

# Essays in Threshold Regression: Theory and Application

by

Chaoyi Chen

A Thesis  
presented to  
The University of Guelph

In partial fulfilment of requirements  
for the degree of  
Doctor of Philosophy  
in  
Economics

Guelph, Ontario, Canada

©Chaoyi Chen, November, 2019

## ABSTRACT

### ESSAYS IN THRESHOLD REGRESSION: THEORY AND APPLICATION

Chaoyi Chen  
University of Guelph, 2019

Advisors:  
Thanasis Stengos, Yiguo Sun

This thesis consists of three essays in the threshold regression model regarding both theory and application.

Chapter 1 investigates the linear index threshold regression model with endogeneity. We propose a two-step GMM estimation method to estimate the model, which allows both the threshold variable and regressors to be endogenous. We show the consistency of the GMM estimator and derive the asymptotic distribution of the GMM estimator for weakly dependent data. We suggest a test of the exogeneity null hypothesis for both the threshold and the slope regressors. Monte Carlo simulations are used to assess the finite sample performance of our proposed estimator. Finally, we present an empirical application investigating the threshold effect of a linear index between external debt and public debt on economic growth for developing countries.

In Chapter 2, we compare the finite sample performance of three non-parametric threshold estimators via the Monte Carlo method. Our results indicate that the finite sample performance of the three estimators is not robust to the position of the threshold level along the distribution of the threshold variable, especially when a structural change occurs at the tail part of the distribution.

In Chapter 3, we examine the effect of the Exchange Rate Pass-Through (ERPT) on

the “*rockets and feathers*” hypothesis using a panel of EU-28 countries. Allowing for the existence of an endogenous threshold variable, our empirical findings indicate that the threshold model is better suited to this analysis than the baseline linear adjustment model. This is the case since the latter restricts the threshold to be centered around zero and the dynamic response to cumulative shocks cannot be properly identified. The empirical findings reveal that the threshold variable expressed by the trade-weighted dollar exchange rate index is statistically significant only in the sample above the threshold (high regime). This means that for the net EU exporting countries, fluctuations in the real effective exchange rate of the US against its major EU trading partners does affect the level of pre-tax retail gasoline prices with the relevant elasticity exceeding unity (complete ERPT). Moreover, all the statistical tests reject the null hypothesis that there is no significant threshold and thus an asymmetric adjustment gasoline mechanism prevails.

## ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my advisors Thanasis Stengos and Yiguo Sun, for their excellent and continuous support of my Ph.D. study and related research. This thesis would not have been possible without their excellent advice, guidance, and, most importantly, patience. I would also like to offer my thanks to Alex Maynard for investing time and providing excellent insights and helpful comments.

I want to express my gratitude to all the faculty members and staff in the Department of Economics and Finance. I would especially like to thank James Atsu Amegashie, Laurent Cellarier, Talat S. Genc, Michael Hoy, Ren Kirkegaard, Mei Li, Ross McKittrick, Asha Sadanand, Henry Thille, and Ilias Tsiakas for their contribution to my understanding of economic theory and helpful suggestions on my dissertation.

Lastly, I wish to thank my parents for their full support over the years. I would like to offer my special thanks to Suellen Wang for her love and support. I wish to thank all my friends at the University of Guelph for their love and encouragement throughout my entire Ph.D. education.

# Table of Contents

<b>List of Tables</b>	<b>vii</b>
<b>List of Figures</b>	<b>ix</b>
<b>1 Chapter 1: A GMM Estimator for Linear Index Threshold Model</b>	<b>1</b>
1.1 Introduction . . . . .	1
1.2 The Model . . . . .	5
1.3 Estimation Strategy . . . . .	6
1.4 Asymptotic Results . . . . .	8
1.5 Testing . . . . .	12
1.5.1 Test for Linearity . . . . .	12
1.5.2 Test for Exogeneity . . . . .	13
1.6 Monte Carlo Simulation . . . . .	15
1.7 Empirical Application . . . . .	22
1.8 Conclusion . . . . .	26
Appendices . . . . .	27
1.8.1 Proof of Lemma . . . . .	27
1.8.2 Proof of Theorem . . . . .	29
1.8.3 Finite Sample Performance of the Test for Linearity . . . . .	47
1.8.4 A Heuristic Example to Illustrate the Smoothness of the GMM Estimator . . . . .	48
1.8.4.i The LSE . . . . .	49
1.8.4.ii The SLSE . . . . .	50
1.8.4.iii The GMM . . . . .	52
<b>2 Chapter 2: Monte Carlo Comparison for Nonparametric Threshold Estimators</b>	<b>55</b>
2.1 Introduction . . . . .	55
2.2 Three Nonparametric Threshold Estimators . . . . .	59
2.2.1 Semiparametric M-Estimator . . . . .	59

2.2.2	DKE and IDKE . . . . .	61
2.3	Estimation Difficulties in the Difference Kernel-Type Estimator with Near Boundary $\gamma_0$ . . . . .	64
2.4	Monte Carlo Designs . . . . .	67
2.5	Monte Carlo Results . . . . .	77
2.6	Conclusions . . . . .	82
<b>3</b>	<b>Can Exchange Rate Pass-Through Explain the Asymmetric Gasoline Puzzle? Evidence from a Pooled Panel Threshold Analysis of the EU</b>	<b>84</b>
3.1	Introduction . . . . .	84
3.2	Literature Review . . . . .	88
3.3	Data and Variables . . . . .	93
3.4	Econometric Framework . . . . .	99
3.4.1	The Baseline Linear Model . . . . .	99
3.4.2	The Threshold Model . . . . .	101
3.5	Results and Discussion . . . . .	104
3.5.1	Testing for Cross-section Dependence . . . . .	104
3.5.2	Unit Root and Cointegration Testing . . . . .	106
3.5.3	Empirical Findings . . . . .	107
3.5.3.i	Benchmark Model Results . . . . .	107
3.5.3.ii	Threshold Model Results . . . . .	110
3.5.3.iii	Asymmetric Testing . . . . .	115
3.6	Concluding Remarks . . . . .	122
	Appendices . . . . .	123
	<b>References</b>	<b>125</b>

## List of Tables

1.1	Simulation performance of the GMM and the smoothed least square estimators, $k_1=k_2=0$ (exogenous case) . . . . .	17
1.2	Simulation performance of the GMM and the smoothed least square estimators, $k_1=0.3, k_2=0$ . . . . .	18
1.3	Simulation performance of the GMM and the smoothed least square estimators, $k_1=0.5, k_2=0$ . . . . .	19
1.4	Simulation performance of the GMM and the smoothed least square estimators, $k_1=0, k_2=0.3$ . . . . .	20
1.5	Simulation Performance of the GMM and the smoothed least square estimators, $k_1=0, k_2=0.5$ . . . . .	21
1.6	Estimation and testing results of the linear index threshold solow growth model . . . . .	25
1.7	Data resource . . . . .	26
1.8	Rejection Probabilities of the Linearity Test for the GMM Estimator . . . . .	48
2.1	Simulation results of nonparametric threshold estimators, Data-generating Mechanism 1 (DGP 1). IDKE, integrated difference kernel estimator. . . . .	70
2.2	Simulation results of nonparametric threshold estimators, DGP 2. . . . .	70
2.3	Cont. . . . .	71
2.4	Simulation results of nonparametric threshold estimators, DGP 3. . . . .	72
2.5	Simulation results of nonparametric threshold estimators, DGP 4. . . . .	73
2.6	Simulation results of nonparametric threshold estimators, DGP 5. . . . .	74
2.7	Simulation results of nonparametric threshold estimators, DGP 6. . . . .	75
2.8	Simulation results of nonparametric threshold estimators, DGP 7. . . . .	76
2.9	Estimated convergence rate of the nonparametric threshold estimators. . . . .	77
3.1	Descriptive statistics . . . . .	98
3.2	Cross-dependence test . . . . .	105
3.3	Westerlund ECM panel cointegration tests . . . . .	107
3.4	Baseline and threshold model results . . . . .	112

3.5 Threshold endogenous test results . . . . . 113



## List of Figures

2.1	Average bias with $\gamma_0$ as various quantiles of the threshold variable, DGP 2, $n = 100$ . . . . .	80
2.2	Average bias with $\gamma_0$ as various quantiles of the threshold variable, DGP 2, $n = 300$ . . . . .	81
2.3	Average bias with $\gamma_0$ as various quantiles of the threshold variable, DGP 5, $n = 100$ . . . . .	81
2.4	Average bias with $\gamma_0$ as various quantiles of the threshold variable, DGP 5, $n = 300$ . . . . .	82
3.1	Impulse Response Functions (IRFs) under high regime . . . . .	118
3.2	Impulse Response Functions (IRFs) under low regime . . . . .	119
3.3	Sample coverage broken down by EU country and year . . . . .	124

# Chapter 1

## Chapter 1: A GMM Estimator for Linear Index Threshold Model

### 1.1 Introduction

Parametric threshold regression models are widely used to characterize nonlinearities in economic relationships. There are many applications of threshold regression model in both time series and cross-sectional scenario. Examples include the pricing asymmetry of oil prices and the nonlinear effect of public debt to GDP ratio on the per capita GDP growth. Threshold models allow us to identify the unknown threshold variable and draw inferences. It has been well established that the estimator of the threshold parameter is super-consistent, while the slope regressors estimators converge at the standard square-n rate. However, there are different approaches to obtain the asymptotic distribution of the

threshold parameter. Firstly, this can be derived using a "fixed threshold effect" assumption. Chan (1993) establishes that the threshold parameter estimator converges to a functional of a compound Poisson process. Yet, in that case statistical inference is impossible to implement in practice due to the presence of nuisance parameters in the joint distribution of the covariates. Secondly, using a "diminishing threshold effect" assumption, introduced by Hansen (2000), the limiting distribution involves two independent Brownian motions and is available through simulations. In that case, inference can be carried out fairly easily.

However, both Chan (1993) and Hansen (2000) rely on the crucial exogeneity assumption for both the slope regressors and the threshold variable. Recently, there is a growing interest in threshold models that allows for endogeneity. Using a two-stage least square method, Caner and Hansen (2004) allow for the slope regressors to be endogenous. In the spirit of the sample selection methodology of Heckman (1979), with a joint normality assumption, Kourtellos et al. (2016) allow for an endogenous threshold variable. Seo and Shin (2016) propose a two-step GMM estimator for the dynamic panel threshold model, which also allows for endogeneity. It is worth noticing that the GMM method allows for both fixed and small threshold effects and the rate of convergence for the GMM threshold estimator is not super-consistent. By relaxing the joint normality assumption of Kourtellos et al. (2016), Kourtellos et al. (2017) propose a two-step estimation method based on a nonparametric control function approach to correct for threshold endogeneity. The semi-parametric threshold model separates the threshold effect into two parts, namely, an ex-

ogeneous threshold effect and an endogenous threshold effect. Therefore, with a "small threshold" effect, the convergence rate of the estimator of the threshold variable depends on diminishing rates of both these two effects.

Seo and Linton (2007) consider a more general model than Hansen (2000). In the spirit of Horowitz (1992), they propose a smoothed least square estimator by allowing the threshold to be a linear index of regressors. The linear index threshold regression model can capture the joint threshold effect between two possible threshold variables. Therefore, this model allows empirical researchers to investigate the threshold effect in a broader setting. Yu and Fan (2019) develops the limiting asymptotic results of the least square estimator for the linear index threshold model in both the fixed threshold and the diminishing threshold effect framework. However, both the smoothed least square estimator and the least square estimator rely on the assumption of exogeneity in both the slope regressors and the threshold variables, which may limit the usefulness of these models.

In this paper, we propose a two-step GMM linear index threshold estimator, which allows both threshold variables and the regressors to be endogenous. We also relax the fixed threshold effect assumption by allowing both for fixed and diminishing threshold effects. We develop the estimation strategy and the limiting results for weakly dependent data. The asymptotic distribution is similar to Seo and Shin (2016). They concentrate on the dynamic panel threshold model, whereas we focus on the linear index threshold model. Similar to Seo and Shin (2016), the convergence rate of the threshold estimators

are  $n^{\frac{1}{2}-\alpha}$  and not super-consistent, where  $\alpha$  measures the diminishing rate of the threshold effect. The slope coefficients converge at the usual  $\sqrt{n}$  rate. We further suggest a test of the linear index threshold effect and provide a Hausman type test for the exogeneity of the regressors. The finite performance of the proposed estimator are studied through Monte Carlo Simulations. We compare our estimator with the smoothed least square estimator of Seo and Linton (2007) and we report the average bias, mean square error and the standard deviation of the threshold estimator specifically. Finally, we investigate the threshold effect of a linear index model between external debt and public debt in economic growth for developing countries. We estimate the augmented Solow linear index threshold using both GMM method and Seo and Linton (2007). We find that after correcting for endogeneity, the joint threshold effect becomes insignificant.

The rest of the paper is organized as follows. In section 2, we introduce the linear index threshold model with endogeneity. In section 3, we propose the two-step GMM estimator of the linear index threshold model and in section 4 we derive the asymptotic results. Section 5 provides the inference for the threshold effect and the endogeneity in slope regressors. In section 6 we report the Monte Carlo results for the proposed estimators. Section 7 presents the empirical application, while section 8 concludes the paper. In the appendix we collect the proofs and we present additional evidence for the small sample performance of the proposed linearity test and some additional heuristic arguments for the smoothness of the GMM objective function that we adopt in our analysis.

## 1.2 The Model

Consider the following linear index threshold model suggested by Seo and Linton (2007)

$$y_t = x_t^T \beta + \delta^T \tilde{x}_t I(q_{1t} + q_{2t}^T \psi > 0) + \varepsilon_t,$$

$$t = 1, \dots, n, \quad (1.1)$$

where  $y_t$  is the dependent variable,  $x_t$  is a  $k \times 1$  vector and  $\tilde{x}_t$  is an  $l \times 1$  vector. Also  $q_t = [q_{1t}^T, q_{2t}^T]^T$  is an  $h \times 1$  threshold variable vector. Note that  $x_t, \tilde{x}_t, q_t$  may have common variables. Many models in the previous literature can be viewed as a special case of this model. For example, for the case that  $x_t = \tilde{x}_t, q_{1t}$  is a constant and  $q_{2t}$  is a scalar, the model becomes the threshold model considered by Hansen (2000). If we further assume that  $x_t$  consists of the lagged  $y_t$  and  $q_{2t} = y_{t-d}$ , the model becomes the self-exciting threshold autoregressive (SETAR) model suggested by Tong and Lim (1980).

Similar to Seo and Shin (2016), we allow for both "fixed threshold effect" and the "diminishing threshold effect", of Hansen (2000). That is, we have

$$\delta = \delta_n = \delta_0 n^{-\alpha}, \quad \alpha \in [0, 1/2]. \quad (1.2)$$

Endogeneity is allowed in both the slope regressors ( $E(x_t \varepsilon_t) \neq 0$ ) and the threshold variables ( $E(q_t \varepsilon_t) \neq 0$ ). To fix the endogeneity problem, we need to find an  $m \times 1$  vector of instrumental variables,  $z_t$ , for  $t = 1, \dots, n$ , where  $m \geq k + l + h - 1$ , satisfying the

following orthogonality condition:

$$E(z_t \varepsilon_t) = 0, \quad (1.3)$$

for all  $t = 1, \dots, n$ .

### 1.3 Estimation Strategy

We consider the following moment condition:

$$E(g_t(\theta_n)) = E(z_t \varepsilon_t) = 0, \quad (1.4)$$

where  $\theta_n$  is the true value with  $\theta_n = [\beta_0^T, \delta_n^T, \psi_0^T]^T$ ,  $\delta_n = \delta_0 n^{-\alpha}$ , and

$$g_t(\theta) = z_t [y_t - x_t^T \beta - \delta^T \tilde{x}_t I(q_{1t} + q_{2t}^T \psi > 0)]. \quad (1.5)$$

Naturally, the sample analogue to  $E(g_t(\theta))$  is,

$$g_n(\theta) = \frac{1}{n} \sum_{t=1}^n g_t(\theta). \quad (1.6)$$

Given that the general identification condition hold for  $E(g_t(\theta_n))$ , the GMM estimators

can be obtained as

$$\hat{\theta}^{GMM} = \underset{\theta \in \Theta}{\operatorname{argmin}} Q_n(\theta), \quad (1.7)$$

where

$$Q_n(\theta) = g_n(\theta)^T W_n g_n(\theta) = \left[ \frac{1}{n} \sum_{t=1}^n g_t(\theta) \right]^T W_n \left[ \frac{1}{n} \sum_{t=1}^n g_t(\theta) \right] \quad (1.8)$$

and  $W_n$  is a positive definite matrix with  $W_n \xrightarrow{p} \Omega^{-1}$ , where  $\Omega = E(g_t(\theta_n)g_t(\theta_n)^T)$ .

For a given  $\psi$ , the model is linear in  $\beta$  and  $\delta$ . Since  $Q_n(\theta)$  is not continuous in  $\psi$ , it is more practical to use a grid search empirically.

For a given  $\psi$  and a weight matrix  $W_n$ ,

$$\left( \hat{\beta}_{(\psi)}^T, \hat{\delta}_{(\psi)}^T \right)^T = [\hat{G}(\psi)^T W_n \hat{G}(\psi)]^{-1} \hat{G}(\psi)^T W_n \left[ -\frac{1}{n} \sum z_t y_t \right], \quad (1.9)$$

where  $\hat{G}(\psi) = [\hat{G}_\beta^T, \hat{G}_\delta^T(\psi)]^T$ ,  $\hat{G}_\beta = -\frac{1}{n} \sum z_t x_t^T$  and  $\hat{G}_\delta(\psi) = -\frac{1}{n} \sum z_t \tilde{x}_t^T I(q_{1t} + q_{2t} \psi > 0)$ .

Then, the threshold index estimators can be obtained by

$$\hat{\psi}^{GMM} = \underset{\psi \in \Theta_\psi}{\operatorname{argmin}} Q_n(\psi) = \left[ \frac{1}{n} \sum_{t=1}^n g_t(\hat{\beta}_{(\psi)}, \hat{\delta}_{(\psi)}, \psi) \right]^T W_n \left[ \frac{1}{n} \sum_{t=1}^n g_t(\hat{\beta}_{(\psi)}, \hat{\delta}_{(\psi)}, \psi) \right] \quad (1.10)$$



and

$$(\hat{\beta}^{GMM^T}, \hat{\delta}^{GMM^T})^T = (\hat{\beta}^T(\hat{\psi}), \hat{\delta}^T(\hat{\psi}))^T. \quad (1.11)$$

Therefore, the 2-step method can be obtained as:

Step 1: Estimate the model with  $W_n = I_m$ , where  $I_m$  is an  $m \times m$  identity matrix, and get residual  $\hat{e}$ .

Step 2: Estimate the model with  $W_n = [\frac{1}{n} \sum_{t=1}^n (z_t z_t^T \hat{e}_t^2)]^{-1}$ .

## 1.4 Asymptotic Results

In this section, we develop the asymptotic theory for the GMM estimator of the linear index threshold model. The regularity assumptions required for deriving the limiting results of the proposed estimator are as follows:

Assumption 1:  $\{(X_t, z_t, q_t, \varepsilon_t)\}$  is a sequence of strictly stationary strong mixing random variables with mixing numbers  $\alpha_s$ ,  $s = 1, 2, \dots$  that satisfies  $\alpha_s = o(s^{-\gamma/(\gamma-1)})$  as  $s \rightarrow \infty$  for some  $\gamma \geq 1$ .

Assumption 2: For some  $\eta > 1$ ,  $E\|X_t X_t^T\|^{\eta+\gamma} < \infty$ ,  $E\|z_t \varepsilon_t\|^{\eta+\gamma} < \infty$ ,  $E\|z_t X_t^T\|^{\eta+\gamma} < \infty$ .  $E[z_t X_t^T X_t z_t^T | q_t] > 0$  a.s.

Assumption 3:  $E(\varepsilon_t | \mathcal{F}_{t-1}) = 0$ , where  $\mathcal{F}_{t-1}$  is the information set ( $\sigma$  field) of data realized by time  $t - 1$ ;  $E(z_t \varepsilon_t) = 0$ ,  $t = 1, 2, 3, \dots$ , and  $Var(n^{-1/2} \sum_{t=1}^n z_t \varepsilon_t)$  is a positive definite matrix.

Assumption 4: The true values of  $\beta$  and  $\psi$  are fixed at  $\beta_0$  and  $\psi_0$ . The true  $\delta$  depends on  $n$  such that  $\delta_n = \delta_0 n^{-\alpha}$  for some  $\alpha \in [0, \frac{1}{2})$  and  $\delta_0 \neq 0$ .  $\theta_n = [\beta_0^T, \delta_n^T, \psi_0^T]^T$  is an interior point of  $\Theta = \Theta_\beta \times \Theta_\delta \times \Theta_\psi$ , which is a compact set.  $\Omega = E(g_t(\theta_n)g_t(\theta_n)^T)$  is finite and positive definite.

Assumption 5:  $E(q_{2t}q_{2t}^T)$  is positive definite.

Assumption 6: For all  $\psi \in \Theta_\psi$ , the linear index of the threshold variables,  $v_t(\psi) = q_{1t} + q_{2t}^T \psi$ , has a continuous and bounded density,  $f_{v_t(\psi)}(\cdot)$ , such that  $f_{v_t(\psi)}(0) > 0$ ;  $E(z_t \delta_0^T \tilde{x}_t q_{2t}^T | v_t(\psi) = 0)$  is continuous at  $\psi_0$

Define:

$$\begin{aligned} G_\beta &= -E(z_t x_t^T), \\ G_\delta(\psi) &= -E\{z_t \tilde{x}_t^T I(q_{1t} + q_{2t}^T \psi > 0)\}, \\ G_\psi(\psi) &= -E\left(z_t \delta_0^T \tilde{x}_t q_{2t}^T | v_t(\psi) = 0\right) f_{v_t(\psi)}(0), \end{aligned} \tag{1.12}$$

where  $G_\beta$  is an  $m \times k$  matrix,  $G_\delta(\psi)$  is an  $m \times l$  matrix,  $G_\psi(\psi)$  is an  $m \times (h - 1)$  matrix and  $f_{v_t}(0)$  is the density for  $v_t$  at  $v_t = 0$ .

Assumption 7:  $G = (G_\beta, G_\delta(\psi_0), G_\psi(\psi_0))$ , then  $G$  is a full column rank matrix.

Assumption 1 gives standard conditions on the stochastic process. We can apply the generic uniform law of large numbers of Andrews (1987) to prove the consistency of our estimator. Assumptions 2-4 are regularity assumptions of the generalized method moment method. We assume  $\varepsilon_t$  is the martingale difference sequence. We allow both for fixed

threshold effect and for diminishing threshold effect. If  $\delta_0 = 0$  (no threshold effect),  $\psi$  can not be identified. Assumption 5 corresponds to Assumption 1(d) in Seo and Shin (2016) and Assumption 6 in Yu and Fan (2019). This assumption is required for the asymptotic uniqueness of the GMM estimator. Assumption 6 is a smoothness assumption on the distributions of the threshold variables and their linear index and the conditional moments, which is standard in threshold models. Assumption 7 is the GMM full rank condition .

**Theorem 1** *Under Assumptions 1-7, as  $n \rightarrow \infty$ , we have*

$$\hat{\theta}^{GMM} \xrightarrow{p} \theta_n. \quad (1.13)$$

**Theorem 2** *Under Assumptions 1-7, as  $n \rightarrow \infty$ ,*

$$\begin{bmatrix} \sqrt{n} & 0 & 0 \\ 0 & \sqrt{n} & 0 \\ 0 & 0 & n^{\frac{1}{2}-\alpha} \end{bmatrix} \begin{pmatrix} \hat{\beta} - \beta_0 \\ \hat{\delta} - \delta_n \\ \hat{\psi} - \psi_0 \end{pmatrix} \xrightarrow{d} N\left(0, \left[G^T \Omega^{-1} G\right]^{-1}\right), \quad (1.14)$$

where  $\Omega$  and  $G$  are defined in Assumption 4 and Assumption 7.

The convergence rate for the estimator of the slope parameter is standard  $\sqrt{n}$ . The convergence rate for the threshold variables depends on the unknown  $\alpha$ , which determines the decaying rate of the threshold effect. Intuitively, unlike the smoothed least square of Seo and Linton (2007), where the smoothness results from the objective function, the

smoothness of the GMM estimator relies on the nature of the sample averaging <sup>1</sup>.

$G_\beta$  and  $G_\delta$  can be estimated as

$$\hat{G}_\beta = -\frac{1}{n} \sum_{t=1}^n z_t x_t^T, \quad (1.15)$$

$$\hat{G}_\delta = -\frac{1}{n} \sum_{t=1}^n z_t \tilde{x}_t^T I(q_{1t} + q_{2t}^T \hat{\psi} > 0). \quad (1.16)$$

For  $G_\psi$ , we can estimate it using a standard Nadaraya-Watson kernel estimator,

$$\hat{G}_\psi = -\frac{1}{nh} \sum_{t=1}^n z_t \hat{\delta}_t^T \tilde{x}_t q_{2t}^T K\left(\frac{q_{1t} + q_{2t}^T \hat{\psi}}{b}\right), \quad (1.17)$$

where  $K(\cdot)$  is the second-order kernel function and  $b$  is the bandwidth.

Let  $\hat{\Omega} = \frac{1}{n} \sum_{t=1}^n g_t(\hat{\theta}) g_t^T(\hat{\theta})$ . As  $n \rightarrow \infty$ ,  $\hat{G}$  and  $\hat{\Omega}$  converge in probability respectively to  $G$  and  $\Omega$  following the uniform law of large number, the consistency of the Nadaraya-Watson estimator and the kernel density estimator for  $\alpha$  mixing data (Robinson (1983), Robinson (1986)).

---

<sup>1</sup>We provide a heuristic example in the appendix to explain the smoothness of the GMM and to compare the limiting behaviors among the least square estimator, the smoothed least square estimator, and the GMM estimator.

## 1.5 Testing

### 1.5.1 Test for Linearity

In equation (1.1), the threshold effect disappears under the null hypothesis,  $\delta_n = 0$ . However, due to the presence of unidentified parameters under the null, the natural way to test for nonlinearity is the *Sup*-Wald test, which is of the following form,

$$SupWald = \sup_{\psi \in \Theta_\psi} Wald(\psi), \quad (1.18)$$

where

$$Wald(\psi) = n(R[\hat{\beta}_{(\psi)}^T, \hat{\delta}_{(\psi)}^T]^T)^T [R(\hat{G}(\psi)^T \hat{\Omega}(\hat{\psi})^{-1} \hat{G}(\psi))^{-1} R^T]^{-1} R[\hat{\beta}_{(\psi)}^T, \hat{\delta}_{(\psi)}^T]^T,$$

$$\hat{G}(\psi) = [\hat{G}_\beta, \hat{G}_\delta(\psi)],$$

$$R = [0_{l \times k}, I_{l \times l}].$$

**Theorem 3** *Suppose that  $\inf_{\psi \in \Theta_\psi} |G(\psi)^T \Omega^{-1} G(\psi)|$  is positive, with Assumptions 1, 2, and 6 hold, under the null, we have*

$$\begin{aligned} SupWald &\xrightarrow{d} \sup_{\psi \in \Theta_\psi} V^T \Omega^{-1/2} G(\psi)^T (G(\psi)^T \Omega^{-1} G(\psi))^{-1} R^T (R(G(\psi)^T \Omega^{-1} G(\psi))^{-1} R^T)^{-1} \\ &\times R(G(\psi)^T \Omega^{-1} G(\psi))^{-1} G(\psi) \Omega^{-1/2} V, \end{aligned} \quad (1.19)$$

where  $V \sim N(0, I_l)$  and  $I_l$  is an  $l$  by  $l$  identity matrix.

Thus, the asymptotic distribution of *SupWald* is the supremum of the "chi-square" process and depends upon the covariance function. However, the critical value is non-tabulated in general. Following Hansen (1996), the asymptotic critical values and the  $p$  value can be generated by bootstrapping <sup>2</sup>.

### 1.5.2 Test for Exogeneity

In this section, extending the new Hausman-type test suggested by Kapetanios (2009), we propose a Hausman test to test for the exogeneity of the slope regressors of the linear index threshold model.

Consider the following null hypothesis, for all  $t$ ,

$$H_0 : E(\varepsilon_t | X_t) = 0 \quad (1.20)$$

Given a consistent threshold estimate  $\hat{\psi}$ , let  $\tilde{\theta}(\hat{\psi}) = [\tilde{\beta}(\hat{\psi})^T, \tilde{\delta}(\hat{\psi})^T]^T$  denote the slope estimator with moment condition  $E[X_t \varepsilon_t | \hat{\psi}] = 0$ , and  $\hat{\theta}(\hat{\psi}) = [\hat{\beta}(\hat{\psi})^T, \hat{\delta}(\hat{\psi})^T]^T$  denote the slope estimator with moment condition  $E[z_t \varepsilon_t | \hat{\psi}] = 0$ . Evidently, with conditional homoskedasticity, if there is no endogeneity in regressors, both estimators are consistent and  $\tilde{\theta}(\hat{\psi})$  is more efficient. However, if the slope regressors are endogenous, only  $\hat{\theta}(\hat{\psi})$  is consistent.

---

<sup>2</sup>In the appendix, we provide a small simulation to assess the finite sample performance of the suggested bootstrapping test. The parametric bootstrap algorithm is similar to Hansen (1996).

Therefore, the test statistic is of the form,

$$H = (\hat{\theta}(\hat{\psi}) - \tilde{\theta}(\hat{\psi}))^T [\text{Var}(\hat{\theta}(\hat{\psi})) - \text{Var}(\tilde{\theta}(\hat{\psi}))]^+ (\hat{\theta}(\hat{\psi}) - \tilde{\theta}(\hat{\psi})), \quad (1.21)$$

where ” + ” denotes the Moore-Penrose pseudoinverse.

**Theorem 4** *With assumptions 1-7 and the conditional homoskedasticity ( $E(\varepsilon_t^2 | \mathcal{F}_{t-1}) = \sigma_\varepsilon^2$ ), under the null hypothesis,*

$$H \xrightarrow{d} \chi_{k_1}^2, \quad (1.22)$$

where  $k_1 = \text{rank} \left( \text{Var}(\hat{\theta}(\hat{\psi})) - \text{Var}(\tilde{\theta}(\hat{\psi})) \right)$ .

As Kapetanios (2009) has shown, the asymptotic variances of the test statistic  $\sqrt{n}(\hat{\theta} - \tilde{\theta})$  may be problematic due to the nature of the nonlinear model. Therefore, following Kapetanios (2009), the properties of the asymptotic tests can be improved using bootstrapping.

