# Estimation and Inference of Threshold Regression Models with Measurement Errors

Chong, Terence Tai Leung and Chen, Haiqiang and Wong, Tsz Nga and Yan, Isabel K.

The Chinese University of Hong Kong, Xiamen University, Bank of Canada, City University of Kong Kong

5 November 2015

# Estimation and Inference of Threshold Regression Models with Measurement Errors

Terence Tai-Leung Chong[*], Haiqiang Chen[†], Tsz-Nga Wong[‡] and Isabel Kit-Ming Yan[§]

5/11/2015

**Abstract**

An important assumption underlying standard threshold regression models and their variants in the extant literature is that the threshold variable is perfectly measured. Such an assumption is crucial for consistent estimation of model parameters. This paper provides the first theoretical framework for the estimation and inference of threshold regression models with measurement errors. A new estimation method that reduces the bias of the coefficient estimates and a Hausman-type test to detect the presence of measurement errors are proposed. Monte Carlo evidence is provided and an empirical application is given.

**Keywords:** Threshold Model; Measurement Error; Hausman-type Test.

**JEL Classification**: C12, C22.

[*]Corresponding author, Terence Tai-Leung Chong, Institute of Global Economics and Finance and Department of Economics, The Chinese University of Hong Kong, Hong Kong and Department of International Economics and Trade, Nanjing University, China. E-mail: chong2064@cuhk.edu.hk. We would like to thank Leonard Stefanski, W.K. Li, George Kapetanios and seminar participants at City University of Hong Kong, Lingnan University and International Symposium on Econometric Theory and Applications (SETA) for helpful comments. We also thank Min Chen and Margaret Loo for their able research assistance. All remaining errors are ours.

[†]Wang Yanan Institute for Studies in Economics, Xiamen University; E-mail: hc335@xmu.edu.cn.

[‡]Bank of Canada

[§]Department of Economics and Finance, City University of Hong Kong.

# 1 Introduction

Measurement error is a common problem in economic data. In particular, macroeconomic data on consumption, unemployment, inflation, and variables that are intrinsically unobservable are often subject to measurement errors because of data aggregation or for other reasons. Madansky (1959) shows that the presence of measurement errors results in inconsistent estimation of parameters in a linear model. Amemiya (1985, 1990) and Schennach (2004) investigate the measurement error problem in nonlinear models. A recent study by Xia and Tong (2011) proposes a method based on feature matching to estimate time series models with measurement errors. The aforementioned methods focus on measurement errors in the explanatory variables, however. Thus far, no study in the literature has attempted to explore the problem of measurement error in the context of threshold regression models. In the presence of measurement errors in the threshold variable, the observations cannot be correctly ranked according to their true values, which can render the estimator inconsistent in such models. This paper provides the first theoretical framework for the inference and estimation of a threshold regression model with measurement errors. Empirically, there is an important distinction between measurement errors in explanatory variables and measurement errors in the threshold variable. In the former case, where the measurement error is often assumed i.i.d. additive to the regressors, all observations of the regressors are confounded by the measurement error. As a result, the true model parameters cannot be retrieved from any subsets of the observations. In the latter case, however, the existence of measurement errors may not lead to misclassification of observations.[1] Therefore, one can improve the parameter estimates of a threshold regression model by estimating a subsample where misclassification is unlikely to occur.

The contribution of our paper is twofold. First, we propose a new method that reduces the bias of the parameter estimates in the presence of measurement errors. Second, we develop a Hausman-type test (Hausman, 1978, 2001; Jeong and Maddala, 1991) for measurement errors in the threshold variables. We apply our test to reestimate the growth convergence model of Hansen (2000), using the per capita output and adult literacy rate as threshold variables. Since the data are taken from earlier years, these two variables might suffer from measurement error. Our test results suggest the existence of measurement errors in both threshold variables. We reestimate the model and find that the convergence hypothesis only holds for countries with lower initial per capita output or those with higher adult literacy rate, which differs from Hansen's (2000) results.

The rest of the paper is organized as follows. Section 2 presents the theoretical model and the

[1] For instance, consider the case where the threshold variable is the GDP per capita. It is unlikely that a country with an extremely high GDP per capita will be misclassified as a poor country even if the threshold variable contains measurement errors.

underlying assumptions. Section 3 proposes a new method to reduce the bias of the parameter estimates. A new test for measurement errors is developed in Section 4. Section 5 provides Monte Carlo evidence for our theory. An empirical application is presented in Section 6 and Section 7 concludes the paper. All proofs are relegated to the appendix.

## 2 The Model

Threshold regression models have developed rapidly since the seminal work of Tong and Lim (1980), and Tong (1983): for example, the smooth transition threshold model (STAR) of Chan and Tong (1986); the functional-coefficient autoregressive (FAR) model of Chen and Tsay (1993); the threshold autoregressive heteroscedastic model of Li and Lam (1995) and Li and Li (1996); and the nested threshold autoregressive (NeTAR) model of Astatkie, Watts and Watt (1997), among others. The model was further extended to allow for multiple threshold values in Tsay (1998) and Gonzalo and Pitarakis (2002). More recently, Chen et al. (2012) investigated the statistical properties of threshold estimators in regression models with multiple threshold variables.[2] Hansen (2011) and Tong (2011) review the development of the threshold model in time series analysis since the 1980s.

The aforementioned studies, however, assume that the threshold variable is not error-ridden. If the threshold variable is measured with errors, some observations could be misclassified, and the parameter estimates will be inconsistent. Consider the following threshold regression model:[3]

$$y_i = \beta_1 x_i + (\beta_2 - \beta_1) x_i \Psi_i^0(\gamma_0) + \varepsilon_i, \tag{1}$$

where $y_i$ and $x_i$ denote the dependent variable and the regressors respectively. $\beta_1$ and $\beta_2$ are the pre-shift and post-shift regression slope parameters respectively. $\Psi_i^0(\gamma_0)$ is an indicator function, which equals one when the true threshold variable $z_i^0$ exceeds the threshold $\gamma_0$. That is,

$$\Psi_i^0(\gamma_0) = I\left(z_i^0 > \gamma_0\right). \tag{2}$$

In the presence of measurement errors, the true value of the threshold variable cannot be observed.[4] Instead, we observe

$$z_i = z_i^0 + u_i.$$

[2] The threshold effect is also considered in modelling conditional distributions (see Wong and Li, 2010).

[3] We consider a univariate model for illustration purposes. The extension to multivariate $x_i$ is provided in the appendix.

[4] If regressors are also measured with errors, we can use the projection theorem to rewrite the model as a new model without measurement errors in the regressors. Our main results can still be established based on the transformed model (see Bai et al., 2008).

where $z_i$ represents the observed threshold variable which is contaminated with measurement errors $u_i$. Let $\Psi_i(\gamma)$ be the observed indicator function defined based on the observed threshold variable

$$\Psi_i(\gamma) = I(z_i > \gamma) = I\left(z_i^0 > \gamma - u_i\right) = \Psi_i^0(\gamma - u_i). \tag{3}$$

Using the observed data $\{y_i, x_i, z_i\}_{i=1}^n$, the above threshold model can be estimated by the following profile least squares (LS) estimation method.[5] The threshold value $\gamma_0$ is estimated by minimizing the sum of squared residuals:

$$\widehat{\gamma} = \arg\min_{\gamma\in\Gamma} S_n(\gamma), \tag{4}$$

where

$$S_n(\gamma) = \sum_{i=1}^n \left(y_i - \widehat{\beta}_1(\gamma)x_i - (\widehat{\beta}_2(\gamma) - \widehat{\beta}_1(\gamma))\Psi_i(\gamma)x_i\right)^2. \tag{5}$$

Given the threshold estimate $\widehat{\gamma}$, the estimators for $\beta_1$ and $\beta_2$ are respectively

$$\widehat{\beta}_1(\widehat{\gamma}) = \sum_{i=1}^n x_i y_i (1 - \Psi_i(\widehat{\gamma})) \left(\sum_{i=1}^n x_i^2 (1 - \Psi_i(\widehat{\gamma}))\right)^{-1} \tag{6}$$

and

$$\widehat{\beta}_2(\widehat{\gamma}) = \sum_{i=1}^n x_i y_i \Psi_i(\widehat{\gamma}) \left(\sum_{i=1}^n x_i^2 \Psi_i(\widehat{\gamma})\right)^{-1}. \tag{7}$$

Before proceeding further, we impose the following assumptions on the threshold model and the measurement error $u_i$.

*A1* : $x_i$ *is strictly stationary, ergodic and* $\rho-$*mixing with* $\rho-$*mixing coefficients satisfying* $\sum_{m=1}^{\infty} \rho_m^{1/2} < \infty$.

*A2* : $\varepsilon_i$ *is a martingale difference sequence and* $E(\varepsilon_i|\mathcal{F}_{i-1}, x_i, z_i^0) = 0$, where $\mathcal{F}_{i-1}$ is the past information set of $\{x_i,\ z_i,\ \varepsilon_i\}$;[6] *and* $\sup_i E|\varepsilon_i|^{2+\kappa} < \infty$, *for some* $\kappa > 0$.

*A3* : $E(|x_i|^4) < \infty$ *and* $E(|x_i\varepsilon_i|^4) < \infty$.

*A4* : $z_i^0$ *is strictly stationary and has a continuous distribution* $F(z)$*. Let* $f(z)$ *denote the density function satisfying* $f(\gamma) \leq \overline{f} < \infty$ *for all* $\gamma \in \Gamma$ *and* $f(\gamma_0) > 0$.

*A5* : $u_i$ *is i.i.d. random variable with zero mean and constant variance* $\sigma_u^2$. $u_i$ *is independent of* $\{x_i,\ z_i^0,\ \varepsilon_i\}$.

*A6* : $\delta = \beta_2 - \beta_1 \neq 0$ and $\gamma_0 \in \Gamma = [\underline{\gamma}, \overline{\gamma}]$.

[5] The estimation method for threshold model is similar to that of structural-change models (Hansen, 2000 and Chong, 2001).

[6] For cross-sectional data, the past information set $\mathcal{F}_{i-1}$ is an empty set.

A1 to A4 are standard assumptions in the literature on threshold models. Assumption $A5$ assumes the measurement errors to be independent of the regressors and the threshold variable. This assumption is imposed to simplify the proof. One might relax this assumption slightly to allow for the mis-measured threshold variable to be one of the regressors. Section 5 of this paper also reports the results for the self-exciting threshold autoregressive model (SETAR), where both $z_i$ and $x_i$ are the lags of $y_i$. Assumption $A6$ assumes the presence of the threshold effect and that the true threshold value $\gamma_0$ falls into a proper subset of the threshold variable space.

