STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **15**, March 2026, pp 1611–1631.
Published online in International Academic Press (www.IAPress.org)

Efficient Test for Threshold Regression Models in Short Panel Data

Dounia Bourzik [1,*], Aziz Lmakri [2], Amal Mellouk [3], Abdelhadi Akharif [1]

[1] *LMA, FSTT, Abdelmalek Essaadi University, Tetouan, Morocco*
[2] *AIMCE laboratory, ENSAM, Hassan II University, Casablanca, Morocco*
[3] *Regional Center for Education and Training Trades, Tangier, Morocco*

Abstract In this paper, we propose locally and asymptotically optimal tests (as defined in the Le Cam sense) that are parametric, Gaussian, and adaptive. These tests aim to address the problem of testing the classical regression model against the threshold regression model in short panel data, where n is large and T is small. The foundation of these tests is the Local Asymptotic Normality (LAN) property. We derive the asymptotic relative efficiencies of these tests, specifically in comparison to the Gaussian parametric tests. The results demonstrate that the adaptive tests exhibit higher asymptotic power than the Gaussian tests. Additionally, we conduct simulation studies and analyze real data to evaluate the performance of the suggested tests, and the results confirm their excellent performance.

Keywords Threshold Regression Model, Local Asymptotic Normality, Local Asymptotic Linearity, Panel Data, Gaussian Tests, Adaptive Tests

AMS 2010 subject classifications 62F03, 62G10

DOI: 10.19139/soic-2310-5070-3091

1. Introduction

Several researchers have conducted studies on regression analysis, a statistical method primarily used to examine the relationship between dependent and independent variables. Among the most notable forms of nonlinear regression models is the threshold regression model, which has attracted considerable attention in both statistics and econometrics. The Panel Threshold Regression (PTR) model was initially proposed by [17], who developed a threshold model for non-dynamic panels with individual fixed effects. The Panel Smooth Transition Regression (PSTR) models were introduced by [15], while the Panel Smooth Transition Autoregressive (PSTAR) models were popularized by [13]. These models have found broad applications in various real-life problems, including economics, finance, monetary policy, environmental studies, and medical sciences, and have demonstrated their superiority over classical linear regression models. For instance, [14] showed that applying a PSTR model to French-listed firms from 2009 to 2017 revealed a non-linear relationship between family ownership and firm performance. Similarly, [7] found that the impact of imported technology on industrial employment was non-linear and depended on the level of technology imports. Using a PSTR model on data from developed and developing countries over the period 2000–2019, they demonstrated that threshold effects played a crucial role in determining the employment outcomes of technology imports. In the environmental domain, [11] investigated a non-linear association between the blue economy, renewable energy, and environmental sustainability in the Middle East and North Africa (MENA) region during 2000–2022. Moreover, [32] showed that the finance–growth relationship was non-linear using panel threshold models on annual data for 153 countries from 2011 to 2020.

*Correspondence to: Dounia Bourzik (Email: dounia.bourzik@etu.uae.ac.ma). LMA, FSTT, Abdelmalek Essaadi University, Tetouan, Morocco.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

In this paper, our focus will be on the PTR model, which is defined as follows:

$$y_{it} = \mu + \beta_1 x_{it}\mathbf{1}_{(q_{it}\leq 0)} + \beta_2 x_{it}\mathbf{1}_{(q_{it}>0)} + \varepsilon_{it}, \quad i = 1, \ldots, n;\ t = 1, \ldots, T, \tag{1}$$

where y_{it} is a panel observation for individual i and time t ($i = 1, \ldots, n$ (large n); $t = 1, \ldots, T$ (small T)), $N = nT$ is the sample size. The regression parameters are μ and $(\beta_1, \beta_2)'$. The threshold variable q_{it} and the regressor x_{it} are scalar. The indicator function is denoted as $\mathbf{1}_{(.)}$. The sequence of unobservable random variables $\{\varepsilon_{it}, 1 \leq i \leq n, 1 \leq t \leq T\}$ is an i.i.d. with mean zero, finite variance σ_f^2, and probability density $f : \varepsilon \mapsto f(\varepsilon) := (1/\sigma_f)f_1(\varepsilon/\sigma_f)$ (where $f_1 \in \mathcal{F}_A$).

Several methods have been established in the literature for estimating the parameters of threshold regression models. These methods include the method of least squares, the nonlinear least squares method, and the concentrated simulated maximum likelihood method. Examples of these methods can be found in the works of [17, 15, 13, 9].

Before addressing the problem of estimating the parameters of model (1), it is crucial to determine whether it is indeed a threshold regression model and how to proceed with the test.

Clearly, model (1) reduces to the classical regression model

$$y_{it} = \mu + \beta x_{it} + \varepsilon_{it}, \quad i = 1, \ldots, n;\ t = 1, \ldots, T,$$

if and only if $\beta_1 = \beta_2 = \beta$. The problem we are addressing is related to the detection problem. More specifically, this problem involves the testing of the null hypothesis $\mathcal{H}_0 : \beta_1 = \beta_2 = \beta$ with unspecified μ, β, σ_f^2, and f_1 against the alternative hypothesis $\mathcal{H}_1 : \beta_1 \neq \beta_2$.

An alternative intuitive way of writing (1) is:

$$y_{it} = \mu + \beta_1 x_{it}^- + \beta_2 x_{it}^+ + \varepsilon_{it}, \tag{2}$$

one traditional method to eliminate the scalar parameter μ is to center y_{it}, x_{it}^-, and x_{it}^+. The equation (2) becomes:

$$Y_{it} = \beta_1 X_{it}^- + \beta_2 X_{it}^+ + \varepsilon_{it}, \tag{3}$$

where $Y_{it} = y_{it} - \overline{y}$, $X_{it}^- = x_{it}^- - \overline{x^-}$, and $X_{it}^+ = x_{it}^+ - \overline{x^+}$. $\overline{y} = \frac{1}{N}\sum_{i=1}^{n}\sum_{t=1}^{T} y_{it}$, $\overline{x^-} = \frac{1}{N}\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}^-$, and $\overline{x^+} = \frac{1}{N}\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}^+$ are the means respectively of y_{it}, x_{it}^-, and x_{it}^+.

The main technical tool used in this study is Le Cam's asymptotic theory of statistical experiments and the properties of LAN families. For more information, refer to [21, 22]. This powerful method has been successfully applied to various inference problems. Relevant references on this topic include [6, 1, 25, 12, 28].

In a LAN family, the random vector $\Delta_{f_1}^{(n)}(\theta)$, referred to as a central sequence, is the $\sqrt{n}$-normalized derivative of the logarithm of the likelihood function with respect to the parameter θ. Intuitively, this vector measures both the direction and magnitude in which the logarithm of the likelihood changes when the parameter is locally perturbed. It summarizes all the local information in the data and enables the construction of asymptotically optimal tests. Within this framework, the logarithm of the likelihood ratio can be locally approximated as $\tau^{(n)'}\Delta_{f_1}^{(n)}(\theta) - \frac{1}{2}\tau^{(n)'}\Gamma_{f_1}^{(n)}(\theta)\tau^{(n)}$, where the central sequence $\Delta_{f_1}^{(n)}(\theta)$ follows an asymptotically normal distribution with mean zero under the null hypothesis, mean $\Gamma_{f_1}(\theta)\tau$ under the alternative hypothesis, and covariance matrix $\Gamma_{f_1}(\theta)$ under both hypotheses.

The local asymptotic normality results play a crucial role in this treatment as they provide guidance for constructing parametric tests that are optimal, both locally and asymptotically. Next, we discuss the derivation of Gaussian tests, which are optimal under Gaussian densities and remain valid even under non-Gaussian densities. In parametric models, the density function f of the innovation is predetermined. However, these models are better suited for practical situations where f is unspecified. This motivates us to consider semiparametric models. The fact that f is generally unknown leads to a decrease in efficiency when compared to parametric situations. When this decrease is zero, the semiparametric model is considered adaptive since both parametric and semiparametric bounds coincide for all f. For further details, please refer to [16, 3, 10, 27].

Threshold regression models offer greater flexibility than traditional linear models in capturing complex relationships between variables. Ignoring these threshold effects can lead to misleading conclusions, underscoring the importance of testing the null hypothesis of linearity against the alternative PTR specification to ensure model adequacy. Motivated by this consideration, this study focuses on detecting the presence of threshold effects in regression frameworks. More precisely, it aims to test a classical regression model against an alternative that introduces a threshold component in short panel settings characterized by a large cross-sectional dimension (n) and a small time dimension (T). Building on the local asymptotic normality property, the paper develops locally and asymptotically optimal tests. Two types of procedures are proposed: parametric procedures—applicable when the error density f_1 is specified, with the Gaussian test ($f_1 = f_{\mathcal{N}(0,1)}$) as a particular case—and adaptive procedures designed for situations in which f_1 is unspecified. We derive the asymptotic relative efficiencies of adaptive tests in comparison to Gaussian parametric tests. The methodology is supported through Monte Carlo simulations and an empirical application examining the relationship between urbanization and carbon dioxide emissions, demonstrating the robustness of both the Gaussian and adaptive tests, which also outperform the likelihood ratio tests proposed by [17].