## 1.6 Monte Carlo Simulation

In this section, we investigate the finite sample performance of the GMM estimator. We use the following structure to carry out the simulations:

$$\begin{aligned}
 y_t &= I(q_{1t} + q_{2t} \leq 0) + e_t, \\
 e_t &= 0.1\varepsilon_t + k_1v_{q_{1t}} + k_2v_{q_{2t}}, \\
 q_{1t} &= 0.5q_{1t-1} + v_{q_{1t}}, \\
 q_{2t} &= 0.5q_{2t-1} + v_{q_{2t}},
 \end{aligned} \tag{1.23}$$

where  $v_{q_{1t}}$ ,  $v_{q_{2t}}$  and  $\varepsilon_t$  are independently normally distributed with mean zero and variance one. We let  $q_{1t}$  and  $q_{2t}$  follow an  $AR(1)$  process,  $I(\cdot)$  is the indication function and  $\psi_0 = 1$ . The degree of endogeneity of the threshold variable is controlled by  $k_1$  and  $k_2$ . We use  $q_{1t-1}$  and  $q_{2t-1}$  as the instrument for  $q_{1t}$  and  $q_{2t}$  respectively <sup>3</sup>.

Clearly, this DGP is a simpler version of the general model,  $y_t = x_t^T\beta + \delta^T\tilde{x}_tI(q_{1t} + q_{2t}^T\psi > 0) + e_t$ , with  $\beta = 0$ ,  $\delta = 1$  and  $x_t = \tilde{x}_t = 1$  for all  $t = 1, 2, \dots$ . We estimate the model both with the GMM and the smoothed least square (LS) method of Seo and Linton (2007). For the smoothed LS, we use the same kernel function and the bandwidth choice with the simulations reported in Seo and Linton (2007). We use 2000 replications with sample sizes  $n = 100, 300$  and  $500$  respectively. To investigate the endogeneity in threshold variable, we

---

<sup>3</sup>We do not have weak instrument problem in this design since the  $AR$  coefficient is not a small number. We have an exact identification case.



vary  $k_1$  and  $k_2$  with values 0, 0.3 & 0.5. All simulations are executed in Matlab. For each simulation, we report the average MSE, Bias and the standard deviation of the threshold estimates. Tables 1.1 - 1.5 report the simulation results. Tables 2-3 reports the results with exogenous  $q_{2t}$  and endogenous  $q_{1t}$ . Finally, Tables 4-5 presents the results with exogenous  $q_{1t}$  and endogenous  $q_{2t}$ .

Table 1.1 shows the results with both exogenous threshold variables. For the linear threshold estimate  $\psi$ , the smoothed least square estimator achieves a better performance than the GMM estimator. This results from the super-consistency of the threshold estimate of the smoothed LS. Since the DGP is designed with a fixed threshold effect, the GMM estimator converges at the normal  $\sqrt{n}$  rate, which implies a slower convergence speed than smoothed LS estimator.

Tables 1.2-1.3 report the results with exogenous  $q_{2t}$  and endogenous  $q_{1t}$ . Tables 1.4-1.5 presents the results with exogenous  $q_{1t}$  and endogenous  $q_{2t}$ . Therefore, for both cases, as expected the smoothed least square has an asymptotic bias. The average biases of the smoothed LS estimator are much larger than the GMM estimators for all cases. In addition, the stronger the endogeneity, the larger the average bias. For the GMM estimator, as the sample size increases, all MSEs of the GMM estimator decrease and converge to zero confirming the consistency of the GMM estimator.

Table 1.1: Simulation performance of the GMM and the smoothed least square estimators,  $k_1=k_2=0$  (exogenous case)

MSE						
$n$	GMM			Smoothed LS		
	$\psi$	$\beta$	$\delta$	$\psi$	$\beta$	$\delta$
100	0.0279	0.0007	0.0019	0.0060	0.0009	0.0021
300	0.0070	0.0002	0.0006	0.0013	0.0002	0.0005
500	0.0030	0.0001	0.0003	0.0006	0.0001	0.0003

Bias						
$n$	GMM			Smoothed LS		
	$\psi$	$\beta$	$\delta$	$\psi$	$\beta$	$\delta$
100	-0.0360	0.0100	-0.0195	0.0005	0.0132	-0.0260
300	-0.0053	0.0058	-0.0121	0.0009	0.0057	-0.0115
500	-0.0028	0.0045	-0.0092	0.0006	0.0038	-0.0074

Standard Error						
$n$	GMM			Smoothed LS		
	$\psi$	$\beta$	$\delta$	$\psi$	$\beta$	$\delta$
100	0.1631	0.0255	0.0389	0.0776	0.0276	0.0382
300	0.0835	0.0127	0.0206	0.0363	0.0138	0.0197
500	0.0550	0.0089	0.0142	0.0252	0.0101	0.0144

This table reports the simulation results of the GMM estimator and the smoothed least square estimator for the DGP defined by equation (1.23) with exogenous threshold variables. The first column shows the sample size that the simulation used. The second to the fourth columns report the results of the GMM estimator for  $\psi$ ,  $\beta$  &  $\delta$  respectively. The fifth to the last column show the results of the smoothed LS estimator.

Table 1.2: Simulation performance of the GMM and the smoothed least square estimators,  $k_1=0.3, k_2=0$

MSE						
$n$	GMM			Smoothed LS		
	$\psi$	$\beta$	$\delta$	$\psi$	$\beta$	$\delta$
100	0.0747	0.0281	0.1099	0.1605	0.0269	0.1058
300	0.0209	0.0242	0.0961	0.0501	0.0237	0.0935
500	0.0089	0.0237	0.0940	0.0296	0.0228	0.0904

Bias						
$n$	GMM			Smoothed LS		
	$\psi$	$\beta$	$\delta$	$\psi$	$\beta$	$\delta$
100	0.0606	0.1604	-0.3228	0.3453	0.1586	-0.3192
300	0.0491	0.1538	-0.3079	0.2111	0.1520	-0.3038
500	0.0354	0.1530	-0.3054	0.1657	0.1501	-0.2995

Standard Error						
$n$	GMM			Smoothed LS		
	$\psi$	$\beta$	$\delta$	$\psi$	$\beta$	$\delta$
100	0.2666	0.0487	0.0755	0.2032	0.0415	0.0630
300	0.1360	0.0242	0.0361	0.0741	0.0238	0.0355
500	0.0874	0.0182	0.0267	0.0467	0.0177	0.0268

This table reports the simulation results of the GMM estimator and the smoothed least square estimator for the DGP defined by equation (1.23) with small endogenous effect from  $q_{1t}$ . The first column shows the sample size that the simulation used. The second to the fourth columns report the results of the GMM estimator for  $\psi$ ,  $\beta$  &  $\delta$  respectively. The fifth to the last column show the results of the smoothed LS estimator.

Table 1.3: Simulation performance of the GMM and the smoothed least square estimators,  $k_1=0.5, k_2=0$

MSE						
$n$	GMM			Smoothed LS		
	$\psi$	$\beta$	$\delta$	$\psi$	$\beta$	$\delta$
100	0.2606	0.0753	0.2971	4.9297	0.4247	1.7088
300	0.1087	0.0631	0.2505	4.7556	0.3820	1.5278
500	0.0759	0.0623	0.2491	4.6673	0.3666	1.4681

Bias						
$n$	GMM			Smoothed LS		
	$\psi$	$\beta$	$\delta$	$\psi$	$\beta$	$\delta$
100	0.1510	0.2588	-0.5192	-1.4815	0.5910	-1.1884
300	0.1882	0.2487	-0.4977	-1.3859	0.5549	-1.1114
500	0.1503	0.2480	-0.4975	-1.3964	0.5422	-1.0854

Standard Error						
$n$	GMM			Smoothed LS		
	$\psi$	$\beta$	$\delta$	$\psi$	$\beta$	$\delta$
100	0.4877	0.0914	0.1659	1.6541	0.2747	0.5447
300	0.2707	0.0360	0.0526	1.6841	0.2722	0.5410
500	0.2310	0.0277	0.0395	1.6488	0.2695	0.5386

This table reports the simulation results of the GMM estimator and the smoothed least square estimator for the DGP defined by equation (1.23) with large endogenous effect from  $q_{1t}$ . The first column shows the sample size that the simulation used. The second to the fourth columns report the results of the GMM estimator for  $\psi, \beta$  &  $\delta$  respectively. The fifth to the last column show the results of the smoothed LS estimator.

Table 1.4: Simulation performance of the GMM and the smoothed least square estimators,  $k_1=0$ ,  $k_2=0.3$

MSE						
$n$	GMM			Smoothed LS		
	$\psi$	$\beta$	$\delta$	$\psi$	$\beta$	$\delta$
100	0.0577	0.0283	0.1093	0.0739	0.0278	0.1071
300	0.0121	0.0243	0.0965	0.0324	0.0237	0.0939
500	0.0060	0.0235	0.0933	0.0212	0.0226	0.0897

Bias						
$n$	GMM			Smoothed LS		
	$\psi$	$\beta$	$\delta$	$\psi$	$\beta$	$\delta$
100	-0.1024	0.1622	-0.3240	-0.2476	0.1615	-0.3210
300	-0.0554	0.1539	-0.3086	-0.1728	0.1522	-0.3044
500	-0.0359	0.1521	-0.3043	-0.1417	0.1492	-0.2984

Standard Error						
$n$	GMM			Smoothed LS		
	$\psi$	$\beta$	$\delta$	$\psi$	$\beta$	$\delta$
100	0.2173	0.0449	0.0661	0.1122	0.0411	0.0636
300	0.0951	0.0238	0.0351	0.0504	0.0229	0.0348
500	0.0685	0.0188	0.0272	0.0342	0.0179	0.0268

This table reports the simulation results of the GMM estimator and the smoothed least square estimator for the DGP defined by equation (1.23) with small endogenous effect from  $q_{2t}$ . The first column shows the sample size that the simulation used. The second to the fourth columns report the results of the GMM estimator for  $\psi$ ,  $\beta$  &  $\delta$  respectively. The fifth to the last column show the results of the smoothed LS estimator.

Table 1.5: Simulation Performance of the GMM and the smoothed least square estimators,  $k_1=0$ ,  $k_2=0.5$

MSE						
$n$	GMM			Smoothed LS		
	$\psi$	$\beta$	$\delta$	$\psi$	$\beta$	$\delta$
100	0.1783	0.0660	0.2564	1.7022	0.0496	0.1871
300	0.0666	0.0632	0.2509	1.4350	0.0518	0.2024
500	0.0373	0.0630	0.2505	1.3642	0.0529	0.2090

Bias						
$n$	GMM			Smoothed LS		
	$\psi$	$\beta$	$\delta$	$\psi$	$\beta$	$\delta$
100	-0.2466	0.2484	-0.4966	-1.1965	0.2148	-0.4241
300	-0.1669	0.2487	-0.4981	-1.0766	0.2250	-0.4471
500	-0.1195	0.2496	-0.4989	-1.0327	0.2286	-0.4556

Standard Error						
$n$	GMM			Smoothed LS		
	$\psi$	$\beta$	$\delta$	$\psi$	$\beta$	$\delta$
100	0.3429	0.0658	0.0988	0.5204	0.0587	0.0852
300	0.1969	0.0360	0.0526	0.5255	0.0343	0.0495
500	0.1519	0.0277	0.0401	0.5458	0.0255	0.0372

This table reports the simulation results of the GMM estimator and the smoothed least square estimator for the DGP defined by equation (1.23) with large endogenous effect from  $q_{2t}$ . The first column shows the sample size that the simulation used. The second to the fourth columns report the results of the GMM estimator for  $\psi$ ,  $\beta$  &  $\delta$  respectively. The fifth to the last column show the results of the smoothed LS estimator.

## 1.7 Empirical Application

For many countries, especially certain advanced economies, public debt has been steadily increasing over the past decades, and there is a growing concern about its impact on long-term growth. Therefore, one of the most active areas of research recently has been to test whether debt has a nonlinear effect on growth. To investigate the potential threshold effect of public debt on growth, many researchers have carried out empirical studies to examine its magnitude of this effect and estimate the level beyond which debt will be detrimental to growth (threshold level of debt). By using Hansen (2000) threshold regression model, Cecchetti, M S and Fabrizio (2011), c? and Afonso and Jalles (2013) find that the public debt will have an adverse effect on economic growth when the public debt to GDP ratio exceeds 85%, 77%, and 59% respectively. By correcting for endogeneity in both slope regressors and the threshold variable with a structural threshold regression model of Kourtellos et al. (2016), Kourtellos et al. (2013) fail to find the significant threshold effect for the public debt.

However, the above findings ignore country heterogeneity. Moreover, all results in the literature estimate the threshold effect by assuming the nonlinearity exists only in public debt. For developing countries, it is natural to expect a threshold effect from external debt. For example, Poirson et al. (2002) show that there is a U shape relationship between external debt and growth in developing countries. In contrast to Poirson et al. (2002), Schclarek (2004) fails to detect any nonlinearity in foreign debt on growth for developing

countries.

One of the methodological problems in the past literature is that the model only allows for one threshold variable. Furthermore, most research relies on the homogeneity assumptions in both slope regressors and the threshold variable, which is highly dubious. It may be useful to conjecture that nonlinearity of growth in developing countries could originate from the joint linear threshold effect between both public debt and external debt. As such, we apply the linear index threshold model to investigate this issue. We examine the following linear index threshold Solow growth model:

$$g_t = x_t^T \beta + \delta^T x_t I(d_{1t} + d_{2t} \psi_1 + \psi_2 \leq 0) + \varepsilon_t, \quad (1.24)$$

where  $g_t$  is the growth rate,  $d_{1t}$  is the demeaned public debt to GDP ratio,  $d_{2t}$  is the demeaned external debt to GDP ratio,  $x$  is the Solow controlling set including constant & five Solow variables, namely, initial income per capita, schoolings, investment, population growth, and openness. It also includes public debt to GDP ratio and external debt to GDP ratio. A detailed data resource description of all variables is given in table 1.7. We also account for time fixed effects. We observe that, according to the heavily indebted poor countries (HIPC) initiative, 33 out of 37 HIPC in our dataset are from the Sub-Saharan African area. Therefore, we also include the regional effects with the Latin-American dummy and the Sub-Saharan dummy.

We employ an averaged ten-year period panel data covering 54 developing countries in 1980-1989, 1990- 1999, 2000-2009 and 2010-2016. The growth rate of real per capita



GDP is from PWT 9.0. The public debt and external debt to GDP ratio are from the IMF Historical Public Debt Database and the data bank of the world bank. In this paper, all variables are instrumented by their lagged values. We estimate the model using both smoothed least square method of Seo and Linton (2007) and our proposed GMM method. We test the nonlinearity by using the sup-wald statistic. As suggested by Hansen (1996), we use the bootstrap method to test for the existence of the threshold effect.

We present the results in Table 1.7. The smoothed least square estimate shows the presence of the significant threshold effect at 1% level with the bootstrap  $P$  value equaling 0.0001. It is worth noting that, with all else being equal, higher external debt leads to higher growth if the country is in the low debt regime and lower growth if the country is in the high debt regime. The finding supports for the inverted-U relationship of the external debt with growth. Furthermore, the positive effect of the external debt on growth in low debt regime is insignificant while the adverse impact in high debt regime is significant at 1% level.

Surprisingly, after correcting the endogeneity in both slope regressors and the threshold variables, the nonlinearity result of the GMM method becomes insignificant with the bootstrap  $P$  value equals 0.2727. Therefore, our finding suggests there is little evidence of nonlinearity in the effects of debt on growth and any finding to the contrary may be the result that the effects of possible self-selection or endogeneity by various countries is ignored in how they behave towards their debt obligations. The linear index threshold effect in external debt is found to be endogenous and the main reason of the heterogeneity in the

Table 1.6: Estimation and testing results of the linear index threshold solow growth model

Method	GMM-Index		Smoothed LS		Linear-GMM	Linear-LS
$\psi_1$	0.57		0.90			
$\psi_2$	0.24		0.12			
	Low	High	Low	High		
Constant	-0.0713 (0.8059)	-0.1609 (0.2534)	-0.0135 (0.8844)	0.0608 (0.5344)	-0.1639 (0.1763)	0.0695 (0.1019)
Initial income	-0.0009 (0.8742)	-0.0044 (0.4905)	-0.0017 (0.8234)	-0.0013 (0.8532)	-0.0036 (0.4571)	-0.0033 (0.2108)
Schooling	0.0027 (0.7179)	0.0013 (0.8526)	0.0002 (0.9835)	-0.0039 (0.6108)	0.0021 (0.7109)	0.001 (0.7669)
Investment	0.0101 (0.3219)	-0.0092** (0.0188)	-0.001 (0.9294)	0.0044 (0.4331)	-0.0013 (0.7442)	-0.0011 (0.7541)
Population	-0.0233 (0.8385)	-0.0757 (0.2019)	-0.016 (0.5920)	0.00253 (0.5281)	-0.0732 (0.1434)	0.0153 (0.2630)
Public debt	0.0110 (0.6240)	0.1320 (0.1587)	-0.0300** (0.0191)	-0.0231* (0.0998)	-0.0006 (0.9159)	-0.0068 (0.1172)
External debt	0.0265** (0.0452)	-0.0354 (0.2132)	0.0056 (0.7034)	-0.0899*** (0.0046)	-0.0076 (0.6059)	-0.0033 (0.7117)
Openness	0.0085 (0.2322)	0.0123** (0.0463)	-0.0028 (0.7577)	0.0166*** (0.0068)	0.0075* (0.0665)	0.0090*** (0.0050)
SupWald	29.6561		80.2709			
SupWald Boot <i>P</i> value	0.2727		0.0001***			
Observations	216		216		216	216

This table presents the estimation of the smoothed least square threshold index model of Seo and Linton (2007) and the GMM threshold index model. The first column shows the slope regressors. The second and third column give the results of the GMM method. The fourth and the fifth column report the the results of the smoothed least square method. The last two columns report the GMM and LS results that ignores the presence of a threshold. All variables are instrumented by the lag values. Time dummies and regional dummies are included but not reported. "\*\*\*\*" denotes significantly different from zero at the 1% level, "\*\*\*" denotes significantly different from zero at the 5% level, and "\*\*" denotes significantly different from zero at the 10% level.

debt-growth relationship is not the level of public debt and/or external debt.

Table 1.7: Data resource

Variables	Description
Initial income	Logarithm of real per capita GDP in chain series at 1980, 1990, 2000, 2010 Lagged values correspond to 1975, 1985, 1995 & 2005 Source: PWT 9.0
Population growth rates	Logarithm of average population growth rates plus 0.05 for the periods 1980-1989, 1990-1999, 2000-2009 & 2010-2016 Lagged values correspond to 1975-1979, 1985-1989, 1995-1999 & 2005-2009 Source: PWT 9.0
Investment	Logarithm of average ratios over each period of investment to real GDP per capita for the periods 1980-1989, 1990-1999, 2000-2009 & 2010-2016 Lagged values correspond to 1975-1979, 1985-1989, 1995-1999 & 2005-2009. Source: PWT 9.0
Schooling	Logarithm of average years of male secondary plus tertiary school attainment (25+) in 1980, 1990, 1999 & 2010. Lagged values correspond to 1975, 1985, 1995 & 2005 Source: Barro and Lee (2010).
Openness	Average ratios of exports plus imports to real GDP per capita 1980-1989, 1990-1999, 2000-2009 and 2010-2016, Lagged values correspond to 1975-1979, 1985- 1989, 1995-1999 and 2005-2009 Source: PWT 9.0
Public Debt	Logarithm of public debt to GDP for the periods 1980-1989, 1990-1999, 2000-2009 and 2010-2016, Lagged values correspond to 1975-1979, 1985- 1989, 1995-1999 and 2005-2009 Source: IMF Historical Public Debt Dataset
External Debt	Logarithm of external debt to GDP for the periods 1980-1989, 1990-1999, 2000-2009 and 2010-2016, Lagged values correspond to 1975-1979, 1985- 1989, 1995-1999 and 2005-2009 Source: World bank

## 1.8 Conclusion

In this paper, we propose a GMM estimator for the linear index threshold model. The GMM estimator allows for the endogeneity of the threshold variable as well as the slope regressors. We show the consistency of the GMM estimator and derive the limiting distribution. We study the finite sample performance of the proposed estimator through Monte Carlo simulation. We compare the performance of the GMM estimator with the smoothed least square estimator of Seo and Linton (2007) under both exogenous and endogenous threshold variable design. The simulation results are consistent with the theory. We use the linear index threshold model to investigate the threshold effect of the linear index combined by the public debt and the external debt on the economic growth in developing countries. The nonlinearity testing result of the GMM estimator shows the threshold effect is insignificant.

## Appendix

Throughout the proof, let  $\|\cdot\|$  denote the Euclidean norm. The integral is taken over  $(-\infty, \infty)$  unless specified otherwise. All limits are taken as  $n \rightarrow \infty$ .  $\xrightarrow{a.s.}$ ,  $\xrightarrow{p}$ , and  $\xrightarrow{d}$  denote almost sure convergence, convergence in probability, and convergence in distribution respectively.  $\wedge$  and  $\vee$  denote the minimum and maximum operators.

By definition, we have

$$g(\theta) = E(g_t(\theta)) = E[z_t y_t - z_t x_t^T \beta - z_t \delta^T \tilde{x}_t I(q_{1t} + q_{2t}^T \psi > 0)],$$

and the sample analogue is

$$g_n(\theta) = \frac{1}{n} \sum_{t=1}^n g_t(\theta) = \frac{1}{n} \sum_{t=1}^n [z_t y_t - z_t x_t^T \beta - z_t \delta^T \tilde{x}_t I(q_{1t} + q_{2t}^T \psi > 0)].$$

### 1.8.1 Proof of Lemma

**Lemma 1:** Under assumptions 1 and 2, it can be shown that:

$$\text{Sup}_{\psi \in \Theta_\psi} \left\| \frac{1}{n} \sum_{t=1}^n z_t \tilde{x}_t^T I(q_{1t} + q_{2t}^T \psi > 0) - E[z_t \tilde{x}_t^T I(q_{1t} + q_{2t}^T \psi > 0)] \right\| \xrightarrow{p} 0, \quad (1.8.25)$$

$$\text{Sup}_{\psi \in \Theta_\psi} \left\| \frac{1}{n} \sum_{t=1}^n x_t \tilde{x}_t^T I(q_{1t} + q_{2t}^T \psi > 0) - E[x_t \tilde{x}_t^T I(q_{1t} + q_{2t}^T \psi > 0)] \right\| \xrightarrow{p} 0. \quad (1.8.26)$$

Under Assumptions 1 and 2,  $E\|z_t \tilde{x}_t^T\|$  and  $E\|x_t \tilde{x}_t^T\|$  are bounded. Then, the proof is straightforward by applying Lemma 1 of Seo and Linton (2007).

**Lemma 2:** Under Assumptions 1, 2, and 6, there is a  $C < \infty$  such that for any  $\psi_1, \psi_2 \in \Theta_\psi$ , we have

$$\begin{aligned} \|E\left(X_t\left(I(\psi_1) - I(\psi_2)\right)\right)\| &\leq C\|\psi_1 - \psi_2\|, \\ \|E\left(X_t \varepsilon_t\left(I(\psi_1) - I(\psi_2)\right)\right)\| &\leq C\|\psi_1 - \psi_2\|, \end{aligned} \quad (1.8.27)$$

where  $I(\psi) = I(q_{1t} + q_{2t}^T \psi > 0)$ .

Note that, for any random variable  $w$ , we have

$$\frac{\partial E\left(wI(q_{1t} + q_{2t}^T \psi > 0)\right)}{\partial \psi_i} = E\left(wq_{2it} | v_t(\psi) = 0\right) f_{v_t(\psi)}(0), \quad (1.8.28)$$

where  $v_t$  defines in Assumption 7.

Thus, applying the first-order Taylor approximation, we have,

$$\begin{aligned} \|E\left(X_t\left(I(\psi_1) - I(\psi_2)\right)\right)\| &\leq \|E(X_t q_{2t}^T | v_t(\psi_2) = 0)\| f_{v_t(\psi_2)}(0) \|\psi_1 - \psi_2\| + O(1) \\ \|E\left(X_t \varepsilon_t\left(I(\psi_1) - I(\psi_2)\right)\right)\| &\leq \|E(X_t \varepsilon_t q_{2t}^T | v_t(\psi_2) = 0)\| f_{v_t(\psi_2)}(0) \|\psi_1 - \psi_2\| + O(1). \end{aligned} \quad (1.8.29)$$

Applying Assumptions 2 and 6, we can show that there exists a  $C$  such that  $\|E(X_t q_{2t} | v_t(\psi_2) = 0)\| f_{v_t(\psi_2)}(0) < C < \infty$  and  $\|E(X_t \varepsilon_t q_{2t}^T | v_t(\psi_2) = 0)\| f_{v_t(\psi_2)}(0) < C < \infty$ . This completes the proof of the Lemma.

## 1.8.2 Proof of Theorem

### Proof of Theorem 1:

First, under Assumptions 1, 2, and 3, by applying Lemma 1, we have

$$\begin{aligned} & \sup_{\beta \in \Theta_\beta, \delta \in \Theta_\delta, \psi \in \Theta_\psi} \left\| \frac{1}{n} \sum_{t=1}^n (z_t y_t - z_t x_t^T \beta - z_t \delta^T \tilde{x}_t I(q_{1t} + q_{2t}^T \psi > 0)) \right. \\ & \left. - E[(z_t y_t - z_t x_t^T \beta - z_t \delta^T \tilde{x}_t I(q_{1t} + q_{2t}^T \psi > 0))] \right\| \xrightarrow{p} 0, \end{aligned} \quad (1.8.30)$$

which is

$$\sup_{\theta \in \Theta} \|g_n(\theta) - E(g_t(\theta))\| \xrightarrow{p} 0. \quad (1.8.31)$$

Evidently,  $E(g_t(\theta))$  is continuous in  $\theta$ . Next, we show that  $E(g_t(\theta)) = 0$  iff  $\theta = \theta_n$ .

Note that, applying simple calculations gives

$$\begin{aligned} G_\beta &= -E(z_t x_t^T), \\ G_\delta(\psi) &= -E(z_t \{\tilde{x}_t I(q_{1t} + q_{2t}^T \psi > 0)\}^T), \\ G_\psi(\psi) &= -E\left(z_t \delta^T \tilde{x}_t q_{2t}^T | v_t(\psi) = 0\right) f_{v_t(\psi)}(0), \end{aligned} \quad (1.8.32)$$

Now, suppose  $\beta = \beta_0$ ,  $\delta = \delta_n$  but  $\psi \neq \psi_0$ , we have

$$\begin{aligned} E(g_t(\theta)) &= E(g_t(\beta_0, \delta_n, \psi)) = E(g_t(\beta_0, \delta_n, \psi)) - E(g_t(\beta_0, \delta_n, \psi_0)) \\ &= E\{z_t \tilde{x}_t^T [I(q_{1t} + q_{2t}^T \psi_0 > 0) - I(q_{1t} + q_{2t}^T \psi > 0)]\} \delta_n = [G_\delta(\psi) - G_\delta(\psi_0)] \delta_n. \end{aligned} \tag{1.8.33}$$

Let  $A = \{-q_{2t}^T \psi < q_{1t} < -q_{2t}^T \psi_0\} \cup \{-q_{2t}^T \psi_0 < q_{1t} < -q_{2t}^T \psi\}$ .

Under Assumptions 5 and 6, the set  $A$  has a positive probability. Therefore, we have

$$E[I(q_{1t} + q_{2t}^T \psi_0 > 0) - I(q_{1t} + q_{2t}^T \psi > 0) | A] \neq 0. \tag{1.8.34}$$

Under equation (1.8.34), by Assumption 2 and 4, we obtain

$$E[z_t \tilde{x}_t^T (I(q_{1t} + q_{2t}^T \psi_0 > 0) - I(q_{1t} + q_{2t}^T \psi > 0)) | A] \delta_n \neq 0,$$

which implies the unconditional expectation

$$E[z_t \tilde{x}_t^T (I(q_{1t} + q_{2t}^T \psi_0 > 0) - I(q_{1t} + q_{2t}^T \psi > 0))] \delta_n \neq 0. \tag{1.8.35}$$

If  $\beta \neq \beta_0$  or  $\delta \neq \delta_n$  but  $\psi = \psi_0$ , we have

$$\begin{aligned} E(g_t(\theta)) &= E(g_t(\beta, \delta_n, \psi_0)) = E(g_t(\beta, \delta_n, \psi_0)) - E(g_t(\beta_0, \delta_n, \psi_0)) \\ &= -E(z_t x_t^T)(\beta - \beta_0) = G_\beta(\beta - \beta_0) \neq 0, \end{aligned} \quad (1.8.36)$$

and

$$\begin{aligned} E(g_t(\theta)) &= E(g_t(\beta_0, \delta, \psi_0)) = E(g_t(\beta_0, \delta, \psi_0)) - E(g_t(\beta_0, \delta_n, \psi_0)) \\ &= -E(z_t \{\tilde{x}_t I(q_{1t} + q_{2t}^T \psi_0 > 0)\}^T)(\delta - \delta_n) = G_\delta(\psi_0)(\delta - \delta_n) \neq 0 \end{aligned} \quad (1.8.37)$$

where the inequality follows by Assumption 7.

If  $\beta \neq \beta_0$  or  $\delta \neq \delta_n$  and  $\psi \neq \psi_0$ , with almost same arguments, we have

$$\begin{aligned} E(g_t(\theta)) &= E(g_t(\beta, \delta, \psi)) = E(g_t(\beta, \delta, \psi)) - \underbrace{E(g_t(\beta_0, \delta_n, \psi_0))}_{E(g_t(\theta_n))=0} \\ &= [G_\delta(\psi) - G_\delta(\psi_0)]\delta_n + G_\beta(\beta - \beta_0) + G_\delta(\psi)(\delta - \delta_n) \neq 0. \end{aligned} \quad (1.8.38)$$

Hence, we obtain  $E(g_t(\theta)) = 0$  if and only if  $\theta = \theta_n$ . Therefore,  $Q(\theta) = E(g_t(\theta))^T W E(g_t(\theta))$

has a unique minimum at  $\theta = \theta_n$ , where  $W$  is a positive definite matrix.

Last, we show  $Q_n(\theta)$  converges uniformly in probability to  $Q(\theta)$ .



$$\begin{aligned}
\sup_{\theta \in \Theta} |Q_n(\theta) - Q(\theta)| &= \sup_{\theta \in \Theta} |g_n(\theta)^T W g_n(\theta) - g(\theta)^T W g(\theta)| \\
&= \sup_{\theta \in \Theta} |(g_n(\theta) - g(\theta))^T W (g_n(\theta) - g(\theta)) + 2(g_n(\theta) - g(\theta))^T W g(\theta)| \\
&\leq \sup_{\theta \in \Theta} \{ \|g_n(\theta) - g(\theta)\|^2 \|W\| + 2 \|g_n(\theta) - g(\theta)\| \|W\| \|g(\theta)\| \}. \tag{1.8.39}
\end{aligned}$$

Applying equation (1.8.31), we have  $\sup_{\theta \in \Theta} |Q_n(\theta) - Q(\theta)| \xrightarrow{p} 0$ , which completes our proof by following Theorem 2.1 of Newey and Mcfadden (1994).

### Proof of Theorem 2:

To derive the asymptotic normality with nonsmooth objective function, we follow Theorem 7.1 of Newey and Mcfadden (1994).

