaussian errors, Hansen (2000, Thm. 3) shows that inferences based on the inversion of the likelihood ratio test are asymptotically conservative. Let $(1-\alpha)100\%$ denote the desired asymptotic confidence level and let $c_\alpha = c_\psi(1-\alpha, \widehat{\varphi})$ denote the $(1-\alpha)100^{th}$ percentile of the distribution ψ using the plug-in estimator $\widehat{\varphi}$. Define the confidence region $\widehat{\gamma} = \{\gamma : LR_n(\gamma) \leq \widehat{\eta}^2 c_\alpha)\}$. Given that $\widehat{\eta}^2$ and $\widehat{\varphi}$ are consistent estimates of the nuisance parameters η^2 and φ, Theorem 2 shows that $P(\gamma_0 \in \widehat{\gamma}) \to 1-\alpha$ and hence, $\widehat{\gamma}$ is a regime-specific heteroskedasticity-robust asymptotic $1-\alpha$ confidence region for γ.

Nevertheless, there are a few caveats. First, it is important to emphasize that the confidence intervals are asymptotically valid under the assumption of the shrinking threshold effect. This suggests that the actual coverage may differ from the desired level for large values of the threshold effect and large degrees of endogeneity of the threshold variable. Second, as argued in Caner and Hansen (2004)

https://doi.org/10.1017/S0266466615000067 Published online by Cambridge University Press

the inference on γ critically relies on the local information around the threshold point.[12]

4.3. Slope Parameters

In this section, we investigate the asymptotic distribution of the 2SLS and GMM estimators of the slope parameters in the STR model in (3.16) subject to the constraints in (3.17).

Let $x_i(\gamma_0) = (x_i', \lambda_{1i}(\gamma_0), \lambda_{2i}(\gamma_0))'$ and $z_i(\gamma_0) = (z_i', \lambda_{1i}(\gamma_0), \lambda_{2i}(\gamma_0))'$. Let us define the following matrices

$$
\begin{aligned}
Q_1 &= E(z_i(\gamma_0)z_i(\gamma_0)'I(q_i \le \gamma_0),\\
Q_2 &= E(z_i(\gamma_0)z_i(\gamma_0)'I(q_i > \gamma_0)\\
S_1 &= E(z_i(\gamma_0)x_i(\gamma_0)'I(q_i \le \gamma_0))\\
S_2 &= E(z_i(\gamma_0)x_i(\gamma_0)'I(q_i > \gamma_0))\\
\Sigma_1 &= E(z_i(\gamma_0)z_i(\gamma_0)'e_i^2 I(q_i \le \gamma_0)\\
\Sigma_2 &= E(z_i(\gamma_0)z_i(\gamma_0)'e_i^2 I(q_i > \gamma_0)\\
V_{1,2SLS} &= \left(S_1'Q_1^{-1}S_1\right)^{-1}S_1'Q_1^{-1}\Sigma_1 Q_1^{-1}S_1\left(S_1'Q_1^{-1}S_1\right)^{-1}\\
V_{2,2SLS} &= \left(S_2'Q_2^{-1}S_2\right)^{-1}S_2'Q_2^{-1}\Sigma_2 Q_2^{-1}S_2\left(S_2'Q_2^{-1}S_2\right)^{-1}\\
V_{2SLS} &= diag(V_{1,2SLS}, V_{2,2SLS})\\
Q &= diag(Q_1, Q_2)\\
V_{1,GMM} &= \left(S_1'\Sigma_1^{-1}S_1\right)^{-1}\\
V_{2,GMM} &= \left(S_2'\Sigma_2^{-1}S_2\right)^{-1}\\
V_{GMM} &= diag(V_{1,GMM}, V_{2,GMM})
\end{aligned}
$$

THEOREM 3. *Under Assumptions 1–3*

$$\sqrt{n}\left(\widehat{\beta}_{C2SLS} - \beta\right) \xrightarrow{d} N(0, V_{C2SLS}) \tag{4.37}$$

where

$$
\begin{aligned}
V_{C2SLS} = {} & V_{2SLS} - Q^{-1}R\left(R'Q^{-1}R\right)^{-1}R'V_{2SLS} - V_{2SLS}R\left(R'Q^{-1}R\right)^{-1}R'Q^{-1}\\
& + Q^{-1}R\left(R'Q^{-1}R\right)^{-1}R'V_{2SLS}R\left(R'Q^{-1}R\right)^{-1}R'Q^{-1}.
\end{aligned} \tag{4.38}
$$

THEOREM 4. *Under Assumptions 1–3*

(a)

$$\sqrt{n}\left(\widehat{\beta}_{CGMM} - \beta\right) \xrightarrow{d} N(0, V_{CGMM}) \tag{4.39}$$

where

$$V_{CGMM} = V_{GMM} - V_{GMM}R\left(R'V_{GMM}R\right)^{-1}R'V_{GMM} \tag{4.40}$$

https://doi.org/10.1017/S0266466615000067 Published online by Cambridge University Press

(b)

$$n\widehat{V}_{CGMM} \xrightarrow{p} V_{CGMM} \quad \textbf{(4.41)}$$

5. MONTE CARLO

We proceed below with a simulation that investigates the finite sample performance of our estimators. The data generating mechanism is given by

$$y_i = \beta_1 + \beta_2 x_{1i} + \beta_3 x_{2i} + \left(\delta_1 + \delta_2 x_{1i} + \delta_3 x_{2i}\right) I\{q_i \le \gamma\} + u_i, \quad \textbf{(5.42)}$$

where the threshold variable q_i is given by

$$q_i = 2 + z_{qi} + v_{qi}. \quad \textbf{(5.43)}$$

The threshold parameter is set at the center of the distribution of q_i, hence $\gamma = 2$. The regressor x_{1i} is also endogenous and given by

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

where

$$z_{xi} = (w x_{2i} + (1-w)\varsigma_{zi}) / \sqrt{w^2 + (1-w)^2}, \quad \textbf{(5.44)}$$

and

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

where x_{2i}, ς_{zi}, and ς_{ui} are independent $i.i.d.$ $N(0,1)$ random variables. The degree of endogeneity of the threshold variable is controlled by the correlation coefficient between u_i and v_{qi} given by $c_{qu}/\sqrt{c_{xu}^2 + c_{qu}^2 + (1 - c_{xu} - c_{qu})^2}$. Similarly, the degree of endogeneity of x_{1i} is determined by the correlation between u_i and v_{xi} given by $c_{xu}/\sqrt{c_{xu}^2 + c_{qu}^2 + (1 - c_{xu} - c_{qu})^2}$. We vary δ_3 and fix c_{xu}, $w = 0.5$, $\beta_1 = \beta_2 = 1$, and $\delta_1 = \delta_2 = 0$. c_{qu} varies over the values of 0.05, 0.25, 0.45 that correspond to correlations between q_i and u_i of about 0.07, 0.4, 0.7, respectively. We consider sample sizes of 100, 250, 500, and 1,000 using 1,000 Monte Carlo replications simulations. In unreported exercises we also investigated alternative values of w and c_{xu} and found qualitatively similar results.[13]

We begin by assessing the performance of the STR threshold estimator $\widehat{\gamma}$ and the performance of our proposed confidence interval $\widehat{\gamma}$. Table 1 presents the quantiles of the distribution of the STR estimator for the threshold parameter by varying the threshold effect δ_3 over the values 1, 2, and 3. We see that the performance of the STR estimator for the threshold parameter γ improves as the threshold effect, δ_3, and/or the sample size, n increases. Specifically, the 50th quantile approaches the true threshold parameter, $\gamma = 2$, as the sample size increases and the

https://doi.org/10.1017/S0266466615000067 Published online by Cambridge University Press

TABLE 1. Quantiles of the distribution of the STR threshold estimator $\widehat{\gamma}$

	$\delta_2 = 1$			$\delta_2 = 2$			$\delta_2 = 3$		
Quantile	5th	50th	95th	5th	50th	95th	5th	50th	95th
Sample size	*Low degree of endogeneity*								
100	1.097	1.964	2.842	1.516	1.971	2.483	1.744	1.976	2.203
250	1.352	1.988	2.608	1.824	1.992	2.186	1.900	1.991	2.088
500	1.635	1.997	2.324	1.898	1.996	2.063	1.948	1.996	2.036
1,000	1.819	1.997	2.136	1.958	1.998	2.031	1.977	1.998	2.021
	Medium degree of endogeneity								
100	1.079	1.937	2.856	1.392	1.964	2.485	1.709	1.975	2.223
250	1.223	1.968	2.601	1.776	1.989	2.186	1.894	1.991	2.094
500	1.361	1.988	2.436	1.874	1.995	2.067	1.940	1.995	2.036
1,000	1.640	1.991	2.211	1.942	1.997	2.035	1.973	1.998	2.021
	High degree of endogeneity								
100	1.051	1.924	2.872	1.333	1.954	2.470	1.714	1.973	2.198
250	1.200	1.955	2.552	1.704	1.986	2.183	1.888	1.989	2.096
500	1.332	1.976	2.455	1.855	1.993	2.072	1.939	1.995	2.034
1,000	1.549	1.977	2.235	1.926	1.997	2.037	1.974	1.998	2.022

width of the distribution becomes smaller as δ increases. These results hold for all three degrees of endogeneity.