After this introduction, the rest of the paper is organized as follows. Subsection 2.1 provides the main definitions and assumptions, while Subsection 2.2 establishes the LAN property. In Subsection 3.1, we propose the optimal parametric test (for specified f_1), and in Subsection 3.2, we present the specific case of the Gaussian test. Subsection 3.3 is dedicated to adaptive tests. Asymptotic relative efficiencies with respect to the Gaussian test are derived in Section 4. Section 5 is devoted to validating our theoretical results through numerical simulations using the RStudio program. In Section 6, we apply our proposed tests to a real dataset that investigates the relationship between urbanization and carbon dioxide emissions, using both RStudio and Stata programs. Finally, we provide some conclusions.

2. Local asymptotic normality

2.1. Notations and main technical assumptions

Denote by $\mathbb{P}^{(n)}_{\sigma_f^2,\beta;f_1}$ the probability distribution under the null hypothesis $\beta_1 = \beta_2 = \beta$. Under the alternative, $\mathbb{P}^{(n)}_{\sigma_f^2,\beta_1,\beta_2;f_1}$ is the probability distribution of the observations $Y^{(n)} = \left(Y_1^{(n)\prime}, Y_2^{(n)\prime}, \ldots, Y_n^{(n)\prime}\right)'$, where $Y_i^{(n)} := (Y_{i1}, Y_{i2}, \ldots, Y_{iT})'$ generated by model (3).

The main technical tool used below is local asymptotic normality with respect to $(\sigma_f^2, \beta_1, \beta_2)$, at $(\sigma_f^2, \beta, \beta)$, of the families of distributions

$$\mathcal{P}^{(n)}_{f_1} := \left\{\mathbb{P}^{(n)}_{\sigma_f^2,\beta_1,\beta_2;f_1} : \sigma_f^2 > 0 \text{ and } (\beta_1, \beta_2) \in \mathbb{R}^2\right\}.$$

To establish the properties of a LAN, certain technical assumptions need to be made regarding the density f (**Assumption (A)**) and the asymptotic behaviour of the regressors (**Assumption (B)**). We are listing these assumptions here for clarity.

Assumption (A)

(A.1) $f(x) > 0,\ \forall x \in \mathbb{R};\ \int_{\mathbb{R}} x f(x)dx = 0;\ 0 < \sigma_f^2 := \int_{\mathbb{R}} x^2 f(x)dx < \infty;$

(A.2) f is absolutely continuous on bounded intervals, i.e., there exists f' such that

$$f(b) - f(a) = \int_a^b f'(x)dx \quad \text{for all } a < b,$$

and, letting $\phi_f := -f'/f$, assume that

$$I_\phi(f) := \int_{\mathbb{R}} \phi_f^2(x) f(x)dx \text{ and } J_\phi(f) := \int_{\mathbb{R}} x^2 \phi_f^2(x) f(x)dx \text{ are finite.}$$

Letting

$\phi_f(x) = \sigma_f^{-1}\phi_{f_1}(x/\sigma_f)$, $I_\phi(f) = \sigma_f^{-2} I_\phi(f_1)$, and $J_\phi(f) = J_\phi(f_1)$. Moreover, $\int_{\mathbb{R}} \phi_f(x) f(x) dx = 0$ and $\int_{\mathbb{R}} x\phi_f(x) f(x) dx = 1$.

Denote by $\mathcal{F}_A$ the class of all densities functions satisfying **Assumption (A)**.

Assumption (B)

(B.1) $M_1^{(n)} = \frac{1}{N}\sum_{i=1}^{n}\sum_{t=1}^{T}(X_{it}^-)^2$, $M_2^{(n)} = \frac{1}{N}\sum_{i=1}^{n}\sum_{t=1}^{T}(X_{it}^+)^2$;

(B.2) $K_1^{(n)} = (M_1^{(n)})^{-1/2}$, $K_2^{(n)} = (M_2^{(n)})^{-1/2}$;

(B.3) $M_i^{(n)} \xrightarrow[n\to\infty]{} M_i$, $K_i^{(n)} \xrightarrow[n\to\infty]{} K_i = M_i^{-1/2}, \quad i = 1, 2$;

(B.4) the classical [26] conditions hold:

$$\lim_{n\to\infty} \frac{\max_{1\le i\le n}(X_{it}^-)^2}{\sum_{i=1}^{n}\sum_{t=1}^{T}(X_{it}^-)^2} = 0, \lim_{n\to\infty} \frac{\max_{1\le i\le n}(X_{it}^+)^2}{\sum_{i=1}^{n}\sum_{t=1}^{T}(X_{it}^+)^2} = 0, \quad t = 1, \ldots, T.$$

2.2. Local asymptotic normality

In this subsection, we will derive the local asymptotic normality property for the model (3), with respect to the scale parameter σ_f^2 and the vector of regression parameters of interest $(\beta_1, \beta_2)'$, for a fixed density $f_1 \in \mathcal{F}_A$.

To do this, let $\tau^{(n)} := (\tau_1^{(n)}, \tau_2^{(n)}, \tau_3^{(n)})'$, where $\tau_1^{(n)}, \tau_2^{(n)}$, and $\tau_3^{(n)}$ are three real sequences such that $\tau^{(n)'}\tau^{(n)}$ is uniformly bounded as $n \to \infty$. Let $\theta := (\sigma_f^2, \beta, \beta)'$. We also consider $\theta + n^{-1/2}\gamma^{(n)}\tau^{(n)}$ the sequences of local alternatives characterized by small perturbations, where

$$\gamma^{(n)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & K_1^{(n)} & 0 \\ 0 & 0 & K_2^{(n)} \end{pmatrix}.$$

The test is equivalent to

$$\mathbb{P}_{\theta;f_1}^{(n)} : \tau_2^{(n)} = \tau_3^{(n)} \text{ against } \mathbb{P}_{\theta+n^{-1/2}\gamma^{(n)}\tau^{(n)};f_1}^{(n)} : \tau_2^{(n)} \neq \tau_3^{(n)}.$$

Denote by $\Lambda_{\theta+n^{-1/2}\gamma^{(n)}\tau^{(n)}/\theta;f}^{(n)}$ the logarithm of the likelihood ratio for $\mathbb{P}_{\theta+n^{-1/2}\gamma^{(n)}\tau^{(n)};f}^{(n)}$ against $\mathbb{P}_{\theta;f}^{(n)}$. Then,

$$\Lambda_{\theta+n^{-1/2}\gamma^{(n)}\tau^{(n)}/\theta;f}^{(n)} := \log\left(\frac{L_{\theta+n^{-1/2}\gamma^{(n)}\tau^{(n)};f}}{L_{\theta;f}}\right),$$

where $L_{\theta+n^{-1/2}\gamma^{(n)}\tau^{(n)};f} = \prod_{i=1}^{n}\prod_{t=1}^{T} f(Y_{it} - \beta_1 X_{it}^- - \beta_2 X_{it}^+)$ is the likelihood function under the alternative hypothesis, and $L_{\theta;f} = \prod_{i=1}^{n}\prod_{t=1}^{T} f(Y_{it} - \beta X_{it})$ is the likelihood function under the null hypothesis.

For $i = 1, \ldots, n$ and $t = 1, \ldots, T$, the standardized residual is defined as follows:

$$Z_{it} = Z_{it}(\theta) := \sigma_f^{-1}(Y_{it} - \beta X_{it}).$$

It is important to note that, under the null hypothesis, it is equivalent to $\varepsilon_{it}/\sigma_f$. The following proposition establishes the local asymptotic normality result for a fixed density f_1 with respect to σ_f^2 and the vector of regression parameters of interest $(\beta_1, \beta_2)'$.

Proposition 1

Let Assumption (B) holds. Fix $f_1 \in \mathcal{F}_A$. Then the family $\mathcal{P}_{f_1}^{(n)}$ is LAN (for $n \to \infty$ with T fixed) at any $\theta =$

$(\sigma_f^2, \beta, \beta)'$, with central sequence

$$\Delta_{f_1}^{(n)}(\theta) := \begin{pmatrix} \Delta_{f_1;1}^{(n)}(\theta) \\ \Delta_{f_1;2}^{(n)}(\theta) \\ \Delta_{f_1;3}^{(n)}(\theta) \end{pmatrix} := \frac{n^{-1/2}}{\sigma_f} \begin{pmatrix} \frac{1}{2\sigma_f}\sum_{i=1}^{n}\sum_{t=1}^{T}(\phi_{f_1}(Z_{it})Z_{it}-1) \\ \sum_{i=1}^{n}\sum_{t=1}^{T}\phi_{f_1}(Z_{it})X_{it}^{-}K_1^{(n)} \\ \sum_{i=1}^{n}\sum_{t=1}^{T}\phi_{f_1}(Z_{it})X_{it}^{+}K_2^{(n)} \end{pmatrix}, \tag{4}$$

and information matrix

$$\Gamma_{f_1}^{(n)}(\theta) := \left(\Gamma_{f_1;pq}^{(n)}(\theta)\right)_{1\le p,q\le 3} := \frac{T}{\sigma_f^2}\begin{pmatrix} \frac{1}{4\sigma_f^2}(J_\phi(f_1)-1) & 0 & 0 \\ 0 & I_\phi(f_1) & I_\phi(f_1)C_X^{(n)} \\ 0 & I_\phi(f_1)C_X^{(n)} & I_\phi(f_1) \end{pmatrix}, \tag{5}$$

where $C_X^{(n)} = K_1^{(n)}K_2^{(n)}\overline{X^-X^+}$ and $\overline{X^-X^+} = \frac{1}{N}\sum_{i=1}^{n}\sum_{t=1}^{T}X_{it}^{-}X_{it}^{+}$.