**Lemma 1:** *Under Assumptions* $A1-A6$, we have

$$\widehat{\beta}_1(\widehat{\gamma})-\beta_1=\delta\frac{\sum_{i=1}^n x_i^2\left[\Psi_i(\gamma_0+u_i)-\Psi_i(\max\{\gamma_0+u_i,\widehat{\gamma}\})\right]}{\sum_{i=1}^n x_i^2(1-\Psi_i(\widehat{\gamma}))}+O_p(\frac{1}{\sqrt{n}}) \tag{8}$$

*and*

$$\widehat{\beta}_2(\widehat{\gamma})-\beta_2=-\delta\frac{\sum_{i=1}^n x_i^2\left[\Psi_i(\widehat{\gamma})-\Psi_i(\max\{\gamma_0+u_i,\widehat{\gamma}\})\right]}{\sum_{i=1}^n x_i^2\Psi_i(\widehat{\gamma})}+O_p(\frac{1}{\sqrt{n}}). \tag{9}$$

Lemma 1 shows that the two conventional LS based estimators $\widehat{\beta}_1(\widehat{\gamma})$ and $\widehat{\beta}_2(\widehat{\gamma})$ are inconsistent in the presence of measurement errors even at $\widehat{\gamma}=\gamma_0$. Note that $\Psi_i(\gamma_0+u_i)-\Psi_i(\max\{\gamma_0+u_i,\widehat{\gamma}\})=I(z_i>(\gamma_0+u_i))-I(z_i>\max\{\gamma_0+u_i,\widehat{\gamma}\})$ is always non-negative. Thus, when $\delta>0$, $\widehat{\beta}_1(\widehat{\gamma})$ will be biased upward and $\widehat{\beta}_2(\widehat{\gamma})$ will be biased downward, and vice versa.

# 3 The New Estimator

In this section, we propose a method that reduces the bias of the estimates of $\beta_1$ and $\beta_2$. For a fixed $\lambda\in(0,1/2)$, define $\widetilde{\beta}_1(\lambda)$ and $\widetilde{\beta}_2(\lambda)$ as the pre- and post-shift OLS estimators using the lowest $[\lambda n]$ ordered observations and the highest $[\lambda n]$ ordered observations in a sample with $n$ observations. Denote $\underline{\gamma_\lambda}$ and $\overline{\gamma_\lambda}$ as the empirical lower and upper $\lambda$-quantiles of the threshold variable $z_i$. We have

$$\widetilde{\beta}_1(\lambda)=\sum_{i=1}^n x_iy_i\left(1-\Psi_i\left(\underline{\gamma_\lambda}\right)\right)\left(\sum_{i=1}^n x_i^2\left(1-\Psi_i\left(\underline{\gamma_\lambda}\right)\right)\right)^{-1}, \tag{10}$$

$$\widetilde{\beta}_2(\lambda)=\sum_{i=1}^n x_iy_i\Psi_i\left(\overline{\gamma_\lambda}\right)\left(\sum_{i=1}^n x_i^2\Psi_i\left(\overline{\gamma_\lambda}\right)\right)^{-1}. \tag{11}$$

Note that $\widetilde{\beta}_1(\lambda)$ and $\widetilde{\beta}_2(\lambda)$ can be considered as weighted OLS estimators of $\beta_1$ and $\beta_2$ respectively with zero weight given to the middle range.

**Lemma 2:** *Under Assumptions* $A1 - A6$, we have

$$\widetilde{\beta}_1(\lambda) - \beta_1 = \delta \frac{\sum_{i=1}^n x_i^2 \left[\Psi_i\left(\gamma_0 + u_i\right) - \Psi_i\left(\max\left\{\gamma_0 + u_i, \underline{\gamma_\lambda}\right\}\right)\right]}{\sum_{i=1}^n x_i^2 \left(1 - \Psi_i\left(\underline{\gamma_\lambda}\right)\right)} + O_p(\frac{1}{\sqrt{n}}) \tag{12}$$

*and*

$$\widetilde{\beta}_2(\lambda) - \beta_2 = -\delta \frac{\sum_{i=1}^n x_i^2 \left[\Psi_i\left(\overline{\gamma_\lambda}\right) - \Psi_i\left(\max\left\{\gamma_0 + u_i, \overline{\gamma_\lambda}\right\}\right)\right]}{\sum_{i=1}^n x_i^2 \Psi_i\left(\overline{\gamma_\lambda}\right)} + O_p(\frac{1}{\sqrt{n}}). \tag{13}$$

Lemma 2 shows that the estimators $\widetilde{\beta}_1(\lambda)$ and $\widetilde{\beta}_2(\lambda)$ are inconsistent in the presence of measurement error. The bias term is related to the distribution of the measurement error $u_i$ and the cutting values $\underline{\gamma_\lambda}$ and $\overline{\gamma_\lambda}$.[7] The following theorem shows that the new estimators have smaller bias than conventional LS estimators given the existence of measurement errors.

Theorem 1: Under Assumptions A1 to A6, for any $\widehat{\gamma} \in (\underline{\gamma_\lambda}, \overline{\gamma_\lambda})$, $\widetilde{\beta}_1(\lambda)$ and $\widetilde{\beta}_2(\lambda)$ are less biased than $\widehat{\beta}_1\left(\widehat{\gamma}\right)$ and $\widehat{\beta}_2\left(\widehat{\gamma}\right)$ when $\sigma_u^2 > 0$.

Note that Theorem 1 is established under the condition that $\widehat{\gamma} \in (\underline{\gamma_\lambda}, \overline{\gamma_\lambda})$. This assumption is automatically satisfied if we define $\Gamma = (\underline{\gamma_\lambda}, \overline{\gamma_\lambda})$ in Equation (6). In the literature, $\lambda$ is commonly set at 15% to ensure enough observations in the extreme regimes.

Note also that in the absence of measurement error, i.e., $\sigma_u^2 = 0$, the bias terms in Equations (10), (11), (14) and (15) are all zeros, implying that both the full-sample estimators $\widehat{\beta}_1\left(\widehat{\gamma}\right)$ and $\widehat{\beta}_2\left(\widehat{\gamma}\right)$ and the new estimators $\widetilde{\beta}_1(\lambda)$ and $\widetilde{\beta}_2(\lambda)$ are consistent. Since, however, $\widetilde{\beta}_1(\lambda)$ and $\widetilde{\beta}_2(\lambda)$ only use the data from a subsample, they will be less efficient than $\widehat{\beta}_1\left(\widehat{\gamma}\right)$ and $\widehat{\beta}_2\left(\widehat{\gamma}\right)$. On the other hand, if there is any measurement error, $\widetilde{\beta}_1(\lambda)$ and $\widetilde{\beta}_2(\lambda)$ will be less biased. Theoretically, given $\widehat{\gamma} \in (\underline{\gamma_\lambda}, \overline{\gamma_\lambda})$, it can be shown that $\widehat{\beta}_1\left(\widehat{\gamma}\right)$ and $\widehat{\beta}_2\left(\widehat{\gamma}\right)$ are the limits of $\widetilde{\beta}_1(\lambda)$ and $\widetilde{\beta}_2(\lambda)$ as $\lambda \to 0.5$. A larger $\lambda$ improves the efficiency of the estimate, but it will also increase the bias caused by the measurement error. As the problem of low efficiency may be as serious compared to a high bias, the new estimators should perform better than the traditional estimators in terms of mean square errors (MSE). In Section 5, we show that under different model settings, $\widetilde{\beta}_1(\lambda)$ and $\widetilde{\beta}_2(\lambda)$ have a smaller MSE than $\widehat{\beta}_1\left(\widehat{\gamma}\right)$ and $\widehat{\beta}_2\left(\widehat{\gamma}\right)$ in most cases.

[7] In some cases where the measurement error follows a truncated or bounded distribution, we may have $\Pr(u_i \leq \underline{\gamma_\lambda} - \gamma_0) = 0$ and $\Pr(u_i \geq \overline{\gamma_\lambda} - \gamma_0) = 0$, implying zero bias terms in Equations (14) and (15).

# 4 Test Statistic

In this section we develop a test for measurement errors. Consider the following null hypothesis of no measurement error:

$$\begin{aligned} H_0 &: \sigma_u^2 = 0, \\ H_1 &: \sigma_u^2 > 0. \end{aligned}$$

Recall that $\sigma_u^2$ refers to the variance of the measurement error term $u_i$. Let $\widehat{\varepsilon}_i = y_i - \widehat{\beta}_1(\widehat{\gamma})\, x_i - (\widehat{\beta}_2(\widehat{\gamma}) - \widehat{\beta}_1(\widehat{\gamma}))\Psi_i(\widehat{\gamma})\, x_i$ and define the following sample moments:

$$\begin{aligned} \widehat{M}_1(\gamma) &= \frac{\sum_{i=1}^n x_i^2 I(z_i \leq \gamma)}{n}, \\ \widehat{M}_2(\gamma) &= \frac{\sum_{i=1}^n x_i^2 I(z_i > \gamma)}{n}. \end{aligned}$$

and

$$\begin{aligned} \widehat{\Omega}_{11}(\gamma_1, \gamma_2) &= \frac{\sum_{i=1}^n x_i^2 I(z_i \leq \gamma_1)\widehat{\varepsilon}_i^2 I(z_i \leq \gamma_2)}{n}, \\ \widehat{\Omega}_{12}(\gamma_1, \gamma_2) &= \frac{\sum_{i=1}^n x_i^2 I(z_i \leq \gamma_1)\widehat{\varepsilon}_i^2 I(z_i > \gamma_2)}{n}, \\ \widehat{\Omega}_{22}(\gamma_1, \gamma_2) &= \frac{\sum_{i=1}^n x_i^2 I(z_i > \gamma_1)\widehat{\varepsilon}_i^2 I(z_i > \gamma_2)}{n}. \end{aligned}$$

Using the above sample moments, we construct a Hausman-type test for measurement error in the threshold variable, defined as

$$T(\lambda) = n \begin{pmatrix} \widehat{\beta}_1(\widehat{\gamma}) - \widetilde{\beta}_1(\lambda) \\ \widehat{\beta}_2(\widehat{\gamma}) - \widetilde{\beta}_2(\lambda) \end{pmatrix}' \widehat{\Pi}^{-1} \begin{pmatrix} \widehat{\beta}_1(\widehat{\gamma}) - \widetilde{\beta}_1(\lambda) \\ \widehat{\beta}_2(\widehat{\gamma}) - \widetilde{\beta}_2(\lambda) \end{pmatrix} \tag{14}$$

where

$$\widehat{\Pi} = \widehat{Var} \begin{pmatrix} \sqrt{n}(\widehat{\beta}_1(\widehat{\gamma}) - \widetilde{\beta}_1(\lambda)) \\ \sqrt{n}(\widehat{\beta}_2(\widehat{\gamma}) - \widetilde{\beta}_2(\lambda)) \end{pmatrix} = \begin{pmatrix} \widehat{\Pi}_{11}, \widehat{\Pi}_{12} \\ \widehat{\Pi}_{12}', \widehat{\Pi}_{22} \end{pmatrix}$$