Table 2 provides the finite sample coverage of the nominal 90% confidence interval for the threshold parameter γ by varying the threshold effect δ_3 over the values $1, 2, 3, 4$, and 5. We find that the coverage probability increases with either the size of the threshold effect or the sample size and becomes conservative for larger values. In particular, while for a small threshold effect the asymptotic coverage is lower than the nominal coverage, as expected for a larger threshold effect the coverage becomes conservative for all three degrees of endogeneity even for a small sample size.

Next, we proceed to assess the performance of the GMM slope estimators $\widehat{\delta}_3$ and $\widehat{\beta}_3$ as well as the performance of their confidence intervals. Theorems 3 and 4 show that we can approximate the distribution of these slope estimators by the conventional normal approximation, which implies that we can construct conventional asymptotic confidence intervals based on the normal approximation as if γ were known with certainty. Consistent with theory, Tables 3 and 4 show that the slope coefficient of the upper regime, $\widehat{\beta}_3$, and the threshold effect, $\widehat{\delta}$, respectively, are centered on the corresponding true values as the sample size increases.

TABLE 2. Nominal 90% confidence interval coverage for γ

δ_3	1	2	3	4	5
Sample size			*Low degree of endogeneity*		
50	0.81	0.82	0.83	0.85	0.85
100	0.91	0.92	0.93	0.94	0.94
250	0.97	0.97	0.97	0.98	0.98
500	1.00	1.00	0.99	0.99	1.00
1,000	1.00	1.00	1.00	1.00	1.00
			Medium degree of endogeneity		
50	0.73	0.78	0.82	0.84	0.84
100	0.81	0.89	0.92	0.92	0.93
250	0.92	0.95	0.97	0.98	0.98
500	0.98	0.99	0.99	0.99	0.99
1,000	0.99	1.00	1.00	1.00	1.00
			High degree of endogeneity		
50	0.67	0.75	0.81	0.82	0.84
100	0.76	0.84	0.89	0.93	0.95
250	0.85	0.95	0.97	0.99	0.99
500	0.91	0.98	0.99	1.00	1.00
1,000	0.94	1.00	1.00	1.00	1.00

Finally, Table 5 presents the finite sample coverage of the nominal 95% confidence intervals for the slope coefficients β_3 and δ_3. The coverage for δ_3 improves for larger values of the threshold effect and sample size and becomes close to the nominal coverage. Interestingly, the coverage of δ_3 is not affected by the degree of endogeneity and is better than the coverage of β_3. In contrast, while the coverage of δ_3 also improves with either the size of the threshold effect or the sample size, it remains below the nominal coverage even for large sample sizes for higher degrees of endogeneity.[14]

6. CONCLUSION

The main contribution of this paper is to propose a threshold regression model that allows for the endogeneity of the threshold variable as well as the slope regressors and develop a limiting distribution theory for cross-section or time series observations. Our approach utilizes regime-specific inverse-Mills ratio terms as the control functions for the conditional expectations and estimates the threshold parameter using a two-step concentrated least squares method and the slope parameters using a 2SLS or a GMM method. Using an asymptotic framework that relies on the assumption of the asymptotically diminishing

https://doi.org/10.1017/S0266466615000067 Published online by Cambridge University Press

TABLE 3. Quantiles of the distribution of the GMM estimator for the slope coefficient $\widehat{\beta}_3$

	$\delta_3 = 1$			$\delta_3 = 2$			$\delta_3 = 3$		
Quantile	5th	50th	95th	5th	50th	95th	5th	50th	95th
Sample size	*Low degree of endogeneity*								
100	0.636	1.020	1.432	0.678	1.022	1.374	0.693	1.014	1.340
250	0.792	1.000	1.249	0.805	0.996	1.213	0.808	1.000	1.211
500	0.869	1.003	1.171	0.876	1.002	1.141	0.876	1.001	1.138
1,000	0.903	1.004	1.104	0.906	1.002	1.097	0.909	1.002	1.095
	Medium degree of endogeneity								
100	0.676	1.052	1.468	0.685	1.042	1.434	0.697	1.019	1.380
250	0.794	1.020	1.279	0.802	1.000	1.221	0.816	0.999	1.208
500	0.875	1.015	1.225	0.880	1.004	1.155	0.881	1.003	1.143
1,000	0.911	1.015	1.158	0.909	1.003	1.102	0.911	1.002	1.094
	High degree of endogeneity								
100	0.680	1.076	1.483	0.703	1.048	1.491	0.708	1.024	1.403
250	0.813	1.045	1.308	0.814	1.017	1.238	0.818	1.002	1.209
500	0.882	1.032	1.250	0.880	1.011	1.169	0.877	1.005	1.143
1,000	0.919	1.025	1.186	0.911	1.008	1.109	0.907	1.003	1.097

threshold effect, we obtain a useful asymptotic distribution of the threshold parameter. One implication of using regime-specific inverse-Mills ratio terms is that the errors are regime-specific heteroskedastic and hence, the distribution of the threshold estimator involves two independent Brownian motions with two different scales. We show that these scale parameters are estimable and by numerically inverting the likelihood ratio we obtain confidence intervals, which are asymptotically conservative. Another implication of our approach is that the estimates cannot be analyzed using results obtained regime-by-regime because it involves restrictions across the two regimes. To overcome this problem, we recast the structural threshold regression as an unconstrained threshold regression subject to restrictions and exploit the relationship between constrained and unconstrained estimation problems. This allows us to decompose the sum of squared errors into two separable regime-specific terms and obtain the slope estimators using a minimum distance estimation method. We show that when the constraints are valid, the rate of convergence of the threshold estimator is not improved relative to the unconstrained problem and as such the asymptotic distribution of the threshold estimator in the unconstrained optimization problem is equivalent to the distribution of the threshold estimator in the constrained problem.

TABLE 4. Quantiles of the distribution of the GMM estimator for the threshold effect $\widehat{\delta_3}$

	$\delta_3 = 1$			$\delta_3 = 2$			$\delta_3 = 3$		
Quantile	5th	50th	95th	5th	50th	95th	5th	50th	95th
Sample size	*Low degree of endogeneity*								
100	0.382	0.980	1.411	1.487	1.971	2.385	2.565	2.981	3.385
250	0.687	0.996	1.247	1.753	1.993	2.244	2.761	2.998	3.246
500	0.794	0.998	1.166	1.825	1.996	2.163	2.829	3.000	3.164
1,000	0.876	0.995	1.117	1.881	1.997	2.115	2.881	3.000	3.111
	Medium degree of endogeneity								
100	0.338	0.930	1.372	1.439	1.956	2.370	2.558	2.966	3.365
250	0.621	0.972	1.225	1.735	1.986	2.228	2.759	2.991	3.226
500	0.725	0.979	1.155	1.822	1.992	2.153	2.833	2.997	3.153
1,000	0.823	0.979	1.108	1.881	1.991	2.112	2.886	2.994	3.116
	High degree of endogeneity								
100	0.396	0.898	1.309	1.423	1.930	2.329	2.572	2.973	3.341
250	0.619	0.938	1.181	1.719	1.970	2.202	2.769	2.985	3.205
500	0.707	0.952	1.123	1.819	1.988	2.140	2.852	2.996	3.138
1,000	0.788	0.960	1.096	1.883	1.988	2.098	2.895	2.994	3.102

We are also hopeful that the methods in this paper will find immediate application to questions with broad policy significance and highlight the importance of allowing for the endogeneity of the threshold variable in practice. For example, Kourtellos, Stengos, and Tan (2013) revisit an important and timely question popularized by Reinhart and Rogoff (2010) over whether there exists a threshold level in the public debt-to-GDP ratio over which the level of public debt has negative effects on long-run growth. Using a large set of alternative theories for possible heterogeneity in the debt-growth relationship, Kourtellos, Stengos, and Tan found strong evidence for threshold effects based on democracy, as a proxy for institutional quality, in the effect of debt on growth.