First, by central limit theorem (CLT), we have  $\sqrt{n}g_n(\theta_n) \xrightarrow{d} N(0, \Omega)$ , where  $\Omega = E(g_t(\theta_n)g_t(\theta_n)^T)$ .

Now, let  $k_n$  be a  $k + l + h - 1$  dimensional diagonal matrix whose first  $k + l$  diagonals are ones and the other element is  $n^\alpha$ ,  $W_n \xrightarrow{p} W = \Omega^{-1}$ , and

$$\begin{aligned}
D_n &= k_n^{-1} G^T W_n g_n(\theta_n), \\
H &= k_n^{-1} G^T W G k_n^{-1}, \\
R(\theta) &= \left( \frac{Q_n(\theta) - Q_n(\theta_n) - Q(\theta) - D_n^T(\theta - \theta_n)}{\|\theta - \theta_n\|} \right). \tag{1.8.40}
\end{aligned}$$

Next, we show the stochastic differentiability condition hold. That is, for any  $\gamma_n \rightarrow 0$ , we have

$$\mathit{Sup}_{\|\theta - \theta_n\| \leq \gamma_n} \left| \frac{\sqrt{n}R(\theta)}{1 + \sqrt{n}\|\theta - \theta_n\|} \right| = o_p(1). \quad (1.8.41)$$

Define

$$\varepsilon_n(\theta) = \frac{g_n(\theta) - g_n(\theta_n) - g(\theta)}{1 + \sqrt{n}\|\theta - \theta_n\|} \quad (1.8.42)$$

For  $\gamma_n \rightarrow 0$  and  $U = \{\|\theta - \theta_n\| \leq \gamma_n\}$ ,  $\mathit{Sup}_{\theta \in U} \{\sqrt{n}\|\varepsilon_n(\theta)\|\} \xrightarrow{p} o_p(1)$  if empirical process  $\sqrt{n}(g_n(\theta) - g(\theta))$  is stochastically equicontinuous. Note that  $g_t(\theta)$  is linear in  $\beta$  and  $\delta$ , which are bounded by the Assumption 4. Therefore, we only need to check the stochastic equicontinuity of the empirical process  $\frac{1}{\sqrt{n}} \sum_{t=1}^n [z_t \tilde{x}_t I(q_{1t} + q_{2t}^T \psi > 0) - E\{z_t \tilde{x}_t I(q_{1t} + q_{2t}^T \psi > 0)\}]$ . Let  $F = (\|z_t \tilde{x}_t\|) \sup_{\|\psi - \psi_0\| \leq \gamma_n} I(q_{1t} > -q_{2t}^T \psi \wedge -q_{2t}^T \psi_0)$  be the envelope function. Since the indicator functions of half intervals constitute a *type I* class or a Vapnik Chervonenkis (VC) class, by Assumptions 1, 2, and 4, the stochastic equicontinuity follows the Theorem 1 of Andrews (1994) and the Theorem 2.14.1 of Vaart and Wellner (2000). Evidently,  $\varepsilon_n(\theta_n) = 0$ .

Following proof of Theorem 7.2 of Newey and Mcfadden (1994), we decompose  $\left| \frac{\sqrt{n}R(\theta)}{1 + \sqrt{n}\|\theta - \theta_n\|} \right|$

into 5 terms,

$$\left| \frac{\sqrt{n}R(\theta)}{1 + \sqrt{n}\|\theta - \theta_n\|} \right| \leq \sum_{j=1}^5 r_{nj}(\theta), \quad (1.8.43)$$

where

$$r_{n1}(\theta) = \frac{\sqrt{n}(2\sqrt{n}\|\theta - \theta_n\| + \|\theta - \theta_n\|^2) \left| \varepsilon_n(\theta)^T W_n \varepsilon_n(\theta) \right|}{\|\theta - \theta_n\|(1 + \sqrt{n}\|\theta - \theta_n\|)}, \quad (1.8.44)$$

$$r_{n2}(\theta) = \frac{\sqrt{n} \left| [g(\theta) - Gk_n^{-1}(\theta - \theta_n)]^T W_n g_n(\theta_n) \right|}{\|\theta - \theta_n\|(1 + \sqrt{n}\|\theta - \theta_n\|)}, \quad (1.8.45)$$

$$r_{n3}(\theta) = \frac{n \left| [g(\theta) + g_n(\theta_n)]^T W_n \varepsilon_n(\theta) \right|}{(1 + \sqrt{n}\|\theta - \theta_n\|)}, \quad (1.8.46)$$

$$r_{n4}(\theta) = \frac{\sqrt{n} \left| g(\theta)^T W_n \varepsilon_n(\theta) \right|}{\|\theta - \theta_n\|}, \quad (1.8.47)$$

$$r_{n5}(\theta) = \frac{\sqrt{n} \left| g(\theta)^T [W_n - W] g(\theta) \right|}{\|\theta - \theta_n\|(1 + \sqrt{n}\|\theta - \theta_n\|)}. \quad (1.8.48)$$

By the consistency of  $\theta$  and  $\text{Sup}_{\theta \in U} \{\sqrt{n}\|\varepsilon_n(\theta)\|\} \xrightarrow{p} o_p(1)$ , we have

$$\text{Sup}_{\theta \in U} \{r_{n1}(\theta)\} = \text{Sup}_{\theta \in U} \left\{ \frac{(2 + \frac{\|\theta - \theta_n\|}{\sqrt{n}}) \left| (\sqrt{n}\varepsilon_n(\theta))^T W_n \sqrt{n}\varepsilon_n(\theta) \right|}{(1 + \sqrt{n}\|\theta - \theta_n\|)} \right\} = o_p(1). \quad (1.8.49)$$

Next, note that, by the differentiability of  $g(\theta)$ , we can show

$$\begin{aligned} \text{Sup}_{\theta \in U} \left\{ \frac{\|\sqrt{n}g(\theta)\|}{(1 + \sqrt{n}\|\theta - \theta_n\|)} \right\} &\leq \text{Sup}_{\theta \in U} \left\{ \frac{\|g(\theta)\|}{\|\theta - \theta_n\|} \right\} \\ &\leq \text{Sup}_{\theta \in U} \left\{ \frac{\|g(\theta_n) + Gk_n^{-1}(\theta - \theta_n) + o(\|\theta - \theta_n\|)\|}{\|\theta - \theta_n\|} \right\} = O(1), \end{aligned} \quad (1.8.50)$$

$$\begin{aligned} \text{Sup}_{\theta \in U} \left\{ \frac{\|g(\theta) - Gk_n^{-1}(\theta - \theta_n)\|}{\|\theta - \theta_n\|(1 + \sqrt{n}\|\theta - \theta_n\|)} \right\} &\leq \text{Sup}_{\theta \in U} \left\{ \frac{\|g(\theta) - Gk_n^{-1}(\theta - \theta_n)\|}{\|\theta - \theta_n\|} \right\} \\ &= \text{Sup}_{\theta \in U} \left\{ \frac{\|g(\theta) - g(\theta_n) - Gk_n^{-1}(\theta - \theta_n)\|}{\|\theta - \theta_n\|} \right\} = o(1) \end{aligned} \quad (1.8.51)$$

Therefore, by Cauchy-Schwarz inequality, we have

$$\text{Sup}_{\theta \in U} \{r_{n2}(\theta)\} = \text{Sup}_{\theta \in U} \left\{ \frac{|[g(\theta) - Gk_n^{-1}(\theta - \theta_n)]^T W_n \sqrt{n}g_n(\theta_n)|}{\|\theta - \theta_n\|(1 + \sqrt{n}\|\theta - \theta_n\|)} \right\} = o_p(1), \quad (1.8.52)$$

$$\text{Sup}_{\theta \in U} \{r_{n3}(\theta)\} \leq \text{Sup}_{\theta \in U} \left\{ \frac{\sqrt{n}\|g(\theta) + g_n(\theta_n)\| \|W_n\| \|\sqrt{n}\varepsilon_n(\theta)\|}{(1 + \sqrt{n}\|\theta - \theta_n\|)} \right\} = o_p(1) \quad (1.8.53)$$

$$\text{Sup}_{\theta \in U} \{r_{n4}(\theta)\} \leq \text{Sup}_{\theta \in U} \left\{ \frac{\|g(\theta)\|}{\|\theta - \theta_n\|} \|W_n\| \|\sqrt{n}\|\varepsilon_n(\theta)\| \right\} = o_p(1), \quad (1.8.54)$$

$$\text{Sup}_{\theta \in U} \{r_{n5}(\theta)\} \leq \text{Sup}_{\theta \in U} \left\{ \frac{\|g(\theta)\|}{\|\theta - \theta_n\|} \|W_n - W\| \frac{\|g(\theta)\|}{\|\theta - \theta_n\|} \right\} = o_p(1). \quad (1.8.55)$$

To sum up, we obtain

$$\text{Sup}_{\theta \in U} \left\{ \left| \frac{\sqrt{n}R(\theta)}{1 + \sqrt{n}\|\theta - \theta_n\|} \right| \right\} \leq \sum_{j=1}^5 \text{Sup}_{\theta \in U} \{r_{nj}(\theta)\} \leq \sum_{j=1}^5 o_p(1) = o_p(1). \quad (1.8.56)$$

Next, we show  $k_n^{-1}(\hat{\theta} - \theta_n) = O_p(n^{-1/2})$ , where  $\hat{\theta} = \underset{\theta \in \Theta}{\text{argmin}} Q_n(\theta)$ .

By Taylor expansion, we have

$$Q(\theta) = Q(\theta_n) + (\theta - \theta_n)^T H(\theta - \theta_n) + o(\|\theta - \theta_n\|^2), \quad (1.8.57)$$

where  $Q(\theta)$  achieves minimum at  $\theta = \theta_n$ , and  $H$  is positive definite.

Above implies we can find a constant  $C > 0$  such that

$$(\theta - \theta_n)^T H(\theta - \theta_n) + o(\|\theta - \theta_n\|^2) \geq C\|\theta - \theta_n\|^2 \geq \lambda_{\min}(H)\|\theta - \theta_n\|^2 \quad (1.8.58)$$

where  $\lambda_{\min}(H)$  is the smallest eigenvalue of  $H$ .

Therefore,

$$Q(\theta) - Q(\theta_n) \geq C\|\theta - \theta_n\|^2. \quad (1.8.59)$$

Since  $Q_n(\hat{\theta}) \leq \sup_{\theta \in \Theta} Q_n(\theta) + o_p(n^{-1})$ , we have

$$0 \geq Q_n(\hat{\theta}) - Q_n(\theta_n) - o_p(n^{-1}) = Q(\hat{\theta}) - Q(\theta_n) + D_n^T(\hat{\theta} - \theta_n) + \|\hat{\theta} - \theta_n\|R(\hat{\theta}) - o_p(n^{-1}). \quad (1.8.60)$$

For  $\theta \in U$ , we have  $|R(\theta)| = (1 + \sqrt{n}\|\theta - \theta_n\|)o_p(n^{-1/2})$ .

Therefore,

$$\begin{aligned} 0 &\geq C\|\hat{\theta} - \theta_n\|^2 + \|G^T W_n g_n(\theta_n)\| \|k_n^{-1}(\hat{\theta} - \theta_n)\| \\ &\quad - \|\hat{\theta} - \theta_n\|(1 + \sqrt{n}\|\hat{\theta} - \theta_n\|)o_p(n^{-1/2}) - o_p(n^{-1}) \end{aligned} \quad (1.8.61)$$

By the fact that  $G^T W_n g_n(\theta_n) \xrightarrow{p} O_p(n^{-1/2})$ , we have

$$\begin{aligned} 0 &\geq C\|\hat{\theta} - \theta_n\|^2 + O_p(n^{-1/2}) \|k_n^{-1}(\hat{\theta} - \theta_n)\| - \|\hat{\theta} - \theta_n\|o_p(n^{-1/2}) - \|\hat{\theta} - \theta_n\|^2 o_p(1) - o_p(n^{-1}) \\ &\geq [C + o_p(1)]\|\hat{\theta} - \theta_n\|^2 + O_p(n^{-1/2})\|k_n^{-1}(\hat{\theta} - \theta_n)\| - o_p(n^{-1}). \end{aligned} \quad (1.8.62)$$

Since  $C + o_p(1)$  is bounded away from zero, we can show

$$\|\hat{\theta} - \theta_n\|^2 + O_p(n^{-1/2})\|k_n^{-1}(\hat{\theta} - \theta_n)\| \leq o_p(n^{-1}). \quad (1.8.63)$$

Hence, we have

$$\begin{aligned} [ \|k_n^{-1}(\hat{\theta} - \theta_n)\| + O_p(n^{-1/2}) ]^2 &\leq \|(\hat{\theta} - \theta_n)\|^2 + O_p(n^{-1/2})\|k_n^{-1}(\hat{\theta} - \theta_n)\| + O_p(n^{-1}) \leq O_p(n^{-1}). \end{aligned} \quad (1.8.64)$$

Taking square root for both sides yields,

$$\left| \left| k_n^{-1}(\hat{\theta} - \theta_n) \right| + O_p(n^{-1/2}) \right| = O_p(n^{-1/2}). \quad (1.8.65)$$

Therefore, by the triangular inequality, we have

$$\left| \left| k_n^{-1}(\hat{\theta} - \theta_n) \right| \right| \leq \left| \left| k_n^{-1}(\hat{\theta} - \theta_n) \right| + O_p(n^{-1/2}) \right| + \left| -O_p(n^{-1/2}) \right| = O_p(n^{-1/2}), \quad (1.8.66)$$

which completes the proof that  $k_n^{-1}(\hat{\theta} - \theta_n)$  is  $\sqrt{n}$  consistent.

Next, let

$$\tilde{\theta} = \theta_n - [k_n^{-1}G^T W G k_n^{-1}]^{-1} (k_n^{-1}G^T W_n) g_n(\theta_n). \quad (1.8.67)$$

Therefore, with  $W = \Omega^{-1}$  and  $W_n \xrightarrow{p} W$ , we have

$$\sqrt{n}[k_n^{-1}(\tilde{\theta} - \theta_n)] = -[G^T W G]^{-1} (G^T W_n) \sqrt{n} g_n(\theta_n) \xrightarrow{d} N(0, [G^T \Omega^{-1} G]^{-1}). \quad (1.8.68)$$

Since

$$Q_n(\theta) - Q_n(\theta_n) \approx 2D_n^T(\theta - \theta_n) + (\theta - \theta_n)^T H(\theta - \theta_n), \quad (1.8.69)$$

we have

$$Q_n(\theta) - Q_n(\theta_n) = (\theta - \theta_n)^T k_n^{-1} G^T W G k_n^{-1} (\theta - \theta_n) + 2(\theta - \theta_n) k_n^{-1} (G^T W_n) g_n(\theta_n) + o_p(n^{-1}). \quad (1.8.70)$$

By the definition of  $\tilde{\theta}$ , we have

$$-(G^T W G) [k_n^{-1} (\tilde{\theta} - \theta_n)] = (G^T W_n) g_n(\theta_n). \quad (1.8.71)$$

Therefore, we obtain

$$Q_n(\theta) - Q_n(\theta_n) = (\theta - \theta_n)^T k_n^{-1} G^T W G k_n^{-1} (\theta - \theta_n) - 2(\theta - \theta_n) k_n^{-1} G^T W G k_n^{-1} (\tilde{\theta} - \theta_n) + o_p(n^{-1}). \quad (1.8.72)$$

Similarly, we can show

$$\begin{aligned} Q_n(\tilde{\theta}) - Q_n(\theta_n) &= (\tilde{\theta} - \theta_n)^T k_n^{-1} G^T W G k_n^{-1} (\tilde{\theta} - \theta_n) + 2(\tilde{\theta} - \theta_n)^T k_n^{-1} (G^T W_n) g_n(\theta_n) + o_p(n^{-1}) \\ &= (\tilde{\theta} - \theta_n)^T k_n^{-1} G^T W G k_n^{-1} (\tilde{\theta} - \theta_n) - 2(\tilde{\theta} - \theta_n)^T k_n^{-1} G^T W G k_n^{-1} (\tilde{\theta} - \theta_n) + o_p(n^{-1}) \\ &= -(\tilde{\theta} - \theta_n)^T k_n^{-1} G^T W G k_n^{-1} (\tilde{\theta} - \theta_n) + o_p(n^{-1}). \end{aligned} \quad (1.8.73)$$

Since  $\tilde{\theta} \in \Theta$ , we have

$$Q_n(\tilde{\theta}) - Q_n(\hat{\theta}) = Q_n(\tilde{\theta}) - Q_n(\theta_n) - (Q_n(\hat{\theta}) - Q_n(\theta_n)) = o_p(n^{-1}). \quad (1.8.74)$$



Thus,

$$\begin{aligned}
& -(\tilde{\theta} - \theta_n)^T k_n^{-1} G^T W G k_n^{-1} (\tilde{\theta} - \theta_n) - (\hat{\theta} - \theta_n)^T k_n^{-1} G^T W G k_n^{-1} (\hat{\theta} - \theta_n) \\
& + 2(\hat{\theta} - \theta_n)^T k_n^{-1} G^T W G k_n^{-1} (\tilde{\theta} - \theta_n) = o_p(n^{-1}),
\end{aligned} \tag{1.8.75}$$

which implies

$$-(k_n^{-1}(\hat{\theta} - \theta_n) - k_n^{-1}(\tilde{\theta} - \theta_n))^T (k_n^{-1} G^T W G k_n^{-1}) (k_n^{-1}(\hat{\theta} - \theta_n) - k_n^{-1}(\tilde{\theta} - \theta_n)) = o_p(n^{-1}). \tag{1.8.76}$$

Note that  $G^T W G$  is positive definite, therefore, we can find a constant  $C \geq 0$  such that

$$-C \|k_n^{-1}(\hat{\theta} - \theta_n) - k_n^{-1}(\tilde{\theta} - \theta_n)\|^2 = o_p(n^{-1}). \tag{1.8.77}$$

Hence,

$$\|k_n^{-1}(\hat{\theta} - \theta_n) - k_n^{-1}(\tilde{\theta} - \theta_n)\| = o_p(n^{-1/2}), \tag{1.8.78}$$

which implies

$$\sqrt{n} k_n^{-1} \|\hat{\theta} - \tilde{\theta}\| \xrightarrow{p} 0. \tag{1.8.79}$$

Following

$$\sqrt{nk_n^{-1}}(\tilde{\theta} - \theta_n) \xrightarrow{d} N(0, (G^T \Omega^{-1} G)^{-1}), \quad (1.8.80)$$

we obtain

$$\sqrt{nk_n^{-1}}(\hat{\theta} - \theta_n) \xrightarrow{d} N(0, (G^T \Omega^{-1} G)^{-1}), \quad (1.8.81)$$

which completes the proof of this Theorem.

**Proof of Theorem 3:**

For a fixed  $\psi \in \Theta_\psi$ ,

$$\begin{aligned} & \begin{pmatrix} \hat{\beta}(\psi) - \beta_0 \\ \hat{\delta}(\psi) - \delta_n \end{pmatrix} \\ &= \left( \hat{G}(\psi)^T \hat{\Omega}(\hat{\psi})^{-1} \hat{G}(\psi) \right)^{-1} \hat{G}(\psi)^T \hat{\Omega}(\hat{\psi})^{-1} \left( g_n(\theta_n) + \frac{1}{n} \sum_{t=1}^n z_t \delta_n^T \tilde{x}_t (I(\psi_0) - I(\psi)) \right). \end{aligned} \quad (1.8.82)$$

Under the null,  $\delta_n = 0$ , we have

$$\hat{\delta}(\psi) = R \left( \hat{G}(\psi)^T \hat{\Omega}(\hat{\psi})^{-1} \hat{G}(\psi) \right)^{-1} \hat{G}(\psi)^T \hat{\Omega}(\hat{\psi})^{-1} g_n(\theta_n), \quad (1.8.83)$$

where  $R$  is defined in (1.19).

First, by applying lemma 1, it is straightforward to show that  $\hat{G}(\psi) \xrightarrow{p} G(\psi)$  uniformly in  $\psi \in \Theta_\psi$ .

Next, we show that  $\hat{\Omega}(\hat{\psi}) \xrightarrow{p} \Omega$ .

Simple calculation shows

$$\hat{\varepsilon}_t = \varepsilon_t - \left( \hat{\beta} - \beta_0 \right)^T x_t - \delta_n^T \tilde{x}_t \left( I(\hat{\psi}) - I(\psi_0) \right) - \left( \hat{\delta} - \delta_n \right)^T \tilde{x}_t I(\hat{\psi}). \quad (1.8.84)$$

Hence,

$$\begin{aligned} \hat{\Omega}(\hat{\psi}) &= \frac{1}{n} \sum_{t=1}^n z_t z_t^T \varepsilon_t^2 \\ &= -\frac{2}{n} \sum_{t=1}^n z_t z_t^T \varepsilon_t x_t^T \left( \hat{\beta} - \beta_0 \right) \\ &\quad - \frac{2}{n} \sum_{t=1}^n z_t z_t^T \varepsilon_t \tilde{x}_t^T I(\hat{\psi}) \left( \hat{\delta} - \delta_n \right) \\ &\quad - \frac{2}{n} \sum_{t=1}^n z_t z_t^T \varepsilon_t \delta_n^T \tilde{x}_t \left( I(\hat{\psi}) - I(\psi_0) \right) \\ &\quad + \frac{1}{n} \sum_{t=1}^n z_t z_t^T \left( \hat{\beta} - \beta_0 \right)^T x_t x_t^T \left( \hat{\beta} - \beta_0 \right) \\ &\quad + \frac{1}{n} \sum_{t=1}^n z_t z_t^T \delta_n^T \tilde{x}_t \tilde{x}_t^T \delta_n \left( I(\hat{\psi}) - I(\psi_0) \right) \\ &\quad + \frac{1}{n} \sum_{t=1}^n z_t z_t^T \left( \hat{\delta} - \delta_n \right)^T \tilde{x}_t \tilde{x}_t^T \left( \hat{\delta} - \delta_n \right) I(\hat{\psi}). \end{aligned} \quad (1.8.85)$$

For the first term, by the boundedness assumption and the consistency of  $\hat{\beta}$ , we have

$$\frac{1}{n} \left\| \sum_{t=1}^n z_t z_t^T \varepsilon_t x_t^T \left( \hat{\beta} - \beta_0 \right) \right\| \leq \frac{1}{n} \sum_{t=1}^n \|z_t\|^2 \|\varepsilon_t\| \|x_t\| \|\hat{\beta} - \beta_0\| \xrightarrow{p} 0. \quad (1.8.86)$$

Similarly, we can show

$$\begin{aligned}
\frac{1}{n} \left\| \sum_{t=1}^n z_t z_t^T \varepsilon_t \tilde{x}_t^T I(\hat{\psi}) (\hat{\delta} - \delta_n) \right\| &\leq \frac{1}{n} \sum_{t=1}^n \|z_t\|^2 \|\varepsilon_t\| \|\tilde{x}_t\| \|\hat{\delta} - \delta_n\| \xrightarrow{p} 0 \\
\frac{1}{n} \left\| \sum_{t=1}^n z_t z_t^T (\hat{\beta} - \beta_0)^T x_t x_t^T (\hat{\beta} - \beta_0) \right\| &\leq \frac{1}{n} \sum_{t=1}^n \|z_t\|^2 \|x_t\|^2 \|\hat{\beta} - \beta_0\|^2 \xrightarrow{p} 0 \\
\frac{1}{n} \left\| \sum_{t=1}^n z_t z_t^T (\hat{\delta} - \delta_n)^T \tilde{x}_t \tilde{x}_t^T (\hat{\delta} - \delta_n) I(\hat{\psi}) \right\| &\leq \frac{1}{n} \sum_{t=1}^n \|z_t\|^2 \|\tilde{x}_t\| \|\hat{\delta} - \delta_n\|^2 \xrightarrow{p} 0.
\end{aligned} \tag{1.8.87}$$

Next, under Assumptions 1, 2, and 3 and applying Lemma 1, we obtain

$$\left\| \frac{1}{n} \sum_{t=1}^n z_t z_t^T \varepsilon_t \delta_n^T \tilde{x}_t \left( I(\hat{\psi}) - I(\psi_0) \right) - E \left( z_t z_t^T \varepsilon_t \delta_n^T \tilde{x}_t \left( I(\hat{\psi}) - I(\psi_0) \right) \right) \right\| \xrightarrow{p} 0. \tag{1.8.88}$$

Applying Lemma 2, we have,

$$E \left( \left\| z_t z_t^T \varepsilon_t \delta_n^T \tilde{x}_t \left( I(\hat{\psi}) - I(\psi_0) \right) \right\| \right) \leq C \|\hat{\psi} - \psi_0\| \xrightarrow{p} 0, \tag{1.8.89}$$

where  $C < \infty$  and  $\|\hat{\psi} - \psi_0\| \xrightarrow{p} 0$ .

Similarly, we can show,

$$\begin{aligned}
\frac{1}{n} \left\| \sum_{t=1}^n z_t z_t^T \delta_n^T \tilde{x}_t \tilde{x}_t^T \delta_n \left( I(\hat{\psi}) - I(\psi_0) \right) - E \left( z_t z_t^T \delta_n^T \tilde{x}_t \tilde{x}_t^T \delta_n \left( I(\hat{\psi}) - I(\psi_0) \right) \right) \right\| &\xrightarrow{p} 0 \\
E \left( \left\| z_t z_t^T \delta_n^T \tilde{x}_t \tilde{x}_t^T \delta_n \left( I(\hat{\psi}) - I(\psi_0) \right) \right\| \right) &\xrightarrow{p} 0
\end{aligned} \tag{1.8.90}$$

Note that  $\frac{1}{n} \sum_{t=1}^n z_t z_t^T \varepsilon_t^2 \xrightarrow{a.s.} \Omega$ . To sum up, we have  $\hat{\Omega}(\hat{\psi}) \xrightarrow{p} \Omega$ . Then, applying the

continuous mapping theorem on equation (1.8.83), we have,

$$\sqrt{n}\hat{\delta}(\psi) \xrightarrow{d} R(G(\psi)^T \Omega^{-1} G(\psi))^{-1} G(\psi)^T \Omega^{-1/2} V, \quad (1.8.91)$$

where  $R$  and  $V$  are defined in (1.19). This completes the proof of this Theorem.

#### Proof of Theorem 4

Let  $\theta_0 = [\beta_0^T, \delta_n^T]^T$ . By Theorem 2, we have

$$\sqrt{n}(\hat{\theta}(\hat{\psi}) - \theta_0) - \sqrt{n}(\hat{\theta}(\psi_0) - \theta_0) = o_p(1), \quad (1.8.92)$$

$$\sqrt{n}(\bar{\theta}(\hat{\psi}) - \theta_0) - \sqrt{n}(\bar{\theta}(\psi_0) - \theta_0) = o_p(1). \quad (1.8.93)$$

For the purposes of this Theorem, we assume knowledge of  $\psi_0$ . Therefore, we can show that

$$\begin{aligned} & \begin{pmatrix} \sqrt{n}(\hat{\theta}(\psi_0) - \theta_0) \\ \sqrt{n}(\bar{\theta}(\psi_0) - \theta_0) \end{pmatrix} \\ &= \begin{bmatrix} \left( \hat{G}(\psi_0)^T \hat{\Omega}(\psi_0)^{-1} \hat{G}(\psi_0) \right)^{-1} \hat{G}(\psi_0)^T \hat{\Omega}(\psi_0)^{-1} & 0 \\ 0 & \left( \tilde{G}(\psi_0)^T \tilde{\Omega}(\psi_0)^{-1} \tilde{G}(\psi_0) \right)^{-1} \tilde{G}(\psi_0)^T \tilde{\Omega}(\psi_0)^{-1} \end{bmatrix} \\ & \times \begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{i=1}^n z_t \varepsilon_t \\ \frac{1}{\sqrt{n}} \sum_{i=1}^n x_t \varepsilon_t \end{pmatrix}, \end{aligned} \quad (1.8.94)$$

where

$$\tilde{G}(\psi_0) = \left[ -\frac{1}{n} \sum_{t=1}^n x_t x_t^T, -\frac{1}{n} \sum_{t=1}^n x_t \tilde{x}_t^T I(\psi_0) \right], \quad (1.8.95)$$

$$\tilde{\Omega}(\psi_0) = \frac{1}{n} \sum_{i=1}^n x_t x_t^T \tilde{\varepsilon}_t,$$

$$\tilde{\varepsilon}_t = y_t - \tilde{\theta}(\psi_0)^T (x_t, \tilde{x}_t^T I(\psi_0)).$$

Following the proof of Lemma 1 of Kapetanios (2009), we can show

$$\begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{i=1}^n z_t \varepsilon_t \\ \frac{1}{\sqrt{n}} \sum_{i=1}^n x_t \varepsilon_t \end{pmatrix} \xrightarrow{d} N \left( 0, \begin{pmatrix} \Omega & \Omega_{zx} \\ \Omega_{xy} & \Omega_{xx} \end{pmatrix} \right), \quad (1.8.96)$$

where  $\Omega = E(z_t z_t^T \varepsilon_t^2)$ ,  $\Omega_{zx} = E(z_t x_t^T \varepsilon_t^2)$ ,  $\Omega_{xy} = E(x_t z_t^T \varepsilon_t^2)$ , and  $\Omega_{xx} = E(x_t x_t^T \varepsilon_t^2)$ .

Next, similar to the proof of  $\hat{\Omega}(\hat{\psi}) \xrightarrow{p} \Omega$ , it is straightforward to show  $\hat{\Omega}(\psi_0) \xrightarrow{p} \Omega$  and  $\tilde{\Omega}(\psi_0) \xrightarrow{p} \Omega_{xx}$ . By Lemma 1, we have  $\hat{G}(\psi_0) \xrightarrow{p} G(\psi_0)$  and  $\tilde{G}(\psi_0) \xrightarrow{p} G_{xx}(\psi_0)$ , where  $G_{xx}(\psi_0) = [-E(x_t x_t^T), -E(x_t \tilde{x}_t^T I(\psi_0))]$ .

Therefore, under the null hypothesis of no endogeneity in regressors, we have

$$\begin{pmatrix} \sqrt{n}(\hat{\theta}(\psi_0) - \theta_0) \\ \sqrt{n}(\tilde{\theta}(\psi_0) - \theta_0) \end{pmatrix} \xrightarrow{d} N \left( 0, \begin{bmatrix} \Upsilon_{zx} & \Upsilon_{zx} G(\psi_0)^T \Omega^{-1} \Omega_{zx} \Omega_{xx}^{-1} G_{xx}(\psi_0) \Upsilon_{xx} \\ \Upsilon_{xx} G_{xx}(\psi_0)^T \Omega_{xx}^{-1} \Omega_{zx} \Omega^{-1} G(\psi_0) \Upsilon_{zx} & \Upsilon_{xx} \end{bmatrix} \right) \quad (1.8.97)$$

where  $\Upsilon_{zx} = \left(G(\psi_0)^T \Omega^{-1} G(\psi_0)\right)^{-1}$  and  $\Upsilon_{xx} = \left(G_{xx}(\psi_0)^T \Omega_{xx}^{-1} G_{xx}(\psi_0)\right)^{-1}$ .