More precisely, for any $\theta^{(n)} := ((\sigma_f{}^{(n)})^2, \beta^{(n)}, \beta^{(n)})'$ such that $n^{1/2}((\sigma_f{}^{(n)})^2-\sigma_f^2)$, $n^{1/2}(K_1^{(n)})^{-1}(\beta^{(n)}-\beta)$, and $n^{1/2}(K_2^{(n)})^{-1}\left(\beta^{(n)}-\beta\right)$ are $O(1)$, and for any bounded sequence $\tau^{(n)} \in \mathbb{R}^3$, we have, under $\mathbb{P}_{\theta^{(n)};f_1}^{(n)}$, as $n\to\infty$,

$$\Lambda_{\theta^{(n)}+n^{-1/2}\gamma^{(n)}\tau^{(n)}/\theta^{(n)};f_1}^{(n)} := \log\left(\frac{d\mathbb{P}_{\theta^{(n)}+n^{-1/2}\gamma^{(n)}\tau^{(n)};f_1}^{(n)}}{d\mathbb{P}_{\theta^{(n)};f_1}^{(n)}}\right) = \tau^{(n)\prime}\Delta_{f_1}^{(n)}(\theta^{(n)}) - \frac{1}{2}\tau^{(n)\prime}\Gamma_{f_1}^{(n)}(\theta)\tau^{(n)} + o_p(1),$$

and

$$(\Gamma_{f_1}^{(n)}(\theta))^{-1/2}\Delta_{f_1}^{(n)}(\theta^{(n)}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I_{3\times 3}).$$

Proof of Proposition 1

The proof consists of checking that the six conditions (Conditions 1.2 to 1.7) in Lemma 1 of [33] are satisfied, uniformly in the vicinity of $(\sigma_f^2, \beta, \beta)$. This is straightforward for all but one of them, Condition 1.2, on which we concentrate here. That condition actually follows (see Lemma 2 Swensen) if we manage to establish the quadratic mean differentiability, in the neighborhood of any $(\sigma_f^2, \beta, \beta)$, of

$$\left(\sigma_f^2, \beta_1, \beta_2\right) \mapsto q^{\frac{1}{2}}_{\sigma_f^2,\beta_1,\beta_2;f_1}(Y) = \left[\frac{1}{\sigma_f}f_1\left(\frac{1}{\sigma_f}\left(Y-\beta_1X^- - \beta_2X^+\right)\right)\right]^{\frac{1}{2}},$$

with X^- and $X^+\in\mathbb{R}$. This last is established using the following Lemma.

Lemma 1

Let Assumption (B) holds and fix $f_1 \in \mathcal{F}_A$. Define, for $Y\in\mathbb{R}$,

$$D_{\sigma_f^2}q^{\frac{1}{2}}_{\sigma_f^2,\beta,\beta;f_1}(Y) = \frac{1}{4\sigma_f^2}q^{\frac{1}{2}}_{\sigma_f^2,\beta,\beta;f_1}(Y)\left(\left(\frac{Y-\beta X}{\sigma_f}\right)\phi_{f_1}\left(\frac{Y-\beta X}{\sigma_f}\right)-1\right),$$

$$D_{\beta_1}q^{\frac{1}{2}}_{\sigma_f^2,\beta_1,\beta_2;f_1}(Y)|_{\beta_1=\beta_2=\beta} = \frac{1}{2\sigma_f}q^{\frac{1}{2}}_{\sigma_f^2,\beta,\beta;f_1}(Y)\phi_{f_1}\left(\frac{Y-\beta X}{\sigma_f}\right)X^-,$$

$$D_{\beta_2}q^{\frac{1}{2}}_{\sigma_f^2,\beta_1,\beta_2;f_1}(Y)|_{\beta_1=\beta_2=\beta} = \frac{1}{2\sigma_f}q^{\frac{1}{2}}_{\sigma_f^2,\beta,\beta;f_1}(Y)\phi_{f_1}\left(\frac{Y-\beta X}{\sigma_f}\right)X^+.$$

Then, as $s, r,$ and $v\to 0$,

1. $\int_{\mathbb{R}} \left[q^{\frac{1}{2}}_{\sigma_f^2+s,\beta+r,\beta;f_1}(Y) - q^{\frac{1}{2}}_{\sigma_f^2+s,\beta,\beta;f_1}(Y) - rD_{\beta_1} q^{\frac{1}{2}}_{\sigma_f^2+s,\beta_1,\beta_2;f_1}(Y)|_{\beta_1=\beta_2=\beta} \right]^2 dY = o\left(r^2\right),$
2. $\int_{\mathbb{R}} \left[q^{\frac{1}{2}}_{\sigma_f^2+s,\beta,\beta+v;f_1}(Y) - q^{\frac{1}{2}}_{\sigma_f^2+s,\beta,\beta;f_1}(Y) - vD_{\beta_2} q^{\frac{1}{2}}_{\sigma_f^2+s,\beta_1,\beta_2;f_1}(Y)|_{\beta_1=\beta_2=\beta} \right]^2 dY = o\left(v^2\right),$
3. $\int_{\mathbb{R}} \left[q^{\frac{1}{2}}_{\sigma_f^2+s,\beta,\beta;f_1}(Y) - q^{\frac{1}{2}}_{\sigma_f^2,\beta,\beta;f_1}(Y) - sD_{\sigma_f^2} q^{\frac{1}{2}}_{\sigma_f^2,\beta,\beta;f_1}(Y) \right]^2 dY = o\left(s^2\right),$
4. $\int_{\mathbb{R}} \left[q^{\frac{1}{2}}_{\sigma_f^2+s,\beta+r,\beta+v;f_1}(Y) - q^{\frac{1}{2}}_{\sigma_f^2,\beta,\beta;f_1}(Y) - \begin{pmatrix} s \\ r \\ v \end{pmatrix}' \begin{pmatrix} D_{\sigma_f^2} q^{\frac{1}{2}}_{\sigma_f^2,\beta,\beta;f_1}(Y) \\ D_{\beta_1} q^{\frac{1}{2}}_{\sigma_f^2,\beta_1,\beta_2;f_1}(Y)|_{\beta_1=\beta_2=\beta} \\ D_{\beta_2} q^{\frac{1}{2}}_{\sigma_f^2,\beta_1,\beta_2;f_1}(Y)|_{\beta_1=\beta_2=\beta} \end{pmatrix} \right]^2 dY =$
$o\left(\left\| \begin{pmatrix} s \\ r \\ v \end{pmatrix} \right\|^2 \right).$

Proof of Lemma 1

1. Let $\zeta := Y - \beta X$. Then the part (1) takes the form

$$\int_{\mathbb{R}} \left[\left(\frac{1}{\sqrt{\sigma_f^2+s}} \right)^{\frac{1}{2}} f_1^{\frac{1}{2}} \left(\frac{\zeta - rX^-}{\sqrt{\sigma_f^2+s}} \right) - \left(\frac{1}{\sqrt{\sigma_f^2+s}} \right)^{\frac{1}{2}} f_1^{\frac{1}{2}} \left(\frac{\zeta}{\sqrt{\sigma_f^2+s}} \right) \right.$$
$$\left. - r \frac{1}{2\sqrt{\sigma_f^2+s}} q^{\frac{1}{2}}_{\sigma_f^2+s,\beta,\beta;f_1}(\zeta+\beta X) \phi_{f_1} \left(\frac{\zeta}{\sqrt{\sigma_f^2+s}} \right) X^- \right]^2 d\zeta = o\left(r^2\right),$$

is equivalent to

$$\int_{\mathbb{R}} \left[f^{\frac{1}{2}}(\zeta - rX^-) - f^{\frac{1}{2}}(\zeta) - \frac{r}{2} f^{\frac{1}{2}}(\zeta)\phi_f(\zeta) X^- \right]^2 d\zeta = o\left(r^2\right),$$

which is equivalent to

$$\int_{\mathbb{R}} r^2 \left[\frac{f^{\frac{1}{2}}(\zeta - rX^-) - f^{\frac{1}{2}}(\zeta)}{r} + \frac{1}{2} \frac{f'(\zeta)}{f^{\frac{1}{2}}(\zeta)} X^- \right]^2 d\zeta = o\left(r^2\right),$$

hence, for proving that, it is sufficient to prove that

$$\lim_{r \to 0} \int_{\mathbb{R}} \left[\frac{f^{\frac{1}{2}}(\zeta - rX^-) - f^{\frac{1}{2}}(\zeta)}{r} + \frac{1}{2} \frac{f'(\zeta)}{f^{\frac{1}{2}}(\zeta)} X^- \right]^2 d\zeta = 0.$$