In terms of future work, a useful extension to our approach is to consider a nonparametric approach using the integrated difference kernel estimator along the lines of Yu and Phillips (2014). A challenging aspect of this problem is to relax the *i.i.d.* assumption and allow for stationary and ergodic data. A further extension with practical importance is to consider the issue of modeling the uncertainty that arises in choosing the true threshold variable from a set of significant threshold variables. This situation often arises in empirical work when different theories imply different threshold variables as a source of heterogeneity. One possible way to deal with this problem is to generalize existing model averaging methods that

TABLE 5. Nominal 95% confidence interval coverage for the slope coefficients

	Coverage for β_3					Coverage for δ_3				
δ_3	1	2	3	4	5	1	2	3	4	5
Sample size	*Low degree of endogeneity*									
50	0.80	0.84	0.87	0.88	0.89	0.80	0.84	0.87	0.88	0.89
100	0.83	0.88	0.91	0.92	0.92	0.83	0.88	0.91	0.92	0.92
250	0.91	0.93	0.93	0.94	0.94	0.91	0.93	0.93	0.94	0.94
500	0.91	0.93	0.94	0.93	0.93	0.91	0.93	0.94	0.93	0.93
1,000	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94
	Medium degree of endogeneity									
50	0.78	0.80	0.83	0.85	0.86	0.79	0.83	0.86	0.88	0.89
100	0.81	0.84	0.89	0.89	0.90	0.81	0.86	0.90	0.91	0.92
250	0.82	0.89	0.91	0.91	0.91	0.86	0.92	0.94	0.94	0.94
500	0.83	0.92	0.93	0.93	0.93	0.85	0.92	0.93	0.93	0.93
1,000	0.83	0.92	0.93	0.93	0.93	0.86	0.93	0.94	0.94	0.94
	High degree of endogeneity									
50	0.74	0.75	0.80	0.82	0.83	0.79	0.81	0.84	0.87	0.87
100	0.76	0.80	0.84	0.86	0.86	0.78	0.83	0.89	0.90	0.90
250	0.77	0.86	0.88	0.89	0.89	0.83	0.90	0.93	0.93	0.94
500	0.78	0.89	0.91	0.92	0.91	0.82	0.92	0.94	0.94	0.94
1,000	0.76	0.89	0.90	0.91	0.90	0.79	0.93	0.94	0.94	0.94

apply to linear models (e.g., Brock and Durlauf, 2001; Hansen, 2007) to threshold regression. Finally, it would also be potentially useful to link STR with the treatment effects literature; see, for example, Yu (2014b) who makes the connection between regression discontinuity design and threshold regression.

NOTES

1. This finding is also related to Gonzalo and Pitarakis (2002) who establish that the single threshold parameter estimator obtained from a misspecified two regime model is consistent when they ignore additional thresholds.

2. More recently, in the context of the multiple-regime threshold autoregressive model, Li and Ling (2012) revisit the theory of Chan and propose a numerical approach to simulate the limiting distribution of the estimated threshold based on a simulation of a related compound Poisson process.

3. Note that our analysis excludes the special cases of (i) the continuous threshold model (see Chan and Tsay, 1998; Hansen, 2000) and (ii) the threshold model where q_i is an element of x_i. However, the general framework of this paper is expected to carry over to these cases.

4. We make Assumption 1.7 to simplify the exposition of the problem. One could allow dependence between v_{xi} and v_{qi} by assuming that $E(v_{xi}|\mathcal{F}_{i-1}, v_{qi})$ is a linear function of v_{qi}, which implies the need for an additional inverse Mills ratio term in each regime.

https://doi.org/10.1017/S0266466615000067 Published online by Cambridge University Press

5. Conditional on γ, estimation in each regime mirrors the Heckman (1979) sample selection bias correction model, the Heckit model.

6. Given that the problem of interest is the constrained estimation, we evaluate the slope parameters on $\widehat{\gamma}$ rather than $\widetilde{\gamma}$. As we will show in Lemma 1, in Section 4, the limit distribution of the unconstrained and constrained threshold estimators is the same and hence the effect of the restriction is asymptotically negligible.

7. Note that although it is not immediately obvious, the constrained estimators $\widehat{\beta}_{C2SLS}$ in (3.24) and $\widehat{\beta}_{CGMM}$ in (3.25) are algebraically equivalent to $\widehat{\theta}_{2SLS}$ in (3.14) and $\widehat{\theta}_{GMM}$ in (3.15), respectively.

8. Note that if we further impose the constraints (3.17) then $\delta_n = (\delta'_{xn}, \kappa_n, -\kappa_n)' = cn^{-\alpha} \to 0$, where $c = (c'_\delta, c_\kappa, -c_\kappa)'$.

9. It is important to emphasize that our theory requires that the reduced form predicted values $\widehat{g}_{xi}$, $\widehat{\lambda}_{1i}(\gamma)$, and $\widehat{\lambda}_{2i}(\gamma)$ are consistent for the true reduced form conditional means $g_{xi}(\gamma)$, $\lambda_{1i}(\gamma)$, and $\lambda_{2i}(\gamma)$, respectively. Assumptions 1 and 2 are sufficient for our theory.

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

11. One difference between the two distributions is that in Bai (1997) the distribution features discontinuity in both D and Ω, which results in two different shifts and scales.

12. Yu and Zhao (2013) study the asymptotic behavior of the LS estimate of the threshold parameter when the density of the threshold variable is neither continuous nor bounded from zero and infinity.

13. In the Internet Appendix we provide complete simulation results including an experiment that assumes a threshold regression model that allows for an endogenous threshold variable but retains the assumption of an exogenous slope variable. The results are similar.

14. To improve the coverage rates, one can use a Bonferroni-type approach along the lines of Hansen (2000) and Caner and Hansen (2004).

REFERENCES

Acemoglu, A., S. Johnson, & J. Robinson (2001) The colonial origins of comparative development: An empirical investigation. *American Economic Review* 91, 1369–1401.

Bai, M. (1997) Estimation of a change point in multiple regression models. *The Review of Economics and Statistics* 79, 551–563.

Boldea, O., A.R. Hall, & S. Han (2012) Asymptotic distribution theory for break point estimators in models estimated via 2SLS. *Econometric Reviews* 31, 1–33.

Brock, W. & S.D. Durlauf (2001) Growth empirics and reality. *World Bank Economic Review* 15, 229–272.

Caner, M. & B. Hansen (2004) Instrumental variable estimation of a threshold model. *Econometric Theory* 20, 813–843.

Chan, K.S. (1993) Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Annals of Statistics* 21, 520–533.

Chan, K.S. & R.S. Tsay (1998) Limiting properties of the least squares estimator of a continuous threshold autoregressive model. *Biometrika* 85, 413–426.

Frankel, J. & D. Romer (1999) Does trade cause growth? *American Economic Review* 89, 379–399.

Gonzalo, J. & J.-Y. Pitarakis (2002) Estimation and model selection based inference in single and multiple threshold models. *Journal of Econometrics* 110, 319–352.

Gonzalo, J. & M. Wolf (2005) Subsampling inference in threshold autoregressive models. *Journal of Econometrics* 127, 201–224.

Hall, A.R., S. Han, & O. Boldea (2012) Inference regarding multiple structural changes in linear models with endogenous regressors. *Journal of Econometrics* 170, 281–302.

Hansen, B.E. (1996) Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64, 413–430.

Hansen, B.E. (2000) Sample splitting and threshold estimation. *Econometrica* 68, 575–603.

Hansen, B. (2007) Least squares model averaging. *Econometrica* 75, 1175–1189.
Heckman, J. (1979) Sample selection bias as a specification error. *Econometrica* 47, 153–161.
Kim, J. & J. Pollard (1990) Cube root asymptotics. *Annals of Statistics* 18, 191–219.
Kourtellos, A., T. Stengos, & C.M. Tan (2013) The effect of public debt on growth in multiple regimes. *Journal of Macroeconomics* 38, 35–43.
Kourtellos, A., T. Stengos, & C.M. Tan (2014) Supplementary Internet Appendix for "Structural Threshold Regression". Mimeo, The University of Cyprus.
Li, D. & S. Ling (2012) On the least squares estimation of multiple regime threshold autoregressive models. *Journal of Econometrics* 167, 240–253.
Li, Q. & J. Wooldridge (2002) Semiparametric estimation of partially linear models for dependent data with generated regressors. *Econometric Theory* 18, 625–645.
Papageorgiou, C. (2002) Trade as a threshold variable for multiple regimes. *Economics Letters* 77, 85–91.
Perron, P. & Z. Qu (2006) Estimating restricted structural change models. *Journal of Econometrics* 134, 372–399.
Reinhart, C.M. & K.S. Rogoff (2010) Growth in time of debt. *American Economic Review Papers and Proceedings* 100, 573–578.
Seo, M.H. & O. Linton (2007) A smoothed least squares estimator for threshold regression models. *Journal of Econometrics* 141, 704–735.
Stryhn, H. (1996) The location of the maximum of asymmetric two-sided Brownian motion with triangular drift. *Statistics and Probability Letters* 29, 279–284.
Tan, C.M. (2010) No one true path: Uncovering the interplay between geography, institutions, and fractionalization in economic development. *Journal of Applied Econometrics* 25, 1100–1127.
van der Vaart, A.W. (1998) *Asymptotic Statistics*. Cambridge University Press.
Yu, P. (2012) Likelihood estimation and inference in threshold regression. *Journal of Econometrics* 167, 274–294.
Yu, P. (2013a) Adaptive estimation of the threshold point in threshold regression. *Journal of Econometrics*. Forthcoming.
Yu, P. (2013b) Inconsistency of 2SLS estimators in threshold regression with endogeneity. *Economics Letters* 120, 532–536.
Yu, P. (2014a) Bootstrap in threshold regression. *Econometric Theory* 30, 676–714.
Yu, P. (2014b) Understanding estimators of treatment effects in regression discontinuity designs. *Econometric Reviews*. doi:10.1080/07474938.2013.833831.
Yu, P. & C.B. Phillips (2014) Threshold Regression with Endogeneity. Mimeo, The University of Hong Kong.
Yu, P. & Y. Zhao (2013) Asymptotics for threshold regression under general conditions. *Econometrics Journal* 16, 430–462.