This implies

$$\sqrt{n} \left( \hat{\theta}(\psi_0) - \tilde{\theta}(\psi_0) \right) \xrightarrow{d} N \left( 0, V \right), \quad (1.8.98)$$

where

$$V = \Upsilon_{zx} + \Upsilon_{xx} - \Upsilon_{zx} G(\psi_0)^T \Omega^{-1} \Omega_{zx} \Omega_{xx}^{-1} G_{xx}(\psi_0) \Upsilon_{xx} - \Upsilon_{xx} G_{xx}(\psi_0)^T \Omega_{xx}^{-1} \Omega_{zx} \Omega^{-1} G(\psi_0) \Upsilon_{zx} \quad (1.8.99)$$

$$= \text{Var}(\hat{\theta}(\psi_0)) + \text{Var}(\tilde{\theta}(\psi_0)) - 2\text{Cov}(\hat{\theta}(\psi_0), \tilde{\theta}(\psi_0)).$$

Evidently, with conditional homoskedasticity,

$$\begin{aligned} \Upsilon_{xx} &= \sigma_\varepsilon^2 G_{xx}(\psi_0)^{-1} \\ \Upsilon_{zx} &= \sigma_\varepsilon^2 \left( G(\psi_0)_{xx}^T (\psi_0)^{-1} G(\psi_0) \right)^{-1}, \end{aligned} \quad (1.8.100)$$

which implies  $\tilde{\theta}(\psi_0)$  is more efficient. Following Hausman (1978),  $V = \text{Var}(\hat{\theta}(\psi_0)) - \text{Var}(\tilde{\theta}(\psi_0))$ . This completes our proof of this Theorem.

### 1.8.3 Finite Sample Performance of the Test for Linearity

To assess the finite sample performance of the linearity test, we use a model similar to (1.23),

$$y_t = bI(q_{1t} + q_{2t} \leq 0) + \varepsilon_t, \quad (1.8.101)$$

$$q_{1t} = 0.5q_{1t-1} + v_{q1t},$$

$$q_{2t} = 0.5q_{2t-1} + v_{q2t},$$

where  $v_{q1t}$ ,  $v_{q2t}$  and  $\varepsilon_t$  are independently normally distributed with mean zero and variance one.

The simulations are done for five sample sizes,  $n = 50$ ,  $n = 100$ ,  $n = 200$ ,  $n = 300$ ,  $n = 500$ , and five threshold effects,  $b = 0$ ,  $b = 0.2$ ,  $b = 0.5$ ,  $b = 0.8$ ,  $b = 1$ . We report the results in Table 8. The replication number is 2000. Throughout the analysis, we use a significance level of 5%. As expected, size is approaching to 5% as sample size increases. Power is increasing in  $b$ , and increasing in  $n$ .



Table 1.8: Rejection Probabilities of the Linearity Test for the GMM Estimator

	Sample Size				
	$n = 50$	$n = 100$	$n = 200$	$n = 300$	$n = 500$
$b = 0$	0.0955	0.077	0.0705	0.0645	0.0565
$b = 0.2$	0.088	0.1304	0.1959	0.2749	0.4313
$b = 0.5$	0.2699	0.5112	0.8446	0.9615	0.999
$b = 0.8$	0.5972	0.9115	0.9975	0.9995	0.9995
$b = 1$	0.8186	0.9945	0.9995	0.9995	0.9995

This table presents the rejection rate of the linearity test for the GMM estimator. The first column gives the different settings of the sample splittings. With  $b = 0$ , there is no threshold effect. Higher value of  $b$  gives higher degree of the threshold effect.

## 1.8.4 A Heuristic Example to Illustrate the Smoothness of the GMM Estimator

To provide more intuition for the Theorem 1.14, we use a simple example to explain the smoothness of the GMM and how the smoothness determines the asymptotic normality. Furthermore, this section also aims to provide some background on the different asymptotic forms of the least square estimator (LSE), the smoothed least square estimator (SLSE), and the GMM estimator (GMM). The model considered is defined as follows,

$$y_i = I(q_{1i} + q_{2i}\psi_0 \leq 0) + \varepsilon_i,$$

where  $q_{1i}, q_{2i} \sim U[0, 1]$ , and  $\varepsilon_i \sim N(0, \sigma^2)$ .

Hence, in this example, we assume all threshold variables are exogenous, and the threshold effect is fixed.

### 1.8.4.i The LSE

As shown in Yu (2015), the LSE can be obtained as,

$$\hat{\psi} = \arg \min_{\psi \in \Theta_\psi} S_n(\psi),$$

where  $S_n(\psi) = \frac{1}{n} \sum_{i=1}^n (I(\psi_0) + \varepsilon_i - I(\psi))^2$  and  $I(\psi) = I(q_{1i} + q_{2i}\psi \leq 0)$ .

Assuming the knowledge of the consistency and the convergence rate, let  $\psi = \psi_0 + \frac{v}{n}$ .

Following Yu and Fan (2019), we can show the centered process as,

$$D_n^{LSE}(v) = S_n(\psi) - S_n(\psi_0) = n^{-1} \sum_{i=1}^n (I(\psi_0 + \frac{v}{n}) - I(\psi_0))^2 + 2n^{-1} \sum_{i=1}^n (I(\psi_0 + \frac{v}{n}) - I(\psi_0))\varepsilon_i.$$

This implies,

$$n(\psi - \psi_0) = \arg \min_v nD_n^{LSE}(v) = \begin{cases} \sum_{i=1}^{N_{1n}(|v|)} \bar{z}_{1i}, & \text{if } v \leq 0 \\ \sum_{i=1}^{N_{2n}(v)} \bar{z}_{2i}, & \text{if } v > 0 \end{cases},$$

where  $N_{1n}(|v|) = \sum_{i=1}^n (I(\frac{v}{n} \leq -\frac{q_{1i}+q_{2i}\psi_0}{q_{2i}} \leq 0))$ ,  $N_{2n}(v) = \sum_{i=1}^n (I(0 \leq -\frac{q_{1i}+q_{2i}\psi_0}{q_{2i}} \leq \frac{v}{n}))$ ,

$\bar{z}_{1i} = 1 + 2\varepsilon_i$ , and  $\bar{z}_{2i} = 1 - 2\varepsilon_i$

Note that for any finite number  $v$ ,  $N_{2n}(v) \sim B(n, P_n(v))$  where  $B(., .)$  is a binomial process,  $P_n(v) = F(0) - F(\frac{v}{n}) \approx f(0)\frac{v}{n}$ , where  $F(.)$  and  $f(.)$  are CDF and PDF of  $(-\frac{q_{1i}+q_{2i}\psi_0}{q_{2i}})$  respectively. Let  $\lambda = nP_n(v)$ . Hence,  $\lambda \rightarrow f_z(0)v$ . As  $n \rightarrow \infty$ ,  $P_n(v) \rightarrow 0$ , which implies  $N_{2n}(v) \rightarrow N_2(v)$ . Similarly, we have  $N_{1n}(|v|) \rightarrow N_1(|v|)$ , where

$N_1(|v|), N_2(v)$  are two independent Poisson process with intensity  $f_z(0)$ .

As a result,

$$n(\hat{\psi} - \psi_0) \xrightarrow{d} \arg \min_v D^{LSE}(v), \quad (1.8.102)$$

where  $D^{LSE}(v)$  is a compound Poisson process with the form,

$$D^{LSE}(v) = \begin{cases} \sum_{i=1}^{N_1(|v|)} z_{1i}, & \text{if } v \leq 0 \\ \sum_{i=1}^{N_2(v)} z_{2i}, & \text{if } v > 0 \end{cases},$$

where  $z_{1i} = \lim_{\Delta \uparrow 0} \bar{z}_{1i} I(\Delta \leq -\frac{q_{1i} + q_{2i}\psi}{q_{2i}} \leq 0)$ , and  $z_{2i} = \lim_{\Delta \downarrow 0} \bar{z}_{2i} I(0 \leq -\frac{q_{1i} + q_{2i}\psi}{q_{2i}} \leq \Delta)$ .

#### 1.8.4.ii The SLSE

Following Seo and Linton (2007), the SLSE can be obtain as,

$$\hat{\psi}^{SLSE} = \arg \min_{\psi \in \Theta_\psi} S_n^{SLS}(\psi),$$

where  $S_n^{SLS}(\psi) = \frac{1}{n} \sum_{i=1}^n (y_i - K(\psi, \sigma_n))^2$ ,  $K(\psi, \sigma_n) = K(\frac{q_{1i} + q_{2i}\psi}{\sigma_n})$ ,  $K(\cdot)$  is a kernel function as defined in Assumption 3 of Seo and Linton (2007), and  $\sigma_n$  is the bandwidth parameter.

Note that, unlike the LSE, the objective function in this case is smoothed in  $\psi$ . Hence, we can apply the standard first order Taylor series to obtain the asymptotic normality.

By simple calculation, we have,

$$T_n(\psi, \sigma_n) = \frac{\partial S_n^{SLS}(\psi)}{\partial \psi} = -\frac{2}{n} \sum_{i=1}^n I(\psi_0) K'(\psi, \sigma_n) \frac{q_{2i}}{\sigma_n} + \frac{2}{n} \sum_{i=1}^n K(\psi, \sigma_n) K'(\psi, \sigma_n) \frac{q_{2i}}{\sigma_n} \quad (1.8.103)$$

$$-\frac{2}{n} \sum_{i=1}^n K'(\psi, \sigma_n) \frac{q_{2i}}{\sigma_n} \varepsilon_i = A(\psi) + B(\psi) + C(\psi),$$

where  $K'(\psi, \cdot) = \frac{\partial K(\cdot, \cdot)}{\partial \psi}$ .

First, by Assumption 3(b) of Seo and Linton (2007), we can show,

$$\sigma_n^{-h} A(\psi_0) \xrightarrow{p} \sigma_n^{-h} E(I(\psi_0) K'(\psi_0, \sigma_n) \frac{q_{2i}}{\sigma_n}) = O(1),$$

where  $h$  defines  $h^{th}$  order kernel.

This implies, as long as  $\sqrt{n\sigma_n}\sigma_n^{-h} \rightarrow 0$ , we have  $\sqrt{n\sigma_n}A(\psi_0) = o_p(1)$ . Similarly, we can show  $\sqrt{n\sigma_n}B(\psi_0) \xrightarrow{p} 0$ .

Next, similar to the proof of Lemma 3 of Seo and Linton (2007), we obtain,

$$\sqrt{n\sigma_n}C(\psi_0) \xrightarrow{d} N(0, V^\psi),$$

where  $V^\psi = 4Var(K'(\psi_0, \sigma_n)q_{2i}\varepsilon_i)$ .

Hence, we have,

$$\sqrt{n\sigma_n}T_n(\psi_0, \sigma_n) \xrightarrow{d} N(0, V^\psi).$$

Then, by the first order Taylor expansion,

$$T_n(\hat{\psi}^{SLSE}, \sigma_n) = T_n(\psi_0, \sigma_n) + Q_n(\tilde{\psi}, \sigma_n)(\hat{\psi}^{SLSE} - \psi_0) = 0,$$

where  $Q_n(\psi) = \frac{\partial T_n(\dots)}{\partial \psi}$ , and  $\tilde{\psi}$  is between  $\hat{\psi}^{SLSE}$  and  $\psi_0$ .

As a result, this provides the asymptotic normality,

$$\sqrt{n\sigma_n}(\hat{\psi}^{SLSE} - \psi_0) \xrightarrow{d} N(0, Q^{-1}V^\psi Q^{-1}),$$

where  $Q = K'(0)E(q_{2i}^2|z_i = 0)f_z(0)$ ,  $z_i = q_{1i} + q_{2i}\psi_0$ , and  $f_z(\cdot)$  is the density of  $z_i$ .

### 1.8.4.iii The GMM

Consider the moment condition  $E(q_{2i}\varepsilon_i) = 0$  for all  $i = 1, \dots, n$ . Therefore, the GMM estimator can be obtained as,

$$\hat{\psi}^{GMM} = \arg \min_{\psi \in \Theta_\psi} S_n^{GMM}(\psi),$$

where  $S_n^{GMM} = [\frac{1}{n} \sum_{i=1}^n q_{2i}(I(\psi_0) + \varepsilon_i - I(\psi))]^2$ .

Note that, similar to the LSE, the objective function is non-smooth in  $\psi$ . Now, assuming the knowledge of the consistency and the converge rate <sup>4</sup>, let  $\psi = \psi_0 + \frac{v}{n^{1/2}}$ . Hence, the

---

<sup>4</sup>The example is designed with a fixed threshold effect. Hence, the theoretical convergence rate of threshold estimator is  $\sqrt{n}$ .

centered process can be shown as,

$$D_n^{GMM}(v) = S_n^{GMM}(\psi) - S_n^{GMM}(\psi_0) = n^{-2}[\sum_{i=1}^n q_{2i}(I(\psi_0) - I(\psi_0 + \frac{v}{n^{1/2}}))]^2 \\ + 2n^{-2} \sum_{i=1}^n q_{2i}(I(\psi_0) - I(\psi_0 + \frac{v}{n^{1/2}})) \sum_{i=1}^n q_{2i}\varepsilon_i.$$

Note that, by comparing  $D_n^{LSE}$  with  $D_n^{GMM}$ , it is obvious that the second term is quite different. For the  $D_n^{LSE}$ , the sum of error cannot be isolated from  $v$ . As a result, we cannot directly apply the central limit theorem (CLT)<sup>5</sup>. By contrast, for the  $D_n^{GMM}$ , the CLT can be applied to  $\sum_{i=1}^n q_{2i}\varepsilon_i$  as long as the multiplier is bounded. The reason comes from the nature of the sample averaging condition.

This implies,

$$n^{1/2}(\psi - \psi_0) = \arg \min_v nD_n^{GMM}(v) = \arg \min_v n^{-1}[\sum_{i=1}^n q_{2i}(I(\psi_0) - I(\psi_0 + \frac{v}{n^{1/2}}))]^2 \\ + 2n^{-1/2} \sum_{i=1}^n q_{2i}(I(\psi_0) - I(\psi_0 + \frac{v}{n^{1/2}})) \frac{1}{n^{1/2}} \sum_{i=1}^n q_{2i}\varepsilon_i = \arg \min_v A_n^{GMM}(v) + B_n^{GMM}(v).$$

Then, by the Glivenko-Cantelli theorem, for any  $v$ ,

$$A_n^{GMM}(v) \xrightarrow{p} n[E(q_{2i}(I(\psi_0) - I(\psi_0 + \frac{v}{n^{1/2}})))]^2 = G_\psi(\psi_0)^2 v^2,$$

where  $G_\psi(\psi_0) = \frac{dE(q_{2i}I(\psi))}{d\psi} |_{\psi=\psi_0}$ .

---

<sup>5</sup>With diminishing threshold framework, the functional central limit theorem can be applied to  $D_n^{LSE}$ , which leads to a limiting distribution formed by a two-sided Brownian motion (Hansen (2000)). Yu and Phillips (2018a) explains on how compound Poisson process can be approximated by two-sided Brownian motion.

Similarly, we can show that,

$$n^{-1/2} \sum_{i=1}^n q_{2i} (I(\psi_0) - I(\psi_0 + \frac{v}{n^{1/2}})) \xrightarrow{p} n^{1/2} E(q_{2i} (I(\psi_0) - I(\psi_0 + \frac{v}{n^{1/2}}))) = G_\psi(\psi_0)v.$$

Hence, by applying the CLT and the continuous mapping Theorem,

$$B^{GMM}(v) \xrightarrow{d} G_\psi(\psi_0)vN(0, \Omega),$$

where  $\Omega = Var(q_{2i}\varepsilon_i)$ .

This follows that,

$$n^{1/2}(\hat{\psi}^{GMM} - \psi) \xrightarrow{d} \hat{v} = \arg \min_v [2G_\psi(\psi_0)^2v^2 + 2G_\psi(\psi_0)vW],$$

where  $W \sim N(0, \Omega)$ .

Obviously,  $\hat{v} = -W/G_\psi(\psi_0)$ . This provides the asymptotic normality,

$$n^{1/2}(\hat{\psi}^{GMM} - \psi_0) \sim N\left(0, \left(G_\psi(\psi_0)\Omega^{-1}G_\psi(\psi_0)\right)^{-1}\right).$$

## **Chapter 2**

# **Chapter 2: Monte Carlo Comparison for Nonparametric Threshold Estimators**

### **2.1 Introduction**

Popularly used to describe structural changes in economic relationships, threshold models have seen many applications, especially in macro fields (e.g., Hansen (2011); Potter (1995)). Typical examples include the nonlinearity in public debt to GDP ratio (e.g., Afonso and Jalles (2013); Cecchetti, Mohanty and Zampolli (2011); Caner et al. (2010)). A number of threshold estimators for threshold models have been proposed in the literature, and the asymptotic results of these estimators can be categorized into two groups based on different assumptions. The first group is based on the “fixed threshold effect” assumption. The second group imposes a “diminishing threshold effect” assumption introduced



by Hansen (2000). For example, it is well known that, for the least-squares estimator, the threshold estimator is super-consistent with the convergence rate  $n$  under the “fixed threshold effect” assumption and  $n^{1-2\alpha}$  under the “diminishing threshold effect” assumption, respectively, where  $\alpha$  measures the diminishing rate of the threshold effect.

The asymptotic theory and statistical inference have been well developed for the least-squares estimator exogenous regressors and exogenous threshold variable (e.g., Chan (1993); Hansen (2000); Seo and Linton (2007),). Recently, there has been a growing interest in studying threshold models with endogenous regressors and/or a threshold variable. Extending the framework of Hansen (2000), Caner and Hansen (2004) applied the two-step least-squares method to estimate threshold models with endogenous slope regressors. In the spirit of the sample selection technique of Heckman (1979), imposing the joint normality assumption, Kourtellos et al. (2016) explored the case that both the threshold variable and slope regressors are endogenous. The work in Seo and Shin (2016) proposed a two-step GMM estimator for a dynamic panel threshold model with fixed effects, which allows endogeneity in both the slope regressors and threshold variable. It is worth noticing that the GMM method allows both a fixed and diminishing threshold effect, and the convergence rate for the GMM threshold estimator is not super-consistent. By relaxing the joint normality assumption of Kourtellos et al. (2016), Kourtellos et al. (2017) proposed a two-step least square estimator based on a nonparametric control function approach to correct the threshold endogeneity. The semiparametric threshold model separates the threshold effect into two parts, namely the exogenous threshold effect and endogenous threshold bias-correction

term. Therefore, with a “small threshold” effect, the convergence rate for the threshold variable depends on diminishing rates of the threshold effect and the bias-correction term.

However, few studies have worked on the estimation and statistical inference of threshold estimators based on nonparametric estimation methods, which do not rely on the least square method. The work in Delgado and Hidalgo (2000) suggested a difference kernel estimator (or DKE), which depends on a chosen point. The convergence rate of Delgado and Hidalgo (2000) DKE is  $nh^{d-1}$ , which depends on both the bandwidth,  $h$ , and the dimensionality of regressors in their threshold model,  $d \geq 1$ . Built upon the method of Delgado and Hidalgo (2000), Yu and Phillips (2018b) introduced an integrated difference kernel estimator (or IDKE). The work in Yu and Phillips (2018b) argued that the IDKE can be applied to the case with the endogenous threshold variable. The convergence rate of the IDKE is not related to either the bandwidth or the dimensionality of regressors and is super-consistent with the rate  $n$ . Using recently-developed discrete smoothing methods, Henderson et al. (2015) introduced a semiparametric M-estimator of a nonparametric threshold regression model. The threshold estimator of Henderson et al. (2015) can be estimated at the rate  $\sqrt{n/h}$  ( $h$  is the bandwidth), which is faster than the parametric convergence rate of  $\sqrt{n}$ . One may notice that the aforementioned convergence rate is the same as that of the smoothed least squares estimator in Seo and Linton (2007). However, they are entirely different. The work in Henderson et al. (2015) focussed on the nonparametric threshold model, and their proposed estimator was based on a non-smooth objective function. On the contrary, Seo and Linton (2007) worked on a linear threshold model, and the

proposed estimator was based on a smooth objective function with the indicator function replaced by a CDF-type smooth function.

With many applications and simulations available for comparing the parametric threshold estimators in the literature, little guidance is available for researchers to apply as to the choice of nonparametric threshold estimators. Moreover, to avoid the boundary effect of the threshold estimator, most simulations are designed deliberately with the true threshold level chosen at the middle point of the threshold variable distribution, which can be highly doubted in reality. Therefore, the purpose of this paper is to carefully compare the three nonparametric threshold estimators mentioned above using the Monte Carlo method. More importantly, we consider the case that the true threshold level is not only at the middle, but also at the two tails of the threshold variable distribution.

The rest of the paper is organized as follows. In Section 2.2, we briefly review the estimation procedure of three nonparametric threshold estimators such as DKE, IDKE and the M-estimator, where threshold models have exogenous regressors and a threshold variable. In Section 2.3, we illustrate the possible theoretical reason for the conjecture of the poor finite sample performance of the difference kernel-type estimators. Section 2.4 presents the design of the Monte Carlo simulations. Section 2.5 reports the finite sample performance. Section 2.6 concludes.

## 2.2 Three Nonparametric Threshold Estimators

In this paper, we aim to compare the finite sample performance of three nonparametric threshold estimators: Henderson et al. (2015) the semiparametric M-estimator, Delgado and Hidalgo (2000) the difference kernel estimator (DKE) and Yu and Phillips (2018b) the integrated difference kernel estimator (IDKE).

Following Henderson et al. (2015), we consider a generalized threshold regression model:

$$y_i = \alpha_0(X_i) + \beta_0 I\{q_i > \gamma_0\} + \varepsilon_i, \quad (2.2.1)$$

for  $i = 1, \dots, n$ , where  $\alpha_0(\cdot)$  is an unknown smooth function,  $X_i$  is a vector of  $d$  regressors,  $q_i$  is the threshold variable,  $\gamma_0$  is the threshold level,  $I(\cdot)$  is the indicator function and  $\beta_0$  measures the jump size of the regression function at  $q > \gamma$ . Furthermore,  $X_i$  and  $q_i$  are both exogenous and may have a common variable.

### 2.2.1 Semiparametric M-Estimator

If  $\gamma_0$  is known a priori, Model (2.2.1) is known as a partially linear model. The conventional method to estimate the unknown  $\gamma_0$  is minimizing the sum of squared errors, which can be iterated by the grid search. Therefore, Henderson et al. (2015) suggested the semiparametric M-estimator of the nonparametric threshold model, which can be obtained in

three steps.

In Step 1, given  $(\beta, \gamma)$ , Model (2.2.1) becomes a standard nonparametric model. Therefore, we can obtain the Nadaraya–Watson (NW) estimator of  $\alpha_0(x)$  at an interior point,  $x$ , i.e.,

$$\hat{\alpha}(x; \beta, \gamma) = \underset{\alpha \in \Theta_\alpha}{\operatorname{argmin}} n^{-1} \sum_{i=1}^n [y_i - \alpha - \beta I\{q_i > \gamma\}]^2 K_h(X_i - x), \quad (2.2.2)$$

where  $K_h(X_i - x) = h^{-d} \prod_{j=1}^d k(\frac{X_{ij} - x_j}{h})$ ,  $X_i = [X_{i,1}, \dots, X_{i,d}]'$ ,  $x = [x_1, \dots, x_d]'$ ,  $k(\cdot)$  is a second order kernel function,  $h$  is the bandwidth and  $d$  is the dimension of  $x$ .

In Step 2, given  $\gamma$ , Model (2.2.1) becomes a partially linear model. Then,  $\beta_0$  can be estimated as:

$$\hat{\beta}(\gamma) = \underset{\beta \in \Theta_\beta}{\operatorname{argmin}} n^{-1} \sum_{i=1}^n [y_i - \hat{\alpha}(X_i; \beta, \gamma) - \beta I\{q_i > \gamma\}]^2 \hat{f}_h^2(X_i), \quad (2.2.3)$$

where  $\hat{f}_h(X_i) = n^{-1} \sum_{i=1}^n K_h(X_i - x)$  works as the weighting function.

The work in Henderson et al. (2015) shows that  $\hat{\beta}(\gamma)$  has the following mathematical expression:

$$\begin{aligned} \hat{\beta}(\gamma) &= \left[ n^{-1} \sum_{i=1}^n \left[ \sum_{j=1}^n K_h(X_i - X_j) (I_i - I_j) \right]^2 \right]^{-1} \\ &\times n^{-1} \sum_{i=1}^n \left[ \sum_{j=1}^n K_h(X_i - X_j) (I_i - I_j) \sum_{j=1}^n K_h(X_i - X_j) (y_i - y_j) \right], \end{aligned} \quad (2.2.4)$$

where we denote  $I_i = I(q_i > \gamma)$ .

In Step 3, we can estimate the threshold level  $\gamma_0$  by solving the following optimization problem,

$$\hat{\gamma} = \arg \min_{\gamma \in \Theta_\gamma} \left| n^{-1} \sum_{i=1}^n \left[ y_i - \hat{\alpha}(X_i; \beta(\gamma), \gamma) - \hat{\beta}(\gamma) I\{q_i > \gamma\} \right] w(X_i) \right|, \quad (2.2.5)$$

where  $w(\cdot)$  is a weighting function and is application dependent.

As mentioned in Section 2.1, the convergence rate of the threshold estimator of Henderson et al. (2015) is  $\sqrt{n/h}$ , which explodes faster than the usual parametric  $\sqrt{n}$  rate. However, the unknown function  $\alpha_0(\cdot)$  and the jump size  $\beta_0$  converge at standard nonparametric rates of  $\sqrt{nh^d}$  and  $\sqrt{nh}$ , respectively.

### 2.2.2 DKE and IDKE

Instead of using the absolute value of the weighted average of the sum of errors as the objective function, Delgado and Hidalgo (2000) considered using the difference between  $\hat{E}[y|x_0, q = \gamma^-]$  and  $\hat{E}[y|x_0, q = \gamma^+]$  as the objective function. Ideally, the closer  $\gamma$  approaches the true value, the larger the absolute value of the above difference should be. As a result, we are able to estimate the threshold level by choosing  $\gamma$ , which gives the most considerable gap between the two one-sided expectations. Therefore, the difference kernel estimator (DKE) can be obtained by:

$$\hat{\gamma}^{DKE} = \arg \max_{\gamma \in \Theta_\gamma} \left( \frac{1}{n} \sum_{i=1}^n y_i K_{h,i}^{\gamma^-} - \frac{1}{n} \sum_{i=1}^n y_i K_{h,i}^{\gamma^+} \right)^2 \quad (2.2.6)$$

where we have:

$$K_{h,i}^{\gamma+} = K_h(X_i - x_0) \cdot k_h^+(q_i - \gamma),$$

$$K_{h,i}^{\gamma-} = K_h(X_i - x_0) \cdot k_h^-(q_i - \gamma),$$

if  $q_i$  is not part of  $X_i$ , and

$$K_{h,i}^{\gamma+} = K_h(X_{1i} - x_{10}) \cdot k_h^+(q_i - \gamma),$$

$$K_{h,i}^{\gamma-} = K_h(X_{1i} - x_{10}) \cdot k_h^-(q_i - \gamma),$$

if  $q_i$  is part of  $X_i$ , i.e.,  $X_i = [X'_{1i}, q_i]'$ , and  $x_0 = [x'_{10}, q_0]'$ . Furthermore,  $k_h^{+/-}(\cdot)$  is the one-sided kernel function with:

$$k_h^+(q_i - \gamma) = k\left(\frac{q_i - \gamma}{h}\right)I(q_i > \gamma),$$

$$k_h^-(q_i - \gamma) = k\left(\frac{q_i - \gamma}{h}\right)I(q_i \leq \gamma),$$

and  $k(\cdot)$  is a second order kernel function.

Obviously, it is reasonable to expect that the DKE estimator is sensitive to the choice of  $x_0$ . Furthermore, the DKE suffers the curse of dimensionality problem as the convergence rate of the DKE,  $nh^{d-1}$ , depends on the dimension of the regressor. To fix these potential weaknesses, ? proposed an integrated difference kernel estimator, which allows  $\hat{\gamma}$  not to rely on the single choice in  $x_0$ , but the expectation of all  $X$ . The  $\hat{\gamma}^{IDKE}$  can be derived as

follows:

$$\hat{\gamma}^{IDKE} = \arg \max_{\gamma \in \Theta_\gamma} n^{-1} \sum_{i=1}^n \left( \frac{1}{n-1} \sum_{j=1, j \neq i}^n y_j K_{h,ij}^{\gamma^-} - \frac{1}{n-1} \sum_{j=1, j \neq i}^n y_j K_{h,ij}^{\gamma^+} \right)^2, \quad (2.2.7)$$

where:

$$K_{h,ij}^{\gamma^+} = K_h(X_i - x_j) \cdot k_h^+(q_i - \gamma),$$

$$K_{h,i}^{\gamma^-} = K_h(X_i - x_j) \cdot k_h^-(q_i - \gamma),$$

if  $q_i$  is not part of  $X_i$ , and

$$K_{h,i}^{\gamma^+} = K_h(X_{1i} - x_{1j}) \cdot k_h^+(q_i - \gamma),$$

$$K_{h,i}^{\gamma^-} = K_h(X_{1i} - x_{1j}) \cdot k_h^-(q_i - \gamma),$$

if  $q_i$  is part of  $X_i$ , i.e.,  $X_i = [X'_{1i}, q_i]'$ , and  $x_j = [x'_{1j}, q_j]'$ .  $k_h^{+/-}(\cdot)$  is defined the same as above.

The IDKE is super-consistent with convergence rate  $n$ . The work in Yu and Phillips (2018b) showed that IDKE is consistent even if the threshold variable is endogenous. They explain that the role of the instruments of the endogenous regressors and the endogenous threshold variable is improving only the efficiency of the IDKE.



## 2.3 Estimation Difficulties in the Difference Kernel-Type Estimator with Near Boundary $\gamma_0$

In this section, we use a simple version of Model (2.2.1) to explain the estimation difficulties of the difference kernel-type estimators when  $\gamma_0$  lies at the tails of the threshold variable distribution. This estimation difficulty motivates us to investigate the position effect of the true threshold level on the finite sample performance. Specifically, we consider the true model as:

$$y_i = I(X_i \geq \gamma_0), \quad (2.3.1)$$

where  $X_i$  is randomly drawn from a uniform distribution over the interval of  $[-0.5, 0.5]$  for  $i = 1, \dots, n$ .