and

$$
\begin{aligned}
\widehat{\Pi}_{11} &= \widehat{M}_1(\underline{\gamma_\lambda})^{-1}\widehat{\Omega}_{11}(\underline{\gamma_\lambda},\underline{\gamma_\lambda})\widehat{M}_1(\underline{\gamma_\lambda})^{-1} + \widehat{M}_1(\widehat{\gamma})^{-1}\widehat{\Omega}_{11}(\widehat{\gamma},\widehat{\gamma})\widehat{M}_1(\widehat{\gamma})^{-1} \\
&\quad -\widehat{M}_1(\underline{\gamma_\lambda})^{-1}\widehat{\Omega}_{11}(\underline{\gamma_\lambda},\widehat{\gamma})\widehat{M}_1(\widehat{\gamma})^{-1} - \widehat{M}_1(\widehat{\gamma})^{-1}\widehat{\Omega}_{11}(\widehat{\gamma},\underline{\gamma_\lambda})\widehat{M}_1(\underline{\gamma_\lambda})^{-1}, \\
\widehat{\Pi}_{12} &= \widehat{M}_1(\underline{\gamma_\lambda})^{-1}\widehat{\Omega}_{12}(\underline{\gamma_\lambda},\overline{\gamma_\lambda})\widehat{M}_2(\overline{\gamma_\lambda})^{-1} + \widehat{M}_1(\widehat{\gamma})^{-1}\widehat{\Omega}_{12}(\widehat{\gamma},\widehat{\gamma})\widehat{M}_2(\widehat{\gamma})^{-1} \\
&\quad -\widehat{M}_1(\underline{\gamma_\lambda})^{-1}\widehat{\Omega}_{12}(\underline{\gamma_\lambda},\widehat{\gamma})\widehat{M}_2(\widehat{\gamma})^{-1} - \widehat{M}_1(\widehat{\gamma})^{-1}\widehat{\Omega}_{12}(\widehat{\gamma},\overline{\gamma_\lambda})\widehat{M}_2(\overline{\gamma_\lambda})^{-1}, \\
\widehat{\Pi}_{22} &= \widehat{M}_2(\overline{\gamma_\lambda})^{-1}\widehat{\Omega}_{22}(\overline{\gamma_\lambda},\overline{\gamma_\lambda})\widehat{M}_2(\overline{\gamma_\lambda})^{-1} + \widehat{M}_2(\widehat{\gamma})^{-1}\widehat{\Omega}_{22}(\widehat{\gamma},\widehat{\gamma})\widehat{M}_2(\widehat{\gamma})^{-1} \\
&\quad -\widehat{M}_2(\overline{\gamma_\lambda})^{-1}\widehat{\Omega}_{22}(\overline{\gamma_\lambda},\widehat{\gamma})\widehat{M}_2(\widehat{\gamma})^{-1} - \widehat{M}_2(\widehat{\gamma})^{-1}\widehat{\Omega}_{22}(\widehat{\gamma},\overline{\gamma_\lambda})\widehat{M}_2(\overline{\gamma_\lambda})^{-1}.
\end{aligned}
$$

The following theorem establishes the null asymptotic distribution for the test statistic.

Theorem 2: Under Assumptions A1 to A6, for $\gamma_0 \in (\underline{\gamma_\lambda}, \overline{\gamma_\lambda})$, we have $T(\lambda) \Rightarrow \chi_2^2$ under $H_0 : \sigma_u^2 = 0$.

This result can easily be extended to models with $k$ regressors, where the corresponding degree of freedom is $2k$. The asymptotic null distribution is not affected by the choice of $\lambda$ under the null hypothesis if the condition $\gamma_0 \in (\underline{\gamma_\lambda}, \overline{\gamma_\lambda})$ is satisfied. The value of $\lambda$, however, might affect the precision of the parameter estimates and the power of the test in finite samples. In the following simulations, we consider different $\lambda's$ to examine their impact on estimation and testing. Note that, under the alternative hypothesis, $\widehat{\beta}_1(\widehat{\gamma}) - \widetilde{\beta}_1(\lambda)$ and $\widehat{\beta}_2(\widehat{\gamma}) - \widetilde{\beta}_2(\lambda)$ are $O_p(1)$, and $T(\lambda)$ diverges to infinity. Therefore, the test is consistent.

# 5 Monte Carlo Simulations

## Experiment 1: Estimation performance of the new estimators

To demonstrate that the use of the restricted sample (the sample excluding the middle $100(1-2\lambda)\%$ of the ordered observations) can reduce the estimation bias, we examine the finite sample performance of the new estimator.[8] We consider the following four data generating processes:

DGP 1: Mean-shift model

$$y_i = \beta_1 I(z_i^0 \leq \gamma_0) + \beta_2 I\left(z_i^0 > \gamma_0\right) + \varepsilon_i, \qquad i = 1, 2, ..., n. \tag{15}$$

DGP 2: Univariate threshold regression model with conditional heteroskedasticity

[8] All simulations are programmed in R. The code can be obtained from the authors upon request.

$$y_i = \beta_1 x_i I(z_i^0 \leq \gamma_0) + \beta_2 x_i I\left(z_i^0 > \gamma_0\right) + (0.2x_i)\varepsilon_i, \qquad i = 1, 2, ..., n. \tag{16}$$

DGP 3: Threshold autoregressive model (TAR)

$$y_i = (\beta_{11} y_{i-1} + \beta_{12} y_{i-2}) I(z_i^0 \leq \gamma_0) + (\beta_{21} y_{i-1} + \beta_{22} y_{i-2}) I\left(z_i^0 > \gamma_0\right) + \varepsilon_i, \qquad i = 1, 2, ..., n. \tag{17}$$

DGP 4: Threshold regression model where the threshold variable is one of the regressors

$$y_i = (\beta_{11} x_i + \beta_{12} z_i^0) I(z_i^0 \leq \gamma_0) + (\beta_{21} x_i + \beta_{22} z_i^0) I\left(z_i^0 > \gamma_0\right) + \varepsilon_i, \qquad i = 1, 2, ..., n. \tag{18}$$

The observed threshold variable is specified as $z_i = z_i^0 + u_i$, where $z_i^0 \sim i.i.d.\ N(10, 1)$ and the measurement error $u_i \sim i.i.d.\ N(0, \sigma_u^2)$. $x_i$ follows an $i.i.d.\ U(0, 10)$ distribution. $\varepsilon_i \sim i.i.d.$ $N(0, 1)$. $u_i$, $z_i^0$, $x_i$ and $\varepsilon_i$ are independent of each other. Let $\gamma_0 = 10$, $\beta_1 = 1$, $\beta_2 = 2$, $\beta_{11} = 0.5$, $\beta_{12} = 0.1$, $\beta_{21} = -0.5$ and $\beta_{22} = -0.1$.

Note that for DGP 4, in the estimated model, the second regressor is $z_i$, which is affected by the measurement error. Thus, it violates Assumption $A5$.

For all cases, we replicate the simulations with $n = 1000, 400$ or $200$ (sample size) and $R = 1000$ (number of replications).

Table 1 reports the mean squared errors (MSE) of the estimators. For each sample size, we study the cases for $\lambda = 0.1, 0.2$, $0.3$, $0.4$ and $0.45$. The MSE of $\widetilde{\beta}_1(\lambda)$ and $\widetilde{\beta}_2(\lambda)$ are reported in the first four rows of each panel, and the last row reports the MSE of the conventional LS estimators $\widehat{\beta}_1(\widehat{\gamma})$ and $\widehat{\beta}_2(\widehat{\gamma})$ from the full sample.[9] The simulation shows that, for most cases, as the sample size increases, $\widetilde{\beta}_1(\lambda)$ and $\widetilde{\beta}_2(\lambda)$ have smaller MSE than the conventional LS estimators $\widehat{\beta}_1(\widehat{\gamma})$ and $\widehat{\beta}_2(\widehat{\gamma})$ in the presence of measurement errors, which is consistent with Theorem 1.[10]

**Table 1**: Performance of the estimators ($\sigma_u^2 = 0.25$)

[9]For DGP3 and DGP4, $MSE_1 = MSE(\widehat{\beta}_{11}) + MSE(\widehat{\beta}_{21})$ and $MSE_2 = MSE(\widehat{\beta}_{12}) + MSE(\widehat{\beta}_{22})$.

[10]In empirical studies, an important question is to find the optimal value for $\lambda$. A possible solution is to use the leave-one-out cross-validation approach, i.e., choose the value of $\lambda$ providing the best out-of-sample prediction as the optimal one. We thank the referee for pointing out this issue.

| $n = 1000$ | $DGP1$ | | $DGP2$ | | $DGP3$ | | $DGP4$ | |
|---|---|---|---|---|---|---|---|---|
| $\lambda$ | $MSE_1$ | $MSE_2$ | $MSE_1$ | $MSE_2$ | $MSE_1$ | $MSE_2$ | $MSE_1$ | $MSE_2$ |
| 0.45 | .0164 | .0162 | .0144 | .0145 | .0338 | .0052 | .0126 | .0016 |
| 0.4 | .0106 | .0100 | .0081 | .0082 | .0200 | .0051 | .0057 | .0008 |
| 0.3 | .0046 | .0052 | .0020 | .0021 | .0089 | .0058 | .0014 | .0004 |
| 0.2 | .0051 | .0052 | .0006 | .0006 | .0078 | .0075 | .0013 | .0005 |
| 0.1 | .0091 | .0099 | .0007 | .0006 | .0153 | .0146 | .0023 | .0010 |
| $\widehat{\beta}(\widehat{\gamma})$ | .0201 | .0208 | .0203 | .0205 | .0407 | .0053 | .0207 | .0034 |
| $n = 400$ | $DGP1$ | | $DGP2$ | | $DGP3$ | | $DGP4$ | |
| $\lambda$ | $MSE_1$ | $MSE_2$ | $MSE_1$ | $MSE_2$ | $MSE_1$ | $MSE_2$ | $MSE_1$ | $MSE_2$ |
| 0.45 | .0205 | .0204 | .0153 | .0156 | .0415 | .0110 | .0193 | .0024 |
| 0.4 | .0146 | .0150 | .0094 | .0095 | .0284 | .0113 | .0096 | .0015 |
| 0.3 | .0107 | .0101 | .0028 | .0031 | .0188 | .0148 | .0036 | .0009 |
| 0.2 | .0125 | .0121 | .0014 | .0013 | .0197 | .0198 | .0030 | .0011 |
| 0.1 | .0275 | .0245 | .0017 | .0019 | .0408 | .0383 | .0063 | .0024 |
| $\widehat{\beta}(\widehat{\gamma})$ | .0223 | .0230 | .0192 | .0189 | .0430 | .0114 | .0204 | .0060 |
| $n = 200$ | $DGP1$ | | $DGP2$ | | $DGP3$ | | $DGP4$ | |
| $\lambda$ | $MSE_1$ | $MSE_2$ | $MSE_1$ | $MSE_2$ | $MSE_1$ | $MSE_2$ | $MSE_1$ | $MSE_2$ |
| 0.45 | .0267 | .0298 | .0171 | .0187 | .0301 | .0280 | .0231 | .0033 |
| 0.4 | .0208 | .0220 | .0110 | .0116 | .0306 | .0313 | .0185 | .0030 |
| 0.3 | .0193 | .0190 | .0046 | .0041 | .0306 | .0313 | .0185 | .0030 |
| 0.2 | .0245 | .0247 | .0025 | .0027 | .0403 | .0405 | .0112 | .0029 |
| 0.1 | .0481 | .0456 | .0038 | .0037 | .0836 | .0912 | .0140 | .0047 |
| $\widehat{\beta}(\widehat{\gamma})$ | .0288 | .0280 | .0184 | .0185 | .0370 | .0365 | .0410 | .0078 |

## Experiment 2: The size and power of the test

We study the size and power of the test statistic in this subsection. The data generating processes are the same with Experiment 1.