APPENDIX

The Model in Matrix Notation

Define the matrix $G(\gamma)$ by stacking $g_i(\gamma)$. Define also the regime-specific matrices $G_\gamma(\gamma)$ and $G_\perp(\gamma)$ by stacking $g_i(\gamma)I(q_i \leq \gamma)$ and $g_i(\gamma)I(q_i > \gamma)$, respectively. Let Y and e be the stacked vectors of y_i and e_i, respectively. Then we can write (3.16) as follows

$$Y = G_\gamma(\gamma)\beta_1 + G_\perp(\gamma)\beta_2 + e = G^*(\gamma)\beta + e, \tag{A.1}$$

where $G^*(\gamma) = (G_\gamma(\gamma), G_\perp(\gamma))$ and $\beta = (\beta_1', \beta_2')'$. By defining $\delta_n = \beta_1 - \beta_2$ we can also write

$$Y = G(\gamma_0)\beta_2 + G_0(\gamma_0)\delta_n + e, \tag{A.2}$$

where $G(\gamma_0)$ and $G_0(\gamma_0)$ are the matrices $G(\gamma)$ and $G_\gamma(\gamma)$ evaluated at γ_0.

Recall that $\widehat{x}_i = \widehat{g}_{xi}$ so that $\widehat{X} = \widehat{G}_x$. Let $\widehat{X}_\gamma$ be the stacked matrix of $\widehat{x}_i' I(q_i \le \gamma)$ and let $\widehat{\lambda}_{1,\gamma}(\gamma)$ and $\widehat{\lambda}_{2,\gamma}(\gamma)$ be the stacked vectors of $\widehat{\lambda}_{1,i}(\gamma) I(q_i \le \gamma)$ and $\widehat{\lambda}_{2,i}(\gamma) I(q_i \le \gamma)$, respectively. Then we can define the $n \times (p+2)$ matrix $\widehat{X}_\gamma(\gamma) = \left(\widehat{X}_\gamma, \widehat{\lambda}_{1,\gamma}(\gamma), \widehat{\lambda}_{2,\gamma}(\gamma)\right)$ and its orthogonal matrix $\widehat{X}_\perp(\gamma) = \left(\widehat{X}_\perp, \widehat{\lambda}_{1,\perp}(\gamma), \widehat{\lambda}_{2,\perp}(\gamma)\right)$. Note that $\widehat{X}_\gamma(\gamma) = \widehat{G}_\gamma(\gamma)$ and $\widehat{X}_\gamma(\perp) = \widehat{G}_\gamma(\perp)$. We can then define the projections $P_\gamma(\gamma) = \widehat{X}_\gamma(\gamma)\left(\widehat{X}_\gamma(\gamma)'\widehat{X}_\gamma(\gamma)\right)^{-1}\widehat{X}_\gamma(\gamma)'$, $P_\perp(\gamma) = \widehat{X}_\perp(\gamma)\left(\widehat{X}_\perp(\gamma)'\widehat{X}_\perp(\gamma)\right)^{-1}\widehat{X}_\perp(\gamma)'$, and $P^*(\gamma) = \widehat{X}^*(\gamma)\left(\widehat{X}^*(\gamma)'\widehat{X}^*(\gamma)\right)^{-1}\widehat{X}^*(\gamma)'$, where $\widehat{X}^*(\gamma) = \left(\widehat{X}_\gamma(\gamma), \widehat{X}_\perp(\gamma)\right)$ such that $P^*(\gamma) = P_\gamma(\gamma) + P_\perp(\gamma)$.

Define the estimation errors from the reduced form estimations $\widehat{r}_{xi} = g_{xi} - \widehat{g}_{xi}$, $\widehat{r}_{\lambda_{1i}} = \lambda_{1i}(\gamma) - \widehat{\lambda}_{1i}(\gamma)$, and $\widehat{r}_{\lambda_{2i}} = \lambda_{2i}(\gamma) - \widehat{\lambda}_{2i}(\gamma)$. Define $\widehat{r}_i = \left(\widehat{r}_{xi}', \widehat{r}_{\lambda 1 i}, \widehat{r}_{\lambda 2 i}\right)'$. Then the second stage residual of the unconstrained model in equation (A.1), $\widetilde{e}_i = \widehat{r}_i'\beta + e_i$ and its vector form $\widetilde{e} = \widehat{r}\beta + e$.

Recall that $\bar{g}_i = \sup_{\gamma\in\Gamma}|g_i(\gamma)|$, $\lambda_{1i}(\gamma) \equiv \lambda_1\left(\gamma - z_i'\pi_q\right)$, and $\lambda_{2i}(\gamma) \equiv \lambda_2\left(\gamma - z_i'\pi_q\right)$. Let $\bar{g}_i = \left(g_{xi}', \bar{\lambda}_{1i}, \bar{\lambda}_{2i}\right)'$ and $\widehat{\bar{g}}_i = \left(\widehat{g}_{xi}', \widehat{\bar{\lambda}}_{1i}, \widehat{\bar{\lambda}}_{2i}\right)'$, where $\bar{\lambda}_{1i} = \sup_{\gamma\in\Gamma}|\lambda_1\left(\gamma - z_i'\pi_q\right)|$, $\bar{\lambda}_{2i} = \sup_{\gamma\in\Gamma}|\lambda_2\left(\gamma - z_i'\pi_q\right)|$, $\widehat{\bar{\lambda}}_{1i} = \sup_{\gamma\in\Gamma}|\lambda_1\left(\gamma - z_i'\widehat{\pi}_q\right)|$, and $\widehat{\bar{\lambda}}_{2i} = \sup_{\gamma\in\Gamma}|\lambda_2\left(\gamma - z_i'\widehat{\pi}_q\right)|$.

Proof of Proposition 1. The proof proceeds as follows. First, we show that $\widetilde{\gamma}$ is consistent for the unconstrained problem following the proof strategy of Caner and Hansen (2004). Then, we show that the same estimator has to be consistent for the constrained problem.

Given that $G(\gamma) = \widehat{G}(\gamma) + \widehat{r}$ and $\widehat{G}(\gamma) = \widehat{X}(\gamma)$ is in the span of $\widehat{X}^*(\gamma)$ then $(I - P^*(\gamma))G(\gamma) = (I - P^*(\gamma))\widehat{r}$ and

$$(I - P^*(\gamma))Y = (I - P^*(\gamma))\left(G(\gamma_0)\beta + G_0(\gamma_0)\delta_n + \widetilde{e}\right)$$

Then

$$\begin{aligned} S_n^U(\gamma) &= Y'(I - P^*(\gamma))Y \\ &= \left(n^{-\alpha}c'G_0(\gamma_0)' + \widetilde{e}'\right)(I - P^*(\gamma))\left(G_0(\gamma_0)n^{-\alpha}c + \widetilde{e}\right) \\ &= \left(n^{-\alpha}c'G_0(\gamma_0)' + \widetilde{e}'\right)\left(G_0(\gamma_0)n^{-\alpha}c + \widetilde{e}\right) \\ &\quad -\left(n^{-\alpha}c'G_0(\gamma_0)' + \widetilde{e}'\right)P^*(\gamma)\left(G_0(\gamma_0)n^{-\alpha}c + \widetilde{e}\right) \end{aligned} \tag{A.3}$$

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

$$\begin{aligned} S_n^{*U}(\gamma) &= n^{-1+2\alpha}\left(n^{-\alpha}c'G_0(\gamma_0)' + \widetilde{e}'\right)P^*(\gamma)\left(G_0(\gamma_0)n^{-\alpha}c + \widetilde{e}\right) \\ &= n^{-1+2\alpha}\widetilde{e}'P^*(\gamma)\widetilde{e} + 2n^{-1+\alpha}c'G_0(\gamma_0)'P^*(\gamma)\widetilde{e} + n^{-1}c'G_0(\gamma_0)'P^*(\gamma)G_0(\gamma_0)c \end{aligned}$$

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 I.A.3 we can show that for all $\gamma \in \Gamma$,

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

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

https://doi.org/10.1017/S0266466615000067 Published online by Cambridge University Press

$$n^{-1+\alpha}c_\delta' G_0(\gamma_0)' P_\gamma(\gamma)\widetilde{e} = n^{\alpha-1/2}\left(\frac{1}{n}G_0(\gamma_0)'\widehat{X}_0(\gamma)\right)\left(\frac{1}{n}\widehat{X}_\gamma(\gamma)'\widehat{X}_\gamma(\gamma)\right)^{-1} \times \left(\frac{1}{\sqrt{n}}\widehat{X}_\gamma(\gamma)'\widetilde{e}\right) \xrightarrow{p} 0$$

So we obtain,

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

Let

$$M_0(\gamma_0,\gamma) = \begin{pmatrix} E\left(g_{xi}g_{xi}' I(q_i \le \gamma_0)\right) & E\left(g_{xi}\lambda_{1i}(\gamma_0) I(q_i \le \gamma_0)\right) & E\left(\lambda_{2,i}(\gamma_0) g_{xi} I(q_i \le \gamma_0)\right) \\ E\left(\lambda_{1i}(\gamma) g_{xi}' I(q_i \le \gamma_0)\right) & E\left(\lambda_{1i}(\gamma_0)\lambda_{1i}(\gamma) I(q_i \le \gamma_0)\right) & E\left(\lambda_{2i}(\gamma_0)\lambda_{1i}(\gamma) I(q_i \le \gamma_0)\right) \\ E\left(\lambda_{2i}(\gamma) g_{xi}' I(q_i \le \gamma_0)\right) & E\left(\lambda_{1i}(\gamma_0)\lambda_{2i}(\gamma) I(q_i \le \gamma_0)\right) & E\left(\lambda_{2i}(\gamma_0)\lambda_{2i}(\gamma) I(q_i \le \gamma_0)\right) \end{pmatrix}$$