The model above can be regarded as Model (2.2.1) with  $\alpha_0(x) \equiv 0$ ,  $\beta_0 = 1$ , and  $\varepsilon_i = 0$  for all  $i = 1, \dots, n$ . Therefore, the DKE is based on the objective function:

$$\hat{Q}_n(\gamma)^{DKE} = \left[ \frac{1}{n} \sum_{i=0}^n k \left( \frac{X_i - \gamma}{h} \right) I(X_i < \gamma) y_i - \frac{1}{n} \sum_{i=0}^n k \left( \frac{X_i - \gamma}{h} \right) I(X_i \geq \gamma) y_i \right]^2. \quad (2.3.2)$$

Letting  $u_x = (X_i - \gamma) / h$  and applying the change of variables, we have the probability limit of  $\hat{Q}_n(\gamma)$  equal to:

$$\begin{aligned}
Q_n(\gamma)^{DKE} &= h^2 \left[ \int_{\frac{-0.5-\gamma}{h}}^{\frac{0.5-\gamma}{h}} k(u_x) I(u_x < 0) I\left(u_x \geq \frac{\gamma_0 - \gamma}{h}\right) du_x \right. \\
&\quad \left. - \int_{\frac{-0.5-\gamma}{h}}^{\frac{0.5-\gamma}{h}} k(u) I(u_x \geq 0) I\left(u_x \geq \frac{\gamma_0 - \gamma}{h}\right) du_x \right]^2, \tag{2.3.3}
\end{aligned}$$

where  $h$  is the bandwidth.

If  $\gamma < \gamma_0$ , we obtain:

$$Q_n(\gamma)^{DKE} = h^2 \left[ \int_{\frac{-0.5-\gamma}{h}}^{\frac{0.5-\gamma}{h}} k(u_x) du_x \right]^2, \tag{2.3.4}$$

and:

$$\frac{\partial Q_n(\gamma)^{DKE}}{\partial \gamma} = 2h \left( \int_{\frac{\gamma_0 - \gamma}{h}}^{\frac{0.5-\gamma}{h}} k(u_x) du_x \right) \left[ k\left(\frac{\gamma_0 - \gamma}{h}\right) - k\left(\frac{0.5 - \gamma}{h}\right) \right] > 0, \tag{2.3.5}$$

where the positive sign follows for all  $\gamma_0 < 0.5$  for any second-order kernel function with a bell shape.

It is worth noting that as  $\gamma_0$  approaches 0.5 from the left side, the difference between  $k\left(\frac{\gamma_0 - \gamma}{h}\right) - k\left(\frac{0.5 - \gamma}{h}\right)$  becomes smaller. As a result, for all  $\gamma$ , the above derivative goes to zero, which makes the objective function flat and leads to the estimation difficulty.

Similarly, if  $\gamma > \gamma_0$ , we have:

$$Q_n(\gamma)^{DKE} = h^2 \left( \int_{\frac{\gamma_0 - \gamma}{h}}^0 k(u_x) du_x - \int_0^{\frac{0.5 - \gamma}{h}} k(u_x) du_x \right)^2, \tag{2.3.6}$$

and:

$$\frac{\partial Q_n(\gamma)^{DKE}}{\partial \gamma} = 2h \left( \int_{\frac{\gamma_0 - \gamma}{h}}^0 k(u_x) du_x - \int_0^{\frac{0.5 - \gamma}{h}} k(u_x) du_x \right) \left[ k\left(\frac{\gamma_0 - \gamma}{h}\right) + k\left(\frac{0.5 - \gamma}{h}\right) \right] < 0, \quad (2.3.7)$$

where the negative sign follows for all  $\gamma_0 > -0.5$  for any second-order kernel function with a bell shape.

Therefore, we observe that as  $\gamma_0$  approaches  $-0.5$  from the right side, for all  $\gamma$ , the difference between  $\int_{\frac{\gamma_0 - \gamma}{h}}^0 k(u_x) du_x - \int_0^{\frac{0.5 - \gamma}{h}} k(u_x) du_x$  becomes smaller, which makes the derivative go to zero, and this results in a flat objective function.

In summary, the DKE is asymptotically consistent with  $\gamma_0 \in (-0.5, 0.5)$ . However, it is reasonable to suspect that DKE may have poor finite performance with the true threshold level lying at the tails of the threshold variable distribution due to the estimation difficulty of the flat objective function.

Next, we assume that there are additional covariates,  $Z_i$ , which are randomly drawn from uniform distribution over the interval of  $[-0.5, 0.5]$ , for all  $i = 1, \dots, n$ , and  $\{X_i\}$  and  $\{Z_i\}$  are independent. Therefore, the probability limit of the objective function of the IDKE is (with the same bandwidth) :

$$\begin{aligned} & Q_n(\gamma)^{IDKE} \\ = & h^4 \int_{-0.5}^{0.5} \left[ \int_{\frac{-0.5 - \gamma}{h}}^{\frac{0.5 - \gamma}{h}} k(u_z) k(u_x) I(u_x < 0) I(u_x \geq \frac{\gamma_0 - \gamma}{h}) du_x du_z \right. \\ & \left. - \int_{\frac{-0.5 - z_0}{h}}^{\frac{0.5 - z_0}{h}} \int_{\frac{-0.5 - \gamma}{h}}^{\frac{0.5 - \gamma}{h}} k(u_z) k(u_x) I(u_x \geq 0) I(u_x \geq \frac{\gamma_0 - \gamma}{h}) du_x du_z \right]^2 dz_0 \quad (2.3.8) \end{aligned}$$

where  $u_z = \frac{Z_i - z_0}{h}$ .

Note that:

$$\frac{\partial Q_n(\gamma)^{IDKE}}{\partial \gamma} = h^2 \int_{-0.5}^{0.5} \left( \int_{\frac{-0.5-z_0}{h}}^{\frac{0.5-z_0}{h}} k(u_z) du_z \right)^2 dz_0 \frac{\partial Q_n(\gamma)^{DKE}}{\partial \gamma}. \quad (2.3.9)$$

Consequently, in this typical example,  $\frac{\partial Q_n(\gamma)^{IDKE}}{\partial \gamma}$  can be interpreted as a rescaled  $\frac{\partial Q_n(\gamma)^{DKE}}{\partial \gamma}$ , which implies the IDKE will suffer the same boundary problem as the DKE estimator.

## 2.4 Monte Carlo Designs

To assess the finite sample performance of the three nonparametric threshold estimators, we consider seven data-generating mechanisms, which are similar to those studied in Henderson et al. (2015); Yu and Phillips (2018b).

- DGP 1:

$$y_i = 2I(x_i \geq \gamma_0) + \varepsilon_i \quad (2.4.1)$$

- DGP 2:

$$y_i = x_i + 2I(x_i \geq \gamma_0) + \varepsilon_i \quad (2.4.2)$$

- DGP 3:

$$y_i = \sin(x_i) + 2I(x_i \geq \gamma_0) + \varepsilon_i \quad (2.4.3)$$

- DGP 4:

$$y_i = x_i^2 + 2I(x_i \geq \gamma_0) + \varepsilon_i \quad (2.4.4)$$

- DGP 5:

$$y_i = x_{1i} + x_{2i} + x_{3i} + 2I(x_{1i} \geq \gamma_0) + \varepsilon_i \quad (2.4.5)$$

- DGP 6:

$$y_i = x_{1i}^2 + x_{2i}x_{3i} + 2I(x_{1i} \geq \gamma_0) + \varepsilon_i \quad (2.4.6)$$

- DGP 7:

$$y_i = \sin(x_{1i}) + \cos(x_{2i}) + \sin(x_{3i}) + 2I(x_{1i} \geq \gamma_0) + \varepsilon_i \quad (2.4.7)$$

where  $x_i$  is randomly drawn from a uniform distribution over the interval of  $[-0.5, 0.5]$  for all  $i = 1, \dots, n$ ,<sup>1</sup> and  $\varepsilon_i$  is randomly drawn from the  $N(0, 1)$  distribution. All DGPs are based on the fixed threshold effect framework of Chan (1993) with both the exogenous threshold variable and exogenous regressors.

DGPs 1–4 are univariate threshold models. More specifically, DGPs 1–2 are typical linear threshold models. DGPs 3–4 are nonlinear threshold models modelling the periodicity and the quadraticity, respectively. DGPs 5–7 are multivariate threshold models. DGP 5 characterizes the multivariate linear threshold model. DGPs 6–7 are nonlinear threshold models extending DGPs 3–4 to multivariate specifications.

---

<sup>1</sup>With the uniform distribution, the intensity of the Poisson process would not change with the change in the true threshold location. Therefore, the limiting distribution of both the DKE and the IDKE is not affected given  $\gamma_0$  is not on the boundary of  $\Theta_\gamma$ .

To examine the position effect of the true threshold level on the finite sample performance, we set  $\gamma_0$  at different segments of the threshold variable distribution. Specifically, we set the true threshold,  $\gamma_0$ , as the  $p^{\text{th}}$  quantile of the threshold variable with  $p = 25, 50$  and  $75$  to place the true threshold level to the left tail, middle and the right tail of the threshold variable distribution, respectively.

We set  $x_0 = x^{\text{max}}$  for the DKE estimate of Delgado and Hidalgo (2000), where  $x^{\text{max}}$  is the data with the greatest empirical density among all generated  $x'_i$ 's for each simulation of each DGP.<sup>2</sup> We use the rule of thumb bandwidth,  $h = C\hat{\sigma}_x n^{-1/(d+4)}$ , where  $C = \frac{4}{d+2} \frac{1}{d+4}$ ,  $d$  is the dimension of  $x_i$  and  $\hat{\sigma}_x$  is the sample standard deviation of  $\{x_i\}$ . We use the Gaussian kernel function. As suggested by Yu and Phillips (2018b), we use the one-sided rescaled Epanechnikov kernel with  $k^-(q, 0) = \frac{3}{4}(1 - q^2)I(q < 0)$  and  $k^+(q, 0) = k^-(-q, 0)$  to estimate the DKE and the IDKE.

We repeat 2000 times for each simulation.<sup>3</sup> We set the sample size  $n = 100, 300$  and  $500$ . For each simulation, we report the average bias, mean squared error (or MSE) and the standard deviation (or stdev) of the threshold estimates. Tables 2.1–2.8 contain the details of the simulation results. Table 2.9 shows the realized convergence rate of the semi-parametric M-estimator of Henderson et al. (2015) and IDKE of Yu and Phillips (2018b).

<sup>2</sup>The theoretical density should be the same for all  $x$  due to the uniform distribution. The reason we use the data-driven choice of  $x_0$  is because we do not know the true density in reality.

<sup>3</sup>All programming is finished in Matlab.

Table 2.1: Simulation results of nonparametric threshold estimators, Data-generating Mechanism 1 (DGP 1). IDKE, integrated difference kernel estimator.

$\gamma_0$ Is the 25 <sup>th</sup> Quantile of the Threshold Variable									
$n$	Bias			MSE			Stdev		
	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE
100	0.0336	0.2705	0.0679	0.0144	0.0913	0.0225	0.1152	0.1345	0.1338
300	0.0015	0.2929	0.0870	0.0006	0.0986	0.0308	0.0241	0.1133	0.1525
500	0.0002	0.2632	0.1530	0.0001	0.0920	0.0544	0.0097	0.1509	0.1760
$\gamma_0$ Is the 50 <sup>th</sup> Quantile of the Threshold Variable									
$n$	Bias			MSE			Stdev		
	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE
100	0.0056	-0.0346	-0.0183	0.0084	0.0154	0.0012	0.0916	0.1191	0.0288
300	0.0007	-0.0346	-0.0083	0.0009	0.0209	0.0002	0.0302	0.1406	0.0126
500	0.0008	-0.0347	-0.0055	0.0003	0.0233	0.0001	0.0166	0.1488	0.0080
$\gamma_0$ Is the 75 <sup>th</sup> Quantile of the Threshold Variable									
$n$	Bias			MSE			Stdev		
	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE
100	-0.0397	-0.2485	-0.0666	0.0163	0.1082	0.0087	0.1215	0.2156	0.0650
300	-0.0028	-0.2590	-0.0377	0.0009	0.1143	0.0029	0.0299	0.2174	0.0391
500	-0.0004	-0.2841	-0.0287	0.0001	0.1288	0.0018	0.0118	0.2193	0.0308

This table reports the simulation results of three estimators, the semiparametric M-estimator of Henderson et al. (2015), the DKE of Delgado and Hidalgo (2000) and the IDKE of Yu and Phillips (2018b) for the simple jump function defined as Equation (2.4.1). The first column gives the sample size that the simulation used. The third to fifth columns report the average bias. The sixth to eighth columns give the mean squared errors of the threshold estimates. The last three columns present the standard deviations.

Table 2.2: Simulation results of nonparametric threshold estimators, DGP 2.

$\gamma_0$ Is the 25 <sup>th</sup> Quantile of the Threshold Variable									
$n$	Bias			MSE			Stdev		
	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE
100	0.0359	0.2272	0.0813	0.0154	0.0823	0.0250	0.1190	0.1752	0.1357
300	0.0053	0.2680	0.1019	0.0020	0.0954	0.0324	0.0442	0.1536	0.1485
500	0.0002	0.2632	0.1530	0.0001	0.0920	0.0544	0.0097	0.1509	0.1760

Table 2.3: Cont.

$\gamma_0$ Is the 50 <sup>th</sup> Quantile of the Threshold Variable									
$n$	Bias			MSE			Stdev		
	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE
100	-0.0008	-0.0246	-0.0151	0.0082	0.0122	0.0009	0.0907	0.1077	0.0257
300	0.0002	-0.0147	-0.0067	0.0009	0.0130	0.0002	0.0306	0.1130	0.0107
500	0.0002	-0.0131	-0.0044	0.0000	0.0154	0.0001	0.0068	0.1233	0.0073
$\gamma_0$ Is the 75 <sup>th</sup> Quantile of the Threshold Variable									
$n$	Bias			MSE			Stdev		
	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE
100	-0.0307	-0.2465	-0.1031	0.0119	0.1049	0.0159	0.1048	0.2101	0.0730
300	-0.0059	-0.2564	-0.0786	0.0023	0.1009	0.0086	0.0477	0.1876	0.0494
500	-0.0008	-0.2651	-0.0699	0.0003	0.1060	0.0065	0.0177	0.1891	0.0397

This table reports the simulation results of three estimators, the semiparametric M-estimator of Henderson et al. (2015), the DKE of Delgado and Hidalgo (2000) and the IDKE of Yu and Phillips (2018b) for the univariate linear threshold model defined as Equation (2.4.2). The first column gives the sample size that the simulation used. The third to fifth columns report the average bias. The sixth to eighth columns give the mean squared errors of the threshold estimates. The last three columns present the standard deviations.



Table 2.4: Simulation results of nonparametric threshold estimators, DGP 3.

$\gamma_0$ Is the 25 <sup>th</sup> Quantile of the Threshold Variable									
$n$	Bias			MSE			Stdev		
	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE
100	0.0303	0.2211	0.0785	0.0128	0.0791	0.0233	0.1092	0.1739	0.1310
300	0.0022	0.2725	0.1137	0.0014	0.0980	0.0373	0.0376	0.1541	0.1561
500	0.0005	0.2694	0.1570	0.0002	0.0961	0.0546	0.0131	0.1535	0.1730
$\gamma_0$ Is the 50 <sup>th</sup> Quantile of the Threshold Variable									
$n$	Bias			MSE			Stdev		
	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE
100	0.0017	-0.0236	-0.0137	0.0073	0.0111	0.0008	0.0852	0.1027	0.0257
300	0.0002	-0.0220	-0.0061	0.0004	0.0132	0.0001	0.0196	0.1128	0.0101
500	-0.0003	-0.0114	-0.0041	0.0001	0.0149	0.0001	0.0112	0.1215	0.0067
$\gamma_0$ Is the 75 <sup>th</sup> Quantile of the Threshold Variable									
$n$	Bias			MSE			Stdev		
	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE
100	-0.0358	-0.2471	-0.1036	0.0160	0.1031	0.0160	0.1212	0.2051	0.0725
300	-0.0027	-0.2592	-0.0822	0.0013	0.1041	0.0091	0.0360	0.1924	0.0482
500	-0.0007	-0.2637	-0.0686	0.0004	0.1031	0.0065	0.0203	0.1832	0.0422

This table reports the simulation results of three estimators, the semiparametric M-estimator of Henderson et al. (2015), the DKE of Delgado and Hidalgo (2000) and the IDKE of Yu and Phillips (2018b) for the univariate threshold periodic model defined as Equation (2.4.3). The first column gives the sample size that the simulation used. The third to fifth report propose the average bias. The sixth to eighth columns give the mean squared errors of the threshold estimates. The last three columns present the standard deviations.

Table 2.5: Simulation results of nonparametric threshold estimators, DGP 4.

$\gamma_0$ Is the 25 <sup>th</sup> Quantile of the Threshold Variable									
$n$	Bias			MSE			Stdev		
	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE
100	0.0371	0.2754	0.1038	0.0168	0.0922	0.0348	0.1242	0.1278	0.1551
300	0.0065	0.2817	0.1479	0.0030	0.0921	0.0526	0.0545	0.1131	0.1754
500	0.0010	0.2884	0.2146	0.0005	0.0974	0.0794	0.0221	0.1196	0.1826
$\gamma_0$ Is the 50 <sup>th</sup> Quantile of the Threshold Variable									
$n$	Bias			MSE			Stdev		
	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE
100	0.0050	-0.0324	-0.0173	0.0086	0.0156	0.0016	0.0930	0.1205	0.0355
300	-0.0010	-0.0408	-0.0071	0.0012	0.0212	0.0002	0.0341	0.1400	0.0135
500	0.0000	-0.0340	-0.0051	0.0000	0.0222	0.0001	0.0038	0.1451	0.0086
$\gamma_0$ Is the 75 <sup>th</sup> Quantile of the Threshold Variable									
$n$	Bias			MSE			Stdev		
	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE
100	-0.0378	-0.2562	-0.0694	0.0157	0.1105	0.0089	0.1196	0.2120	0.0640
300	-0.0025	-0.2622	-0.0445	0.0007	0.1131	0.0037	0.0266	0.2107	0.0411
500	-0.0007	-0.2709	-0.0358	0.0004	0.1162	0.0024	0.0203	0.2070	0.0334

This table reports the simulation results of three estimators, the semiparametric M-estimator of Henderson et al. (2015), the DKE of Delgado and Hidalgo (2000) and the IDKE of Yu and Phillips (2018b) for the univariate threshold quadratic model defined as Equation (2.4.4). The first column gives the sample size that the simulation used. The third to fifth report propose the average bias. The sixth to eighth columns give the mean squared errors of the threshold estimates. The last three columns present the standard deviations.

Table 2.6: Simulation results of nonparametric threshold estimators, DGP 5.

$\gamma_0$ Is the 25 <sup>th</sup> Quantile of the Threshold Variable									
$n$	Bias			MSE			Stdev		
	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE
100	0.0141	0.2560	0.0751	0.0060	0.1005	0.0213	0.0762	0.1871	0.1253
300	0.0005	0.2587	0.0421	0.0006	0.0970	0.0104	0.0253	0.1733	0.0931
500	0.0000	0.2696	0.0333	0.0000	0.0977	0.0085	0.0038	0.1583	0.0862
$\gamma_0$ Is the 50 <sup>th</sup> Quantile of the Threshold Variable									
$n$	Bias			MSE			Stdev		
	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE
100	-0.0035	-0.0232	-0.0167	0.0050	0.0248	0.0014	0.0710	0.1559	0.0335
300	0.0000	-0.0176	-0.0082	0.0001	0.0205	0.0003	0.0118	0.1420	0.0136
500	0.0001	-0.0330	-0.0057	0.0000	0.0222	0.0001	0.0041	0.1452	0.0106
$\gamma_0$ Is the 75 <sup>th</sup> Quantile of the Threshold Variable									
$n$	Bias			MSE			Stdev		
	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE
100	-0.0203	-0.2778	-0.1173	0.0085	0.1239	0.0212	0.0900	0.2161	0.0864
300	-0.0007	-0.2878	-0.0958	0.0002	0.1256	0.0133	0.0154	0.2069	0.0639
500	0.0000	-0.2883	-0.0944	0.0000	0.1253	0.0119	0.0035	0.2056	0.0544

This table reports the simulation results of three estimators, the semiparametric M-estimator of Henderson et al. (2015), the DKE of Delgado and Hidalgo (2000) and the IDKE of Yu and Phillips (2018b) for the multivariate linear threshold model defined as Equation (2.4.5). The first column gives the sample size that the simulation used. The third to fifth report propose the average bias. The sixth to eighth columns give the mean squared errors of the threshold estimates. The last three columns present the standard deviations.

Table 2.7: Simulation results of nonparametric threshold estimators, DGP 6.

$\gamma_0$ Is the 25 <sup>th</sup> Quantile of the Threshold Variable									
$n$	Bias			MSE			Stdev		
	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE
100	0.0197	0.2495	0.0704	0.0082	0.0972	0.0188	0.0882	0.1871	0.1177
300	0.0002	0.2652	0.0364	0.0001	0.0997	0.0094	0.0114	0.1714	0.0898
500	0.0000	0.2738	0.0297	0.0000	0.1003	0.0074	0.0032	0.1594	0.0807
$\gamma_0$ Is the 50 <sup>th</sup> Quantile of the Threshold Variable									
$n$	Bias			MSE			Stdev		
	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE
100	0.0019	-0.0107	-0.0158	0.0051	0.0242	0.0013	0.0711	0.1553	0.0323
300	-0.0004	-0.0251	-0.0074	0.0002	0.0216	0.0002	0.0138	0.1450	0.0125
500	0.0001	-0.0280	-0.0054	0.0000	0.0210	0.0001	0.0036	0.1422	0.0094
$\gamma_0$ Is the 75 <sup>th</sup> Quantile of the Threshold Variable									
$n$	Bias			MSE			Stdev		
	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE
100	-0.0184	-0.2709	-0.1164	0.0082	0.1177	0.0207	0.0886	0.2105	0.0846
300	-0.0007	-0.2717	-0.0975	0.0004	0.1157	0.0131	0.0194	0.2048	0.0600
500	0.0002	-0.2647	-0.0889	0.0000	0.1080	0.0104	0.0042	0.1949	0.0497

This table reports the simulation results of three estimators, the semiparametric M-estimator of Henderson et al. (2015), the DKE of Delgado and Hidalgo (2000) and the IDKE of Yu and Phillips (2018b) for the multivariate threshold quadratic model defined as Equation (2.4.6). The first column gives the sample size that the simulation used. The third to fifth columns report the average bias. The sixth to eighth columns give the mean squared errors of the threshold estimates. The last three columns present the standard deviations.

Table 2.8: Simulation results of nonparametric threshold estimators, DGP 7.

$\gamma_0$ Is the 25 <sup>th</sup> Quantile of the Threshold Variable									
$n$	Bias			MSE			Stdev		
	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE
100	0.0207	0.2936	0.1292	0.0097	0.1086	0.0419	0.0964	0.1498	0.1588
300	0.0005	0.2915	0.1275	0.0003	0.1031	0.0393	0.0168	0.1347	0.1517
500	0.0003	0.2947	0.1378	0.0001	0.1048	0.0427	0.0105	0.1341	0.1542
$\gamma_0$ Is the 50 <sup>th</sup> Quantile of the Threshold Variable									
$n$	Bias			MSE			Stdev		
	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE
100	-0.0034	0.0004	-0.0373	0.0051	0.0265	0.0074	0.0716	0.1630	0.0778
300	0.0013	0.0049	-0.0366	0.0003	0.0229	0.0029	0.0178	0.1514	0.0398
500	0.0003	0.0077	-0.0315	0.0001	0.0180	0.0019	0.0081	0.1339	0.0294
$\gamma_0$ Is the 75 <sup>th</sup> Quantile of the Threshold Variable									
$n$	Bias			MSE			Stdev		
	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE	Semi-M	DKE	IDKE
100	-0.0244	-0.2830	-0.2242	0.0106	0.1137	0.0575	0.0998	0.1834	0.0849
300	0.0000	-0.2798	-0.2068	0.0001	0.1074	0.0457	0.0084	0.1708	0.0539
500	0.0000	-0.2823	-0.1963	0.0000	0.1039	0.0403	0.0036	0.1558	0.0424

This table reports the simulation results of three estimators, the semiparametric M-estimator of Henderson et al. (2015), the DKE of Delgado and Hidalgo (2000) and the IDKE of Yu and Phillips (2018b) for the multivariate threshold periodic model defined as Equation (2.4.7). The first column gives the sample size that the simulation used. The third to fifth columns report the average bias. The sixth to eighth columns give the mean squared errors of the threshold estimates. The last three columns present the standard deviations.

Table 2.9: Estimated convergence rate of the nonparametric threshold estimators.

<b>Semiparametric M-Estimator of Henderson et al. (2015)</b>							
	<b>DGP 1</b>	<b>DGP 2</b>	<b>DGP 3</b>	<b>DGP 4</b>	<b>DGP 5</b>	<b>DGP 6</b>	<b>DGP 7</b>
$p = 25$	-1.235	-1.202	-1.209	-1.280	-1.224	-1.347	-1.307
$p = 50$	-1.162	-1.195	-1.171	-1.234	-1.349	-1.335	-1.347
$p = 75$	-1.215	-1.251	-1.203	-1.205	-1.227	-1.234	-1.331
<b>IDKE of Yu and Phillips (2018b)</b>							
	<b>DGP 1</b>	<b>DGP 2</b>	<b>DGP 3</b>	<b>DGP 4</b>	<b>DGP 5</b>	<b>DGP 6</b>	<b>DGP 7</b>
$p = 25$	-2.207	-2.126	-2.164	-2.541	-1.556	-1.436	-2.557
$p = 50$	-1.352	-1.287	-1.305	-1.335	-1.428	-1.348	-1.982
$p = 75$	-1.758	-1.949	-1.966	-1.757	-1.876	-2.115	-2.626

This table reports the realized convergence rates of the semiparametric M-estimator of Henderson et al. (2015) and the IDKE of Yu and Phillips (2018b). The realized convergence rates are shown as the coefficient estimate by regressing the logarithm of  $RMSE$  on the logarithm of the sample size for each DGP. Samples sizes used are  $n = 100, 200, 300, 400, 500, 600$  and  $700$ .

## 2.5 Monte Carlo Results

For the semi-parametric M-estimator introduced by Henderson et al. (2015), our results show that the performance was slightly affected by the position of the true threshold level. Meanwhile, as the sample size increased, this position effect gradually vanished.<sup>4</sup> In addition, we observed that the bias was smaller for multivariate models than univariate models. Using the bandwidth as defined in Section 2.4, which behaved roughly as  $O(n^{-1/5})$  for univariate models and  $O(n^{-1/7})$  for multivariate models, the theoretical convergence rates were  $O(n^{-1.2})$  and  $O(n^{-1.14})$  accordingly. From Table 2.9, the super-consistency was con-

<sup>4</sup>With  $n = 100$ , the bias, MSE and standard deviation were larger with  $\gamma_0$  placed at two tails and  $\gamma_0$  placed at the median point. However, with  $n = 500$ , there was no apparent difference between tail position  $\gamma_0$  estimation and the median position  $\gamma_0$  estimation.

firmed with the estimated convergence rate of  $\hat{\gamma}$ . Consistent with the theory, the realized convergence rate decreased as the dimension increased. It is quite interesting that, for almost all univariate models, the realized convergence rate of  $\hat{\gamma}$  was faster when  $\gamma_0$  was at the left- and right-tail position than when  $\gamma_0$  was at the median position. However, for multivariate models, the realized rates seemed to be stable with the position of  $\gamma_0$ .

For the DKE, as we conjectured, it was severely affected by the position of the true threshold value for all DGPs, which may result from the estimation difficulties, as we argued in Section 2.3. Furthermore, even with the middle-positioned  $\gamma_0$ , the bias still showed a non-decreasing pattern with the increasing sample size under some multivariate specifications.<sup>5</sup> Intuitively, this may result from the choice of  $x_0$ , which distorts the result by providing useless information. According to the comment in the Supplementary Material of Yu and Phillips (2018b), the choice of  $x_0$  is crucial in identifying the DKE estimator. On the one hand, the optimal  $x_0$  should make  $[E(y|x_0, q = \gamma_0^-) - E(y|x_0, q = \gamma_0^+)]^2$  as large as possible. On the other hand, one needs the conditional density  $f(x_0|q = \gamma_0)$  to be large enough to provide sufficient information. Therefore, theoretically, with a uniform distribution and univariate linear threshold model as in DGP2, the ideal  $x_0$  should be at the middle of its distribution with the value of zero. However, in the simulation, we set  $x_0$  equal to the value with the largest empirical density, which may appear at the two tails. This may lead  $[E(y|x_0, q = \gamma_0^-) - E(y|x_0, q = \gamma_0^+)]^2$  to approach zero. Moreover, with the multivariate and nonlinear specification, we can expect more distortion involved. As a

---

<sup>5</sup>For example, in Table 2.7, the bias monotonically increases with the in sample size.

result, the DKE performs the worst among all three competitors for all DGPs.

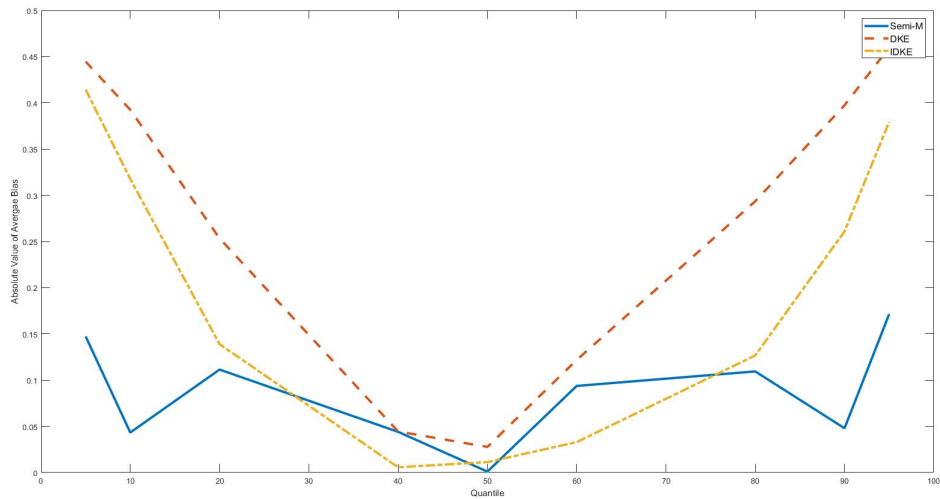
For the IDKE, our results reveal several features. Firstly, the IDKE was affected by the position of the actual threshold value. The influence was not as substantial as the DKE. Indeed, the integration allowed more local information to be used and alleviated the possible distortion due to the choice of  $x_0$ . Surprisingly, unlike the DKE, this position effect seemed to be asymmetric for the IDKE. For most of the DGPs, we observed that the absolute value of the average bias and MSE was larger with the left-tailed  $\gamma_0$  than the right-tailed  $\gamma_0$ . The theoretical convergence rate of the IDKE estimator,  $n$ , is not related to either the bandwidth or the dimension, which is faster than the semi-parametric M-estimator of Henderson et al. (2015). This is consistent with our realized convergence rates, which are shown in Table 2.9. Moreover, for all DGPs, the realized convergence rates were faster with two-sided tailed  $\gamma_0$  than the median  $\gamma_0$ .