We have

$$\begin{aligned} \lim_{r \to 0} \frac{f^{\frac{1}{2}}(\zeta - rX^-) - f^{\frac{1}{2}}(\zeta)}{r} &= \lim_{r \to 0} \frac{f^{\frac{1}{2}}(\zeta - rX^-) - f^{\frac{1}{2}}(\zeta)}{-rX^-} \times \frac{-rX^-}{r} \\ &= \left(f^{\frac{1}{2}}(\zeta) \right)' \times \left(-X^- \right) \\ &= -\frac{1}{2} \frac{f'(\zeta)}{f^{\frac{1}{2}}(\zeta)} X^-, \end{aligned}$$

and

$$\begin{aligned}\int_{\zeta=-\infty}^{+\infty}\left[\frac{f^{\frac{1}{2}}(\zeta-rX^{-})-f^{\frac{1}{2}}(\zeta)}{r}\right]^2 d\zeta &= \int_{\zeta=-\infty}^{+\infty}\frac{1}{r^2}\left[\int_{t=\zeta}^{\zeta-rX^{-}}\frac{1}{2}f'(t)f^{\frac{-1}{2}}(t)dt\right]^2 d\zeta\\ &\leq \frac{-rX^{-}}{r^2}\int_{\zeta=-\infty}^{+\infty}\int_{t=\zeta}^{\zeta-rX^{-}}\left[\frac{1}{2}f'(t)f^{\frac{-1}{2}}(t)\right]^2 dt d\zeta\\ &\leq \frac{-rX^{-}}{r^2}\int_{t=-\infty}^{+\infty}\int_{\zeta=t+rX^{-}}^{t}\left[\frac{1}{2}f'(t)f^{\frac{-1}{2}}(t)\right]^2 d\zeta dt\\ &\leq \left[\frac{-rX^{-}}{r}\right]^2\int_{t=-\infty}^{+\infty}\left[\frac{1}{2}f'(t)f^{\frac{-1}{2}}(t)\right]^2 dt\\ &\leq (-X^{-})^2\int_{t=-\infty}^{+\infty}\left[\frac{1}{2}f'(t)f^{\frac{-1}{2}}(t)\right]^2 dt\\ &\leq \int_{\mathbb{R}}\left[\frac{-1}{2}f'(t)f^{\frac{-1}{2}}(t)X^{-}\right]^2 dt.\end{aligned}$$

This completes the proof of part (1) of Lemma 1.

2. The proof follows similarly to part (1).
3. The problem here reduces to the classical case of linear models considered by [33].
4. The result here follows from (1), (2), and (3) above. This completes the proof of Lemma 1.

□

□

Based on convergence from $\overline{X^{-}X^{+}}$ to $\mu_{X^{-}X^{+}}$ as $n\to\infty$ and (B.3)-subsequences, the information matrix $\Gamma_{f_1}^{(n)}(\theta)$ converges to

$$\Gamma_{f_1}(\theta) := \frac{T}{\sigma_f^2}\begin{pmatrix}\frac{1}{4\sigma_f^2}(J_\phi(f_1)-1) & 0 & 0\\ 0 & I_\phi(f_1) & I_\phi(f_1)\mu_{C_X}\\ 0 & I_\phi(f_1)\mu_{C_X} & I_\phi(f_1)\end{pmatrix}, \tag{6}$$

where $\mu_{C_X} = K_1K_2\mu_{X^{-}X^{+}}$.

In the case of a Gaussian distribution (where $f_1 = f_{\mathcal{N}}$; standardized normal density $\mathcal{N}(0,1)$), this is considered an exceptional case. However, $\phi_{f_1}(x)$, $J_\phi(f_1)$, and $I_\phi(f_1)$ are reduced to x, 3, and 1, respectively. It is easy to confirm that equations (4), (5), and (6) also simplify to

$$\Delta_{\mathcal{N}}^{(n)}(\theta) := \frac{n^{-1/2}}{\sigma}\begin{pmatrix}\frac{1}{2\sigma}\sum_{i=1}^{n}\sum_{t=1}^{T}(Z_{it}^2-1)\\ \sum_{i=1}^{n}\sum_{t=1}^{T}Z_{it}X_{it}^{-}K_1^{(n)}\\ \sum_{i=1}^{n}\sum_{t=1}^{T}Z_{it}X_{it}^{+}K_2^{(n)}\end{pmatrix},$$

$$\Gamma_{\mathcal{N}}^{(n)}(\theta) := \frac{T}{\sigma^2}\begin{pmatrix}\frac{1}{2\sigma^2} & 0 & 0\\ 0 & 1 & C_X^{(n)}\\ 0 & C_X^{(n)} & 1\end{pmatrix},$$

and

$$\Gamma_{\mathcal{N}}(\theta) := \frac{T}{\sigma^2}\begin{pmatrix}\frac{1}{2\sigma^2} & 0 & 0\\ 0 & 1 & \mu_{C_X}\\ 0 & \mu_{C_X} & 1\end{pmatrix}.$$

The result of Proposition 1 allows us to construct parametric tests that are asymptotically optimal under a specified f_1. It is important to note that these tests are only valid under the specified f_1. Afterward, we will propose tests such as Gaussian and adaptive tests, which are valid under general densities.

3. Locally asymptotically optimal tests

We are interested in testing the null hypothesis $\beta_1 = \beta_2 = \beta$ in model (3). This model includes an unspecified error density $f_1 \in \mathcal{F}_A$, which is a semiparametric hypothesis. The null hypothesis can be formally expressed as:

$$\mathcal{H}_0^{(n)} := \bigcup_{g_1 \in \mathcal{F}_A} \mathcal{H}_0^{(n)}(g_1) := \bigcup_{g_1 \in \mathcal{F}_A} \bigcup_{\sigma_g^2 > 0} \bigcup_{\beta \in \mathbb{R}} \left\{ \mathbb{P}^{(n)}_{\sigma_g^2, \beta; g_1} \right\}.$$

Parametric alternatives takes the form (for fixed density $f_1 \in \mathcal{F}_A$)

$$\mathcal{H}_1^{(n)}(f_1) := \bigcup_{\sigma_f^2 > 0} \bigcup_{\beta_1 \in \mathbb{R}} \bigcup_{\beta_2 \in \mathbb{R}} \left\{ \mathbb{P}^{(n)}_{\sigma_f^2, \beta_1, \beta_2; f_1} \right\}.$$

The parameter σ_f^2 is a nuisance parameter, while $(\beta_1, \beta_2)'$ is the vector of regression parameters of interest. Before addressing the semiparametric hypothesis $\mathcal{H}_0^{(n)}$ (unspecified density), let's first examine the parametric problem of testing $\mathcal{H}_0^{(n)}(f_1)$ (with f_1 specified) against $\mathcal{H}_1^{(n)}(f_1)$.

3.1. *Optimal parametric tests*

As mentioned earlier, the Le Cam theory of LAN experiments permits the construction of tests that are locally asymptotically most stringent. The fundamental concept is the weak convergence of the sequence of local experiments to the Gaussian shift model. For a comprehensive discussion on locally asymptotically optimal testing in LAN families, please refer to [37, 21, 36].

In this subsection, we construct locally asymptotically optimal tests (i.e., the most stringent tests) in the presence of a nuisance parameter for testing the null hypothesis $\beta_1 = \beta_2 = \beta$ in model (3). We assume that the innovation density f_1 specified. The main consequence of the LAN results is that, for each $\theta = (\sigma_f^2, \beta, \beta)'$, and for given $f_1 \in \mathcal{F}_A$, the sequences of local experiments

$$\xi_{f_1}^{(n)}(\theta) := \left\{ \mathbb{P}^{(n)}_{\theta + n^{-1/2}\gamma^{(n)}\tau; f_1} \mid \tau \in \mathbb{R}^3 \right\},$$

converge weakly to the Gaussian shift experiments ($\Gamma_{f_1}(\theta)$ given in (6))

$$\xi_{f_1}(\theta) := \left\{ \mathcal{N}\left(\Gamma_{f_1}(\theta)\tau, \Gamma_{f_1}(\theta)\right) \mid \tau \in \mathbb{R}^3 \right\},$$

while the central sequence $\Delta_{f_1}^{(n)}(\theta)$ under $\mathbb{P}^{(n)}_{\theta + n^{-1/2}\gamma^{(n)}\tau; f_1}$ converges in distribution to the Gaussian vector $\Delta = (\Delta_1, \Delta_2, \Delta_3)' \sim \mathcal{N}\left(\Gamma_{f_1}(\theta)\tau, \Gamma_{f_1}(\theta)\right)$. The classical theory of hypothesis testing in Gaussian shifts (see Section 11.9 in [21]) provides the general form for locally asymptotically most stringent tests of hypotheses in LAN models. In this case, the null hypothesis $\mathcal{H}_0^{(n)}(f_1) := \bigcup_{\sigma_f^2 > 0} \bigcup_{\beta \in \mathbb{R}} \left\{ \mathbb{P}^{(n)}_{\sigma_f^2, \beta; f_1} \right\}$ and the local alternative $\mathcal{H}_1^{(n)}(f_1)$ can be expressed as:

$$\mathcal{H}_0^{(n)}(f_1) : \tau \in \mathcal{M}(\Omega) \ \text{ against } \ \mathcal{H}_1^{(n)}(f_1) : \tau \notin \mathcal{M}(\Omega),$$

where $\mathcal{M}(\Omega)$ is the linear subspace of dimension 2 of $\mathbb{R}^3$ generated by the matrix $\Omega' := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}$. Such tests, should be based on

$$Q_{f_1}^{(n)}(\theta) := \Delta_{f_1}^{(n)\prime}(\theta) \left[\Gamma_{f_1}^{(n)^{-1}}(\theta) - \Omega(\Omega' \Gamma_{f_1}^{(n)}(\theta) \Omega)^{-1} \Omega' \right] \Delta_{f_1}^{(n)}(\theta).$$