Table 2a reports the size of the test under the null of no measurement error for different DGP. When the sample size is large, the rejection rates are close to the asymptotic $\alpha$ for all cases.

Table 2b reports the power of the test in the presence of measurement errors ($\sigma_u^2 = 0.25$). Note that the power performance is closely related to the estimation accuracy. The rejection rate approaches one as the sample size increases. Meanwhile, the power performance depends nonlinearly on the value of $\lambda$, which is consistent to our findings in Experiment 1.

For DGP4, where the threshold variable is one of the regressors, the size remains unaffected, but the power is much improved. In this case, measurement errors exist in both the threshold variable and the regressor under the alternative hypothesis. This causes further bias of the estimates and may enlarge the value of the test statistic.

**Table 2a**: Size of the test ($\sigma_u^2 = 0$)

| $n = 1000$ | $DGP1$ | | | $DGP2$ | | | $DGP3$ | | | $DGP4$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\lambda$ | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 |
| 0.45 | .079 | .034 | .006 | .087 | .037 | .006 | .088 | .041 | .010 | .100 | .052 | .014 |
| 0.4 | .094 | .044 | .007 | .104 | .045 | .008 | .087 | .041 | .009 | .101 | .062 | .013 |
| 0.3 | .106 | .056 | .005 | .110 | .053 | .011 | .092 | .045 | .007 | .104 | .053 | .008 |
| 0.2 | .095 | .053 | .015 | .091 | .041 | .007 | .094 | .044 | .004 | .116 | .062 | .009 |
| 0.1 | .103 | .048 | .010 | .107 | .061 | .009 | .106 | .049 | .010 | .103 | .048 | .010 |
| $n = 400$ | $DGP1$ | | | $DGP2$ | | | $DGP3$ | | | $DGP4$ | | |
| $\lambda$ | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 |
| 0.45 | .149 | .109 | .066 | .119 | .068 | .042 | .120 | .079 | .047 | .096 | .051 | .028 |
| 0.4 | .099 | .048 | .026 | .091 | .039 | .006 | .075 | .032 | .006 | .093 | .041 | .005 |
| 0.3 | .090 | .037 | .006 | .111 | .057 | .014 | .086 | .034 | .005 | .107 | .058 | .010 |
| 0.2 | .093 | .047 | .008 | .089 | .045 | .011 | .105 | .053 | .009 | .096 | .050 | .009 |
| 0.1 | .120 | .059 | .009 | .077 | .028 | .003 | .094 | .042 | .004 | .104 | .044 | .006 |
| $n = 200$ | $DGP1$ | | | $DGP2$ | | | $DGP3$ | | | $DGP4$ | | |
| $\lambda$ | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 |
| 0.45 | .273 | .231 | .167 | .214 | .183 | .170 | .239 | .212 | .162 | .209 | .182 | .173 |
| 0.4 | .117 | .063 | .026 | .071 | .029 | .004 | .095 | .053 | .019 | .079 | .033 | .006 |
| 0.3 | .098 | .040 | .006 | .100 | .054 | .006 | .091 | .041 | .008 | .084 | .040 | .005 |
| 0.2 | .100 | .052 | .008 | .094 | .041 | .004 | .103 | .044 | .008 | .104 | .050 | .009 |
| 0.1 | .095 | .041 | .009 | .087 | .037 | .000 | .082 | .034 | .006 | .089 | .043 | .007 |

**Table 2b**: Power of the test ($\sigma_u^2 = 0.25$)

| $n = 1000$ | $DGP1$ | | | $DGP2$ | | | $DGP3$ | | | $DGP4$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\lambda$ | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 |
| 0.45 | .614 | .559 | .495 | .996 | .987 | .925 | .575 | .522 | .481 | .999 | .996 | .951 |
| 0.4 | .838 | .721 | .444 | 1.0 | 1.0 | .999 | .662 | .504 | .294 | 1.0 | 1.0 | 1.0 |
| 0.3 | .928 | .881 | .690 | 1.0 | 1.0 | 1.0 | .855 | .763 | .509 | 1.0 | 1.0 | 1.0 |
| 0.2 | .842 | .767 | .549 | .996 | .995 | .984 | .815 | .712 | .453 | 1.0 | 1.0 | 1.0 |
| 0.1 | .542 | .405 | .199 | .963 | .929 | .756 | .513 | .383 | .150 | 1.0 | 1.0 | 1.0 |
| $n = 400$ | $DGP1$ | | | $DGP2$ | | | $DGP3$ | | | $DGP4$ | | |
| $\lambda$ | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 |
| 0.45 | .640 | .601 | .546 | .651 | .511 | .301 | .631 | .601 | .547 | .804 | .717 | .606 |
| 0.4 | .380 | .321 | .244 | .913 | .825 | .518 | .397 | .344 | .279 | .975 | .938 | .796 |
| 0.3 | .435 | .314 | .119 | .949 | .897 | .699 | .361 | .234 | .074 | 1.0 | 1.0 | .992 |
| 0.2 | .402 | .260 | .093 | .849 | .735 | .457 | .335 | .201 | .056 | 1.0 | .999 | .998 |
| 0.1 | .228 | .141 | .039 | .592 | .435 | .172 | .193 | .105 | .018 | 1.0 | .999 | .990 |
| $n = 200$ | $DGP1$ | | | $DGP2$ | | | $DGP3$ | | | $DGP4$ | | |
| $\lambda$ | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 |
| 0.45 | .654 | .614 | .525 | .480 | .437 | .415 | .682 | .624 | .523 | .749 | .734 | .721 |
| 0.4 | .418 | .383 | .295 | .502 | .335 | .126 | .423 | .388 | .303 | .622 | .489 | .277 |
| 0.3 | .225 | .164 | .079 | .632 | .460 | .163 | .209 | .123 | .063 | .946 | .884 | .665 |
| 0.2 | .191 | .098 | .024 | .474 | .324 | .098 | .148 | .063 | .010 | .966 | .935 | .772 |
| 0.1 | .141 | .067 | .016 | .298 | .152 | .018 | .119 | .056 | .005 | .920 | .825 | .467 |

# 6 Empirical Application

Hansen (2000) examines the convergence hypothesis by analyzing the relationship between economic growth and the initial endowment of different countries. The baseline model is as follows:

$$y_{i,85-60} = \begin{cases} \alpha_1 + \beta_1 \ln(Y/L)_{i,1960} + \pi_1 \ln(I/Y)_i + \mu_1 \ln(n_i + g + \delta) \\ +\varphi_1 \ln(school)_i + e_i, & \text{if } z_i \leq \gamma \\ \alpha_2 + \beta_2 \ln(Y/L)_{i,1960} + \pi_2 \ln(I/Y)_i + \mu_2 \ln(n_i + g + \delta) \\ +\varphi_2 \ln(school)_i + e_i, & \text{if } z_i > \gamma \end{cases}$$

For country $i$, $(Y/L)_{i,t}$ denotes the real GDP per member of the population aged 15 to 64 in year $t$; $y_{i,85-60} = \ln(Y/L)_{i,1985} - \ln(Y/L)_{i,1960}$ is defined as the difference of per capita real GDP between 1960 and 1985; $(I/Y)_i$ refers to the average of investment to GDP ratio over the period 1960 to 1985; $n_i$ is the average of the working-age population growth rate over the sample period;

$(school)_i$ is the average of the fraction of working-age population enrolled in secondary school over the sample period. A negative value for $\beta$ in the regression provides evidence of convergence. We set $g+\delta = 0.05$, where $g$ is the growth rate of technology and $\delta$ is the depreciation rate. The threshold variables are the per capita output $Y/L$ in 1960 and the adult literacy rate in 1960. We also allow for heteroskedasticity in the error term. Hansen (2000) provides estimation and testing results, assuming no measurement error in the threshold variable. The threshold value estimated for initial per capita $Y/L$ is 863 with a 95% confidence interval $[594, 1794]$, and the estimated threshold value for adult literacy is 45% with a 95% confidence interval $[19\%, 57\%]$. The bootstrapping p-values of the Sup-LM test statistics for testing the presence of threshold effect are 0.088 and 0.214 respectively. Our point estimates for threshold values are very close to those obtained in Hansen (2000), where the first estimate is 877 and the second is 45%. The minor difference could be owed to the difference in the grid size.

In the model, the initial endowment is proxied by the per capita output, or the adult literacy rate measured in the 1960s. The use of proxies is likely to give rise to measurement errors, especially when the data are taken from early years. We apply the test developed in Section 4 with $\lambda = 0.15$ to test for measurement error in per capita output and adult literacy rate. When the per capita output is used as a threshold variable, the test statistic value is 47.02 and the p-value is smaller than 0.01. When the adult literacy rate is used as the threshold variable, the test statistic is 45.69 and the p-value is smaller than 0.01. Therefore, we reject the null hypothesis of no measurement error in the threshold variable in both cases at the 5% significance level.[11]

Tables 3a and 3b report the estimation results with per capita output and adult literacy rate as the threshold variables respectively. The first two columns report the results from the standard threshold model, and the last two columns report the results from the extreme regimes after the middle observations have been dropped. The heteroskedasticity-consistent standard errors are reported in parentheses.

[11] We also examine the test results by setting $\lambda$ as 0.1, 0.2 or 0.3. The null hypotheses are still rejected for all cases.

**Table 3a:** Coefficient Estimations for Per Capita Output $y_{i,60}$

| | Traditional Method ($\widehat{\gamma} = 877$) | | New Method ($\lambda = 0.15$) | |
|---|---|---|---|---|
| | $y_{i,60} \leq 877$ | $y_{i,60} > 877$ | $y_{i,60} \leq 777$ | $y_{i,60} > 6527$ |
| *Intercept* | $4.31^*$ (1.62) | $3.66^*$ (1.61) | $4.77^*$ (1.32) | $-1.49$ (1.61) |
| $\ln(Y/L)_{i,1960}$ | $-0.65^*$ (0.21) | $-0.32^*$ (0.06) | $-0.79^*$ (0.16) | $-0.066$ (0.16) |
| $\ln(I/Y)_i$ | $0.23^*$ (0.071) | $0.49^*$ (0.144) | $0.31^*$ (0.075) | $0.47^*$ (0.082) |
| $\ln(n_i + g + \delta)$ | $-0.29$ (0.33) | $-0.49^*$ (0.25) | $-0.43$ (0.41) | $-1.43^*$ (0.15) |
| $\ln(school)_i$ | 0.02 (0.097) | $0.35^*$ (0.09) | $-0.03$ (0.09) | $0.31^*$ (0.091) |

**Table 3b:** Coefficient Estimations for Adult Literacy Rate $lr_{i,60}$

| | Traditional Method ($\widehat{\gamma} = 45.02$) | | New Method ($\lambda = 0.15$) | |
|---|---|---|---|---|
| | $lr_{i,60} \leq 45.02$ | $lr_{i,60} > 45.02$ | $lr_{i,60} \leq 15$ | $lr_{i,60} > 98$ |
| *Intercept* | 2.09 (1.87) | $4.31^*$ (0.96) | $5.41^*$ (2.33) | 2.64 (2.03) |
| $\ln(Y/L)_{i,1960}$ | $-0.12$ (0.16) | $-0.39^*$ (0.06) | $-0.26$ (0.25) | $-0.41^*$ (0.19) |
| $\ln(I/Y)_i$ | 0.17 (0.21) | $0.83^*$ (0.13) | $-0.11$ (0.23) | 0.25 (0.18) |
| $\ln(n_i + g + \delta)$ | $-0.39$ (0.51) | $-0.42$ (0.27) | 0.43 (0.39) | $-0.81^*$ (0.37) |
| $\ln(school)_i$ | $0.45^*$ (0.11) | 0.095 (0.13) | $0.66^*$ (0.11) | 0.11 (0.14) |

In Table 3a, the estimated coefficients for $\ln(Y/L)_{i,1960}$ are significantly negative in the model using the full sample, which supports the convergence hypothesis. After the middle observations have been dropped, however, only the regime with lower per capita output supports the convergence hypothesis. The result is different from that of Hansen (2000). In Table 3b, our result shows that the convergence hypothesis holds only for countries with higher adult literacy rates, which corroborates Hansen's finding (2000).