In summary, the simulation results give some evidence that the finite sample performances were affected by the position of the true threshold level for all three nonparametric threshold estimators. However, this effect was heterogeneous. The position effect least influenced the semi-M estimator of Henderson et al. (2015). Meanwhile, the difference kernel-type estimators were severely distorted by the tailed  $\gamma_0$ , which confirms our conjecture made in Section 3. Furthermore, our results show that the position of the true threshold level also affects the realized convergence rate. We also found, for the semi-M estimator of Henderson et al. (2015) and the IDKE estimator, the tail distortion tended to be reduced in multivariate models.



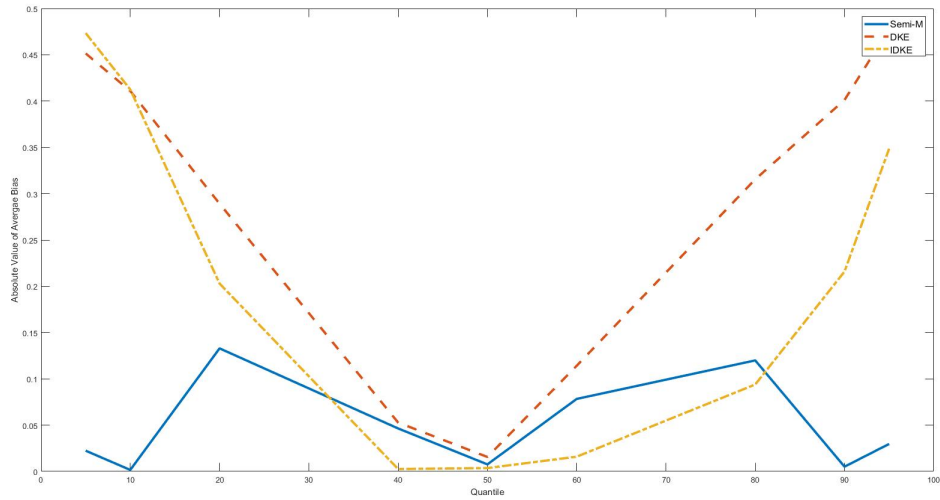
As a robustness check of our findings, Figures 2.1–2.4 show the simulation results of DGP 2 and DGP 5 with  $\gamma_0$  taking different positions along the threshold variable distribution. It is obvious that, for all figures, semi-M had lower average bias in absolute value than difference kernel-type estimators with tail  $\gamma_0$ . Furthermore, we found the gap between the average bias of the semi-M estimator and the average bias of the difference kernel-type estimators to drop greatly with  $\gamma_0$  approaching the middle position of the threshold variable distribution.

Figure 2.1: Average bias with  $\gamma_0$  as various quantiles of the threshold variable, DGP 2,  $n = 100$ .



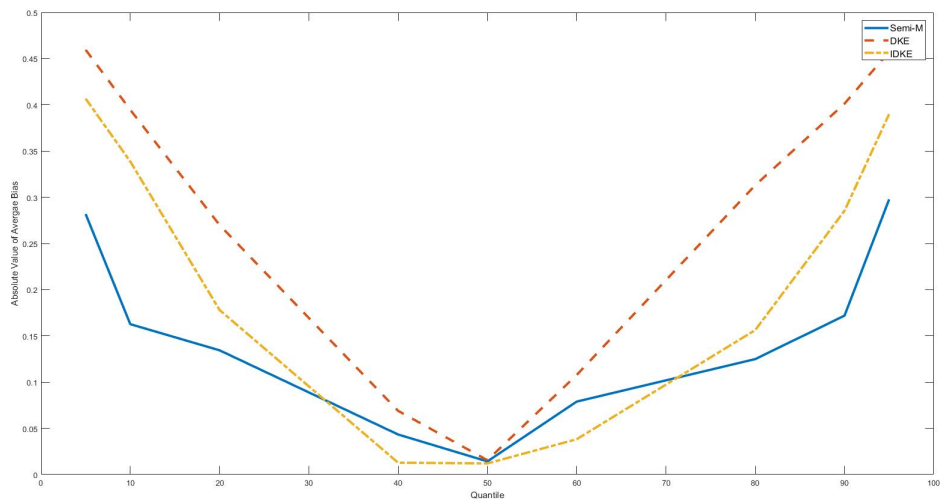
This figure shows absolute values of the average bias with the true threshold level being several quantiles of the threshold variable ( $5^{th}$ ,  $10^{th}$ ,  $20^{th}$ ,  $40^{th}$ ,  $50^{th}$ ,  $60^{th}$ ,  $80^{th}$ ,  $90^{th}$ ,  $95^{th}$ ). The simulation is based on DGP 2. The sample size is 100.

Figure 2.2: Average bias with  $\gamma_0$  as various quantiles of the threshold variable, DGP 2,  $n = 300$ .



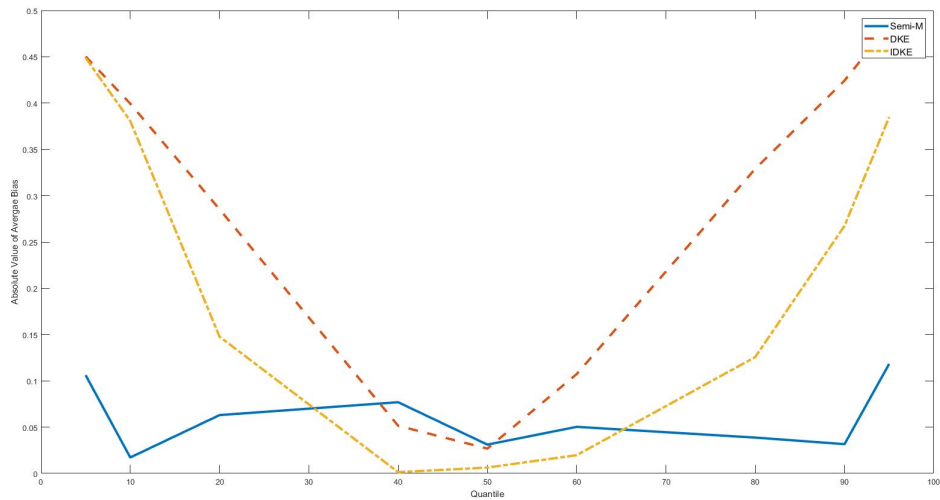
This figure shows absolute values of the average bias with the true threshold level being several quantiles of the threshold variable ( $5^{th}$ ,  $10^{th}$ ,  $20^{th}$ ,  $40^{th}$ ,  $50^{th}$ ,  $60^{th}$ ,  $80^{th}$ ,  $90^{th}$ ,  $95^{th}$ ). The simulation is based on DGP 2. The sample size is 300.

Figure 2.3: Average bias with  $\gamma_0$  as various quantiles of the threshold variable, DGP 5,  $n = 100$ .



This figure shows absolute values of the average bias with the true threshold level being several quantiles of the threshold variable ( $5^{th}$ ,  $10^{th}$ ,  $20^{th}$ ,  $40^{th}$ ,  $50^{th}$ ,  $60^{th}$ ,  $80^{th}$ ,  $90^{th}$ ,  $95^{th}$ ). The simulation is based on DGP 5. The sample size is 100.

Figure 2.4: Average bias with  $\gamma_0$  as various quantiles of the threshold variable, DGP 5,  $n = 300$ .



This figure shows absolute values of the average bias with the true threshold level being several quantiles of the threshold variable ( $5^{th}$ ,  $10^{th}$ ,  $20^{th}$ ,  $40^{th}$ ,  $50^{th}$ ,  $60^{th}$ ,  $80^{th}$ ,  $90^{th}$ ,  $95^{th}$ ). The simulation is based on DGP 5. The sample size is 300.

## 2.6 Conclusions

In this paper, we evaluated the finite sample performance of three non-parametric threshold estimators and identified the relationship between the performances of different estimators and the position of the true threshold level with Monte Carlo methods.

The study shows, with all three estimators affected by the tail position of the true threshold value, that the semi-M estimator of Henderson et al. (2015) outperformed DKE and IDKE for roughly all DGPs considered in the paper. Interestingly, there appears to be some evidence that the distortion can be reduced if there are other covariates besides the threshold variable for the semi-M estimator and the IDKE. Consistent with the theory, we find that the realized convergence rates support the super-consistency in the threshold estimate for all three estimators. However, we find that the realized convergence rates are also affected by the position of the true threshold value. We therefore conclude that, in applied works, using the difference kernel-type estimation, researchers must be careful when the threshold estimate is at the left-tail or the right-tail of the threshold variable distribution.

## **Chapter 3**

### **Can Exchange Rate Pass-Through**

### **Explain the Asymmetric Gasoline**

### **Puzzle? Evidence from a Pooled Panel**

### **Threshold Analysis of the EU**

#### **3.1 Introduction**

ERPT, namely the change in import prices resulting from an exchange rate shock, is an important topic in Economics that has received significant attention from the researchers within the last twenty years (see for example Goldberg and Campa (2006a); Gopinath et al. (2010); Ceglowski (2010); Devereux and Yetman (2010); Brun-Aguerre et al. (2012); Auer

and Schoenle (2016)).

From an international economics perspective, a key question is to what extent the exchange-rate fluctuations are passed-through to the prices of imported goods (Fabra and Reguant (2014)). Exchange rate fluctuations between dollar and other currencies play a crucial role in determining the transmission pricing mechanism in commodity markets including oil industry as well (Galeotti et al. (2003)). As a consequence, the estimation of sensitivity (elasticity) of local-currency import prices (i.e gasoline prices) to changes in local-currency price of foreign currency known as ERPT is of paramount importance for controlling the transmission of inflation between countries, testing the law of one price and the existence of Purchasing Power Parity (Goldberg and Knetter (1997); Goldberg and Campa (2006a); Krugman (1986); Helpman and Krugman (1987)).

Within the last years there is a plethora of studies in the Industrial Organization (IO) literature investigating the existence of gasoline price asymmetry with controversial results. Most of these studies apply cointegration techniques by utilizing an asymmetric (vector) error-correction model ( Borenstein et al. (1997); Eckert (2002); Galeotti et al. (2003); Deltas (2008); Polemis (2011); Wlazlowski et al. (2012); Greenwood-Nimmo and Shin (2013); Bumpass et al. (2015); Kristoufek and Lunackova (2015); Blair et al. (2017); Eleftheriou et al. (2018)), while others rely on non-parametric methods (Godby et al. (2000); Mann (2016); Polemis and Tsionas (2016); Bagnai and Ospina (2018)) in order to uncover the existence of price asymmetries. The asymmetric price adjustment mechanism has also been examined on a theoretical ground as well. Theories of asymmetric price

adjustment identify possible causes of asymmetry in a number of reasons such as inter alia tacit collusion (Radchenko (2005)), inventory capacity and hoarding (Borenstein and Shepard (1996)), and consumer search (Johnson (2002)).

Despite the rich body of literature, existing studies fail to explain the role of exchange rate fluctuations in determining the causes of the asymmetric gasoline adjustment path (commonly known as *rockets and feathers* hypothesis).<sup>1</sup> In particular, past studies have been methodologically restrictive in the sense that the retail gasoline short-run responses, given an input (crude) cost shock, were attributed to crude oil fluctuations. However, “these studies would therefore be biased these studies would therefore be misspecified if mark-up rules were actually described by an alternative relationship, as would be the case if, for example, price asymmetries were instead triggered by a minimum absolute increase in crude cost” (Godby et al. (2000)). Specifically, the authors argue that this is a possibility, not that it is the usual case and try to estimate a TAR to investigate this possibility, but do not find any evidence of asymmetric pricing in the Canadian market.

Using several possible exchange rate-retail price relationships, we attempt to determine whether an asymmetric pricing pattern in the weekly data for 28 EU countries can be explained by the ERPT mechanism. This approach traces the effects of the exchange rate on the coefficient of each regressor (marginal response) over the sample. In this case, the trade-weighted dollar exchange rate index acts as a threshold variable in order to capture the marginal effect of a given variable as an unknown function of an observable covariate,

---

<sup>1</sup>This means that prices increase rapidly in response to cost increases (like a rocket) but fall only slowly in response to cost decreases (like a feather)

introducing heterogeneity. Subsequently, the EU-28 countries will be sorted according to their level of international competitiveness toward the US economy placing them into net exporters (high regime countries) and net importers (low regime countries) respectively. This happens since a rise of the exchange rate index tends to increase the value of the US imports and lower the value of the exports. Therefore, EU countries increase their exports to the US compared to their imports (net exporters). The opposite mechanism is triggered when the relevant index decreases. The contribution of this paper is three-fold. First, it goes beyond the existing literature in that it uses a particularly long panel of EU-28 countries at a weekly basis. Second, in contrast to the existing empirical studies which assume that the variables are not correlated across the panel dimension (cross sectional independence) we perform appropriate cointegration techniques in order to deal with this issue. The latter may arise due to common unobserved effects generated by changes in the European legislation (i.e taxation, currency regulatory restrictions, import quotas, etc). Third and foremost, it is the first study to our knowledge that tries to examine the impact of the ERPT on asymmetric gasoline price adjustment. Moreover, the application of the dynamic panel GMM threshold model developed by Seo and Shin (2016) constitutes an additional novelty of this paper. Previous studies assume the threshold to be zero. However, it is possible that this might not be the case for the European gasoline market as a whole. It may be possible that the threshold lies at some positive value or it may be that the asymmetric behaviour is not triggered until a certain change in input price is felt in some fixed time period (Godby et al. (2000)). Using the GMM threshold model allows us to test for possible asymmetric



gasoline pricing mechanism triggered by exchange rate fluctuations.

In this study, we employ a pooled panel threshold model within an error correction framework and allowing for the presence of an endogenous threshold variable to investigate the following research questions: Is there evidence of short-run gasoline asymmetric pricing in the EU-28 as a whole over the sample period? Does the ERPT mechanism constitute a possible cause of gasoline asymmetric adjustment? Are there any non-linear effects in *rockets and feathers* hypothesis? Asymmetric pricing is tested for the net retail unleaded EU gasoline markets. The empirical findings confirm the superiority of the threshold model compared to the baseline linear specifications, while attributing the asymmetric gasoline adjustment mechanism to ERPT.

The rest of the paper is organized as follows. Section 2 provides a comprehensive survey to the ERPT literature. Section 3 describes the data while Section 4 presents the empirical models (baseline and threshold model) estimated in this paper and discusses econometric issues. Section 5 reports the estimation results, Section 6 concludes the paper.

## **3.2 Literature Review**

The literature on ERPT starts with the seminal paper of Kreinin (1977) who uses an experimental approach to estimating the degree of ERPT in six OECD countries (US, Japan, Canada, Germany, Belgium and Italy). He finds an incomplete ERPT for all the sample countries except for Italy (100%). This is attributed to factors such as the different level of market power prevailing in each country or the ability of the importing country to influence

the world price due to its relatively large size.

However, the majority of the empirical studies regarding ERPT use linear econometric models (i.e log linear, error correction models, VAR, etc) dealing with stationarity and cointegration properties where the dependent (exogenous) variable is the import price regressed on several control/predetermined variables such as exporter's cost, competing prices, income (GDP), and nominal exchange rate between the importing and the exporting country (see for example Woo and Hooper (1984); Hooper and Mann (1989)). The coefficient of the estimated nominal exchange rate variable denotes the elasticity of domestic/importing prices to variations in the exchange rate referred to as the pass-through coefficient. All of these studies consent that the ERPT in the US is incomplete ranging from 50-60%, where the rest (50-40%) of the exchange rate change is offset by changes in the markup (Goldberg and Knetter (1997)). One possible explanation for such asymmetric pass-through is that firms adjust their markups to accommodate the local market environment ( Krugman (1986); Helpman and Krugman (1987)). The study of Feenstra (1989), sheds some light on the explanation of the incomplete ERPT by linking the latter to the presence of imperfect competition. Feenstra (1989) uses a log-linear model and quarterly data over the period 1974:1 to 1987:1 for the U.S. imports of Japanese cars, compact trucks and heavy motorcycles to find that there is a symmetric response of import prices to changes in the bilateral exchange rate and an import tariff.

A number of past studies also investigate the extent of ERPT using disaggregated industry level data. More specifically, Dornbusch (1987) uses two-digit industry level data

to link the incomplete ERPT with micro-economic factors (i.e market concentration, product homogeneity, market shares). Yang (1997), uses monthly data for the 87 (three and four-digit SIC) manufacturing sectors over the period from 1980:12 to 1991:12 in order to estimate the speed of ERPT in the US industry sector. He adopts a two-stage procedure, in which the ERPT elasticities are estimated through a typical log linear model expressed in first differences and these estimates are regressed against several independent variables (costs, market power, market concentration, etc). His findings suggest that ERPT is asymmetric and varies across industries. The degree of pass-through is positively (negatively) correlated to product differentiation, (elasticity of marginal cost). Subsequent work by Taylor (2000) argues that the responsiveness of ERPT depends positively on the level of inflation in a sense that low ERPT in low inflation countries comes as a result of the low inflation environment.

Other studies such as Hfner and Schrder (2002); Choudhri et al. (2005); Hahn (2003); Bailliu and Fujii (2004); Gagnon and Ihrig (2004); Faruquee (2006); Goldberg and Campa (2006*a,b*) have tried to explore the impact of ERPT on import prices and core inflation in the euro zone area or a number of European Monetary Union (EMU) countries by applying standard econometric techniques (log linear models, ECMs and VARs) with controversial results about the rate and the causes of the adjustment.

In an interesting study, Campa and Mnguez (2006), investigate the ERPT into the import prices of twelve EMU countries originating outside the eurozone area. They use monthly time series data over the period 1989:1 to 2001:3 for thirteen different product

categories for each country. They argue that in the short-run, ERPT is incomplete since the estimated pass-rate coefficients (elasticities) are in their vast majority less than one ( $\gamma < 1$ ). However, the same conclusion does not hold in the long run where it is reported a symmetric ERPT. McCarthy (2000) also examines the speed of ERPT on producer and consumer prices for nine selected industrialized countries. He estimates a parsimonious VAR model including variables such as oil price inflation, output gap, nominal exchange rate, import price inflation, consumer and producer price inflation, short-term interest rate and money growth. His results confirm the aforementioned literature suggesting an incomplete ERPT due to market distortions (lack of effective competition).

Subsequent work by Gopinath et al. (2010) investigates the ERPT by developing a dynamic currency choice model. They use monthly time series (at a country level) and panel data (at industry level) on the US import prices for dollar and non-dollar goods over the period 1994-2005 to find that there is a large difference in the pass-through between the two pricing categories. The econometric methodology is based on (fixed effects) OLS estimators employing standard pass-through regression models appeared in first differences. These findings have also been corroborated by the studies of Bhattacharya et al. (2008), Ceglowski (2010), Devereux and Yetman (2010), and Brun-Aguerre et al. (2012).

The impact of market structure on the ERPT nexus is more evident in the recent study of Auer and Schoenle (2016). The authors use annual firm-level data on standard ERPT regression analysis over the period 1994-2005 for the thirty four largest trading partners of the US. They argue that market share affects the rate at which firms react to changing

competitor prices.

Earlier work by Al-Abri and Goodwin (2009) and Aleem and Lahiani (2014) stands apart from those discussed above in that it uses nonlinear econometric methodology. Al-Abri and Goodwin (2009) use a threshold cointegration model (TAR) in order to reveal the determinants of the ERPT in sixteen OECD countries and five categories of imported goods (Food and agricultural products, energy, raw materials, manufacturing, and non-manufacturing). The authors use quarterly data spanning the period 1975:1 to 2002:2 to support that in their non-linear model the import prices respond faster and by a larger degree to nominal exchange rate fluctuations than in the standard log linear models. On the other hand, Aleem and Lahiani (2014) rely on the flexible threshold vector autoregression model (TVAR) to investigate the degree of ERPT rate in Mexico by utilizing monthly seasonally adjusted data from 1994:1 to 2009:11. They find that domestic prices react strongly to a positive one unit exchange rate shock only above the threshold level of the rate of inflation.

Although the issue of ERPT into domestic prices is well documented in the literature, there are few studies focusing on products that are relatively homogeneous and priced in an international market known as *commodities* (i.e petroleum prices, agriculture products, precious metals, etc).

Yanagisawa (2012) uses weekly data for the Japan over the period January 2012 to February 2013 and ECM techniques in order to investigate the ERPT into domestic oil price. He decomposes the pass through structure of gasoline price into two distinct features comprising of the dollar and the exchange rate factor. It is worth mentioning that this

study considers the issue of the “*numeraire*” currency (dollar) for the ERPT into commodity pricing. He finds an incomplete but rather symmetric of the pass-through rate of the dollar factor, a premise also supported by the empirical literature. The opposite result is confirmed when the pass-through of the exchange rate factor is taken into account.

Finally, Akelik and Ogunc (2016) examine the degree of ERPT to domestic fuel prices at different oil market segments in Turkey over the period 2004-2014. They use monthly data and VAR methodology to depict that the ERPT to domestic gasoline prices is considerably fast and just one third of a change in crude oil price is reflected to the gasoline prices. This is attributed to the significant share of taxation on retail prices. On the other hand, they argue that the impact of oil prices on transport services takes a longer time compared to other domestic prices, suggesting that a 10% change in the international crude oil prices is associated with a 0.42% change in consumer inflation at the end of one year.

All in all the majority of the above ERPT papers treat the exchange rate as a cost shifter. They have no distinction between the change in the price of the product and change in the exchange rate. The reason is that the product typically does not have an international price denominated in a specific currency.

### **3.3 Data and Variables**

We use a large unbalanced panel dataset of weekly observations spanning the period from January 1994 to January 2015. The primary sample includes all 28 European Union countries, but the coverage for each country varies, largely because of differences in accession

dates into the EU. All variables are in their natural logarithms expressed in real terms and deflated by the Harmonised Consumer Price index provided by Eurostat. Input cost price (i.e Brent crude oil price) measured in dollars per barrel is taken from the USA Department of Energy (EIA).<sup>2</sup> It is worth mentioning that, the coverage period for the tax-inclusive gasoline price (price at the pump) is more limited than the coverage period for the pre-tax (net) retail gasoline price. The data coverage for the gasoline price series is shown in Table 3.6 in the Appendix.

Pre-tax gasoline retail prices expressed in local currencies are obtained from the Weekly Oil Bulletin.<sup>3</sup> It is worth mentioning that pre-tax prices are used to avoid the possibility that countries with heterogeneous excise tax levels (e.g Italy and Estonia) experience very different percentage responses to one percent change in the underlying marginal cost, solely because the fixed amount of the excise tax moves up the origin of the retail price.

The exchange rate effect is quantified by two indicators: a) The Dollar trade-weighted exchange rate index (1997=100) which is drawn directly from the Federal Reserve Bank of St. Louis, and b) The nominal effective Euro trade-weighted exchange rate index obtained by the European Central Bank. The first term is the change in the trade-weighted value of the dollar (or the consumption weighted dollar exchange rate), and the second term is the change in the number of units of local currency to the dollar.

Specifically the Dollar trade-weighted exchange rate index (commonly known as *broad* index) is the weighted average of the foreign exchange value of the U.S. dollar against the

---

<sup>2</sup>[https://www.eia.gov/dnav/pet/pet\\_pri\\_spt\\_s1\\_d.htm](https://www.eia.gov/dnav/pet/pet_pri_spt_s1_d.htm)

<sup>3</sup><http://ec.europa.eu/energy/en/data-analysis/weekly-oil-bulletin>

currencies of a broad group of major U.S. trading partners (Federal Reserve Bank of St. Louis - FRED (2017))<sup>4</sup>. This index, which will act as the endogenous threshold variable in our model, is used to determine the U.S. dollar purchasing value, and to summarize the effects of dollar appreciation and depreciation against foreign currencies. When the value of the dollar increases, imports to the U.S. become less expensive while exports to other countries become more expensive. In other words, if the index rises (decreases), *ceteris paribus*, the purchasing power of the US dollar also rises (decreases) which will reduce (increase) the cost of imports but will undermine (enhance) the competitiveness of the US exports.<sup>5</sup> Alternatively, if this index rises (decreases), the value of the EU (and of the other foreign countries as well) exports (imports) to the US also rises (decreases) constituting the EU countries as net exporters (importers).

One could argue that ranking or splitting countries according to their exports/imports to and from the USA seems arbitrary. The reason is that many EU countries are not really dependent on the USA, mainly the smaller ones that are much more dependent on exports within the EU (Greece, Portugal, etc). However, the broad index was introduced by the U.S. Federal Reserve Board in 1998 in response to the implementation of the euro (which replaced many of the foreign currencies that were previously used in the earlier index) and to more accurately reflect current U.S. trade patterns. The Federal Reserve selected

---

<sup>4</sup>This index includes the Euro Area, Canada, Japan, Mexico, China, United Kingdom, Taiwan, Korea, Singapore, Hong Kong, Malaysia, Brazil, Switzerland, Thailand, Philippines, Australia, Indonesia, India, Israel, Saudi Arabia, Russia, Sweden, Argentina, Venezuela, Chile and Colombia

<sup>5</sup>Trade-weighted dollar index places importance (weight) to currencies most widely used in international trade, over comparing the value of the U.S. dollar to all foreign currencies. Since the currencies are weighted differently, changes in each currency will have a unique effect on the trade-weighted dollar and their corresponding indexes



26 currencies to use in the broad index, anticipating the adoption of the euro by eleven countries of the European Union (EU). It is noteworthy that when the broad index was introduced, U.S. trade with the 26 represented economies accounted for over 90% of the total U.S. imports and exports ( Federal Reserve Bank of St. Louis - FRED (2017)).

The second exchange rate factor can be represented by the inclusion of the nominal effective Euro trade-weighted exchange rate index. The latter denotes a geometric weighted average of the bilateral exchange rates of the euro against the currencies of a selection of trading partners. More specifically, this indicator is computed against a group of 42 partner countries (EER-42), accounting for roughly 90% of total euro area manufacturing trade in 1999-2001. It is worth mentioning that a fixed weighting scheme is employed in these computations. According to the ECB, the scheme is based on manufacturing trade and takes into account so-called third-market effects, (i.e. competition faced by euro area products in a partner country from products of a third country). This index was first constructed in 1999 and the first update of the weights took place in 2004. Moreover, the overall trade weights underpinning the EER-42 index are updated every five years. Similarly with the other exchange rate index, the interpretation of this indicator is straightforward. In particular, if the index rises (decreases), *ceteris paribus*, the euro appreciates (depreciates) against its major trading countries resulting in a reduction (increase) of the exports (imports).

Based on the above considerations, we argue that the ERPT specifications differ from the standard specifications provided by the IO literature (see among others Galeotti et al, 2003; Deltas, 2008; Polemis and Tsionas, 2017) in the following ways. First, all prices are

in logs and coefficient estimates denote elasticities since there is no other meaningful way to jointly estimate the model involving series from different countries in different units. Second, the retail prices are in local currency, and not in euros. Pre-tax prices are used to avoid the possibility that countries with very different (fixed amount) excise tax levels experience very different percentage responses to one percent change in the underlying marginal cost, solely because the fixed amount of the excise tax moves up the origin of the retail price. Third, the input price is the real price of crude oil (i.e., the price deflated by the US dollar price index). The deflator that we used is the trade-weighted value, but we have also used the consumption-weighted values as a robustness check. Fourth, we have included two exchange rate terms that will be treated in exactly the same way as we treat input prices, i.e., we will have the lags, and in the asymmetric model we will distinguish between positive and negative changes. Note that the two exchange rate terms will be treated in exactly the same way as we treat input prices (i.e., we have the lags, and in the asymmetric model we distinguish between positive and negative changes). They may also be in the co-integration vector, but an alternative is to have the co-integration vector be in a common currency (e.g., euros, under the premise that in the long run pass-through is equal to one).<sup>6</sup> Specifically, the first term is the change in the trade-weighted value of the dollar (or the consumption weighted dollar exchange rate), and the second term is the change in the number of units of local currency to the dollar. These changes will be differences in the

---

<sup>6</sup>In such cases, the basic equation becomes  $\Delta \ln(R_{j,t}^{lc}) = \alpha_j + b_{0,j} \Delta \ln(C_t^r) + b_{1,j} \Delta \ln(C_{t-1}^r) + b_{0,j}^{W\$} \Delta \ln(X_t^{W\$}) + b_{0,j}^{W\$} \Delta \ln(X_{t-1}^{W\$}) + b_{0,j}^{lc/\$} \Delta \ln(X_t^{lc/\$}) + b_{0,j}^{lc/\$} \Delta \ln(X_{t-1}^{lc/\$}) + c_{1,j} \Delta \ln(R_{j,t-1}^{lc}) + d_j [\ln(R_{j,t-1}^{lc}) - k_j - m_j^r \ln(C_{t-1}^r) - m_j^{W\$} \ln(X_{t-1}^{W\$}) - m_j^{lc/\$} \ln(X_{t-1}^{lc/\$}) + \varepsilon_{j,t}]$

Table 3.1: Descriptive statistics

Variables	Observations	Mean	Std	Min	Max
<i>GNP</i>	22645	6.038	0.416	4.908	6.758
<i>GNP<sup>lc</sup></i>	22645	7.345	1.838	4.536	13.60
<i>ln(Brent)</i>	30218	3.704	0.746	2.245	4.949
<i>ln(BrentR)</i>	30218	-0.927	0.778	-2.488	0.386
<i>ln(DolrTWXin)</i>	30218	4.681	0.0906	4.489	4.869
<i>ln(LCtoUSD)</i>	22622	1.091	1.978	-1.241	7.746

*GNP* is the net retail price of gasoline, *GNP<sup>lc</sup>* is the net retail price of gasoline in local currency, *Brent* is the Brent crude oil price, *BrentR* is the Brent crude oil price in trade-weighted real dollars, *DolrTWXin* is the trade-weighted dollar exchange rate index, *LCtoUSD* denotes the units of local currency to USD dollar.

log values of the corresponding variables.<sup>7</sup> Finally, our approach allows for an endogenous treatment of all the regressors and the threshold variable at the same time, contrary to the threshold autoregressive model of Godby et al. (2000).

Table 3.1, provides a complete description of the variables (expressed in natural logarithms) included in this study. As it is evident over the sample period, net retail gasoline prices (not including taxes) averaged 6 dollars per gallon. As it is expected the retail gasoline prices and crude oil fluctuations follow a similar pattern. Specifically, gasoline prices have been rising slightly over the examined period, with a drift of 0.08 cents per week. Regarding the short run price fluctuations it is important to note that the standard deviation of net retail prices (expressed in Euros) is smaller than that of crude oil (Brent) and spot gasoline price (New York) suggesting the existence of a *dampening* effect in the gasoline market ( Polemis and Tsionas (2016); Deltas (2008)). In other words, retail gasoline prices are relatively sticky and do not fully transmit short run fluctuations in the input prices.

<sup>7</sup>We have also estimated the two separate models, using just one exchange rate index in each model but the results were not satisfactory.

## 3.4 Econometric Framework

In this section, we describe the baseline linear one step error correction model (symmetric and asymmetric) that will be contrasted with the pooled panel GMM threshold model developed by Seo and Shin (2016) that accounts for the inclusion of endogenous regressors.<sup>8</sup>

In order to check for the validity of the threshold model we also used the Threshold Error Correction Model (TR), which follows the methodology of Hansen (1999, 2000) in an error correction framework.