Through mathematical calculations, the statistical of the test can be expressed as follows:

$$\begin{aligned} Q_{f_1}^{(n)}(\theta) &= \frac{1}{2\left(\Gamma_{f_1;22}^{(n)}(\theta) - \Gamma_{f_1;23}^{(n)}(\theta)\right)} \left(\Delta_{f_1;2}^{(n)}(\theta) - \Delta_{f_1;3}^{(n)}(\theta)\right)^2, \\ &= \frac{\sigma_f^2}{2TI_\phi(f_1)(1 - C_X^{(n)})} \left(\Delta_{f_1;2}^{(n)}(\theta) - \Delta_{f_1;3}^{(n)}(\theta)\right)^2, \\ &= \frac{1}{2NI_\phi(f_1)(1 - C_X^{(n)})} \left[\sum_{i=1}^{n} \sum_{t=1}^{T} \phi_{f_1}(Z_{it}) \left(X_{it}^- K_1^{(n)} - X_{it}^+ K_2^{(n)}\right)\right]^2. \end{aligned} \tag{7}$$

Remark 1
The statistic $Q_{f_1}^{(n)}(\theta)$ can be written as the quadratic form

$$\begin{aligned} Q_{f_1}^{(n)}(\theta) &= \Delta_{f_1}^{(n)\prime}(\theta)\Gamma_{f_1}^{(n)^{-1/2}}(\theta)\left[I_{3\times 3} - \Gamma_{f_1}^{(n)^{1/2}}(\theta)\Omega(\Omega'\Gamma_{f_1}^{(n)}(\theta)\Omega)^{-1}\Omega'\Gamma_{f_1}^{(n)^{1/2}}(\theta)\right]\Gamma_{f_1}^{(n)^{-1/2}}(\theta)\Delta_{f_1}^{(n)}(\theta), \\ &= \left\|A(\theta)\,\Gamma_{f_1}^{(n)^{-1/2}}(\theta)\Delta_{f_1}^{(n)}(\theta)\right\|^2, \end{aligned}$$

where $A(\theta) = I_{3\times 3} - \Gamma_{f_1}^{(n)^{1/2}}(\theta)\Omega(\Omega'\Gamma_{f_1}^{(n)}(\theta)\Omega)^{-1}\Omega'\Gamma_{f_1}^{(n)^{1/2}}(\theta)$ is an idempotent symmetric matrix.

The proposed test mentioned above is still unsatisfactory because it involves the parameter θ, which is unspecified under the null hypothesis $\mathcal{H}_0^{(n)}(f_1)$. To address this issue, we introduce an estimate $\hat{\theta} := \hat{\theta}^{(n)} := (\hat{\sigma}_n^2, \hat{\beta}^{(n)}, \hat{\beta}^{(n)})'$ of θ satisfying the following assumptions:
Assumption (C) The estimate $\hat{\theta}^{(n)}$ is such that

(C.1) $\sqrt{n}$-consistent, i.e., for all $f_1 \in \mathcal{F}_A$ and all $\varepsilon > 0$, there exist $c := c(f_1, \theta, \varepsilon)$ and $N := N(f_1, \theta, \varepsilon)$ such that under $\mathbb{P}_{\sigma_f^2,\beta;f_1}^{(n)}$, we have

$$P\left(\sqrt{n}\left\|\hat{\theta}^{(n)} - \theta\right\| > c\right) < \varepsilon \quad \forall n \geq N.$$

(C.2) Locally asymptotically discrete, i.e., the number of possible values of $\hat{\theta}^{(n)}$ in intervals of the form $(\theta - n^{-1/2}\xi, \theta + n^{-1/2}\xi)$, $\xi > 0$, is eventually bounded, as $n \to \infty$, under $\mathbb{P}_{\sigma_f^2,\beta;f_1}^{(n)}$.

Note that several estimates, such as the maximum likelihood estimates, the Yule-Walker estimates, the M-estimates, and the least square estimates, satisfy the condition (C.1) on the rate of convergence in probability. Part (C.2) has little practical implications.

The proposition below demonstrates that replacing $\hat{\theta}^{(n)}$ with θ does not affect the asymptotic behaviour of the test statistic (7).

Proposition 2
Suppose that Assumptions (A) and (B) hold, denote by $\hat{\theta}^{(n)}$ a deterministic sequence satisfying Assumptions (C). Then, under $\mathbb{P}_{\sigma_f^2,\beta;f_1}^{(n)}$, as $n \to \infty$, we have

(a)

$$\Delta_{f_1}^{(n)}(\hat{\theta}^{(n)}) - \Delta_{f_1}^{(n)}(\theta) = -n^{1/2}\Gamma_{f_1}^{(n)}(\theta)\gamma^{(n)^{-1}}(\hat{\theta}^{(n)} - \theta) + o_P(1).$$

(b)

$$\left(\Delta_{f_1;2}^{(n)}(\hat{\theta}^{(n)}) - \Delta_{f_1;3}^{(n)}(\hat{\theta}^{(n)})\right) = \left(\Delta_{f_1;2}^{(n)}(\theta) - \Delta_{f_1;3}^{(n)}(\theta)\right) + o_P(1).$$

Proof

(a) Asymptotic local linearity implies:

$$\Delta^{(n)}_{f_1}(\theta + n^{-1/2}\gamma^{(n)}\tau^{(n)}) - \Delta^{(n)}_{f_1}(\theta) = -\Gamma^{(n)}_{f_1}(\theta)\tau^{(n)} + o_P(1).$$

If we choose $\tau^{(n)} = n^{1/2}{\gamma^{(n)}}^{-1}(\hat{\theta}^{(n)} - \theta)$, we will have (a).

(b) Letting $1_2 = (1,1)'$. From (a), under $\mathbb{P}^{(n)}_{\sigma^2_f,\beta;f_1}$, as $n \to \infty$,

$$\begin{aligned}\Delta^{(n)}_{f_1;2,3}(\hat{\theta}^{(n)}) - \Delta^{(n)}_{f_1;2,3}(\theta) &= -n^{1/2}\Gamma^{(n)}_{f_1;2,3}(\theta){\gamma^{(n)}_{2,3}}^{-1}1_2(\hat{\beta}^{(n)} - \beta) + o_P(1) \\ &= -\frac{n^{1/2}T}{\sigma^2_f} I_\phi(f_1) \begin{pmatrix} 1 & C^{(n)}_X \\ C^{(n)}_X & 1 \end{pmatrix} \begin{pmatrix} {K^{(n)}}^{-1} & 0 \\ 0 & {K^{(n)}}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\beta}^{(n)} - \beta \\ \hat{\beta}^{(n)} - \beta \end{pmatrix} \\ &\quad + o_P(1),\end{aligned}$$

then

$$\left(\Delta^{(n)}_{f_1;2}(\hat{\theta}^{(n)}) - \Delta^{(n)}_{f_1;2}(\theta)\right) - \left(\Delta^{(n)}_{f_1;3}(\hat{\theta}^{(n)}) - \Delta^{(n)}_{f_1;3}(\theta)\right) = 0 + o_P(1),$$

$$\left(\Delta^{(n)}_{f_1;2}(\hat{\theta}^{(n)}) - \Delta^{(n)}_{f_1;3}(\hat{\theta}^{(n)})\right) = \left(\Delta^{(n)}_{f_1;2}(\theta) - \Delta^{(n)}_{f_1;3}(\theta)\right) + o_P(1).$$

□

The test based on (7) is locally asymptotically most stringent. To be more precise, applying Le Cam's Third Lemma yields the following result.