# 7 Conclusion

It is well documented in the literature that the presence of measurement errors causes inconsistent estimation of model parameters. This paper examines the case of a threshold regression model with measurement errors. It is shown that measurement errors in the threshold variable may not lead to misclassification of observations, as the indicator variable for classifying observations may absorb some of the errors. If observations in the two extremes of the threshold spectrum have a lower probability of being misclassified, the estimates obtained from the full-sample will differ from those from a less contaminated subsample in the presence of measurement errors.

This paper develops a new test for the presence of measurement error in the threshold variable. Our test is based on the estimation difference between two estimators; the first assigns equal weight to each observation, and the second assigns zero weight to highly contaminated observations. Under the null hypothesis of no measurement error, both estimators are consistent, but the second estimator is less efficient. Under the alternative hypothesis, both estimators are inconsistent, but the second estimator is less biased. Our test statistic is shown to have an asymptotic Chi-square distribution. Monte Carlo evidence shows that the new test has good performance in terms of size and power. This paper also contributes to the literature by developing a new estimation method for reducing the bias of parameter estimates in the presence of measurement errors. Significant improvement in the parameter estimates is found by estimating a subsample with observations that are less likely to suffer from measurement errors. For future research in this line, one could extend our analysis to models with multiple regimes (Bai et al., 2008) and multiple threshold variables (Chen et al., 2012).

# References


1. Amemiya, Y. (1985). Instrumental Variable Estimator for the Nonlinear Errors in Variables Model. *Journal of Econometrics* 28(3), 273-289.

2. Amemiya, Y. (1990). Two Stage Instrumental Variable Estimators for the Nonlinear Errors-in-Variables Model. *Journal of Econometrics* 44(3), 311-332.

3. Armstrong, B. (1985). Measurement Error in Generalized Linear Models, *Communications in Statistics: Simulation and Computation* 14(3), 529-544.

4. Astatkie, T., D. G. Watts and W. E. Watt (1997). Nested Threshold Autoregressive (NeTAR) Models. *International Journal of Forecasting* 13(1), 105-116.

5. Bai, J., H. Chen, T. T. L. Chong and X. Wang (2008). Generic Consistency of the Break-Point Estimator under Specification Errors in a Multiple-Break Model. Econometrics Journal 11(2), 287-307.

6. Chan, K. S. and H. Tong (1986). On Estimating Thresholds in Autoregressive Models. *Journal of Time Series Analysis* 7(3), 179-190.

7. Chen, R. and S. Tsay (1993). Functional-Coefficient Autoregressive Models. *Journal of the American Statistical Association* 88, 298-308.

8. Chen, H., T. T. L. Chong and J. Bai (2012). Theory and Applications of TAR Model with Two Threshold Variables, *Econometric Reviews* 31(2), 142–170.

9. Chong, T. T. L. (2001). Structural Change in AR(1) Models. *Econometric Theory* 17(1), 87-155.

10. Chong, T. T. L. (2003). Generic Consistency of the Break-Point Estimator under Specification Errors. *Econometrics Journal* 6(1), 167-192.

11. Gonzalo, J. and J. Pitarakis (2002). Estimation and Model Selection Based Inference in Single and Multiple Threshold Models. *Journal of Econometrics* 110(2), 319-352.

12. Hansen, B. E. (2000). Sample Splitting and Threshold Estimation. *Econometrica* 68(3), 575-603.

13. Hansen, B.E. (2011). Threshold Autoregression in Economics. *Statistics and Its Interface*, 4, 123-127.

14. Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica* 46(6), 1251-1271.

15. Hausman, J. A. (2001). Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left. *Journal of Economic Perspectives* 15(4), 57-67.

16. Jeong, J. and G. S. Maddala (1991). Measurement Errors and Tests for Rationality. *Journal of Business and Economic Statistics* 9(4), 431-439.

17. Madansky, A. (1959). The Fitting of Straight Lines when Both Variables are subject to Error. *Journal of the American Statistical Association* 54, 173-205.

18. Li, C. W. and W. K. Li (1996). On a Double-threshold Autoregressive Heteroscedastic Time Series Model. *Journal of Applied Econometrics* 11(3), 253-274.

19. Li, W. K. and K. Lam (1995). Modeling Asymmetry in Stock Returns by a Threshold Autoregressive Conditional Heteroscedastic Model. *The Statistician* 44(3), 333-341.

20. Newey, W. K. and K. D. West (1987). A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* 55(3), 703-708.

21. Schennach, S. M. (2004). Estimation of Nonlinear Models with Measurement Error. *Econometrica* 72(1), 33-75.

22. Tong, H. and K.S. Lim (1980). Threshold Autoregression, Limit Cycles and Cyclical Data. *Journal of the Royal Statistical Society, Series B* 42(3), 245-292.

23. Tong, H. (1983). *Threshold Models in Nonlinear Time Series Analysis: Lecture Notes in Statistics* 21. Berlin: Springer.

24. Tong, H. (2011). Threshold Models in Time Series Analysis-30 Years on. *Statistics and Its Interface* 4(2), 107–118.

25. Tsay, R. S. (1998). Testing and Modeling Multivariate Threshold Models. *Journal of the American Statistical Association* 93, 1188-1202.

26. Wong, S.T. and W.K. Li (2010). A Threshold Approach for Peaks-over-threshold Modelling Using Maximum Product of Spacings. *Statistica Sinica* 20, 1257-1572.

27. Xia, Y. and Tong, H. (2011) Feature Matching in Time Series Modeling (with discussion), Rejoinder. *Statistical Science* 26(1), 21-61.

# Appendix: Mathematical Proofs

**Proof of Lemma 1:**

By plugging the true model

$$y_i = \beta_1 x_i + \delta x_i \Psi_i^0 (\gamma_0) + \varepsilon_i,$$

into Equation (8), and using $\Psi_i^0 (\gamma_0) = \Psi_i (\gamma_0 + u_i)$, under Assumptions A1-A6, we have

$$\begin{aligned}
\widehat{\beta}_1 (\widehat{\gamma}) &= \sum_{i=1}^{n} x_i y_i (1 - \Psi_i (\widehat{\gamma})) \left( \sum_{i=1}^{n} x_i^2 (1 - \Psi_i (\widehat{\gamma})) \right)^{-1} \\
&= \beta_1 + \delta \frac{\sum_{i=1}^n x_i^2 \Psi_i^0 (\gamma_0) (1 - \Psi_i (\widehat{\gamma}))}{\sum_{i=1}^n x_i^2 (1 - \Psi_i (\widehat{\gamma}))} + \frac{\sum_{i=1}^n x_i \varepsilon_i (1 - \Psi_i (\widehat{\gamma}))}{\sum_{i=1}^n x_i^2 (1 - \Psi_i (\widehat{\gamma}))} \\
&= \beta_1 + \delta \frac{\sum_{i=1}^n x_i^2 \Psi_i (\gamma_0 + u_i) (1 - \Psi_i (\widehat{\gamma}))}{\sum_{i=1}^n x_i^2 (1 - \Psi_i (\widehat{\gamma}))} + O_p(\frac{1}{\sqrt{n}}) \\
&= \beta_1 + \delta \frac{\sum_{i=1}^n x_i^2 [\Psi_i (\gamma_0 + u_i) - \Psi_i (\max \{\gamma_0 + u_i, \widehat{\gamma}\})]}{\sum_{i=1}^n x_i^2 (1 - \Psi_i (\widehat{\gamma}))} + O_p(\frac{1}{\sqrt{n}}).
\end{aligned}$$

Similarly, we can show that

$$\widehat{\beta}_2(\widehat{\gamma}) - \beta_2 = -\delta \frac{\sum_{i=1}^n x_i^2 [\Psi_i (\widehat{\gamma}) - \Psi_i (\max \{\gamma_0 + u_i, \widehat{\gamma}\})]}{\sum_{i=1}^n x_i^2 \Psi_i (\widehat{\gamma})} + O_p(\frac{1}{\sqrt{n}}).$$

**Proof of Lemma 2:**

The proof is similar to that of Lemma 1. By plugging the true model

$$y_i = \beta_1 x_i + \delta x_i \Psi_i^0 (\gamma_0) + \varepsilon_i,$$

into Equation (12), under Assumptions A1-A6, we have

$$\begin{aligned}
\widetilde{\beta}_1(\lambda) &= \sum_{i=1}^{n} x_i y_i \left(1-\Psi_i\left(\underline{\gamma_\lambda}\right)\right)\left(\sum_{i=1}^{n} x_i^2\left(1-\Psi_i\left(\underline{\gamma_\lambda}\right)\right)\right)^{-1} \\
&= \beta_1+\delta\frac{\sum_{i=1}^{n} x_i^2\Psi_i^0\left(\gamma_0\right)\left(1-\Psi_i\left(\underline{\gamma_\lambda}\right)\right)}{\sum_{i=1}^{n} x_i^2\left(1-\Psi_i\left(\underline{\gamma_\lambda}\right)\right)}+\frac{\sum_{i=1}^{n} x_i\varepsilon_i\left(1-\Psi_i\left(\underline{\gamma_\lambda}\right)\right)}{\sum_{i=1}^{n} x_i^2\left(1-\Psi_i\left(\underline{\gamma_\lambda}\right)\right)} \\
&= \beta_1+\delta\frac{\sum_{i=1}^{n} x_i^2\Psi_i\left(\gamma_0+u_i\right)\left(1-\Psi_i\left(\underline{\gamma_\lambda}\right)\right)}{\sum_{i=1}^{n} x_i^2\left(1-\Psi_i\left(\underline{\gamma_\lambda}\right)\right)}+O_p(\frac{1}{\sqrt{n}}) \\
&= \beta_1+\delta\frac{\sum_{i=1}^{n} x_i^2\left[\Psi_i\left(\gamma_0+u_i\right)-\Psi_i\left(\max\left\{\gamma_0+u_i,\underline{\gamma_\lambda}\right\}\right)\right]}{\sum_{i=1}^{n} x_i^2\left(1-\Psi_i\left(\underline{\gamma_\lambda}\right)\right)}+O_p(\frac{1}{\sqrt{n}}).
\end{aligned}$$