### 3.4.1 The Baseline Linear Model

The baseline model follows the estimation approach in Deltas (2008). We estimate first symmetric and asymmetric error correction models (ECMs) at the country level. The basic symmetric error correction model is of the following form:

$$\begin{aligned} \Delta \ln(R_{j,t}^{lc}) = & \alpha_j + \sum_{l=0}^L b_{l,j} \Delta \ln(C_{t-l}^r) + \sum_{l=1}^L c_{l,j} \Delta \ln(R_{t-l}^{lc}) + \sum_{l=0}^L d_{l,j} \Delta \ln(X_{t-l}^{W\$}) \\ & + \sum_{l=0}^L e_{l,j} \Delta \ln(X_{t-l}^{lc/\$}) + z_j \left[ \ln(R_{j,t-1} - k_j - m_j \ln(C_{t-1})) \right] + \varepsilon_{j,t}, \end{aligned} \quad (3.4.1)$$

where  $R_{j,t}^{lc}$  is the retail price of gasoline in country  $j$  and week  $t$  in local currency,  $C_t^r$  is the price of crude oil (common to every country) in trade-weighted real dollars (the price

---

<sup>8</sup>Other methods that deal with endogeneity are the Structural Threshold Error Correction Model (STR), described in Kourtellis et al. (2016) and the Semiparametric Structural Threshold Error Correction Model (SMSTR), developed by Kourtellis et al. (2017). However, in our case given the nature of our panel data set to conserve space we only report the results from the GMM model of Seo and Shin (2016) as this is the most appropriate method to use. However, the empirical results of the other models are available upon request.

in dollars divided by the trade-weighted dollar index),  $X_t^{W\$}$  is the trade-weighted dollar exchange rate index,  $X_t^{lc/\$}$  is the exchange rate of local currency units per dollar,  $R_{j,t}$  is the retail price of gasoline in country  $j$  and week  $t$  in Euros,  $C_{j,t}$  is the price of crude oil (common for every country) in dollars. The dependent variable  $\Delta \ln(R_{j,t}^{lc})$  denotes the change in the log retail price in local currency from week  $t - 1$  to week  $t$  in country  $j$  and similarly for other difference terms. Note that in our models, all prices are in natural logarithms and coefficient estimates denote elasticities since there is no other meaningful way to jointly estimate the models involving series from different countries in different units.<sup>9</sup>

When estimating this regression in one step, the error correction term is multiplied out yielding the linear regression of the form:

$$\begin{aligned} \Delta \ln(R_{j,t}^{lc}) = & \alpha_j - k_j z_j + \sum_{l=0}^L b_{l,j} \Delta \ln(C_{t-l}^r) + \sum_{l=1}^L c_{l,j} \Delta \ln(R_{t-l}^{lc}) \\ & + \sum_{l=0}^L d_{l,j} \Delta \ln(X_{t-l}^{W\$}) + \sum_{l=0}^L e_{l,j} \Delta \ln(X_{t-l}^{lc/\$}) + z_j \ln(R_{j,t-1}) \\ & - z_j m_j \ln(C_{t-1}) + \varepsilon_{j,t}. \end{aligned} \quad (3.4.2)$$

It is worth mentioning that the regression constant is a composite term each component of which is not separately identified in the onestep regression. However, this is not important for assessing the price dynamics or for performing simulations of the retail price

---

<sup>9</sup>We also used the US dollar price index with the consumption-weighted values being a robustness check. However, the empirical results did not pose any significant differences.

response to upstream price changes.<sup>10</sup>

Lastly, the asymmetric model is given by the following equation:

$$\begin{aligned}
\Delta \ln(R_{j,t}^{lc}) = & \alpha_j + \sum_{l=0}^L b_{l,j}^+ \Delta \ln(C_{t-l}^{r+}) + \sum_{l=0}^L b_{l,j}^- \Delta \ln(C_{t-l}^{r-}) + \sum_{l=1}^L c_{l,j}^+ \Delta \ln(R_{t-l}^+) + \sum_{l=1}^L c_{l,j}^- \Delta \ln(R_{t-l}^-) \\
& + \sum_{l=0}^L b_{l,j}^+ \Delta \ln(X_{t-l}^{W\&+}) + \sum_{l=0}^L b_{l,j}^- \Delta \ln(X_{t-l}^{W\$-}) + \sum_{l=0}^L b_{l,j}^+ \Delta \ln(X_{t-l}^{lc/\$+}) + \sum_{l=0}^L b_{l,j}^- \Delta \ln(X_{t-l}^{lc/\$-}) \\
& + d_j \left[ \ln(R_{j,t-1}) - m_j \ln(C_{t-1}) \right] + \varepsilon_{j,t}, \tag{3.4.3}
\end{aligned}$$

where  $\Delta \ln(C_{t-l}^+) = \Delta C_{t-l}$  if  $\Delta C_{t-l} \geq 0$  and zero otherwise,  $\Delta \ln(C_{t-l}^-) = \Delta C_{t-l}$  if  $\Delta C_{t-l} < 0$  and zero otherwise, and similarly for the expressions involving lagged changes in the retail prices and the two exchange rate term. This can be estimated using both the one step and the two step approaches outlined above.

### 3.4.2 The Threshold Model

We use the novel pooled panel GMM threshold method of Seo and Shin (2016). The latter studies a dynamic threshold panel data model, which allows both regressors and threshold effect to be endogenous. Seo and Shin (2016) propose first-difference GMM (FD-GMM) estimators and derive their limiting behaviors based on Newey and Mcfadden (1994), which allows both the fixed threshold effect and the diminishing threshold effect asymptotic framework of Hansen (2000). In order to check for the presence of a threshold

<sup>10</sup>The basic symmetric ECM (see Eq. 3.4.1) can also be estimated in two steps. In order to check the validity of the results, we also ran the other way and found similar results. Due to space competition the results are available upon request.

effect, they rely on bootstrap-based testing procedure.

One could also resort alternatively to a semiparametric specification using local smoothers or splines/series to capture possible turning points. However such methods involve bandwidth choices and they do not lend themselves to estimating sharp turning points/thresholds as it is the case in the threshold model that we adopt in a fully interactive way (Polemis and Stengos (2018); Kourtellos et al. (2016)). Moreover, one important advantage of this methodology is that it avoids the ad hoc, subjective pre-selection of threshold values which has been a major critique of previous studies (Christie (2012)). In contrast to a simple case where the sample is split according to a known pre-assigned threshold value, the method that we use first tests for the presence of such a threshold and then estimates it (see for example Hansen (2000); Caner and Hansen (2004); Kourtellos et al. (2016)). In principle, one can test for additional sample splits, something that we did but and we were not able to detect.

Based on the above, Equation 3.4.1 can be cast in terms of threshold regression model that can be expressed as follows:

$$\begin{aligned}\Delta \ln(R_t^{lc}) &= \alpha_0 + \alpha_1^T \Omega_t + v_t + \varepsilon_t, X_t^{W\$} \leq \gamma, \\ \ln(R_t^{lc}) &= \alpha_0 + \alpha_2^T \Omega_t + v_t + \varepsilon_t, X_t^{W\$} > \gamma,\end{aligned}\tag{3.4.4}$$

where we suppress the country index  $j$  and only use time as subscript.  $X_t^{W\$}$  is the threshold variable,  $\gamma$  is the threshold level and  $\Omega_t$  is a  $d_x \times 1$  vector expressed in first differences

containing all the regressors of the model in a compact form, including also all the lags ( $C_t, R_{t-1}^{lc}, X_t^{W\$}$  and  $X_t^{lc/\$}$ ), while  $\alpha_1$  and  $\alpha_2$  are regime specific coefficients. Moreover,  $\alpha_0$  is the country fixed effect that control for differences across the cross-sectional element (i.e. taxation level, demand and supply characteristics, gasoline market structure), capturing individual heterogeneity. We also include the relevant year (time) fixed effect ( $v_t$ ) which captures the co-movement of the series due to external shocks (Polemis and Stengos (2018)). Finally,  $\varepsilon_t$  denotes the idiosyncratic i.i.d error term.

For concreteness, the above two equations can be integrated into one as follows:

$$\Delta \ln(R_t^{lc}) = \alpha_0 + \alpha_2^T \Omega_t + v_t + \delta^T \Omega_t I(q_t \leq \gamma) + \varepsilon_t, \quad (3.4.5)$$

where  $\delta = \alpha_1 - \alpha_2$ ,  $q_t$  represents the scalar endogenous threshold variable ( $X_t^{W\$}$ ) that splits the sample into two different groups (low and high regime).  $I(\cdot)$  is the indication function denoting the regime defined by the threshold variable and the threshold level  $\gamma$  (sample split value). The indication function takes the value one when the condition in the parenthesis is satisfied and zero otherwise.<sup>11</sup>

We estimate Eq. (3.4.5) using the GMM method of Seo and Shin (2016) as fully described in Asimakopoulos and Karavias (2016). The latter which uses Arellano and Bond (1991) type instruments is more advanced than other threshold methods such as Hansen (1999) and Kremer et al. (2013). This is attributed to the fact that it allows for endogeneity in both the regressors and the threshold variable (Seo and Shin (2016)). The potential

---

<sup>11</sup>The choice of lag length  $p = 2$  is chosen by Akaike's selection Information Criterion (AIC).



endogeneity problem is associated with exchange rate fluctuations in asymmetric gasoline pricing mechanism. While there remains debate in the literature whether fluctuations in the exchange rate drives asymmetric gasoline pricing mechanism or gasoline price asymmetry drives exchange rate volatility, the fact is that the potential for endogeneity exists. As a consequence this model fully incorporates this issue by allowing the exchange rate factor variable (trade-weighted dollar exchange rate index) to be endogenously determined.

## **3.5 Results and Discussion**

This section presents the results of the threshold models along with the benchmark linear specifications (symmetric and asymmetric). In addition, we offer a comparative discussion between the threshold effects and the static panel fixed effects linear specification benchmark models, while we firstly check for the existence of cross-section dependency and stationarity properties of our sample variables by using second generation tests for unit roots.

### **3.5.1 Testing for Cross-section Dependence**

One of the additional complications that arise when dealing with panel data compared to the pure time-series case, is the possibility that the variables or the random disturbances are correlated across the panel dimension. The early literature on unit root and cointegration tests adopted the assumption of no cross-sectional dependence. However, it is common

Table 3.2: Cross-dependence test

Variable	CD test	<i>P</i> value	Correlation	Absolute (Correlation)
$\ln(GNP)$	459.72***	0.000	0.963	0.963
$\ln(GNP^{lc})$	194.70***	0.000	0.456	0.780
$\ln(Brent)$	643.95***	0.000	1.000	1.000
$\ln(BrentR)$	627.60***	0.000	1.000	1.000
$\ln(DolrTWXin)$	627.60***	0.000	1.000	1.000
$\ln(LCtoUSD)$	253.12***	0.000	0.539	0.645

Notes: Under the null hypothesis of cross-sectional independence the CD statistic is distributed as a two-tailed standard normal. Results are based on the test of Pesaran (2004). The p-values are for a one-sided test based on the normal distribution. Correlation and Absolute (correlation) are the average (absolute) value of the off-diagonal elements of the cross-sectional correlation matrix of residuals.  $GNP$ , is the net retail price of gasoline in Euros,  $GNP^{lc}$ , is the net retail price of gasoline in local currency,  $Brent$  is the Brent crude oil price in USD,  $BrentR$  is the Brent crude oil price in trade-weighted real dollars,  $DolrTWXin$  is the trade-weighted dollar exchange rate index,  $LCtoUSD$  denotes the units of local currency to USD dollar. All variables are expressed in natural logarithms. Significant at \*\*\* 1% level of statistical significance.

for macro-level data to violate this assumption which will result in low power and size distortions of tests that assume cross-section independence ( Polemis and Stengos (2018)).

We use the cross-section dependence test proposed by Pesaran (2004). The test is based on the estimation of the linear panel model of the form:

$$y_{it} - \alpha_i + \beta_i^T x_{it} + \mu_{it}, \quad i = 1, \dots, N; \quad T = 1, \dots, T, \quad (3.5.1)$$

where  $T$  and  $N$  are the time and panel dimensions respectively,  $\alpha_i$  is the provincial-specific intercept, and  $x_{it}$  is a  $k \times 1$  vector of regressors, and  $\mu_{it}$  is the random disturbance term.

The null hypothesis in both tests assumes the existence of cross-section correlation:  $Cov(\mu_{it}, \mu_{jt}) = 0$  for all  $t$  and for all  $i \neq j$ . This is tested against the alternative hypothesis that  $Cov(\mu_{it}, \mu_{jt}) \neq 0$  for at least one pair of  $i$  and  $j$ . The Pesaran (2004) test is a type of Lagrange-Multiplier test that is based on the errors obtained from estimating Eq. 3.5.1 by

the OLS method. If the relevant test strongly rejects the null hypothesis of cross-section independence for all the models then we proceed to test for unit roots using tests that are robust to cross-section dependence (the so-called second generation tests for unit roots in panel data). We carry out the first part of the empirical analysis by examining the presence of cross-section dependence. We use the cross-section dependence test (CD test) proposed by Pesaran (2004).

As it is evident from Table 3.5 the relevant test strongly rejects the null hypothesis ( $p$ -value =0.000) of cross-section independence for all the variables. In light of this evidence, we proceed to test for unit roots using tests that are robust to cross-section dependence.

### **3.5.2 Unit Root and Cointegration Testing**

To examine the stationarity properties of the variables in our models we use the second generation unit root tests for unbalanced panel-data proposed by Breitung and Das (2005) and Pesaran (2007). The test results suggest that all the sample variables are integrated of order one (I-1).<sup>12</sup>

In order to investigate whether a long-run equilibrium relationship exists among the variables in our models we implement four cointegration tests proposed by Westerlund (2007) that allow for cross-section dependence. The results of the tests are presented in Table 3.5.2. The results indicate that the null hypothesis of no cointegration can be rejected in most of the cases, revealing that there is a structural relationship between the sample

---

<sup>12</sup>Due to space limitation the results of the unit root testing are available from the authors on request.

Table 3.3: Westerlund ECM panel cointegration tests

Equation	Statistic			
	$G_\tau$	$G_\alpha$	$P_\tau$	$P_\alpha$
$\ln(GNP) = f(\ln(Brent))$	-4.955*** (0.000)	-45.127*** (0.000)	-26.008*** (0.000)	-43.164*** (0.000)
$\ln(GNP^{lc}) = f(\ln(BrentR))$	-2.870*** (0.000)	-17.690*** (0.000)	-8.747 (0.998)	-6.020 (0.995)
$\ln(GNP^{lc}) = f(\ln(DolrTWXin))$	-2.755*** (0.004)	-14.900*** (0.000)	-10.295 (0.849)	-7.307 (0.928)
$\ln(GNP^{lc}) = f(\ln(LCtoUSD))$	-2.432 (0.309)	-16.267*** (0.000)	-13.922*** (0.001)	-15.702*** (0.000)

The test regression was fitted with a constant and trend and four lags and leads. The kernel bandwidth was set according to the rule  $4(T/100)^{2/9}$ . The null hypothesis assumes that there is no co-integration. The numbers in parentheses denote the p-values.  $GNP$ , is the net retail price of gasoline in Euros,  $GNP^{lc}$  is the net retail price of gasoline in local currency,  $Brent$  is the Brent crude oil price in USD,  $BrentR$  is the Brent crude oil price in trade-weighted real dollars,  $DolrTWXin$  is the trade-weighted dollar exchange rate index,  $LCtoUSD$  denotes the units of local currency to USD dollar. All variables are expressed in natural logarithms. Significant at \*\*\* 1% level of statistical significance.

variables.

### 3.5.3 Empirical Findings

This section presents the empirical findings of the baseline (benchmark) model and the TR model along with the necessary asymmetric testing in order to further examine the validation of the ERPT hypothesis on gasoline pricing.

#### 3.5.3.i Benchmark Model Results

In this section we proceed with the exposition of results generated from the benchmark linear specifications that will be contrasted with the threshold model. In this way, we will be able to draw the differences between these results and the traditional benchmark linear specifications in order to focus on issues that were depicted in the threshold model and

are different from the linear baseline one (Polemis and Stengos (2018)). From Table 3.4, it is evident that nearly all of the variables are statistically significant in nearly all of the specifications. However, the relevant signs of most of the regressors entering the linear models (symmetric and asymmetric one) differ drastically revealing that the results are not robust.<sup>13</sup> Specifically, examining the linear asymmetric model, it is evident that the crude oil positive coefficient is larger than its negative counterpart, indicating that the effect of upstream price increases is larger than the price decreases. The relevant estimate for the positive coefficient is equal to 0.29 compared to 0.28 for the negative one. This means that a 10% increase (decrease) of the crude oil will lead on average to a short-run (three week) increase (decrease) of the net retail gasoline price equal to 8.4% (6.5%). In other words, it becomes clearer over the three-week period that retail prices are rising faster than they are falling.

Regarding the exchange rate terms included in the baseline linear model, some interesting results emerge. First, the real effective exchange rate term ( $\Delta \ln(X_t^W)$ ) known also as the “*broad index*” provides an interesting asymmetric pattern since the estimated coefficients when significant alternate their signs. The results point out that when input cost prices (i.e. crude oil prices) are rising, an increase in the broad index initially reduces the level of retail gasoline prices by 5.7% in the short run. This trend is fully reversed two weeks later revealing an increase by 9.8%. In contrast, when crude oil prices are falling, an increase in the broad index leads to an increase in the level of retail prices estimated at 9%

---

<sup>13</sup>We have also estimated the one-step ECM asymmetric model but the results do not differ substantially and therefore not reported here.

approximately in the very short run period (one week). In such a case, the EU countries become net exporters, while the US competitiveness diminishes, since the purchasing power of the US dollar is increased resulting to a decrease in the cost of US imports. However, gasoline prices follow a decreasing rate of return equal to 14% two weeks later. This behaviour is consistent with a fully asymmetric gasoline pricing pattern to ERPT fluctuations. The latter is fully confirmed both by the threshold model results and the impulse response functions described below.

Second, the nominal effective Euro trade-weighted exchange rate effect expressed by  $\Delta \ln(X^{le})$  is positively correlated with the retail gasoline price in all of the specifications of the ECMs. The relevant estimate for the positive coefficient is larger than its negative counterpart (0.49 compared to 0.33). Surprisingly the cointegration-terms (lagged crude oil and retail price) denoting the long-run relationship between the net retail gasoline price and its crude oil marker (Brent crude oil price or New York spot gasoline price) are not statistically significant and not reported in the table. The same finding applies to the two error correction terms representing the speed of adjustment toward the long-run equilibrium. All in all, the empirical findings suggest the absence of short-run and long-run price asymmetry.

Next we apply the necessary linearity tests of the benchmark linear specifications against the non-linear alternative ones given in the threshold model. The tests we use are based on bootstrap critical values of a Wald type heteroskedasticity-consistent test of the null hypothesis against a TR alternative. Specifically all the bootstrapped tests reject linearity in

favour of the threshold model with p-values equal to 0.000 in all cases. As a consequence and in alignment with the aforementioned results, the baseline model does not capture the nonlinear effects of the ERPT mechanism.

### **3.5.3.ii Threshold Model Results**

In this section, we proceed to estimate the threshold model. As it is evident from the inspection of Table 3.4, we find that the optimal threshold level of the ERPT proxied by the trade-weighted dollar exchange rate index is estimated to 4.6232. However, there is a prevailing issue of endogeneity. The latter is associated with the use of the exchange rate term which is treated as an endogenous covariate in our models. This could be explained by the fact that although it has been documented in the literature that exchange rate affects the level of retail gasoline prices (see among others Galeotti et al. (2003); Polemis (2011); Polemis and Fotis (2013)) there is a possibility that the direction of causality might also be reversed. Moreover, it is almost certainly the case that ERPT and upstream pricing adjustment mechanism are not randomly determined among the EU-28 countries throughout the sample period, thus raising the concern that the coefficients of exchange rate and crude oil marker (Brent or New York spot gasoline prices) are biased.

To provide a credible identification strategy that would address this issue and allow interpreting the results in a causal way we followed two approaches. Firstly, we perform the necessary tests to detect endogeneity in the threshold model. The following table depicts the endogeneity test results (see Kourtellos et al. (2017)). It is worth mentioning that, the

proposed test for the endogeneity of the threshold variable ( $X_t^W$ ), is valid regardless of whether the threshold effect is zero or not. Moreover, the test statistic is applicable regardless of whether the regressors are endogenous or exogenous. Under the null hypothesis,  $X_t^W$  is exogenous, while under the alternative hypothesis the threshold variable is endogenous. As it is evident from Table 3.5, the two bootstrap test statistics (White and Homo) reject the null hypothesis. This means that the threshold variable (trade-weighted dollar exchange rate index) is treated as endogenous in our TR model.

In the second stage and after having identified that the threshold variable is endogenous, we rely on the GMM model developed by Seo and Shin (2016).<sup>14</sup> As a consequence, this may lead to biased results. Specifically, the main variable of interest is the trade-weighted dollar exchange rate index. Recall that when entered linearly to the asymmetric model, the coefficients alternated their signs giving an indication of an inconsistent behaviour (see Table 3.4 column 1). On the other hand, the results for the non-linear model with an endogenous threshold do suggest a strong non-linear relationship between retail gasoline prices and exchange rate. The point estimates suggest that the level of real effective exchange rate is positively related to the level of net retail gasoline price. However, it is evident that the trade-weighted dollar exchange rate index is more important in the sample above the

---

<sup>14</sup>We have also used three other panel threshold models namely Threshold Error Correction Model along the lines of Hansen (2000), Structural Threshold Error Correction Model developed by Kourtellos et al. (2016) and Semiparametric Structural Threshold Error Correction Model described in Kourtellos et al. (2017). However, they did not perform well since an (endogenous) threshold variable and endogenous regressors co-exist in the model. Therefore, the analysis relies solely on the GMM model. The results of these models are available upon request.



Table 3.4: Baseline and threshold model results

Method	OLS-baseline model	OLS-baseline model		GMM-Threshold Model	
	(Symmetric)	(Asymmetric)	(Asymmetric)	4,6232	
Threshold Regime	-	-	-	Low	High
Constant	0.0000 (0.9625)	-0.0019 (0.1657)	-	0.0005 (0.9804)	0.0015 (0.6094)
$\Delta \ln(C_t^r)$	0.2914*** (0.0000)	0.2944*** (0.0000)	0.2842*** (0.0000)	0.4612 (0.1999)	0.4132*** (0.0000)
$\Delta \ln(C_{t-1}^r)$	0.1618*** (0.0000)	0.3059*** (0.0000)	0.2099*** (0.0000)	0.1425 (0.6379)	-0.0337 (0.4064)
$\Delta \ln(C_{t-2}^r)$	0.1482*** (0.0000)	0.2439*** (0.0000)	0.1523*** (0.0000)	0.4906** (0.0218)	0.5583*** (0.0000)
$\Delta \ln(X_t^W)$	-0.1145 (0.3663)	-0.5739*** (0.0006)	0.9043*** (0.0006)	2.3203 (0.1947)	2.6203*** (0.0000)
$\Delta \ln(X_{t-1}^W)$	0.0207 (0.8677)	0.0308 (0.8539)	-0.1421 (0.5832)	-0.2669 (0.9262)	-3.3567*** (0.0000)
$\Delta \ln(X_{t-2}^W)$	0.401*** (0.0016)	0.9844*** (0.0000)	-1.4094*** (0.0000)	1.4322 (0.101)	2.249*** (0.0000)
$\Delta \ln(X_t^{lc})$	0.4417*** (0.0000)	0.4923*** (0.0000)	0.3311*** (0.005)	-1.4858 (0.1605)	0.0982 (0.6484)
$\Delta \ln(X_{t-1}^{lc})$	0.3468*** (0.0000)	0.434*** (0.0000)	0.3828*** (0.001)	-0.0092 (0.995)	1.4611*** (0.0000)
$\Delta \ln(X_{t-2}^{lc})$	0.1690*** (0.0004)	0.1564** (0.0139)	0.4773*** (0.0000)	-0.2144 (0.7493)	-0.3303** (0.0349)
$\Delta \ln(R_{t-1}^{lc})$	-0.0952*** (0.0018)	-0.2439*** (0.0000)	-0.0644** (0.0239)	-1.0649*** (0.0000)	0.1018* (0.0778)
$\Delta \ln(R_{t-2}^{lc})$	-0.0466** (0.033)	-0.0643* (0.0961)	-0.032 (0.2529)	-0.0601 (0.5902)	0.0166 (0.7411)
$\ln(R_{t-1})$	-0.0001 (0.5907)	-	-	-0.0004 (0.7487)	-0.0002 (0.6258)
$\ln(C_{t-1})$	0.0098* (0.0647)	-	-	-0.0560 (0.7330)	-0.0230*** (0.0060)
$Error\ Correction_{t-1}$	-	0.0000 (0.9411)	-0.0000 (0.7491)	-	-
Adjusted $R^2$	0.347	0.371	-	-	-
$J$ statistic	-	-	-	1.512	-
$D - W$ Pvalue	0.8439	0.9205	-	0.1380	-
$Sup$ -wald Statistic	-	-	-	46.3847	-
$Sup$ -wald Boot $P$ value	-	-	-	0.0041***	-
Observation	22645	22645	22645	22645	22645

Column one refers to the pooled panel OLS results of the symmetric model (baseline). Column two refers to the pooled panel results of the asymmetric model which is estimated with a two-step ECM procedure. Column three presents the estimations of the dynamic GMM of Seo and Shin (2016). The threshold variable is the trade-weighted dollar exchange rate index,  $X_t^W$ . All variables are instrumented with its lag terms.  $R_t^{lc}$  is the net retail price of gasoline in local currency,  $C_t^r$  is the price of crude oil in trade-weighted real dollars,  $X_t^{W\$}$  is the trade-weighted dollar exchange rate index,  $X_t^{lc/\$}$  is the exchange rate of local currency units per dollar,  $R_{j,t}$  is the net retail price of gasoline in Euros,  $C_{j,t}$  is the price of crude oil in dollars,  $\Delta \ln(R_{j,t}^{lc})$  is the change in the log retail price in local currency from week  $t-1$  to week  $t$  in country  $j$  and similarly for other difference terms.  $D - W$  denotes the Durbin-Watson test for autocorrelation in panel data. All models include time and country fixed effects.\*\*\* Significance at 1%, \*\* Significance at 5%, \* Significance at 10%.

Table 3.5: Threshold endogenous test results

Polynomial	Wald ( <i>White</i> )	Wald ( <i>Homo</i> )	Boot <i>P</i> ( <i>White</i> )	Boot <i>P</i> ( <i>Homo</i> )	GCV
0	15.0911	20.2219	0.0000***	0.0000***	0.00039
1	21.4792	22.1012	0.0000***	0.0101***	0.000379
2	24.8343	24.0289	0.0202**	0.0606*	0.000377

This table presents the endogeneous tests results suggested by Kourtellos et al. (2017) at varying polynomials. Boot (*White*) and Boot (*Homo*) are corresponding bootstrap critical values at 5% significant level. \*\*\* Significance at 1%, \*\* Significance at 5%, \* Significance at 10%.

threshold (high regime) since the relevant coefficient (2.6203) is statistically significant.

This means that a 10% increase (decrease) in the level of exchange rate leads to a 26.2% increase (decrease) in the retail gasoline price in the short-run.

The large (asymmetric) change in the gasoline retail price due to an increase in the level of exchange rate (euro appreciation) might be explained by the finding that exchange rate pass through usually differs from pass through of the international price of gasoline which is expressed in USD. One possible explanation for this discrepancy can be attributed to the fact that changes in exchange rate are stronger or more weakly correlated with economic activity in a country (and hence domestic demand) than changes in gasoline prices in USD. One other possible source of the pass through differences might be contractual since supply contracts (e.g refinery to gas stations, or at other parts of the supply chain) are pegged/indexed to the dollar price of oil. This means that all the charges that the EU refining companies pass on to their clients (e.g wholesalers, large final customers, and unbranded petrol stations) are expressed in USD (i.e premia, base pricing, etc). Therefore, some contracts automatically pass through the dollar price change, because they are simply indexed to the USD price.

This finding gives sufficient evidence that for net EU exporting countries (high regime), fluctuations in the real effective exchange rate of the US against its major EU trading partners does affect the level of net retail gasoline prices and subsequently the asymmetric pricing mechanism. In other words, the ERPT model reveals differences in gasoline pricing changes between net exporters and net importers EU countries. The relevant differences explain the asymmetric gasoline pricing whereby price increases transmit faster than price decreases (“*rockets and feathers*” hypothesis). This findings is fully confirmed by the relevant asymmetric gasoline testing below.

It is also worth mentioning that the magnitude of the relevant elasticity exceeds unity denoting that ERPT is almost complete. This finding runs contrary to the existing studies where the relevant estimated elasticity ranges from 0.4 to 0.6 (see for example Krugman (1986); Helpman and Krugman (1987); Feenstra (1989); Goldberg and Knetter (1997)).

Notably, the other control variables have the expected signs and are all statistically significant for values above the threshold (high regime). Similarly to the linear model, the upstream oil price marker (Brent crude oil price) is positively correlated with the net retail gasoline price as it was expected. The relevant short-run price elasticity is estimated to 0.413. This means that a 10% increase (decrease) of the Brent crude oil price will lead to a short-run increase (decrease) of the net retail gasoline price equal to 4.13%. This pattern does not change since the input price coefficient remains statistically significant even when the number of lags is set to two (0.5583). Regarding, the second exchange rate term for the net exporting countries (high regime), we argue that the relevant coefficients are statically

significant alternating their signs only when one and two lags are present (1.4611 and -0.3303 respectively). Surprisingly the lagged retail price cointegration term ( $\ln(R_{t-1})$ ) is not statistically significant below and above the threshold.

### 3.5.3.iii Asymmetric Testing

Having estimated the GMM we proceed to capture possible asymmetries that arise from differential responses of net retail gasoline price changes to positive and negative fluctuations in the exchange rate.

Equations 3.4.4 above provide a less restrictive formulation than what has been traditionally estimated in the literature. By using the partial or full adjustment models, we impose a threshold at zero. We then proceed to estimate possible asymmetries that arise from differential responses of gasoline price changes ( $\Delta \ln(R_t^{lc})$ ) to positive and negative changes in the exchange rate fluctuations ( $\Delta \ln(X_t^W)$ ). Following the spirit of Godby et al. (2000), we provide a direct test for asymmetric behaviour around the estimated threshold. In this case, the null hypothesis of no asymmetry is expressed as:  $H_0 : \alpha = \delta$ , where  $\alpha = (\alpha_0, \alpha'_1, \dots, \alpha'_k)^T$  and  $\delta = (\delta_0, \delta'_1, \dots, \delta'_k)^T$ , and the alternative hypothesis of asymmetric gasoline pricing can be formulated as  $H_1 : \alpha \neq \delta$ . The test we use is based on bootstrap critical values of a Wald type heteroskedasticity-consistent test of the null hypothesis against the existence of an asymmetric gasoline adjustment mechanism (see for example Hansen (1996); Godby et al. (2000)). In other words failure to reject the null hypothesis implies that there is no significant threshold (no asymmetry).