Proposition 3

Suppose that Assumptions (A), (B), and (C) hold. Then,

(i) $Q^{(n)}_{f_1}(\hat{\theta}^{(n)}) = Q^{(n)}_{f_1}(\theta) + o_P(1)$ is asymptotically chi-square, with 1 degrees of freedom under $\mathbb{P}^{(n)}_{\theta;f_1}$, and asymptotically noncentral chi-square, still with 1 degrees of freedom but with noncentrality parameter $\rho_{f_1} := \frac{T I_\phi(f_1)(1-\mu_{C_X})}{2\sigma^2_f}(\tau_2 - \tau_3)^2$ under $\mathbb{P}^{(n)}_{\theta+n^{-1/2}\gamma^{(n)}\tau;f_1}$;

(ii) the sequence of tests rejecting the null hypothesis $\mathcal{H}^{(n)}_0(f_1)$ whenever $Q^{(n)}_{f_1}(\hat{\theta}^{(n)})$ exceeds the $(1-\alpha)$-quantile of a chi-square distribution with one degree of freedom, i.e. $Q^{(n)}_{f_1}(\hat{\theta}^{(n)}) > \chi^2_{1,1-\alpha}$, is locally asymptotically optimal (most stringent), at asymptotic level α, for $\mathcal{H}^{(n)}_0(f_1)$ against

$$\bigcup_{\sigma^2_f>0}\bigcup_{\beta_1\in\mathbb{R}}\bigcup_{\beta_2\in\mathbb{R}}\left\{\mathbb{P}^{(n)}_{\sigma^2_f,\beta_1,\beta_2;f_1}\right\};$$

(iii) the sequence of tests has asymptotic power $1 - \mathcal{F}\left(\chi^2_{1,1-\alpha};\rho_{f_1}\right)$, at $\mathbb{P}^{(n)}_{\theta+n^{-1/2}\gamma^{(n)}\tau;f_1}$, where $\mathcal{F}(.,\rho_{f_1})$ denotes the noncentral chi-square distribution function with one degree of freedom and noncentrality parameter ρ_{f_1}.

Proof

(i) Follows from:

- Proposition 2(b) implies that, under $\mathbb{P}^{(n)}_{\theta;f_1}$, as $n \to \infty$,

$$\begin{aligned}
&\left(\Delta^{(n)}_{f_1;2}(\hat{\theta}^{(n)}) - \Delta^{(n)}_{f_1;3}(\hat{\theta}^{(n)})\right)^2 - \left(\Delta^{(n)}_{f_1;2}(\theta) - \Delta^{(n)}_{f_1;3}(\theta)\right)^2 \\
&= \left[\left(\Delta^{(n)}_{f_1;2}(\hat{\theta}^{(n)}) - \Delta^{(n)}_{f_1;3}(\hat{\theta}^{(n)})\right) - \left(\Delta^{(n)}_{f_1;2}(\theta) - \Delta^{(n)}_{f_1;3}(\theta)\right)\right] \\
&\quad \times \left[\left(\Delta^{(n)}_{f_1;2}(\hat{\theta}^{(n)}) - \Delta^{(n)}_{f_1;3}(\hat{\theta}^{(n)})\right) + \left(\Delta^{(n)}_{f_1;2}(\theta) - \Delta^{(n)}_{f_1;3}(\theta)\right)\right] \\
&= o_P(1) \times \left[2\left(\Delta^{(n)}_{f_1;2}(\theta) - \Delta^{(n)}_{f_1;3}(\theta)\right) + o_P(1)\right] \\
&= o_P(1).
\end{aligned}$$

- Letting $\hat{\sigma}^2_n$ an estimator of σ^2_f is defined as: $\hat{\sigma}^2_n := \frac{1}{N}\sum_{i=1}^{n}\sum_{t=1}^{T}(\varepsilon_{it}(\hat{\theta}^{(n)}))^2$.

Hence,

$$\begin{aligned}
Q^{(n)}_{f_1}(\hat{\theta}^{(n)}) &= \frac{\hat{\sigma}^2_n}{\sigma^2_f} Q^{(n)}_{f_1}(\theta) + o_P(1), \\
&= Q^{(n)}_{f_1}(\theta) + o_P(1).
\end{aligned}$$

Under $\mathbb{P}^{(n)}_{\theta;f_1}$: $\left(\Delta^{(n)}_{f_1;2}(\theta) - \Delta^{(n)}_{f_1;3}(\theta)\right) \sim \mathcal{N}\left(0, \frac{2TI_\phi(f_1)(1-\mu_{C_X})}{\sigma^2_f}\right)$, then

$$\left[\frac{\sigma_f}{\sqrt{2TI_\phi(f_1)(1-C^{(n)}_X)}}\left(\Delta^{(n)}_{f_1;2}(\theta) - \Delta^{(n)}_{f_1;3}(\theta)\right)\right]^2 = Q^{(n)}_{f_1}(\theta) \sim \chi^2_1.$$

From Le Cam's Third Lemma, we have under $\mathbb{P}^{(n)}_{\theta+n^{-1/2}\gamma^{(n)}\tau;f_1}$:

$$\left(\Delta^{(n)}_{f_1;2}(\theta) - \Delta^{(n)}_{f_1;3}(\theta)\right) \sim \mathcal{N}\left(\frac{TI_\phi(f_1)(1-\mu_{C_X})}{\sigma^2_f}(\tau_2 - \tau_3), \frac{2TI_\phi(f_1)(1-\mu_{C_X})}{\sigma^2_f}\right).$$

So that

$$\frac{\sigma_f}{\sqrt{2TI_\phi(f_1)(1-\mu_{C_X})}}\left(\Delta^{(n)}_{f_1;2}(\theta) - \Delta^{(n)}_{f_1;3}(\theta)\right) \sim \mathcal{N}\left(\frac{\sqrt{TI_\phi(f_1)(1-\mu_{C_X})}(\tau_2-\tau_3)}{\sqrt{2}\sigma_f}, 1\right).$$

Cochran's Theorem leads to

$$\left[\frac{\sigma_f}{\sqrt{2TI_\phi(f_1)(1-C^{(n)}_X)}}\left(\Delta^{(n)}_{f_1;2}(\theta) - \Delta^{(n)}_{f_1;3}(\theta)\right)\right]^2 = Q^{(n)}_{f_1}(\theta) \sim \chi^2_1(\rho_{f_1}),$$

with $\rho_{f_1} = \left(\frac{\sqrt{TI_\phi(f_1)(1-\mu_{C_X})}(\tau_2-\tau_3)}{\sqrt{2}\sigma_f}\right)^2 = \frac{TI_\phi(f_1)(1-\mu_{C_X})}{2\sigma^2_f}(\tau_2-\tau_3)^2$, which gives the desired result under $\mathbb{P}^{(n)}_{\theta+n^{-1/2}\gamma^{(n)}\tau;f_1}$.

(ii) Stringency is a consequence of the weak convergence of local experiments to Gaussian shifts, see [21].

(iii) We know that the power of the test is defined by

$$\begin{aligned}
\text{Prob}\left[\text{rejecting } \mathcal{H}^{(n)}_f(\theta) \;/\; \mathcal{H}^{(n)}_f(\theta + n^{-1/2}\gamma^{(n)}\tau)\right] &= \text{Prob}\left[Q^{(n)}_{f_1}(\theta) > \chi^2_{1,1-\alpha} \;/\; \tau_2 \neq \tau_3\right] \\
&= 1 - \mathcal{F}\left(\chi^2_{1,1-\alpha}; \rho_{f_1}\right).
\end{aligned}$$

□

3.2. Gaussian test

In this subsection, we will construct the Gaussian test $Q_{\mathcal{N}}^{(n)}(\theta)$. The Gaussian central sequences $\Delta_{\mathcal{N};2}^{(n)}(\theta)$ and $\Delta_{\mathcal{N};3}^{(n)}(\theta)$ will allow us to obtain asymptotically optimal tests under $f_1 = f_{\mathcal{N}(0,1)}$. These tests will also efficiently detect panel threshold regression in the parametric Gaussian model, which is characterized by Gaussian disturbances. It is highly desirable to extend the validity of the Gaussian optimal test to general densities g_1 in a broad class of densities.
Define

$$\Delta_{\mathcal{N};2,3}^{(n)}(\theta) := \begin{pmatrix} \Delta_{\mathcal{N};2}^{(n)}(\theta) \\ \Delta_{\mathcal{N};3}^{(n)}(\theta) \end{pmatrix} := \frac{n^{-1/2}}{\sigma} \begin{pmatrix} \sum_{i=1}^{n}\sum_{t=1}^{T} Z_{it} X_{it}^{-} K_1^{(n)} \\ \sum_{i=1}^{n}\sum_{t=1}^{T} Z_{it} X_{it}^{+} K_2^{(n)} \end{pmatrix}.$$

Then, under $\mathbb{P}_{\theta;g_1}^{(n)}$, $\left(\Delta_{\mathcal{N};2}^{(n)}(\theta) - \Delta_{\mathcal{N};3}^{(n)}(\theta)\right)$ is asymptotically normal with zero mean and variance $\frac{2T(1-\mu_{C_X})}{\sigma^2}$.

However, it is clear that, still under $\mathbb{P}_{\theta+n^{-1/2}\gamma^{(n)}\tau;g_1}^{(n)}$, the central sequence $\Delta_{\mathcal{N};2,3}^{(n)}(\theta)$ and the log-likelihood $\Lambda_{\theta+n^{-1/2}\gamma^{(n)}\tau/\theta;g_1}^{(n)}$ are jointly binormal. Therefore, the desired result can be obtained by applying Le Cam's Third Lemma.