Similarly, we can show that

$$\widetilde{\beta}_2(\lambda)-\beta_2=-\delta\frac{\sum_{i=1}^{n} x_i^2\left[\Psi_i\left(\overline{\gamma_\lambda}\right)-\Psi_i\left(\max\left\{\gamma_0+u_i,\overline{\gamma_\lambda}\right\}\right)\right]}{\sum_{i=1}^{n} x_i^2\Psi_i\left(\overline{\gamma_\lambda}\right)}+O_p(\frac{1}{\sqrt{n}}).$$

**Proof of Theorem 1:**

We first prove that $\widetilde{\beta}_1(\lambda)$ is less biased than $\widehat{\beta}_1\left(\widehat{\gamma}\right)$. Based on Lemmas 1 and 2, we only need to prove the following inequality:

$$\left|\frac{\sum_{i=1}^{n} x_i^2\left[\Psi_i\left(\gamma_0+u_i\right)-\Psi_i\left(\max\left\{\gamma_0+u_i,\widehat{\gamma}\right\}\right)\right]}{\sum_{i=1}^{n} x_i^2\left(1-\Psi_i\left(\widehat{\gamma}\right)\right)}\right|>\left|\frac{\sum_{i=1}^{n} x_i^2\left[\Psi_i\left(\gamma_0+u_i\right)-\Psi_i\left(\max\left\{\gamma_0+u_i,\underline{\gamma_\lambda}\right\}\right)\right]}{\sum_{i=1}^{n} x_i^2\left(1-\Psi_i\left(\underline{\gamma_\lambda}\right)\right)}\right|.$$

Note that both $\Psi_i\left(\gamma_0+u_i\right)-\Psi_i\left(\max\left\{\gamma_0+u_i,\widehat{\gamma}\right\}\right)$ and $\Psi_i\left(\gamma_0+u_i\right)-\Psi_i\left(\max\left\{\gamma_0+u_i,\underline{\gamma_\lambda}\right\}\right)$ are non-negative. Thus, it suffices to show that

$$\frac{\sum_{i=1}^{n} x_i^2\left[\Psi_i\left(\gamma_0+u_i\right)-\Psi_i\left(\max\left\{\gamma_0+u_i,\widehat{\gamma}\right\}\right)\right]}{\sum_{i=1}^{n} x_i^2\left(1-\Psi_i\left(\widehat{\gamma}\right)\right)}>\frac{\sum_{i=1}^{n} x_i^2\left[\Psi_i\left(\gamma_0+u_i\right)-\Psi_i\left(\max\left\{\gamma_0+u_i,\underline{\gamma_\lambda}\right\}\right)\right]}{\sum_{i=1}^{n} x_i^2\left(1-\Psi_i\left(\underline{\gamma_\lambda}\right)\right)}. \tag{19}$$

Using the definition of the indicator function $\Psi_i\left(\cdot\right)$ ,the left side of the inequality (19) can be written as

$$\frac{\sum_{i=1}^{n} x_i^2 \left[\Psi_i\left(\gamma_0+u_i\right)-\Psi_i\left(\max\left\{\gamma_0+u_i,\widehat{\gamma}\right\}\right)\right]}{\sum_{i=1}^{n} x_i^2\left(1-\Psi_i\left(\widehat{\gamma}\right)\right)} = \frac{\sum_{i=1}^{n} x_i^2 I(\gamma_0+u_i<z_i\leq\widehat{\gamma})}{\sum_{i=1}^{n} x_i^2 I(z_i\leq\widehat{\gamma})}$$
$$= \frac{\sum_{i=1}^{n} x_i^2 I(\gamma_0+u_i<z_i\leq\widehat{\gamma},\gamma_0+u_i\leq\underline{\gamma_\lambda})+\sum_{i=1}^{n} x_i^2 I(\gamma_0+u_i<z_i\leq\widehat{\gamma},\gamma_0+u_i>\underline{\gamma_\lambda})}{\sum_{i=1}^{n} x_i^2 I(z_i\leq\underline{\gamma_\lambda})+\sum_{i=1}^{n} x_i^2 I(\underline{\gamma_\lambda}<z_i\leq\widehat{\gamma})}.$$

Given $\widehat{\gamma}>\underline{\gamma_\lambda}$, we have $\sum_{i=1}^{n} x_i^2 I(\gamma_0+u_i<z_i\leq\widehat{\gamma},\gamma_0+u_i\leq\underline{\gamma_\lambda})\geq\sum_{i=1}^{n} x_i^2 I(\gamma_0+u_i<z_i\leq\underline{\gamma_\lambda})$ and $\sum_{i=1}^{n} x_i^2 I(\gamma_0+u_i<z_i\leq\widehat{\gamma},\gamma_0+u_i>\underline{\gamma_\lambda})\geq\sum_{i=1}^{n} x_i^2 I(\underline{\gamma_\lambda}<z_i\leq\widehat{\gamma})$.

Thus,

$$\frac{\sum_{i=1}^{n} x_i^2 \left[\Psi_i\left(\gamma_0+u_i\right)-\Psi_i\left(\max\left\{\gamma_0+u_i,\widehat{\gamma}\right\}\right)\right]}{\sum_{i=1}^{n} x_i^2\left(1-\Psi_i\left(\widehat{\gamma}\right)\right)}$$
$$\geq \frac{\sum_{i=1}^{n} x_i^2 I(\gamma_0+u_i<z_i\leq\underline{\gamma_\lambda})+\sum_{i=1}^{n} x_i^2 I(\underline{\gamma_\lambda}<z_i\leq\widehat{\gamma})}{\sum_{i=1}^{n} x_i^2 I(z_i\leq\underline{\gamma_\lambda})+\sum_{i=1}^{n} x_i^2 I(\underline{\gamma_\lambda}<z_i\leq\widehat{\gamma})}$$
$$> \frac{\sum_{i=1}^{n}\left[x_i^2 I(\gamma_0+u_i<z_i\leq\underline{\gamma_\lambda}\right]}{\sum_{i=1}^{n} x_i^2 I(z_i\leq\underline{\gamma_\lambda})}. \tag{20}$$

Using the definition of the indicator function $\Psi_i(\cdot)$, the right side of the inequality (19) can be written as

$$\frac{\sum_{i=1}^{n} x_i^2\left[\Psi_i\left(\gamma_0+u_i\right)-\Psi_i\left(\max\left\{\gamma_0+u_i,\underline{\gamma_\lambda}\right\}\right)\right]}{\sum_{i=1}^{n} x_i^2\left(1-\Psi_i\left(\underline{\gamma_\lambda}\right)\right)} = \frac{\sum_{i=1}^{n}\left[x_i^2 I(\gamma_0+u_i<z_i\leq\underline{\gamma_\lambda}\right]}{\sum_{i=1}^{n} x_i^2 I(z_i\leq\underline{\gamma_\lambda})}. \tag{21}$$

Combining the inequality (20) and the equation (21), we have

$$\frac{\sum_{i=1}^{n} x_i^2 \left[\Psi_i\left(\gamma_0+u_i\right)-\Psi_i\left(\max\left\{\gamma_0+u_i,\widehat{\gamma}\right\}\right)\right]}{\sum_{i=1}^{n} x_i^2\left(1-\Psi_i\left(\widehat{\gamma}\right)\right)} > \frac{\sum_{i=1}^{n} x_i^2\left[\Psi_i\left(\gamma_0+u_i\right)-\Psi_i\left(\max\left\{\gamma_0+u_i,\underline{\gamma_\lambda}\right\}\right)\right]}{\sum_{i=1}^{n} x_i^2\left(1-\Psi_i\left(\underline{\gamma_\lambda}\right)\right)}$$

which completes the proof.

Next, we prove that $\widetilde{\beta}_2(\lambda)$ is less biased than $\widehat{\beta}_2\left(\widehat{\gamma}\right)$. Using Lemmas 1 and 2, we only need to show that

$$\left|\frac{\sum_{i=1}^{n} x_i^2\left[\Psi_i\left(\widehat{\gamma}\right)-\Psi_i\left(\max\left\{\gamma_0+u_i,\widehat{\gamma}\right\}\right)\right]}{\sum_{i=1}^{n} x_i^2\Psi_i\left(\widehat{\gamma}\right)}\right| > \left|\frac{\sum_{i=1}^{n} x_i^2\left[\Psi_i\left(\overline{\gamma_\lambda}\right)-\Psi_i\left(\max\left\{\gamma_0+u_i,\overline{\gamma_\lambda}\right\}\right)\right]}{\sum_{i=1}^{n} x_i^2\Psi_i\left(\overline{\gamma_\lambda}\right)}\right|.$$

Since both $\Psi_i\left(\widehat{\gamma}\right)-\Psi_i\left(\max\left\{\gamma_0+u_i,\widehat{\gamma}\right\}\right)$ and $\Psi_i\left(\overline{\gamma_\lambda}\right)-\Psi_i\left(\max\left\{\gamma_0+u_i,\overline{\gamma_\lambda}\right\}\right)$ are non-negative, we only need to show that

$$\frac{\sum_{i=1}^{n} x_i^2 \left[\Psi_i\left(\widehat{\gamma}\right) - \Psi_i\left(\max\left\{\gamma_0 + u_i, \widehat{\gamma}\right\}\right)\right]}{\sum_{i=1}^{n} x_i^2 \Psi_i\left(\widehat{\gamma}\right)} > \frac{\sum_{i=1}^{n} x_i^2 \left[\Psi_i\left(\overline{\gamma_\lambda}\right) - \Psi_i\left(\max\left\{\gamma_0 + u_i, \overline{\gamma_\lambda}\right\}\right)\right]}{\sum_{i=1}^{n} x_i^2 \Psi_i\left(\overline{\gamma_\lambda}\right)}.$$

Given $\widehat{\gamma} < \overline{\gamma_\lambda}$,we have

$$\begin{aligned}
& \frac{\sum_{i=1}^{n} x_i^2 \left[\Psi_i\left(\widehat{\gamma}\right) - \Psi_i\left(\max\left\{\gamma_0 + u_i, \widehat{\gamma}\right\}\right)\right]}{\sum_{i=1}^{n} x_i^2 \Psi_i\left(\widehat{\gamma}\right)} \\
= & \frac{\sum_{i=1}^{n} x_i^2 I\left(\widehat{\gamma} < z_i < \gamma_0 + u_i\right)}{\sum_{i=1}^{n} x_i^2 I\left(\widehat{\gamma} < z_i\right)} \\
\geq & \frac{\sum_{i=1}^{n} x_i^2 I\left(\overline{\gamma_\lambda} < z_i < \gamma_0 + u_i\right) + \sum_{i=1}^{n} x_i^2 I(\widehat{\gamma} < z_i \leq \overline{\gamma_\lambda})}{\sum_{i=1}^{n} x_i^2 I\left(\overline{\gamma_\lambda} < z_i\right) + \sum_{i=1}^{n} x_i^2 I(\widehat{\gamma} < z_i \leq \overline{\gamma_\lambda})} \\
> & \frac{\sum_{i=1}^{n} x_i^2 I\left(\overline{\gamma_\lambda} < z_i < \gamma_0 + u_i\right)}{\sum_{i=1}^{n} x_i^2 I\left(\overline{\gamma_\lambda} < z_i\right)}
\end{aligned}$$

and

$$\frac{\sum_{i=1}^{n} x_i^2 \left[\Psi_i\left(\overline{\gamma_\lambda}\right) - \Psi_i\left(\max\left\{\gamma_0 + u_i, \overline{\gamma_\lambda}\right\}\right)\right]}{\sum_{i=1}^{n} x_i^2 \Psi_i\left(\overline{\gamma_\lambda}\right)} = \frac{\sum_{i=1}^{n} x_i^2 I\left(\overline{\gamma_\lambda} < z_i < \gamma_0 + u_i\right)}{\sum_{i=1}^{n} x_i^2 I\left(\overline{\gamma_\lambda} < z_i\right)}.$$

Thus, we have

$$\frac{\sum_{i=1}^{n} x_i^2 \left[\Psi_i\left(\widehat{\gamma}\right) - \Psi_i\left(\max\left\{\gamma_0 + u_i, \widehat{\gamma}\right\}\right)\right]}{\sum_{i=1}^{n} x_i^2 \Psi_i\left(\widehat{\gamma}\right)} > \frac{\sum_{i=1}^{n} x_i^2 \left[\Psi_i\left(\overline{\gamma_\lambda}\right) - \Psi_i\left(\max\left\{\gamma_0 + u_i, \overline{\gamma_\lambda}\right\}\right)\right]}{\sum_{i=1}^{n} x_i^2 \Psi_i\left(\overline{\gamma_\lambda}\right)}$$

which completes the proof.