Compute

$$\frac{1}{n}\widehat{X}_\gamma(\gamma)' G_0(\gamma_0) = \begin{pmatrix} \frac{1}{n}\widehat{X}_\gamma' G_{x,0} & \frac{1}{n}\widehat{X}_\gamma' \lambda_{1,0}(\gamma_0) & \frac{1}{n}\widehat{X}_\gamma' \lambda_{2,0}(\gamma_0) \\ \frac{1}{n}\widehat{\lambda}_{1,\gamma}(\gamma)' G_{x,0} & \frac{1}{n}\widehat{\lambda}_{1,\gamma}(\gamma)'\lambda_{1,0}(\gamma_0) & \frac{1}{n}\widehat{\lambda}_{1,\gamma}(\gamma)'\lambda_{2,0}(\gamma_0) \\ \frac{1}{n}\widehat{\lambda}_{2,\gamma}(\gamma)' G_{x,0} & \frac{1}{n}\widehat{\lambda}_{2,\gamma}(\gamma)'\lambda_{1,0}(\gamma_0) & \frac{1}{n}\widehat{\lambda}_{2,\gamma}(\gamma)'\lambda_{2,0}(\gamma_0) \end{pmatrix}$$
$$= \begin{pmatrix} \frac{1}{n}\sum_i g_{xi}'\widehat{x}_i I(q_i \le \gamma_0) & \frac{1}{n}\sum_i \lambda_{1,i}(\gamma_0)\widehat{x}_i I(q_i \le \gamma_0) & \frac{1}{n}\sum_i \lambda_{2,i}(\gamma_0)\widehat{x}_i I(q_i \le \gamma_0) \\ m\frac{1}{n}\sum_i g_{xi}'\widehat{\lambda}_{1i}(\gamma) I(q_i \le \gamma_0) & \frac{1}{n}\sum_i \lambda_{1i}(\gamma_0)\widehat{\lambda}_{1i}(\gamma) I(q_i \le \gamma) & \frac{1}{n}\sum_i \lambda_{2i}(\gamma_0)\widehat{\lambda}_{1i}(\gamma) I(q_i \le \gamma_0) \\ \frac{1}{n}\sum_i g_{xi}'\widehat{\lambda}_{2i}(\gamma) I(q_i \le \gamma_0) & \frac{1}{n}\sum_i \lambda_{1i}(\gamma_0)\widehat{\lambda}_{2i}(\gamma) I(q_i \le \gamma_0) & \frac{1}{n}\sum_i \lambda_{2i}(\gamma_0)\widehat{\lambda}_{2i}(\gamma) I(q_i \le \gamma_0) \end{pmatrix}$$

Note that when $\gamma = \gamma_0$, $M_0(\gamma_0,\gamma_0) = M_0(\gamma_0)$ we obtain

$$\frac{1}{n}G_0(\gamma_0)' P_\gamma(\gamma) G_0(\gamma_0) \to M_0(\gamma_0,\gamma)' M_\gamma(\gamma)^{-1} M_0(\gamma_0,\gamma)$$

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

$$S_n^{*U}(\gamma) \to c' M_0(\gamma_0,\gamma)' M_\gamma(\gamma)^{-1} M_0(\gamma_0,\gamma)c \tag{A.4}$$

by a Glivenko-Cantelli theorem for stationary ergodic processes.

Given the monotonicity of the inverse Mills ratio, $M_0(\gamma_0,\gamma_0+\epsilon) \ge M_0(\gamma_0)$ for any $\epsilon > 0$ with equality at $\gamma = \gamma_0$. To see this note that for $\epsilon > 0$, $\lambda_{1i}(\gamma_0+\epsilon) > \lambda_{1i}(\gamma_0)$ and $\lambda_{2i}(\gamma_0+\epsilon) > \lambda_{2i}(\gamma_0)$. Therefore, we need to show that $S_n^{*U}(\gamma) < M_0(\gamma_0)$ 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 $\gamma \in (\gamma_0,\overline{\gamma}]$.

https://doi.org/10.1017/S0266466615000067 Published online by Cambridge University Press

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

$$M_\varepsilon(\gamma_0+\epsilon) - M_0(\gamma_0) = \int_{\gamma_0}^{\gamma_0+\epsilon} E(g_i(t)g_i(t)'|q=t)f_q(t)dt$$
$$> \inf_{\gamma_0<\gamma\le\gamma_0+\epsilon} Eg_{xi}(\gamma)g'_{xi}(\gamma))|q=\gamma)\left(\int_{\gamma_0}^{\gamma_0+\epsilon} f(\nu)d\nu\right)$$
$$= \inf_{\gamma_0<\gamma\le\gamma_0+\epsilon} D_1(\gamma)\left(\int_{\gamma_0}^{\gamma_0+\epsilon} f(\nu)d\nu\right) > 0$$

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

Given that the conditions of Van der Vaart (1998, Thm. 5.7) are satisfied, the uniform convergence of $S_n^{*U}(\gamma)$, i.e. $\sup_{\gamma\in\Gamma}|S_n^{*U}(\gamma) - S_n^{*U}(\gamma_0)| \xrightarrow{p} 0$ as $n \longrightarrow \infty$, the compactness of Γ, and the fact that $S_n^{*U}(\gamma)$ is uniquely maximized at γ_0, we can have $\sup_{|\gamma-\gamma_0|\ge\epsilon} S_n^{*U}(\gamma) < S_n^{*U}(\gamma_0)$ for every $\epsilon > 0$. Therefore, it follows that the estimator for γ obtained by minimizing the CLS based on the unconstrained projection in equation (3.16), $\widetilde{\gamma} \xrightarrow{p} \gamma_0$.

Denote the constrained sum of squares errors under the optimal split as $S^R(\widehat{\gamma})$ and the constrained sum of squares errors under the true split as $S^R(\gamma_0)$. Assuming the restrictions in equation (3.17) hold we have

$$S_n^R(\widehat{\gamma}) \le S_n^R(\gamma_0) \le S_n^U(\gamma) \tag{A.5}$$

When the threshold estimate is not consistent, we have that

$$S_n^U(\widetilde{\gamma}) \ge S_n^U(\gamma) + C||\beta_0 - \beta||^2 + o_p(1),$$

holds with positive probability, where β_0 is the vector of true slope coefficients. Since $S^U(\widetilde{\gamma}) \le S^R(\widehat{\gamma})$, we also have

$$S_n^R(\widehat{\gamma}) \ge S_n^U(\gamma) + C||\beta_0 - \beta||^2 + o_p(1). \tag{A.6}$$

holds with positive probability. Comparing (A.5) with (A.6) we get a contradiction if the threshold parameter is not consistently estimated. Hence, the constrained estimator $\widehat{\gamma}$ is also consistent from (A.5). This completes the proof. ■

Proof of Lemma 1. Recall that

$$S_n^R(\gamma) = S_n^U(\gamma) + (\vartheta - R'\widehat{\beta})'\left(R'(\widehat{X}^*(\gamma)'\widehat{X}^*(\gamma))^{-1}R\right)^{-1}(\vartheta - R'\widehat{\beta})$$

Then, we can obtain

$$S_n^R(\gamma) - S_n^R(\gamma_0) = \left[S_n^U(\gamma) - S_n^U(\gamma_0)\right]$$
$$+\Big[(\vartheta - R'\widehat{\beta})'\left(R'\left(\widehat{X}^*(\gamma)'\widehat{X}^*(\gamma)\right)^{-1}R\right)^{-1}(\vartheta - R'\widehat{\beta})$$
$$-(\vartheta - R'\widehat{\beta}_0)'(R'(X^*(\gamma_0)'X^*(\gamma_0))^{-1}R)^{-1}(\vartheta - R'\widehat{\beta}_0))\Big],$$

where $\widehat{\beta}_0$ is the $\widehat{\beta}$ evaluated at γ_0. We show that the second term is $o_p(1)$.

https://doi.org/10.1017/S0266466615000067 Published online by Cambridge University Press

Define $\Delta_i(\gamma) = I(q_i \le \gamma) - I(q_i \le \gamma_0)$ and $\widetilde{I} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. Consider the case of $\gamma \le \gamma_0$ for some $\epsilon > 0$. Then,

$$\frac{1}{n}||\widehat{X}^*(\gamma)'\widehat{X}^*(\gamma) - \widehat{X}^*(\gamma_0)'\widehat{X}(\gamma_0)||$$
$$= \frac{1}{n}||\left(\sum_i g_i(\gamma)g_i(\gamma)'\Delta_i(\gamma) - \sum_i g_i(\gamma)\widehat{r}\,'\Delta_i(\gamma) - \sum_i g_i(\gamma)'\widehat{r}_i\Delta_i(\gamma) + \sum_i \widehat{r}_i\widehat{r}_i'\Delta_i(\gamma)\right)\otimes \widetilde{I}||$$
$$\le \sqrt{2}\frac{1}{na_n}\left(tr\left(\sum_i g_i(\gamma_0+\epsilon)g_i(\gamma_0+\epsilon)'\Delta_i(\gamma)a_n\right)^2\right)^{1/2}$$
$$+\sqrt{2}\frac{2}{na_n}\left(tr\left(\sum_i g_i(\gamma_0+\epsilon)\widehat{r}_i'\Delta_i(\gamma)a_n\right)^2\right)^{1/2}$$
$$+\sqrt{2}\frac{1}{na_n}\left(tr\left(\sum_i \widehat{r}_i\widehat{r}_i'\Delta_i(\gamma)a_n\right)^2\right)^{1/2} = o_p(1).$$