The Gaussian test may then be based on a statistic of the form

$$\begin{aligned} Q_{\mathcal{N}}^{(n)}(\theta) &= \frac{1}{2\left(\Gamma_{\mathcal{N};22}^{(n)}(\theta) - \Gamma_{\mathcal{N};23}^{(n)}(\theta)\right)} \left(\Delta_{\mathcal{N};2}^{(n)}(\theta) - \Delta_{\mathcal{N};3}^{(n)}(\theta)\right)^2, \\ &= \frac{1}{2N(1-C_X^{(n)})}\left[\sum_{i=1}^{n}\sum_{t=1}^{T} Z_{it}\left(X_{it}^{-}K_1^{(n)} - X_{it}^{+}K_2^{(n)}\right)\right]^2. \end{aligned} \quad (8)$$

The asymptotic linearity holds for $\Delta_{\mathcal{N};2,3}^{(n)}(\theta)$ not just under $\mathbb{P}_{\theta;f_{\mathcal{N}}}^{(n)}$, but under $\mathbb{P}_{\theta;g_1}^{(n)}$. Then, under $\mathbb{P}_{\theta;g_1}^{(n)}$, and for any bounded sequences $\tau^{(n)} = (\tau_1^{(n)}, \tau_2^{(n)}, \tau_2^{(n)})' \in \mathbb{R}^3$ and $1_2\tau_2^{(n)} := (\tau_2^{(n)}, \tau_2^{(n)})' \in \mathbb{R}^2$, as $n \to \infty$ with T fixed,

$$\Delta_{\mathcal{N};2,3}^{(n)}(\theta + n^{-1/2}\gamma^{(n)}\tau^{(n)}) - \Delta_{\mathcal{N};2,3}^{(n)}(\theta) = -\Gamma_{\mathcal{N},g_1;2,3}^{(n)}(\theta) 1_2 \tau_2^{(n)} + o_P(1), \quad (9)$$

with

$$\Gamma_{\mathcal{N},g_1;2,3}^{(n)}(\theta) = \frac{T}{\sigma\sigma_g}\begin{pmatrix} 1 & C_X^{(n)} \\ C_X^{(n)} & 1 \end{pmatrix}, \text{ such as } C_X^{(n)} = K_1^{(n)}K_2^{(n)}\overline{X^{-}X^{+}}.$$

The next result is immediate from (9). Let Assumption (B) holds, assume that $\hat{\theta}^{(n)}$ satisfies Assumptions (C) and fix $\theta \in \mathbb{R}_+^* \times \mathbb{R}^2$, we have

$$\left(\Delta_{\mathcal{N};2}^{(n)}(\hat{\theta}^{(n)}) - \Delta_{\mathcal{N};3}^{(n)}(\hat{\theta}^{(n)})\right) = \left(\Delta_{\mathcal{N},2}^{(n)}(\theta) - \Delta_{\mathcal{N},3}^{(n)}(\theta)\right) + o_P(1). \quad (10)$$

Proposition 4
Let Assumptions $(A), (B)$ and, (C) hold, for any $g_1 \in \mathcal{F}_A$. Then,

(i) $Q_{\mathcal{N}}^{(n)}(\hat{\theta}^{(n)}) = Q_{\mathcal{N}}^{(n)}(\theta) + o_p(1)$ is asymptotically chi-square, with 1 degrees of freedom under $\mathbb{P}_{\theta;g_1}^{(n)}$, and asymptotically noncentral chi-square, still with 1 degrees of freedom but with noncentrality parameter $\rho_{\mathcal{N}} := \frac{T(1-\mu_{C_X})}{2\sigma_g^2}(\tau_2 - \tau_3)^2$ under $\mathbb{P}_{\theta+n^{-1/2}\gamma^{(n)}\tau;g_1}^{(n)}$;

(ii) the sequence of tests rejecting the null hypothesis $\bigcup_{g_1\in\mathcal{F}_A}\bigcup_{\sigma_g^2>0}\bigcup_{\beta\in\mathbb{R}}\left\{\mathbb{P}_{\sigma_g^2,\beta;g_1}^{(n)}\right\}$ whenever $Q_{\mathcal{N}}^{(n)}(\hat{\theta}^{(n)}) > \chi_{1,1-\alpha}^2$, is locally asymptotically optimal (most stringent), at asymptotic level α, against alternatives of the form

$$\bigcup_{\sigma^2>0}\bigcup_{\beta_1\in\mathbb{R}}\bigcup_{\beta_2\in\mathbb{R}}\left\{\mathbb{P}_{\sigma^2,\beta_1,\beta_2;f_{\mathcal{N}}}^{(n)}\right\};$$

(iii) the sequence of tests has asymptotic power $1 - \mathcal{F}\left(\chi^2_{1,1-\alpha}; \rho_{\mathcal{N}}\right)$, at $\mathbb{P}^{(n)}_{\theta+n^{-1/2}\gamma^{(n)}\tau;g_1}$.

Proof
Let $\hat{\sigma}^2$ an estimator of σ^2, under $\mathbb{P}^{(n)}_{\theta;g_1}$, as $n \to \infty$,

$$\hat{\sigma}^2 = \sigma^2 + o_P(1). \tag{11}$$

Using (10) and (11), one verifies (i). (ii) and (iii) of Proposition 4 follow the same way from Proposition 3. □

3.3. Adaptive test

The approach considered so far has been a parametric one. The central sequences $\Delta^{(n)}_{f_1}(\theta)$ and the information matrices $\Gamma^{(n)}_{f_1}(\theta)$ are associated with a specified f_1. The test described in Proposition 3 has a strong parametric flavor, but its validity and optimality only hold at f_1. Specifying f or f_1 in practice is quite unrealistic. A more reasonable approach is a semiparametric one where f_1 remains completely unspecified. This is our motivation for considering the semiparametric model. Thus, the semiparametrically efficient test achieves asymptotically the parametric efficiency bound at any f_1 and performs asymptotically as well as the parametrically efficient test. In this case, we are referring to the adaptive test.

As $Q^{(n)}_{f_1}(\theta)$ depends on the unknown density f_1 through the score function ϕ_{f_1}, it is natural to consider an appropriate estimation of this function. Let (a_n) and (b_n) be two sequences of positive numbers converging to zero. Consider a kernel $k(.)$ that satisfies the Conditions $\mathbf{K}$ of [30]. Denote by $\hat{f}_{1;nT}$ and $\hat{f}'_{1;nT}$ two functions defined for $x, y_{11}, \ldots, y_{nT} \in \mathbb{R}$, by

$$\hat{f}_{1;nT}\left(x; y_{11}, \ldots, y_{nT}\right) := \frac{1}{Na_n} \sum_{i=1}^{n} \sum_{t=1}^{T} k\left(\frac{x - y_{it}}{a_n}\right),$$

$$\hat{f}'_{1;nT}\left(x; y_{11}, \ldots, y_{nT}\right) := \frac{1}{Na_n^2} \sum_{i=1}^{n} \sum_{t=1}^{T} k'\left(\frac{x - y_{it}}{a_n}\right).$$

Letting

$$\hat{\phi}_{nT}\left(Z_{it}\right) := -\frac{\hat{f}'_{1;nT}\left(Z_{it}; Z_{11}, \ldots, Z_{nT}\right)}{b_n + \hat{f}_{1;nT}\left(Z_{it}; Z_{11}, \ldots, Z_{nT}\right)}.$$

Considering the following estimates of σ^2_f and $I_\phi(f_1)$ respectively, $\hat{\sigma}^2_n := \frac{1}{N} \sum_{i=1}^{n} \sum_{t=1}^{T} (\varepsilon_{it}(\hat{\theta}^{(n)}))^2$ and $\hat{I}_{nT} := \frac{1}{N} \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{\phi}^2_{nT}(Z_{it})$.

Let

$$\hat{\Delta}^{(n)}_{2;3}(\theta) := \begin{pmatrix} \hat{\Delta}^{(n)}_2(\theta) \\ \hat{\Delta}^{(n)}_3(\theta) \end{pmatrix} := \frac{n^{-1/2}}{\hat{\sigma}_n} \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{\phi}_{nT}\left(Z_{it}\right) \begin{pmatrix} X^-_{it} K^{(n)}_1 \\ X^+_{it} K^{(n)}_2 \end{pmatrix},$$

and

$$\hat{\Gamma}^{(n)}_{2;3}(\theta) = \frac{T}{\hat{\sigma}^2_n} \hat{I}_{nT} \begin{pmatrix} 1 & C^{(n)}_X \\ C^{(n)}_X & 1 \end{pmatrix},$$

under $\mathbb{P}^{(n)}_{\theta;g_1}$, $\left(\hat{\Delta}^{(n)}_2(\theta) - \hat{\Delta}^{(n)}_3(\theta)\right)$ is asymptotically normal with zero mean and variance $\frac{2TI_\phi(g_1)(1-\mu_{C_X})}{\sigma^2_g}$.