**Proof of Theorem 2:**

Consider a general threshold regression with multiple regressors

$$y_i = \beta_1' x_i + \left(\beta_2' - \beta_1'\right) x_i I\left(z_i^0 > \gamma_0\right) + \varepsilon_i,$$

where $x_i$ is a $k \times 1$ vector of covariates. When $k = 1$, we have the univariate model given by the equation (1).

The model can be rewritten in matrix form as follows:

$$Y = [I - I^0(\gamma_0)]X'\beta_1 + I^0(\gamma_0)X'\beta_2 + \varepsilon,$$

where

$$I^0(\gamma_0) = diag\left\{\Psi_1^0(\gamma_0), \Psi_2^0(\gamma_0), ..., \Psi_n^0(\gamma_0)\right\}.$$

$\Psi_i^0(\gamma_0)$ is an indicator function defined in the equation (2); $Y = (y_1, y_2, ..., y_n)'$ , $X = (x_1, x_2, ..., x_n)$ and $\varepsilon = (\varepsilon_1,\ \varepsilon_2,\ ...,\ \varepsilon_n)'$. We observe

$$z_i = z_i^0 + u_i.$$

Let

$$\Psi_i(\gamma) = I(z_i > \gamma),$$

and

$$I(\gamma) = diag\{\Psi_1(\gamma), \Psi_2(\gamma), ..., \Psi_n(\gamma)\}.$$

Note that $\Psi_i(\gamma) = I(z_i > \gamma) = I(z_i^0 > \gamma - u_i) = \Psi_i^0(\gamma - u_i)$ and $\Psi_i(\gamma_0 + u_i) = \Psi_i^0(\gamma_0)$, thus, $I(\gamma_0 + u) = I^0(\gamma_0)$.

Given any $\gamma \in (\underline{\gamma_\lambda},\ \overline{\gamma_\lambda})$, the conventional LS estimators for $\beta$ are given by

$$\begin{aligned}\widehat{\beta}_1(\gamma) &= [X(I - I(\gamma))X']^{-1}X[I - I(\gamma)]Y \\ &= [X(I - I(\gamma))X']^{-1}X(I - I(\gamma))[X'\beta_1 + I(\gamma_0 + u)X'\delta + \varepsilon] \\ &= \beta_1 + \varphi_1 + o_p(1), \end{aligned} \tag{22}$$

and

$$\begin{aligned}\widehat{\beta}_2(\gamma) &= (XI(\gamma)X')^{-1}X'I(\gamma)Y \\ &= (XI(\gamma)X')^{-1}XI(\gamma)[X'\beta_2 - I(\gamma_0 + u)X'\delta + \varepsilon] \\ &= \beta_2 - \varphi_2 + o_p(1) \end{aligned} \tag{23}$$

where

$$\varphi_1 = [X(I - I(\gamma))X']^{-1}X(I - I(\gamma))I(\gamma_0 + u)X'\delta$$

and

$$\varphi_2 = (XI(\gamma)X')^{-1}XI(\gamma)I(\gamma_0 + u)X'\delta.$$

Given any $\lambda \in (0, 1/2)$, the new estimators $\widetilde{\beta}_1(\lambda)$ and $\widetilde{\beta}_2(\lambda)$ are

$$\widetilde{\beta}_1(\lambda) = [X(I - I(\underline{\gamma_\lambda}))X']^{-1}X[I - I(\underline{\gamma_\lambda})]Y, \tag{24}$$

and

$$\widetilde{\beta}_2(\lambda) = [XI(\overline{\gamma_\lambda})X']^{-1}XI(\overline{\gamma_\lambda})Y. \tag{25}$$

Under the null, we have $u = 0$ and thus $I(\gamma_0) = I^0(\gamma_0)$. Given the assumption that $\gamma_0 \in (\underline{\gamma_\lambda}, \overline{\gamma_\lambda})$, the equation (24) can be written as

$$\begin{aligned}
\widetilde{\beta}_1(\lambda) &= [X(I - I(\underline{\gamma_\lambda}))X']^{-1}X[I - I(\underline{\gamma_\lambda})]Y \\
&= \beta_1 + [X(I - I(\underline{\gamma_\lambda}))X']^{-1}X[I - I(\underline{\gamma_\lambda})]I(\gamma_0)X'\beta_2 + [X(I - I(\underline{\gamma_\lambda}))X']^{-1}X[I - I(\underline{\gamma_\lambda})]\varepsilon \\
&= \beta_1 + [X(I - I(\underline{\gamma_\lambda}))X']^{-1}X[I - I(\underline{\gamma_\lambda})]\varepsilon.
\end{aligned}$$

The equation (22) can be written as

$$\begin{aligned}
\widehat{\beta}_1(\gamma_0) &= \beta_1 + [X(I - I(\gamma_0))X']^{-1}X(I - I(\gamma_0))I(\gamma_0)X'\delta + (X(I - I(\gamma_0))X')^{-1}X(I - I(\gamma_0))\varepsilon \\
&= \beta_1 + [X(I - I(\gamma_0))X']^{-1}X[I - I(\gamma_0)]\varepsilon.
\end{aligned}$$

Thus,

$$\sqrt{n}(\widetilde{\beta}_1(\lambda) - \widehat{\beta}_1(\gamma_0)) = \sqrt{n}[(X(I - I(\underline{\gamma_\lambda}))X')^{-1}X(I - I(\underline{\gamma_\lambda})) - (X(I - I(\gamma_0))X')^{-1}X(I - I(\gamma_0))]\varepsilon.$$

Similarly, we have

$$\sqrt{n}(\widetilde{\beta}_2(\lambda) - \widehat{\beta}_2(\gamma_0)) = \sqrt{n}[(X(I(\overline{\gamma_\lambda}))X')^{-1}X(I(\overline{\gamma_\lambda})) - (X(I(\gamma_0))X')^{-1}X(I(\gamma_0))]\varepsilon.$$

Before proceeding further, for any $\gamma$, we define the following conditional moment functionals for $x_i$ as

$$\begin{aligned}
M_1(\gamma) &= E\left(x_i x_i' I(z_i \leq \gamma)\right), \\
M_2(\gamma) &= E\left(x_i x_i' I(z_i > \gamma)\right).
\end{aligned}$$

For any $\gamma_1$ and $\gamma_2$ , define the conditional moment matrix for $x_i\varepsilon_i$ as

$$\begin{aligned}
\Omega_{11}(\gamma_1, \gamma_2) &= E(x_i I(z_i \leq \gamma_1)\varepsilon_i\varepsilon_i I(z_i \leq \gamma_2)x_i'), \\
\Omega_{12}(\gamma_1, \gamma_2) &= E(x_i I(z_i \leq \gamma_1)\varepsilon_i\varepsilon_i I(z_i > \gamma_2)x_i'), \\
\Omega_{22}(\gamma_1, \gamma_2) &= E(x_i I(z_i > \gamma_1)\varepsilon_i\varepsilon_i I(z_i > \gamma_2)x_i').
\end{aligned}$$

The corresponding sample moment estimators are defined as

$$\begin{aligned}
\widehat{M}_1(\gamma) &= \frac{X(I - I(\gamma))X'}{n}, \\
\widehat{M}_2(\gamma) &= \frac{XI(\gamma)X'}{n},
\end{aligned}$$

and

$$\begin{aligned}
\widehat{\Omega}_{11}(\gamma_1, \gamma_2) &= \frac{X(I - I(\gamma_1))\widehat{\varepsilon\varepsilon}'(I - I(\gamma_2))X'}{n} \\
\widehat{\Omega}_{12}(\gamma_1, \gamma_2) &= \frac{X(I - I(\gamma_1))\widehat{\varepsilon\varepsilon}'(I(\gamma_2))X'}{n} \\
\widehat{\Omega}_{22}(\gamma_1, \gamma_2) &= \frac{X(I(\gamma_1))\widehat{\varepsilon\varepsilon}'(I(\gamma_2))X'}{n}
\end{aligned}$$

Under Assumptions A1-A6, the law of large number holds and thus $\widehat{M}_1(\gamma) \xrightarrow{p} M_1(\gamma)$, $\widehat{M}_2(\gamma) \xrightarrow{p} M_2(\gamma)$, $\widehat{\Omega}_{ij}(\gamma_1, \gamma_2) \xrightarrow{p} \Omega_{ij}(\gamma_1, \gamma_2)$ for all $i = 1, 2,\ j = 1, 2$.