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

$$\left(\frac{1}{n}\widehat{X}^*(\gamma)'\widehat{X}^*(\gamma)\right)^{-1} = \left(\frac{1}{n}\widehat{X}^*(\gamma_0)'\widehat{X}^*(\gamma_0)\right)^{-1} + o_p(1). \tag{A.7}$$

and

$$\frac{1}{n}(R'(\widehat{X}^*(\gamma)'\widehat{X}^*(\gamma))^{-1}R)^{-1} = \frac{1}{n}(R'(\widehat{X}^*(\gamma_0)'\widehat{X}^*(\gamma_0))^{-1}R)^{-1} + o_p(1). \tag{A.8}$$

Note that $S_n^U(\gamma) - S_n^U(\gamma_0) = o_p(1)$ and $n^{1/2}(\widehat{\beta} - \beta_0) = n^{1/2}(\widehat{\beta}_0 - \beta_0) + o_p(1)$. Then,

$$\begin{aligned}
S_n^R(\gamma) - S_n^R(\gamma_0) &= \left[S_n^U(\gamma) - S_n^U(\gamma_0)\right] \\
&\quad + \Big[(\vartheta - R'\widehat{\beta})'\left(R'\left(\widehat{X}^*(\gamma)'\widehat{X}^*(\gamma)\right)^{-1}R\right)^{-1}(\vartheta - R'\widehat{\beta}) \\
&\quad - (\vartheta - R'\widehat{\beta}_0)'\left(R'\left(\widehat{X}^*(\gamma_0)'\widehat{X}^*(\gamma_0)\right)^{-1}R\right)^{-1}(\vartheta - R\widehat{\beta}_0)\Big] \\
&= \Big[(\vartheta - R'\widehat{\beta})\left(R'\left(\widehat{X}^*(\gamma_0)'\widehat{X}^*(\gamma_0)\right)^{-1}R\right)^{-1}(\vartheta - R'\widehat{\beta}) \\
&\quad - (\vartheta - R'\widehat{\beta}_0)'\left(R'\left(\widehat{X}^*(\gamma_0)'\widehat{X}^*(\gamma_0)\right)^{-1}R\right)^{-1}(\vartheta - R'\widehat{\beta})\Big] + o_p(1) \\
&= n^{1/2}\left(\beta_0 - \widehat{\beta}\right)'R\left(\left(R'\left(\widehat{X}^*(\gamma_0)'\widehat{X}^*(\gamma_0)\right)^{-1}R\right)^{-1}R'n^{1/2}\left(\beta_0 - \widehat{\beta}\right)\right. \\
&\quad - n^{1/2}\left(\beta_0 - \widehat{\beta}_0\right)'R\left(R'\left(\widehat{X}^*(\gamma_0)\widehat{X}^*(\gamma_0)\right)^{-1}R\right)^{-1}R'n^{1/2}\left(\beta_0 - \widehat{\beta}_0\right) \\
&\quad + o_p(1) \\
&= o_p(1).
\end{aligned}$$

This completes the proof. ∎

https://doi.org/10.1017/S0266466615000067 Published online by Cambridge University Press

Proof of Theorem 1. Lemmas I.A.4 and I.A.5 of the Internet Appendix imply $a_n(\widehat{\gamma}-\gamma_0)=\arg\max_v Q_n(v)=O_p(1)$ and $Q_n(v)\Rightarrow Q(v)$, respectively. Given that the limit functional $Q(v)$ is continuous, has a unique maximum, and $\lim_{|v|\to\infty} Q(v)=-\infty$, almost surely, by Kim and Pollard (1990, Thm. 2.7) and Hansen (2000, Thm. 1) we can get $n^{1-2\alpha}(\widehat{\gamma}-\gamma_0)\xrightarrow{d} \arg\max_{-\infty<v<\infty} Q_n(v)$.

Set $\omega=\zeta_1/\mu^2$ and recall that $\mathcal{W}_i(b^2v)=b\mathcal{W}_i(v)$. By making the change of variables $v=(\zeta_1/\mu^2)s$ we can rewrite the asymptotic distribution as follows:

$$\arg\max_{-\infty<v<\infty} Q_n(v) = \begin{cases} \arg\max_{-\infty<v<\infty}\left(-\frac{\zeta_1}{\mu}|s|+2\zeta_1^{1/2}\mathcal{W}_1\left(\left(\zeta_1/\mu^2\right)s\right)\right), & \text{if } s\in[-\bar{v},0] \\ \arg\max_{-\infty<v<\infty}\left(-\frac{\zeta_1}{\mu}|s|+2\zeta_2^{1/2}\mathcal{W}_2\left(\left(\zeta_1/\mu^2\right)s\right)\right), & \text{if } s\in[0,\bar{v}] \end{cases}$$

or equivalently

$$\arg\max_{-\infty<v<\infty} Q_n(v) = \begin{cases} \omega \arg\max_{-\infty<s<\infty}\left(-\frac{1}{2}|s|+\mathcal{W}_1(s)\right), & \text{if } s\in[-\bar{v},0] \\ \omega \arg\max_{-\infty<s<\infty}\left(-\frac{1}{2}|s|+\sqrt{\varphi}\mathcal{W}_2(s)\right), & \text{if } s\in[0,\bar{v}] \end{cases}$$

where $\varphi=\zeta_2/\zeta_1$. Hence,

$$n^{1-2\alpha}\left(\widehat{\gamma}-\gamma_0\right)\xrightarrow{d}\arg\max_{-\infty<v<\infty}\omega\mathcal{T}(s),$$

where

$$\mathcal{T}(s)=\begin{cases} -\frac{1}{2}|s|+\mathcal{W}_1(-s), & \text{if } s\in[-\bar{v},0] \\ -\frac{1}{2}|s|+\sqrt{\varphi}\mathcal{W}_2(s), & \text{if } s\in[0,\bar{v}] \end{cases}$$

■

Proof of Theorem 2. From Lemma I.A.3, equation I.A.13 of the Internet Appendix and Hansen's (2000) Lemma A.12 and Theorem 2 we have $\widetilde{\sigma}_e^2 LR_n(\gamma_0)-Q_n(v)\xrightarrow{p}0$. Then,

$$LR_n(\gamma)=\frac{Q_n(\bar{v})}{\widetilde{\sigma}_e^2}+o_p(1)=\frac{1}{\widetilde{\sigma}_e^2}\sup_{-\infty<v<\infty}Q_n(v)+o_p(1)\xrightarrow{d}\frac{1}{\sigma_e^2}\sup_{-\infty<v<\infty}Q_n(v).$$

Using the change of variables $v=(\zeta_1/\mu^2)s$ the limiting distribution can be rewritten as follows

$$\begin{aligned}
&\frac{1}{\sigma_e^2}\sup_{-\infty<v<\infty}Q_n(v)=\\
&=\frac{1}{\sigma_e^2}\sup_{-\infty<v<\infty}\left(\left(-\mu|v|+2\zeta_1^{1/2}\mathcal{W}_1(v)\right)I(v<0)+\left(-\mu|v|+2\zeta_2^{1/2}\mathcal{W}_2(v)\right)I(v>0)\right)\\
&=\frac{1}{\sigma_e^2}\sup_{-\infty<v<\infty}\left(\left(-|\frac{\zeta_1}{\mu}s|+2\zeta_1^{1/2}\mathcal{W}_1\left(\frac{\zeta_1}{\mu^2}s\right)\right)I(v<0)\right.\\
&\qquad\left.+\left(-|\frac{\zeta_1}{\mu}s|+2\zeta_2^{1/2}\mathcal{W}_2\left(\frac{\zeta_1}{\mu^2}s\right)\right)\right)I(v>0)\Big)\\
&=\frac{\zeta_1}{\sigma_e^2\mu}\sup_{-\infty<v<\infty}\left((-|s|+2\mathcal{W}_1(s))I(v<0)+(-|s|+2\sqrt{\varphi}\mathcal{W}_2(s))I(v>0)\right)\\
&=\eta^2\psi,\ \text{where}\ \ \eta^2=\frac{\zeta_1}{\sigma_e^2\mu_1}.
\end{aligned}$$

Observe that $\psi = 2\max(\psi_1, \psi_2)$, where $\psi_1 = \sup_{s\le 0}(-|s| + 2\mathcal{W}_1(s))$ and $\psi_2 = \sup_{s>0}\left(-|s| + 2\sqrt{\varphi}\mathcal{W}_2(s)\right)$ are independent but not identical exponential distributions with $P(\psi_1 \le x/2) = 1 - e^{-x/2}$ and $P(\psi_2 \le x/2) = 1 - e^{-\sqrt{\varphi}x/2}$, respectively. Hence,