From Table 3.4, we notice that the null hypothesis is strongly rejected with a SupWald Bootstrapped P-value for the GMM equal to 0.0041. In this case, we can safely argue that gasoline asymmetry is present in the EU oil industry. These results are in alignment with some of the empirical studies reported in the literature (see for example Borenstein and Shepard (1996); Deltas (2008); Polemis (2011); Greenwood-Nimmo and Shin (2013); Kristoufek and Lunackova (2015); Polemis and Tsionas (2016)). One possible reason for this behaviour might be attributed to the fact that in such a case, the profit function is inherently asymmetric. If prices are too high, the costs to profit of a sub-optimal level of sales is partly offset by the higher price (and hence profit margin) of each unit sold. But if prices are too low, beyond some point the firm will be selling more units, and each of them at a loss, so that the quantity and price effects on profits reinforce rather than offset each other. We notice that all underlying estimated equations pass a battery of diagnostic tests. Specifically, the reported J-statistic test indicates that the instrument list satisfies the orthogonality conditions in all of the specifications, since the null hypothesis that the over-identifying restrictions are valid cannot be rejected.

We proceed with the sensitivity analysis of the model to large exogenous shocks generated by the volatility of crude oil price over a simulated ten month period under both a high and low regime.<sup>18</sup> For this reason, we opt for a panel VAR model to capture the impact of an exogenous factor (e.g. crude oil price), on the sample variables. It is stressed that for each estimated group, all variables within the panel VAR enter as endogenous. For

exposition purposes, we present a first order  $3 \times 3$  panel VAR model as follows:

$$X_{it} = \mu_i + \mu_{it-1} + \gamma_t + \mu_i + e_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (3.5.2)$$

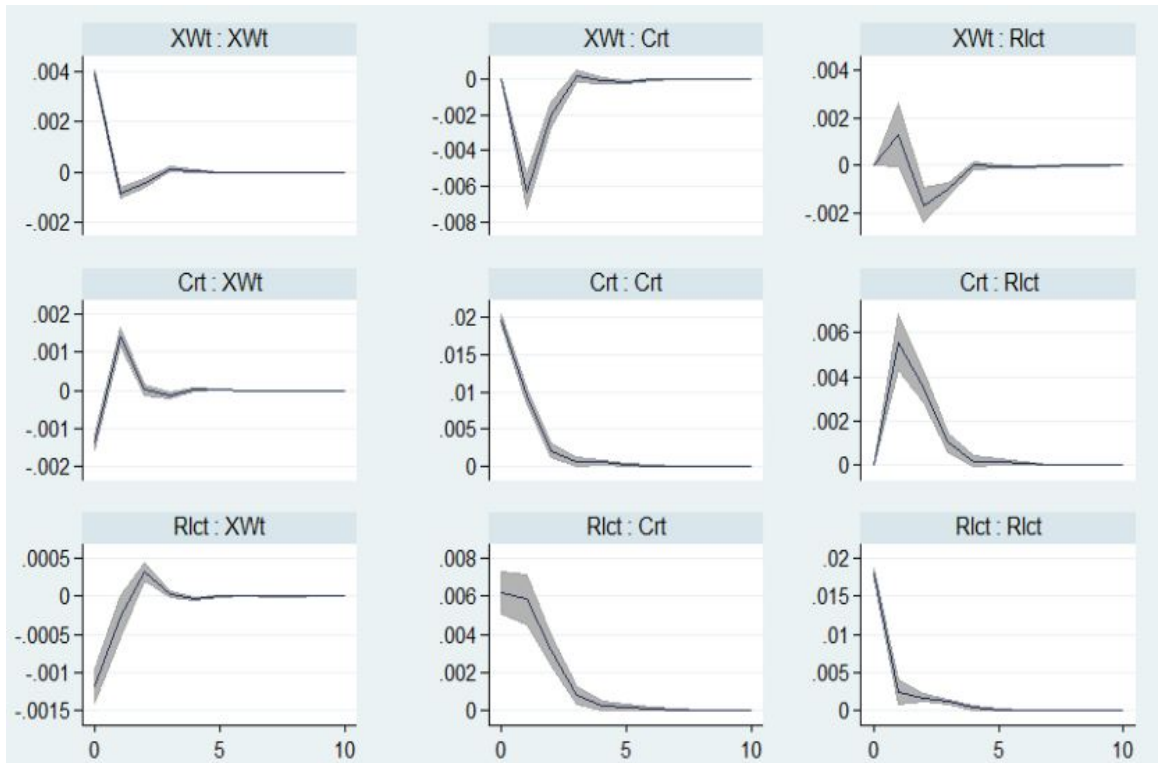
where  $X_{it}$  is a vector of four random variables, that is the net retail price of gasoline in local currency ( $R_t^{lc}$ ), the price of crude oil in tradeweighted real dollars ( $C_t^r$ ), and finally the trade-weighted dollar exchange rate index ( $X_t^W$ ).<sup>15</sup> Thus,  $\Phi$  is  $3 \times 3$  matrix of coefficients,  $\mu_i$  is a vector of  $\mu$  individual effects. The  $\gamma_t$  stands for the time fixed effects and  $\mu_i$  are the state fixed effects that control for differences across regions (e.g. differences in technology used in the production process, economic conditions, environmental legislation). Finally  $e_{it}$  is the idiosyncratic term that is assumed to be i.i.d.

The IRFs derived from the unrestricted panel VAR in the case of the high regime are reported in Figure 3.5.3.iii. This diagram plots the response of each variable in the Panel VAR to its own innovation and to the innovations of the other variables within a 95% confidence interval (CI).

By a careful inspection of the relevant figure one could notice that the response of the threshold variable ( $X_t^W$ ) to its own innovations is downward sloping but almost positive across the whole simulated time period (ten months). Similarly, the response of crude oil price ( $C_t^r$ ), to innovations generated by the broad index is negative in the very short-run

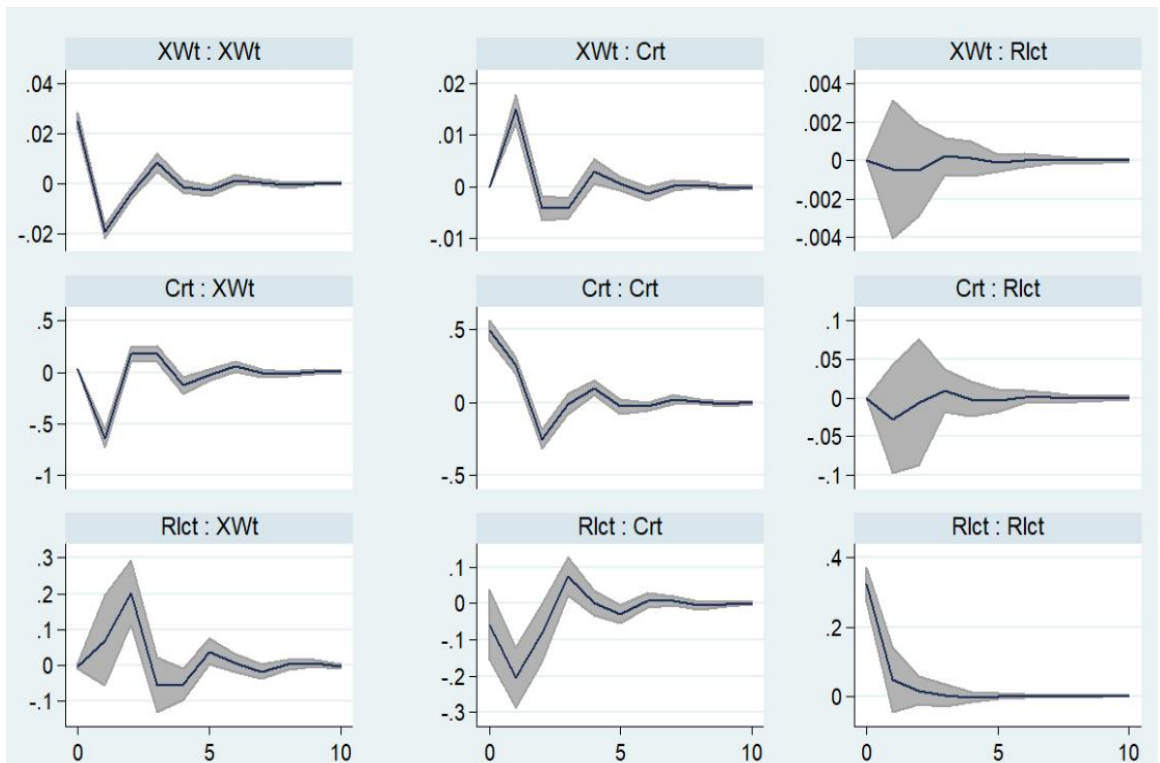
<sup>15</sup>The inclusion of the other exchange rate term variable ( $X^{lc}$ ) in the panel VAR provided results which are available by the authors on request.

Figure 3.1: Impulse Response Functions (IRFs) under high regime



Each row of the diagram shows the response of each variable to a one standard deviation shock of the other variables entered in panel VAR. In order to estimate the IRFs we use the Cholesky decomposition method adjusted by the degrees of freedom.  $R_t^{lc}$  is the net retail price of gasoline in local currency,  $C_t^{[r]}$  is the price of crude oil in trade-weighted real dollars,  $X_t^W$  is the trade-weighted dollar exchange rate index. Standard errors are 5% on each side generated by Monte-Carlo with 500 repetitions. The grey shaded area denote the 95% confidence bands.

Figure 3.2: Impulse Response Functions (IRFs) under low regime



Each row of the diagram shows the response of each variable to a one standard deviation shock of the other variables entered in panel VAR. In order to estimate the IRFs we use the Cholesky decomposition method adjusted by the degrees of freedom.  $R_t^{lc}$  is the net retail price of gasoline in local currency,  $C_t^{[r]}$  is the price of crude oil in trade-weighted real dollars,  $X_t^W$  is the trade-weighted dollar exchange rate index. Standard errors are 5% on each side generated by Monte-Carlo with 500 repetitions. The grey shaded area denote the 95% confidence bands.



(e.g. for the first two simulated months) where it reaches its low turning into positive after the second month and stabilizing thereafter. This finding indicates that the effect of an exogenous disturbance (e.g. exchange rate fluctuation) within the EU countries on the crude oil price is rather than short-lived and dies out after a three weeks period.

However, when we account for the effect of the exchange rate term on the net retail price of gasoline (expressed in local currency) in the high regime, the response pattern has clearly a different behaviour. Specifically, the innovations generated by a one standard deviation shock in the broad index are positive for the first two weeks showing a decreasing rate of return from the second week and thereafter. Subsequently, from the third simulated week and henceforth, the volatility pattern expressed in the very short-run seems to converge to a steady state equilibrium. This outcome reveals that the effect on the retail gasoline prices generated by an exchange rate shock is transitory for the EU countries in the high threshold regime.

We move on with the analysis of the responses of the crude oil price ( $C_t^r$ ) to its own innovation and to the rest variables entered in the panel VAR. As it becomes clear, from the relevant figure (second row), we observe similar findings, as the response of the broad index to crude oil price fluctuations is found to be positive and statistically significant one week after the shock. The cumulative effect reaches its peak (0.15%) during the first month, and then converges toward zero after the end of the fourth month. However, the increasing trend stops within the next month of the peak response.

similar pattern is observed when we analyze the impact of the retail gasoline price to

a one standard deviation shock in the input cost price (crude oil). In other words, we argue that the effect is positive only in the short run (two months at most) and fades away in the medium term (one month after the initial shock). This representation provides a solid illustration of an asymmetric gasoline price response pattern for the sample countries above the threshold.

The discussion now turns to the investigation of the response pattern of the sample variables under the low threshold regime (see Figure 3.5.3.iii). As a general statement, we argue that the responses of the sample variables to innovation shocks exhibit non-linear patterns, which seem to converge in the medium term (one month after the shock). It is worth emphasising that the response of the threshold variable (broad index) to a one standard deviation shock in the crude oil price follows a positive and statistically significant trend for the whole simulated period. Surprisingly, the impact of retail gasoline price to crude oil price fluctuations is negative in the very short-run turning into positive thereafter. However, the confidence bands become wide at the end of the second month making the response of  $R_t^{lc}$  to crude oil price fluctuations insignificant in the medium-term. All in all, the asymmetric response converges one month after the shock.

Finally, if we try to compare the IRFs between the two threshold regimes (high and low), some interesting remarks emerge. First, the response of retail gasoline price to a one standard-deviation shock of the crude oil price is more abrupt in the high threshold regime. However, there is a common convergence trend in both regimes after a three week period. This finding coincides with other studies (see for example Polemis and Tsionas

(2016)) unraveling the absence of a sluggish price-adjustment mechanism. Second, a crude oil shock in both regimes is short-lived. Third, the exchange rate term behaviour to crude oil price innovations, varies significantly under the two regimes. Specifically, in the high threshold regime, the impact in the very-short run period is positive while the opposite holds in the low regime. This result is a clear evidence that asymmetric pricing effect can be attributed to ERPT.

### **3.6 Concluding Remarks**

This paper provides new insights into *rockets and feathers* hypothesis since it tries to investigate the impact of ERPT on asymmetric gasoline pricing mechanism. For this reason we use a large weekly panel of EU-28 countries over the period January 1994 to January 2015. Our pooled panel GMM threshold model follows the spirit of Seo and Shin (2016) and allows for the existence of a threshold effect with endogenous regressors.

In this study we use a bootstrap procedure to test the null hypothesis of a linear (symmetric) formulation against a TR alternative. Moreover, we provide a direct test for asymmetric behaviour around the estimated threshold. The results of the baseline model (expressed in symmetric and asymmetric formulation) compared with the TR model that we use in the present study reveal significant differences in the interpretation of the key variable of interest (real effective exchange rate). This means that the baseline model does not capture the nonlinear effects stemmed from the existence of a threshold according to the bootstrapped  $P$ -values of the relevant linearity tests. As a consequence, the threshold

model is better suited to assess these effects on gasoline price mechanism under the two different regimes of ERPT.

The empirical findings reveal that the threshold variable expressed by the trade-weighted dollar exchange rate index is statistically significant only in the sample above the threshold (high regime). This means that for the net EU exporting countries, fluctuations in the real effective exchange rate of the US against its major EU trading partners does affect the level of pre-tax retail gasoline prices with the relevant elasticity exceeding unity (complete ERPT). Moreover, all the relevant statistical tests reject the null hypothesis that there is no significant threshold and thus an asymmetric adjustment gasoline mechanism prevails.

Lastly, the IRFs results uncover that the responses of all variables to crude oil price innovations (shocks) do not exhibit sizeable effects. Specifically, all shocks take place in the very-short run (i.e. less than two months) and die out (converge) in the medium run (i.e. four months after the initial shock). Nevertheless, none of these shocks has a long lasting effect, since the exchange rate term indices of the sample countries (below and above the threshold) return back to their initial equilibrium positions. The IRFs confirm our previous findings that ERPT can explain better the asymmetric gasoline adjustment mechanism within the EU periphery.

## **Appendix**

Figure 3.3: Sample coverage broken down by EU country and year

Year	AT	BE	BU	CR	CY	CZ	DE	ES	FI	FR	GER	GR	HU	IR	IT	LA	LI	LU	MA	NL	PL	PO	RO	SK	SL	ES	SW	UK	
1994																													
1995																													
1996																													
1997																													
1998																													
1999																													
2000																													
2001																													
2002																													
2003																													
2004																													
2005																													
2006																													
2007																													
2008																													
2009																													
2010																													
2011																													
2012																													
2013																													
2014																													
2015*																													

Data available only for January. Dark shaded areas correspond to the case when both pre-tax and post-tax data are available. Light shaded areas when only pre-tax or post-tax data are available. Unshaded areas when no data are available.

## References

- Afonso, A. and Jalles, J. T. (2013), 'Growth and productivity: The role of government debt', *International Review of Economics Finance* **25**, 384–407.
- Akelik, F. and Ogunc, F. (2016), 'Pass-through of crude oil prices at different stages in turkey', *Central Bank Review* **16**(1), 41–51.
- Al-Abri, A. S. and Goodwin, B. K. (2009), 'Re-examining the exchange rate pass-through into import prices using non-linear estimation techniques: Threshold cointegration', *International Review of Economics Finance* **18**(1), 142–161.
- Aleem, A. and Lahiani, A. (2014), 'A threshold vector autoregression model of exchange rate pass-through in mexico', *Research in International Business and Finance* **30**, 24–33.
- Andrews, D. W. (1994), 'Chapter 37 empirical process methods in econometrics', *Handbook of Econometrics* pp. 2247–2294.
- Andrews, D. W. K. (1987), 'Consistency in nonlinear econometric models: A generic uniform law of large numbers', *Econometrica* **55**(6), 14–65.
- Arellano, M. and Bond, S. (1991), 'Some tests of specification for panel data: Monte carlo evidence and an application to employment equations', *The Review of Economic Studies* **58**(2), 277–297.
- Asimakopoulos, S. and Karavias, Y. (2016), 'The impact of government size on economic growth: A threshold analysis', *Economics Letters* **139**, 65–68.
- Auer, R. A. and Schoenle, R. S. (2016), 'Market structure and exchange rate pass-through', *Journal of International Economics* **98**, 60–77.
- Bagnai, A. and Ospina, C. A. M. (2018), 'Asymmetries, outliers and structural stability in the us gasoline market', *Energy Economics* **69**, 250–260.
- Bailliu, J. and Fujii, E. (2004), 'Exchange rate pass-through and the inflation environment in industrialized countries: An empirical investigation', *SSRN Electronic Journal* .

- Bhattacharya, P. S., Karayalcin, C. A. and Thomakos, D. D. (2008), 'Exchange rate pass-through and relative prices: An industry-level empirical investigation', *Journal of International Money and Finance* **27**(7), 1135–1160.
- Blair, B. F., Campbell, R. C. and Mixon, P. A. (2017), 'Price pass-through in us gasoline markets', *Energy Economics* **65**, 42–49.
- Borenstein, S., Cameron, A. C. and Gilbert, R. (1997), 'Do gasoline prices respond asymmetrically to crude oil price changes?', *The Quarterly Journal of Economics* **112**(1), 305–339.
- Borenstein, S. and Shepard, A. (1996), 'Dynamic pricing in retail gasoline markets', *The RAND Journal of Economics* **27**(3), 429–451.
- Breitung, J. and Das, S. (2005), 'Panel unit root tests under cross-sectional dependence', *Statistica Neerlandica* **59**(4), 414–433.
- Brun-Aguerre, R., Fuertes, A.-M. and Phylaktis, K. (2012), 'Exchange rate pass-through into import prices revisited: What drives it?', *Journal of International Money and Finance* **31**(4), 818–844.
- Bumpass, D., Ginn, V. and Tuttle, M. (2015), 'Retail and wholesale gasoline price adjustments in response to oil price changes', *Energy Economics* **52**, 49–54.
- Campa, J. M. and Mnguez, J. M. G. (2006), 'Differences in exchange rate pass-through in the euro area', *European Economic Review* **50**(1), 121–145.
- Caner, M., Grennes, T. J. and Khler-Geib, F. F. N. (2010), 'Finding the tipping point - when sovereign debt turns bad', *SSRN Electronic Journal* .
- Caner, M. and Hansen, B. E. (2004), 'Instrumental variable estimation of a threshold model', *Econometric Theory* **20**(05), 813–843.
- Cecchetti, S. G., M S, M. and Fabrizio, Z. (2011), 'The real effects of debt', *Bank For International Settlements Working Papers* **352**.
- Cecchetti, S., Mohanty, M. and Zampolli, F. (2011), 'The real effects of debt', *BIS Working Papers* **20**(352).
- Ceglowski, J. (2010), 'Exchange rate pass-through to bilateral import prices', *Journal of International Money and Finance* **29**(8), 1637–1651.
- Chan, K. S. (1993), 'Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model', *The Annals of Statistics* **21**(1), 520–533.
- Choudhri, E. U., Faruquee, H. and Hakura, D. S. (2005), 'Explaining the exchange rate pass-through in different prices', *Journal of International Economics* **65**(2), 349–374.

- Christie, T. (2012), 'The effect of government spending on economic growth: Testing the non-linear hypothesis', *Bulletin of Economic Research* **66**(2), 183–204.
- Delgado, M. A. and Hidalgo, J. (2000), 'Nonparametric inference on structural breaks', *Journal of Econometrics* **96**(1), 113–144.
- Deltas, G. (2008), 'Retail gasoline price dynamics and local market power\*', *The Journal of Industrial Economics* **56**(3), 613–628.
- Devereux, M. B. and Yetman, J. (2010), 'Price adjustment and exchange rate pass-through', *Journal of International Money and Finance* **29**(1), 181–200.
- Dornbusch, R. (1987), 'Exchange rates and prices', *American Economic Review* **77**, 93–106.
- Eckert, A. (2002), 'Retail price cycles and response asymmetry', *Canadian Journal of Economics/Revue Canadienne d'Economie* **35**(1), 52–77.
- Eleftheriou, K., Nijkamp, P. and Polemis, M. L. (2018), 'Asymmetric price adjustments in us gasoline markets: impacts of spatial dependence on the rockets and feathers hypothesis', *Regional Studies* **53**(5), 667–680.
- Fabra, N. and Reguant, M. (2014), 'Pass-through of emissions costs in electricity markets', *American Economic Review* **104**(9), 2872–2899.
- Faruquee, H. (2006), 'Exchange rate pass-through in the euro area', *IMF Staff Papers* **53**(1), 63–88.
- Federal Reserve Bank of St. Louis - FRED (2017), 'Board of governors of the federal reserve system (us), trade weighted u.s. dollar index: Broad [twexb]'. data retrieved from <https://fred.stlouisfed.org/series/TWEXB>.
- Feenstra, R. (1989), 'Symmetric pass-through of tariffs and exchange rates under imperfect competition: An empirical test', *Journal of International Economics* **27**(1-2), 25–45.
- Gagnon, J. E. and Ihrig, J. (2004), 'Monetary policy and exchange rate pass-through', *International Journal of Finance and Economics* **9**(4), 315–338.
- Galeotti, M., Lanza, A. and Manera, M. (2003), 'Rockets and feathers revisited: an international comparison on european gasoline markets', *Energy Economics* **25**(2), 175–190.
- Godby, R., Lintner, A. M., Stengos, T. and Wandschneider, B. (2000), 'Testing for asymmetric pricing in the canadian retail gasoline market', *Energy Economics* **22**(3), 349–368.
- Goldberg, L. and Campa, J. M. (2006a), 'Distribution margins, imported inputs, and the sensitivity of the cpi to exchange rates', *NBER Working Paper* (12121).



- Goldberg, L. and Campa, J. M. (2006b), 'Pass through of exchange rates to consumption prices: What has changed and why?', *NBER Working Paper* (12547).
- Goldberg, P. and Knetter, M. (1997), 'Goods prices and exchange rates: What have we learned?', *Journal of Economic Literature* **35**(3), 1243–1272.
- Gopinath, G., Itskhoki, O. and Rigobon, R. (2010), 'Currency choice and exchange rate pass-through', *American Economic Review* **100**(1), 304–336.
- Greenwood-Nimmo, M. and Shin, Y. (2013), 'Taxation and the asymmetric adjustment of selected retail energy prices in the uk', *Economics Letter* **121**, 411–416.
- Hahn, E. (2003), 'Pass-through of external shocks to euro area inflation', *ECB Working Papers* (243).
- Hansen, B. E. (1996), 'Inference when a nuisance parameter is not identified under the null hypothesis', *Econometrica* **64**(2), 413–430.
- Hansen, B. E. (1999), 'Threshold effects in non-dynamic panels: Estimation, testing, and inference', *Journal of Econometrics* **93**(2), 345–368.
- Hansen, B. E. (2000), 'Sample splitting and threshold estimation', *Econometrica* **68**, 575–603.
- Hansen, B. E. (2011), 'Threshold autoregression in economics', *Statistics and Its Interface* **4**(2), 123–127.
- Hausman, J. A. (1978), 'Specification tests in econometrics', *Econometrica* **46**(6), 1251–1271.
- Heckman, J. J. (1979), 'Sample selection bias as a specification error', *Econometrica* **47**(1), 153–161.
- Helpman, E. and Krugman, P. R. (1987), *Market structure and foreign trade: increasing returns, imperfect competition, and the international economy*, Vol. 1, MIT Press.
- Henderson, D. J., Parmeter, C. F. and Su, L. (2015), 'Nonparametric threshold regression: estimation and inference', *Research Collection School of Economics* pp. 1–62.
- Hooper, P. and Mann, C. L. (1989), 'Exchange rate pass-through in the 1980s: The case of u.s. imports of manufactures', *Brookings Papers on Economic Activity* **20**(1), 297–337.
- Horowitz, J. L. (1992), 'A smoothed maximum score estimator for the binary response model', *Econometrica* **60**(3), 505–531.
- Hfner, F. P. and Schrder, M. (2002), 'Exchange rate pass-through to consumer prices: A european perspective', *SSRN Electronic Journal* .

- Johnson, R. N. (2002), 'Search costs, lags and prices at the pump', *Review of Industrial Organization* **20**(1), 33–50.
- Kapetanios, G. (2009), 'Testing for exogeneity in threshold models', *Econometric Theory* **26**(1), 231–259.
- Kourtellos, A., Stengos, T. and Sun, Y. (2017), 'Endogeneity in semiparametric threshold regression', *SSRN Electronic Journal* .
- Kourtellos, A., Stengos, T. and Tan, C. M. (2013), 'The effect of public debt on growth in multiple regimes', *Journal of Macroeconomics* **38**, 35–43.
- Kourtellos, A., Stengos, T. and Tan, C. M. (2016), 'Structural threshold regression', *Econometric Theory* **32**(4), 827–860.
- Kreinin, M. E. (1977), 'The effect of exchange rate changes on the prices and volume of foreign trade', *International Monetary Fund Staff Papers* **24**(2), 297–329.
- Kremer, M., Lee, J., Robinson, J. and Rostapshova, O. (2013), 'Behavioral biases and firm behavior: Evidence from kenyan retail shops', *American Economic Review* **103**(3), 362–368.
- Kristoufek, L. and Lunackova, P. (2015), 'Rockets and feathers meet joseph: Reinvestigating the oilgasoline asymmetry on the international markets', *Energy Economics* **49**, 1–8.
- Krugman, P. (1986), 'Pricing to market when the exchange rate changes', *NBER Working Paper* (1926).
- Mann, J. (2016), 'Rockets and feathers meet markup margins: Applications to the oil and gasoline industry', *Canadian Journal of Economics/Revue canadienne d'conomique* **49**(2), 772–788.
- Mccarthy, J. (2000), 'Pass-through of exchange rates and import prices to domestic inflation in some industrialized economies', *SSRN Electronic Journal* .
- Newey, W. K. and Mcfadden, D. (1994), 'Chapter 36 large sample estimation and hypothesis testing', *Handbook of Econometrics* pp. 2111–2245.
- Pesaran, H. (2004), 'General diagnostic tests for cross section dependence in panels', *CE-Sifo Working Papers* **69**.
- Pesaran, M. H. (2007), 'A simple panel unit root test in the presence of cross-section dependence', *Journal of Applied Econometrics* **22**(2), 265–312.
- Poirson, H., Ricci, L. A. and Pattillo, C. A. (2002), 'External debt and growth', *IMF Working Papers* **02**(69).

- Polemis, M. L. (2011), 'Competition and price asymmetries in the greek oil sector: an empirical analysis on gasoline market', *Empirical Economics* **43**(2), 789–817.
- Polemis, M. L. and Fotis, P. N. (2013), 'Do gasoline prices respond asymmetrically in the euro zone area? evidence from cointegrated panel data analysis', *Energy Policy* **56**, 425–433.
- Polemis, M. L. and Stengos, T. (2018), 'Does competition prevent industrial pollution? evidence from a panel threshold model', *Business Strategy and the Environment* **28**(1), 98–110.
- Polemis, M. L. and Tsionas, M. G. (2016), 'An alternative semiparametric approach to the modelling of asymmetric gasoline price adjustment', *Energy Economics* **56**, 384–388.
- Potter, S. M. (1995), 'A nonlinear approach to us gnp', *Journal of Applied Econometrics* **10**(2), 109–125.
- Radchenko, S. (2005), 'Oil price volatility and the asymmetric response of gasoline prices to oil price increases and decreases', *Energy Economics* **27**(5), 708–730.
- Robinson, P. M. (1983), 'Nonparametric estimators for time series', *Journal of Time Series Analysis* **4**(3), 185–207.
- Robinson, P. M. (1986), 'On the consistency and finite-sample properties of nonparametric kernel time series regression, autoregression and density estimators', *Annals of the Institute of Statistical Mathematics* **38**(3), 539–549.
- Schclarek, A. (2004), 'Debt and economic growth in developing and industrial countries', *Working Papers, Lund University, Department of Economics* **2005:34**.
- Seo, M. H. and Linton, O. (2007), 'A smoothed least squares estimator for threshold regression models', *Journal of Econometrics* **141**(2), 704–735.
- Seo, M. H. and Shin, Y. (2016), 'Dynamic panels with threshold effect and endogeneity', *Journal of Econometrics* **195**(2), 169–186.
- Taylor, J. B. (2000), 'Low inflation, pass-through, and the pricing power of firms', *European Economic Review* **44**(7), 1389–1408.
- Tong, H. and Lim, K. S. (1980), 'Threshold autoregression, limit cycles and cyclical data', *Journal of the Royal Statistical Society: Series B (Methodological)* **42**(3), 245–268.
- Vaart, A. W. v. d. and Wellner, J. A. (2000), *Weak convergence and empirical processes: with applications to statistics*, Springer.
- Westerlund, J. (2007), 'Testing for error correction in panel data', *Oxford Bulletin of Economics and Statistics* **69**(6), 709–748.

- Wlazlowski, S., Giuliatti, M., Binner, J. and Milas, C. (2012), 'Price transmission in the eu wholesale petroleum markets', *Journal of Business Economic Statistics* **30**(2), 165–172.
- Woo, W. T. and Hooper, P. (1984), 'Exchange rates and the prices of nonfood, nonfuel products', *Brookings Papers on Economic Activity* **1984**(2), 511–536.
- Yanagisawa, A. (2012), 'Structure for pass-through of oil price to gasoline price in japan', *IEEJ Energy Journal* **7**(3).
- Yang, J. (1997), 'Exchange rate pass-through in u.s. manufacturing industries', *Review of Economics and Statistics* **79**(1), 95–104.
- Yu, P. and Fan, X. (2019), 'Threshold regression with a threshold boundary', *Mimeo* .
- Yu, P. and Phillips, P. C. (2018a), 'Threshold regression asymptotics: From the compound poisson process to two-sided brownian motion', *Economics Letters* **172**, 123–126.
- Yu, P. and Phillips, P. C. (2018b), 'Threshold regression with endogeneity', *Journal of Econometrics* **203**(1), 50–68.