Next, we derive the covariance matrix of $\sqrt{n}\left(\widetilde{\beta}_1(\lambda) - \widehat{\beta}_1(\gamma_0)\right)$. Note that

$$\begin{aligned}
&\widehat{Var}\left[\sqrt{n}\left(\widetilde{\beta}_1(\lambda) - \widehat{\beta}_1(\gamma_0)\right)\right] \\
= &[(\frac{X(I - I(\underline{\gamma_\lambda}))X'}{n})^{-1}\frac{X(I - I(\underline{\gamma_\lambda}))\widehat{\varepsilon}}{\sqrt{n}} - (\frac{X(I - I(\gamma_0))X'}{n})^{-1}\frac{X(I - I(\gamma_0))\widehat{\varepsilon}}{\sqrt{n}}] \\
&[(\frac{X(I - I(\underline{\gamma_\lambda}))X'}{n})^{-1}\frac{X(I - I(\underline{\gamma_\lambda}))\widehat{\varepsilon}}{\sqrt{n}} - (\frac{X(I - I(\gamma_0))X'}{n})^{-1}\frac{X(I - I(\gamma_0))\widehat{\varepsilon}}{\sqrt{n}}]' \\
= &(\frac{X(I - I(\underline{\gamma_\lambda}))X'}{n})^{-1}\frac{X(I - I(\underline{\gamma_\lambda}))\widehat{\varepsilon\varepsilon}'(I - I(\underline{\gamma_\lambda}))X'}{n}(\frac{X(I - I(\underline{\gamma_\lambda}))X'}{n})^{-1} \\
&+(\frac{X(I - I(\gamma_0))X'}{n})^{-1}\frac{X(I - I(\gamma_0))\widehat{\varepsilon\varepsilon}'(I - I(\gamma_0))X'}{n}(\frac{X(I - I(\gamma_0))X'}{n})^{-1} \\
&-(\frac{X(I - I(\gamma_0))X'}{n})^{-1}\frac{X(I - I(\gamma_0))\widehat{\varepsilon\varepsilon}'(I - I(\underline{\gamma_\lambda}))X'}{n}(\frac{X(I - I(\underline{\gamma_\lambda}))X'}{n})^{-1} \\
&-(\frac{X(I - I(\underline{\gamma_\lambda}))X'}{n})^{-1}\frac{X(I - I(\underline{\gamma_\lambda}))\widehat{\varepsilon\varepsilon}'(I - I(\gamma_0))X'}{n}(\frac{X(I - I(\gamma_0))X'}{n})^{-1} \\
= &\widehat{M}_1(\underline{\gamma_\lambda}))^{-1}\widehat{\Omega}_{11}(\underline{\gamma_\lambda}, \underline{\gamma_\lambda})\widehat{M}_1(\underline{\gamma_\lambda}))^{-1} + \widehat{M}_1(\gamma_0))^{-1}\widehat{\Omega}_{11}(\gamma_0, \gamma_0)\widehat{M}_1(\gamma_0))^{-1} \\
&-\widehat{M}_1(\underline{\gamma_\lambda}))^{-1}\widehat{\Omega}_{11}(\underline{\gamma_\lambda}, \gamma_0)\widehat{M}_1(\gamma_0))^{-1} - \widehat{M}_1(\gamma_0))^{-1}\widehat{\Omega}_{11}(\gamma_0, \underline{\gamma_\lambda})\widehat{M}_1(\underline{\gamma_\lambda}))^{-1} \\
\equiv &\widehat{\Pi}_{11}(\underline{\gamma_\lambda}, \gamma_0), say.
\end{aligned}$$

Using the convergence results of $\widehat{M}_i$ and $\widehat{\Omega}_{ij}$,

$$
\begin{aligned}
& \widehat{\Pi}_{11}(\underline{\gamma_\lambda}, \gamma_0) \\
& \xrightarrow{p} M_1(\underline{\gamma_\lambda})^{-1}\Omega_{11}(\underline{\gamma_\lambda})M_1(\underline{\gamma_\lambda})^{-1} + M_1(\gamma_0)^{-1}\Omega_{11}(\gamma_0)M_1(\gamma_0)^{-1} \\
& -M_1(\underline{\gamma_\lambda})^{-1}\Omega_{11}(\underline{\gamma_\lambda}, \gamma_0)M_1(\gamma_0)^{-1} - M_1(\gamma_0)^{-1}\Omega_{11}(\gamma_0, \underline{\gamma_\lambda})M_1(\underline{\gamma_\lambda})^{-1} \\
\equiv \; & \Pi_{11}(\underline{\gamma_\lambda}, \gamma_0), say.
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
& \widehat{Var}\left(\sqrt{n}(\widetilde{\beta}_2(\lambda) - \widehat{\beta}_2(\gamma_0))\right) \\
= \; & [(\frac{X(I(\overline{\gamma_\lambda}))X'}{n})^{-1}\frac{X(I(\overline{\gamma_\lambda}))\widehat{\varepsilon}}{\sqrt{n}} - (\frac{X(I(\gamma_0))X'}{n})^{-1}\frac{X(I(\gamma_0))\widehat{\varepsilon}}{\sqrt{n}}] \\
& [(\frac{X(I(\overline{\gamma_\lambda}))X'}{n})^{-1}\frac{X(I(\overline{\gamma_\lambda}))\widehat{\varepsilon}}{\sqrt{n}} - (\frac{X(I(\gamma_0))X'}{n})^{-1}\frac{X(I(\gamma_0))\widehat{\varepsilon}}{\sqrt{n}}]' \\
= \; & \widehat{M}_2(\overline{\gamma_\lambda})^{-1}\widehat{\Omega}_{22}(\overline{\gamma_\lambda}, \overline{\gamma_\lambda})\widehat{M}_2(\overline{\gamma_\lambda})^{-1} + \widehat{M}_2(\gamma_0)^{-1}\widehat{\Omega}_{22}(\gamma_0, \gamma_0)\widehat{M}_2(\gamma_0)^{-1} \\
& -\widehat{M}_2(\overline{\gamma_\lambda})^{-1}\widehat{\Omega}_{22}(\overline{\gamma_\lambda}, \gamma_0)\widehat{M}_2(\gamma_0)^{-1} - \widehat{M}_2(\gamma_0)^{-1}\widehat{\Omega}_{22}(\gamma_0, \overline{\gamma_\lambda})\widehat{M}_2(\overline{\gamma_\lambda})^{-1} \\
\equiv \; & \widehat{\Pi}_{22}(\overline{\gamma_\lambda}, \gamma_0) \xrightarrow{p} \Pi_{22}(\overline{\gamma_\lambda}, \gamma_0).
\end{aligned}
$$

The covariance between $\sqrt{n}(\widehat{\beta}_1(\gamma_0) - \widetilde{\beta}_1(\lambda))$ and $\sqrt{n}(\widehat{\beta}_2(\gamma_0) - \widetilde{\beta}_2(\lambda))$ can be written as

$$
\begin{aligned}
& \widehat{Cov}\left(\sqrt{n}(\widetilde{\beta}_1(\lambda) - \widehat{\beta}_1(\gamma_0)), \sqrt{n}(\widehat{\beta}_2(\gamma_0) - \widetilde{\beta}_2(\lambda))\right) \\
= \; & [(\frac{X(I - I(\underline{\gamma_\lambda}))X'}{n})^{-1}\frac{X(I - I(\underline{\gamma_\lambda}))\widehat{\varepsilon}}{\sqrt{n}} - (\frac{X(I - I(\gamma_0))X'}{n})^{-1}\frac{X(I - I(\gamma_0))\widehat{\varepsilon}}{\sqrt{n}}] \\
& [(\frac{X(I(\overline{\gamma_\lambda}))X'}{n})^{-1}\frac{X(I(\overline{\gamma_\lambda}))\widehat{\varepsilon}}{\sqrt{n}} - (\frac{X(I(\gamma_0))X'}{n})^{-1}\frac{X(I(\gamma_0))\widehat{\varepsilon}}{\sqrt{n}}]' \\
= \; & \widehat{M}_1(\underline{\gamma_\lambda})^{-1}\widehat{\Omega}_{12}(\underline{\gamma_\lambda}, \overline{\gamma_\lambda})\widehat{M}_2(\overline{\gamma_\lambda})^{-1} + \widehat{M}_1(\gamma_0)^{-1}\widehat{\Omega}_{12}(\gamma_0, \gamma_0)\widehat{M}_2(\gamma_0)^{-1} \\
& -\widehat{M}_1(\underline{\gamma_\lambda})^{-1}\widehat{\Omega}_{12}(\underline{\gamma_\lambda}, \gamma_0)\widehat{M}_2(\gamma_0)^{-1} - \widehat{M}_1(\gamma_0))^{-1}\widehat{\Omega}_{12}(\gamma_0, \overline{\gamma_\lambda})\widehat{M}_2(\overline{\gamma_\lambda})^{-1} \\
\equiv \; & \widehat{\Pi}_{12}(\underline{\gamma_\lambda}, \overline{\gamma_\lambda}, \gamma_0) \xrightarrow{p} \Pi_{12}(\underline{\gamma_\lambda}, \overline{\gamma_\lambda}, \gamma_0).
\end{aligned}
$$

Let

$$
\begin{aligned}
\Pi(\underline{\gamma_\lambda}, \overline{\gamma_\lambda}, \gamma_0) &= \begin{pmatrix} \Pi_{11}(\underline{\gamma_\lambda}, \gamma_0), \; \Pi_{12}(\underline{\gamma_\lambda}, \overline{\gamma_\lambda}, \gamma_0) \\ \Pi_{12}(\underline{\gamma_\lambda}, \overline{\gamma_\lambda}, \gamma_0)', \Pi_{22}(\overline{\gamma_\lambda}, \gamma_0) \end{pmatrix}, \\
\widehat{\Pi}(\underline{\gamma_\lambda}, \overline{\gamma_\lambda}, \gamma_0) &= \begin{pmatrix} \widehat{\Pi}_{11}(\underline{\gamma_\lambda}, \gamma_0), \; \widehat{\Pi}_{12}(\underline{\gamma_\lambda}, \overline{\gamma_\lambda}, \gamma_0) \\ \widehat{\Pi}_{12}(\underline{\gamma_\lambda}, \overline{\gamma_\lambda}, \gamma_0)', \widehat{\Pi}_{22}(\overline{\gamma_\lambda}, \gamma_0) \end{pmatrix}.
\end{aligned}
$$

We have

$$\widehat{\Pi}(\underline{\gamma_\lambda}, \overline{\gamma_\lambda}, \gamma_0) \xrightarrow{p} \Pi(\underline{\gamma_\lambda}, \overline{\gamma_\lambda}, \gamma_0).$$

Applying the central limiting theorem for martingale processes, we have

$$\sqrt{n} \begin{pmatrix} \widehat{\beta}_1(\gamma_0) - \widetilde{\beta}_1(\lambda) \\ \widehat{\beta}_2(\gamma_0) - \widetilde{\beta}_2(\lambda) \end{pmatrix} \Rightarrow N(0, \Pi(\underline{\gamma_\lambda}, \overline{\gamma_\lambda}, \gamma_0)).$$

Therefore

$$T_0(\lambda) = n \begin{pmatrix} \widehat{\beta}_1(\gamma_0) - \widetilde{\beta}_1(\lambda) \\ \widehat{\beta}_2(\gamma_0) - \widetilde{\beta}_2(\lambda) \end{pmatrix}' \widehat{\Pi}(\underline{\gamma_\lambda}, \overline{\gamma_\lambda}, \gamma_0)^{-1} \begin{pmatrix} \widehat{\beta}_1(\gamma_0) - \widetilde{\beta}_1(\lambda) \\ \widehat{\beta}_2(\gamma_0) - \widetilde{\beta}_2(\lambda) \end{pmatrix} \xrightarrow{d} \chi^2(2k).$$

Under the null, from the Lemma A.9 of Hansen (2000), we have $\widehat{\gamma} - \gamma_0 = O_p(\frac{1}{n})$ and thus the impact from the estimation is negligible. It follows that

$$\begin{aligned} T(\lambda) &= n \begin{pmatrix} \widehat{\beta}_1(\widehat{\gamma}) - \widetilde{\beta}_1(\lambda) \\ \widehat{\beta}_2(\widehat{\gamma}) - \widetilde{\beta}_2(\lambda) \end{pmatrix}' \widehat{\Pi}(\underline{\gamma_\lambda}, \overline{\gamma_\lambda}, \widehat{\gamma})^{-1} \begin{pmatrix} \widehat{\beta}_1(\widehat{\gamma}) - \widetilde{\beta}_1(\lambda) \\ \widehat{\beta}_2(\widehat{\gamma}) - \widetilde{\beta}_2(\lambda) \end{pmatrix} \\ &= T_0(\lambda) + o_p(1) \Rightarrow \chi^2(2k). \end{aligned}$$