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

It is useful to view the 2SLS and GMM estimators of $\beta = (\beta_1', \beta_2')'$ defined in Section 3.3 as special cases of the following class of estimators. Given consistent weight matrices $\widetilde{W_j} \xrightarrow{p} W_j > 0$, we can define the class of unconstrained GMM estimators

$$\widetilde{\beta}_1 = (\widehat{X}_1'\widehat{Z}_1\widetilde{W}_1\widehat{Z}_1'\widehat{X}_1)^{-1}\widehat{X}_1'\widehat{Z}_1\widetilde{W}_1\widehat{Z}_1'Y, \quad \textbf{(A.9a)}$$

$$\widetilde{\beta}_2 = (\widehat{X}_2'\widehat{Z}_2\widetilde{W}_2\widehat{Z}_2'\widehat{X}_2)^{-1}\widehat{X}_2'\widehat{Z}_2\widetilde{W}_2\widehat{Z}_2'Y. \quad \textbf{(A.9b)}$$

Define $\widetilde{\beta} = (\widetilde{\beta}_1', \widetilde{\beta}_2')'$ and $\widetilde{W} = diag(\widetilde{W_1}, \widetilde{W_2})$. Then, the class of constrained GMM estimators is given as a minimum distance estimator that solves the problem, $\widehat{\beta}_C = \operatorname{argmin}_{R'\beta=\vartheta} J_n(\beta)$, where $J_n(\beta) = n(\widetilde{\beta} - \beta)'\widetilde{W}(\widetilde{\beta} - \beta)$ and consistent weight matrix $\widetilde{W} \xrightarrow{p} W > 0$. This constrained estimator can be computed by

$$\widehat{\beta}_C = \widetilde{\beta} - \widetilde{W}R(R'\widetilde{W}R)^{-1}(R'\widetilde{\beta} - \vartheta). \quad \textbf{(A.10)}$$

Proof of Theorem 3. The unconstrained 2SLS estimators $\widetilde{\beta}_{1,2SLS}$ and $\widetilde{\beta}_{2,2SLS}$ fall in the class of GMM estimators defined in equations (A.9a) and (A.9b) with

$$\widehat{W}_{1,2SLS} = \left(\frac{1}{n}\sum_{i=1}^{n}\widehat{z}_i(\widehat{\gamma})\widehat{z}_i(\widehat{\gamma})'I(q_i \le \widehat{\gamma})\right)^{-1}$$

$$\widehat{W}_{2,2SLS} = \left(\frac{1}{n}\sum_{i=1}^{n}\widehat{z}_i(\widehat{\gamma})\widehat{z}_i(\widehat{\gamma})'I(q_i > \widehat{\gamma})\right)^{-1}$$

replacing $\widetilde{W}_1$ and $\widetilde{W}_2$.

From Hansen (1996, Lemma 1) and the consistency of $\widehat{\gamma}$ we obtain that $\widehat{W}_{1,2SLS} \xrightarrow{p} Q_1^{-1}$ and $\widehat{W}_{2,2SLS} \xrightarrow{p} Q_2^{-1}$. Therefore, the unconstrained 2SLS estimators $\widetilde{\beta}_{1,2SLS}$ and $\widetilde{\beta}_{2,2SLS}$ are asymptotically normal with covariance matrices $V_{1,2SLS}$ and $V_{2,2SLS}$, which are obtained by (I.A.42a) and (I.A.42b) of the Internet Appendix with Q_1^{-1} and Q_2^{-1} replacing W_1 and W_2, respectively. Let $\widetilde{\beta}_{2SLS} = \left(\widetilde{\beta}_{1,2SLS}', \widetilde{\beta}_{2,2SLS}'\right)'$ and $V_{2SLS} = diag(V_{1,2SLS}, V_{2,2SLS})$ then we easily see that

$$\sqrt{n}\left(\widetilde{\beta}_{2SLS} - \beta\right) \Rightarrow N(0, V_{2SLS}).$$

Similarly, the constrained 2SLS estimators $\widehat{\beta}_{C2SLS} = \left(\widehat{\beta}_{1,C2SLS}', \widehat{\beta}_{2,C2SLS}'\right)'$ fall in the class of GMM estimators defined in equation (A.10) of the Internet Appendix with $\widehat{W}_{2SLS} = diag(\widehat{W}_{1,2SLS}, \widehat{W}_{2,2SLS})$ replacing $\widetilde{W}$ and $\widetilde{\beta}_{2SLS} = \left(\widetilde{\beta}_{1,2SLS}', \widetilde{\beta}_{2,2SLS}'\right)'$ replacing $\widetilde{\beta}$.

https://doi.org/10.1017/S0266466615000067 Published online by Cambridge University Press

Therefore, the constrained 2SLS estimator $\widehat{\beta}_{C2SLS}$ is also asymptotically normal with covariance matrix V_{C2SLS} given by the covariance matrix of the class GMM estimator in (I.A.44) by setting $W = Q$ and with V_{C2SLS} and V_{2SLS} replacing V_C and V, respectively. Specifically,

$$\sqrt{n}(\widehat{\beta}_{C2SLS} - \beta) \Rightarrow N(0, V_{C2SLS})$$

where

$$V_{C2SLS} = V_{2SLS} - Q^{-1}R(R'Q^{-1}R)^{-1}R'V_{2SLS} - V_{2SLS}R(R'Q^{-1}R)^{-1}R'Q^{-1} + Q^{-1}R(R'Q^{-1}R)^{-1}R'V_{2SLS}R(R'Q^{-1}R)^{-1}R'Q^{-1}. \blacksquare$$

Proof of Theorem 4. Define

$$\widehat{S}_1(\gamma) = \frac{1}{n}\sum_{i=1}^{n} z_i(\gamma)z_i(\gamma)I(q_i \le \gamma)$$

$$\widehat{S}_2(\gamma) = \frac{1}{n}\sum_{i=1}^{n} z_i(\gamma)z_i(\gamma)'I(q_i > \gamma)$$

$$\widetilde{\Sigma}_{1,GMM}(\gamma) = \frac{1}{n}\sum_{i=1}^{n} z_i(\gamma)z_i(\gamma)'\widetilde{e}^2_{i,2SLS}I(q_i \le \gamma)$$

$$\widetilde{\Sigma}_{2,GMM}(\gamma) = \frac{1}{n}\sum_{i=1}^{n} z_i(\gamma)z_i(\gamma)'\widetilde{e}^2_{i,2SLS}I(q_i > \gamma)$$

$$\widetilde{V}_{1,GMM}(\gamma) = \widehat{S}_1(\gamma)'\widetilde{\Sigma}_1(\gamma)^{-1}\widehat{S}_1(\gamma)$$

$$\widetilde{V}_{2,GMM}(\gamma) = \widehat{S}_2(\gamma)'\widetilde{\Sigma}_2(\gamma)^{-1}\widehat{S}_2(\gamma)$$

$$\widetilde{V}_{GMM}(\gamma) = diag(\widetilde{V}_1(\gamma), \widetilde{V}_2(\gamma))$$

$$\widetilde{V}_{CGMM}(\gamma) = \widetilde{V}_{GMM}(\gamma) - \widetilde{V}_{GMM}(\gamma)R\left(R'\widetilde{V}_{GMM}(\gamma)R\right)^{-1}R'\widetilde{V}_{GMM}(\gamma)$$

Notice that the unconstrained GMM estimators $\widetilde{\beta}_{1,GMM}$ and $\widetilde{\beta}_{2,GMM}$ fall in the class of GMM estimators defined in equations (A.9a) and (A.9b) by replacing $\widetilde{W}_1$ and $\widetilde{W}_2$ with $\widetilde{\Sigma}^{-1}_{1,GMM}(\gamma)$ and $\widetilde{\Sigma}^{-1}_{2,GMM}(\gamma)$, respectively. Similarly, the constrained GMM estimator $\widehat{\beta}_{CGMM}$ falls in the class of GMM estimators defined in equation (A.10) of the Internet Appendix with $\widetilde{W}(\gamma) = \widetilde{V}^{-1}_{GMM}(\gamma)$.

To prove Theorem 4 it is sufficient to show that the following hold uniformly in $\gamma \in \Gamma$:

$$\widetilde{\Sigma}_{1,GMM}(\gamma) \xrightarrow{p} E(z_i(\gamma)z_i(\gamma)'e_i^2 I(q_i \le \gamma) \quad \textbf{(A.11)}$$

$$\widetilde{\Sigma}_{2,GMM}(\gamma) \xrightarrow{p} E(z_i(\gamma)z_i(\gamma)'e_i^2 I(q_i > \gamma) \quad \textbf{(A.12)}$$

$$\widetilde{S}_1(\gamma) \xrightarrow{p} E(z_i(\gamma)x_i(\gamma)'I(q_i \le \gamma)) \quad \textbf{(A.13)}$$

$$\widetilde{S}_2(\gamma) \xrightarrow{p} E(z_i(\gamma)x_i(\gamma)'I(q_i > \gamma)) \quad \textbf{(A.14)}$$

https://doi.org/10.1017/S0266466615000067 Published online by Cambridge University Press

Then, by the consistency of $\widehat{\gamma}$, we get $\widetilde{S}_1(\widehat{\gamma}) \xrightarrow{p} S_1$, $\widetilde{S}_2(\widehat{\gamma}) \xrightarrow{p} S_2$, $n^{-1}\widetilde{\Sigma}_{1,GMM} = \widetilde{\Sigma}_{1,GMM}(\widehat{\gamma}) \xrightarrow{p} \Sigma_{1,GMM}$, and $n^{-1}\widetilde{\Sigma}_{2,GMM} = \widetilde{\Sigma}_{2,GMM}(\widehat{\gamma}) \xrightarrow{p} \Sigma_{2,GMM}$. Theorem 4 follows from Lemma I.A.6 of the Internet Appendix.

We now establish (A.11). Equations (A.13), (A.12), and (A.14) can be proven similarly. Recall that $x_i(\gamma_0) = (x_i', \lambda_{1i}(\gamma_0), \lambda_{2i}(\gamma_0))'$ and $\beta_1 = (\beta_{x1}', \kappa_{11}, \kappa_{12})$ and $\beta_2 = (\beta_{x2}', \kappa_{21}, \kappa_{22})$. Let $\lambda_i(\gamma_0) = (\lambda_{1i}(\gamma_0), \lambda_{2i}(\gamma_0))'$, $\kappa_1 = (\kappa_{11}, \kappa_{12})'$, and $\kappa_2 = (\kappa_{21}, \kappa_{22})'$. Also define, $\widetilde{\delta} = \widetilde{\beta}_1 - \widetilde{\beta}_2$, $\widetilde{\delta}_\kappa = \widetilde{\kappa}_1 - \widetilde{\kappa}_2$, $\widehat{x}_i(\widehat{\gamma}) = (x_i', \widehat{\lambda}_i(\widehat{\gamma})')'$, $x_i^* = x_i^*(\gamma_0) = (x_i(\gamma_0)' I(q_i \le \gamma_0), x_i(\gamma_0)' I(q_i > \gamma_0))'$, $\Delta\widehat{x}(\widehat{\gamma}) = x_i(\gamma_0)(I(q_i \le \widehat{\gamma}) - I(q_i \le \gamma_0))$, and $\Delta\widehat{\lambda}_i(\widehat{\gamma}) = \widehat{\lambda}_i(\widehat{\gamma}) - \lambda_i(\gamma_0)$. Then compute that

$$\widetilde{e}_i = e_i - x_i^{*'}(\widetilde{\beta} - \beta) - \Delta\widehat{x}(\widehat{\gamma})'\widetilde{\delta} - \Delta\widehat{\lambda}_i(\widehat{\gamma})'\kappa_2 - \Delta\widehat{\lambda}_i(\widehat{\gamma})' I(q_i \le \widehat{\gamma})\delta_\kappa$$

Note that the above expression is similar to the one in Caner and Hansen (2004, Thm. 3), with the difference that it includes the fourth and fifth terms due to the presence of the inverse Mill ratio terms.

Then we get

$$\begin{aligned}
&\widetilde{\Sigma}_{1,GMM}(\gamma) - \frac{1}{n}\sum_{i=1}^n z_i(\gamma) z_i(\gamma)' e_i^2 I(q_i \le \gamma) \\
&\quad = -\frac{2}{n}\sum_{i=1}^n z_i(\gamma) z_i(\gamma)' I(q_i \le \gamma) e_i x_i^{*'}(\widetilde{\beta} - \beta) \\
&\quad\quad - \frac{2}{n}\sum_{i=1}^n z_i(\gamma) z_i(\gamma)' I(q_i \le \gamma) e_i \Delta\widehat{x}(\widehat{\gamma})'\widetilde{\delta} \\
&\quad\quad - \frac{2}{n}\sum_{i=1}^n z_i(\gamma) z_i(\gamma)' I(q_i \le \gamma) e_i \Delta\widehat{\lambda}_i(\widehat{\gamma})'\kappa_2 \\
&\quad\quad - \frac{2}{n}\sum_{i=1}^n z_i(\gamma) z_i(\gamma)' I(q_i \le \gamma) e_i \Delta\widehat{\lambda}_i(\widehat{\gamma})' I(q_i \le \widehat{\gamma})\delta_\kappa \\
&\quad\quad + \frac{1}{n}\sum_{i=1}^n z_i(\gamma) z_i(\gamma)' I(q_i \le \gamma)(\widetilde{\beta} - \beta)' x_i^* x_i^{*'}(\widetilde{\beta} - \beta) \\
&\quad\quad + \frac{2}{n}\sum_{i=1}^n z_i(\gamma) z_i(\gamma)' I(q_i \le \gamma)(\widetilde{\beta} - \beta)' x_i^* \Delta\widehat{x}(\widehat{\gamma})'\widetilde{\delta} \\
&\quad\quad + \frac{2}{n}\sum_{i=1}^n z_i(\gamma) z_i(\gamma)' I(q_i \le \gamma)(\widetilde{\beta} - \beta)' x_i^* \Delta\widehat{\lambda}_i(\widehat{\gamma})'\kappa_2 \\
&\quad\quad + \frac{2}{n}\sum_{i=1}^n z_i(\gamma) z_i(\gamma)' I(q_i \le \gamma)(\widetilde{\beta} - \beta)' x_i^* \Delta\widehat{\lambda}_i(\widehat{\gamma})' I(q_i \le \widehat{\gamma})\delta_\kappa \\
&\quad\quad + \frac{1}{n}\sum_{i=1}^n z_i(\gamma) z_i(\gamma)' I(q_i \le \gamma)\widetilde{\delta}' \Delta\widehat{x}(\widehat{\gamma}) \Delta\widehat{x}(\widehat{\gamma})'\widetilde{\delta} \\
&\quad\quad + \frac{2}{n}\sum_{i=1}^n z_i(\gamma) z_i(\gamma)' I(q_i \le \gamma)\widetilde{\delta}' \Delta\widehat{x}(\widehat{\gamma}) \Delta\widehat{\lambda}_i(\widehat{\gamma})'\kappa_2
\end{aligned}$$

https://doi.org/10.1017/S0266466615000067 Published online by Cambridge University Press

$$+\frac{2}{n}\sum_{i=1}^{n} z_i(\gamma)z_i(\gamma)'I(q_i\leq\gamma)\widetilde{\delta}'\Delta\widehat{x}(\widehat{\gamma})\Delta\widehat{\lambda}_i(\widehat{\gamma})'I(q_i\leq\widehat{\gamma})\delta_\kappa$$
$$+\frac{1}{n}\sum_{i=1}^{n} z_i(\gamma)z_i(\gamma)'I(q_i\leq\gamma)\kappa_2'\Delta\widehat{\lambda}_i(\widehat{\gamma})\Delta\widehat{\lambda}_i(\widehat{\gamma})'\kappa_2$$
$$+\frac{2}{n}\sum_{i=1}^{n} z_i(\gamma)z_i(\gamma)'I(q_i\leq\gamma)\kappa_2'\Delta\widehat{\lambda}_i(\widehat{\gamma})\Delta\widehat{\lambda}_i(\widehat{\gamma})'I(q_i\leq\widehat{\gamma})\delta_\kappa$$
$$+\frac{1}{n}\sum_{i=1}^{n} z_i(\gamma)z_i(\gamma)'I(q_i\leq\gamma)\delta_\kappa'\Delta\widehat{\lambda}_i(\widehat{\gamma})\Delta\widehat{\lambda}_i(\widehat{\gamma})'I(q_i\leq\widehat{\gamma})\delta_\kappa$$

All the terms on the right-hand side converge in probability to zero, uniformly in γ because the data have bounded fourth moments, the inverse Mills ratio terms are bounded, $|\widetilde{\beta}-\beta|\xrightarrow{p}0$, $|\widehat{\pi}_q-\pi_q|\xrightarrow{p}0$, and $|\widehat{\gamma}-\gamma|\xrightarrow{p}0$ (hence $|\widehat{\lambda}_i(\widehat{\gamma})-\lambda_i(\gamma)|\xrightarrow{p}0$). To see this we illustrate the first term.

$$\frac{2}{n}\left|\sum_{i=1}^{n} z_i(\gamma)z_i(\gamma)'I(q_i\leq\gamma)e_i x_i^{*'}(\widetilde{\beta}-\beta)\right|\leq\frac{2}{n}\sum_{i=1}^{n}|\bar{z}_i|^2||e_i||\bar{x}_i||\widetilde{\beta}-\beta|\xrightarrow{p}0$$

Therefore, by Hansen (1996, Lemma 1) we obtain uniformly in γ:

$$\sup_{\gamma\in\Gamma}\left|\widetilde{\Sigma}_{1,GMM}(\gamma)-\frac{1}{n}\sum_{i=1}^{n} z_i(\gamma)z_i(\gamma)'e_i^2 I(q_i\leq\gamma)\right|\xrightarrow{p}0.$$

This completes the proof. ∎

https://doi.org/10.1017/S0266466615000067 Published online by Cambridge University Press