Now, we can propose an adaptive test to test the null hypothesis (with unspecified innovation density) $\bigcup_{g_1 \in \mathcal{F}_A} \mathcal{H}^{(n)}_0(g_1) := \bigcup_{g_1 \in \mathcal{F}_A} \bigcup_{\sigma^2_g > 0} \bigcup_{\beta \in \mathbb{R}} \left\{\mathbb{P}^{(n)}_{\sigma^2_g,\beta;g_1}\right\}$ against the local alternative $\mathcal{H}^{(n)}_1(g_1)$. The adaptive test is written in the following form:

$$\hat{Q}^{(n)}(\theta) = \frac{1}{2N\hat{I}_{nT}(1-C_X^{(n)})}\left[\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\phi}_{nT}(Z_{it})\left(X_{it}^{-}K_1^{(n)} - X_{it}^{+}K_2^{(n)}\right)\right]^2. \quad (12)$$

Using the linearity of $\hat{\Delta}_{2;3}^{(n)}(\theta)$ under $\mathbb{P}_{\theta;g_1}^{(n)}$, $g_1 \in \mathcal{F}_A$, for any bounded sequences $\tau^{(n)} = (\tau_1^{(n)}, \tau_2^{(n)}, \tau_2^{(n)})' \in \mathbb{R}^3$ and $1_2\tau_2^{(n)} := (\tau_2^{(n)}, \tau_2^{(n)})' \in \mathbb{R}^2$, as $n \to \infty$ with T fixed,

$$\hat{\Delta}_{2;3}^{(n)}(\theta + n^{-1/2}\gamma^{(n)}\tau^{(n)}) - \hat{\Delta}_{2;3}^{(n)}(\theta) = -\Gamma_{g_1;2,3}^{(n)}(\theta)1_2\tau_2^{(n)} + o_P(1),$$

with

$$\Gamma_{g_1;2,3}^{(n)}(\theta) = \frac{T}{\sigma_g^2}I_\phi(g_1)\begin{pmatrix} 1 & C_X^{(n)} \\ C_X^{(n)} & 1 \end{pmatrix}.$$

Proposition 5
Let Assumptions (A), (B), and (C) hold, and that the sequences (a_n) and (b_n) converging to zero. Then, under $\mathbb{P}_{\theta;g_1}^{(n)}$ and as $n \to \infty$, we have:

(i) $\hat{\Gamma}_{2;3}^{(n)} - \Gamma_{g_1;2,3}^{(n)} = o_P(1)$;
(ii) $\hat{\Delta}_{2;3}^{(n)}(\hat{\theta}^{(n)}) - \Delta_{g_1;2,3}^{(n)}(\hat{\theta}^{(n)}) = o_P(1)$.

Proposition 6
Let Assumptions (A), (B), and (C) hold, and that the sequences (a_n) and (b_n) converging to zero. for any $g_1 \in \mathcal{F}_A$. Then,

(i) $\hat{Q}^{(n)}(\hat{\theta}^{(n)}) = Q_{g_1}^{(n)}(\theta) + o_P(1)$ is asymptotically chi-square, with 1 degrees of freedom under $\mathbb{P}_{\theta;g_1}^{(n)}$, and asymptotically noncentral chi-square, still with 1 degrees of freedom but with noncentrality parameter $\rho := \frac{TI_\phi(g_1)(1-\mu_{C_X})}{2\sigma_g^2}(\tau_2-\tau_3)^2$ under $\mathbb{P}_{\theta+n^{-1/2}\gamma^{(n)}\tau;g_1}^{(n)}$;
(ii) the sequence of tests rejecting the null hypothesis $\bigcup_{g_1\in\mathcal{F}_A}\bigcup_{\sigma_g^2>0}\bigcup_{\beta\in\mathbb{R}}\left\{\mathbb{P}_{\sigma_g^2,\beta;g_1}^{(n)}\right\}$ whenever $\hat{Q}^{(n)}(\hat{\theta}^{(n)}) > \chi_{1,1-\alpha}^2$, is locally asymptotically optimal (most stringent), at asymptotic level α, against alternatives of the form

$$\bigcup_{\sigma_g^2>0}\bigcup_{\beta_1\in\mathbb{R}}\bigcup_{\beta_2\in\mathbb{R}}\left\{\mathbb{P}_{\sigma_g^2,\beta_1,\beta_2;g_1}^{(n)}\right\};$$

(iii) the sequence of tests has asymptotic power $1-\mathcal{F}\left(\chi_{1,1-\alpha}^2;\rho\right)$, at $\mathbb{P}_{\theta+n^{-1/2}\gamma^{(n)}\tau;g_1}^{(n)}$.

The test statistic $\hat{Q}^{(n)}(\hat{\theta}^{(n)})$ thus defines an adaptive test, which is optimal but remains valid under a large class of densities.

4. Asymptotic relative efficiencies

The squared ratios of noncentrality parameters under local alternatives are used to calculate the Asymptotic Relative Efficiencies (AREs) of the adaptive test based on $\hat{Q}^{(n)}$ with respect to the Gaussian test based on $Q_{\mathcal{N}}^{(n)}$. In order to compare the performance of the adaptive and parametric tests, we calculate the AREs for the adaptive tests in comparison to the Gaussian tests under density $g_1 \in \mathcal{F}_A$.

Proposition 7
The AREs under $g_1 \in \mathcal{F}_A$, of the adaptive tests based on $\hat{Q}^{(n)}$ with respect to the Gaussian test based on $Q_{\mathcal{N}}^{(n)}$, when

testing $\mathbb{P}^{(n)}_{\theta;g_1}$ against $\mathbb{P}^{(n)}_{\theta+n^{-1/2}\gamma^{(n)}\tau;g_1}$, are

$$ARE_{g_1}(\hat{Q}^{(n)}/Q^{(n)}_{\mathcal{N}}) = \left(\frac{\rho}{\rho_{\mathcal{N}}}\right)^2 = \left(I_\phi(g_1)\right)^2. \tag{13}$$

Table 1 provides numerical values of (13) for $\hat{Q}^{(n)}$ under various density functions g_1: Normal, Logistic, Double Exponential, Student-t_3, skew-normal $s\mathcal{N}(10)$, and skew-Student $st_5(10)$ (refer to [5] for a definition of skew normal and skew-t densities).

The obtained results are satisfactory. Furthermore, it is evident from Table 1 that the AREs of the suggested adaptive tests are consistently greater than or equal to one when compared to the parametric Gaussian tests across all distributions. In the case of the Normal distribution, the AREs are equal to one. Therefore, the adaptive tests demonstrate superior performance compared to the Gaussian tests for all density functions.

Table 1. Asymptotic relative efficiencies of adaptive test compared to their Gaussian counterpart

Densities g_1	$\mathcal{N}$	l	$\mathcal{D}e$	t_3	$s\mathcal{N}(10)$	$st_5(10)$
$ARE_{g_1}(\hat{Q}^{(n)}/Q^{(n)}_{\mathcal{N}})$	1.0000	1.2026	4.0000	4.0000	9.1063	9.1641

5. Simulation

In order to enhance the interpretation and validity of the theoretical results presented in the previous sections, we conducted a simulation experiment using Rstudio programming. The purpose of this section is to assess the performance of the proposed tests at an asymptotic level of $\alpha = 5\%$. Now, let's evaluate the following model:

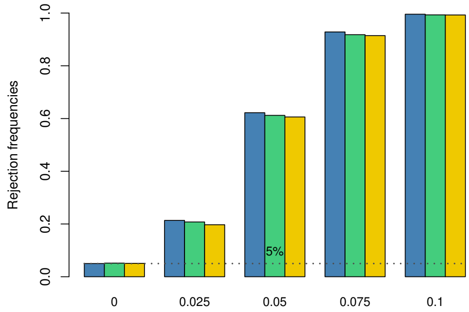
$$Y_{it} = \beta_1 X^-_{it} + \beta_2 X^+_{it} + \varepsilon_{it}, \quad i = 1, \dots, n = 100,\ t = 1, \dots, T = 10, \tag{14}$$

where,

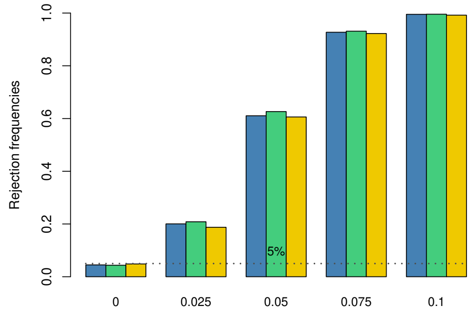
- $\beta_1 = 7$,
- $\beta_2 = \beta_1 + \lambda_2$, $\lambda_2 = 0$ for null hypothesis, and $\lambda_2 = 0.025, 0.05, 0.075, 0.1$ for increasingly several alternatives,
- the x_{it}'s are i.i.d. uniform $(-10, 10)$, $X^-_{it} = x^-_{it} - \overline{x^-}$ and $X^+_{it} = x^+_{it} - \overline{x^+}$,
- the ε_{it}'s are i.i.d. with a symmetric density: Gaussian $(\mathcal{N})$, logistic (l), double exponential $(\mathcal{D}e)$, Student with $\nu = 3$ degrees of freedom (t_3), and an asymmetric density: skew-normal $s\mathcal{N}(\delta)$ and skew-Student $st_5(\delta)$ densities (both with skewness parameter value $\delta = 10$).

To evaluate the performance of the proposed procedures using finite samples, we generated 2500 independent samples of size $N = nT = 1000$ from model (14). For each replication, we conducted the following tests at the asymptotic level $\alpha = 5\%$: the likelihood ratio test for panel threshold effects proposed by [17], the Gaussian test based on $Q^{(n)}_{\mathcal{N}}$ (see Proposition 4), and the adaptive test based on $\hat{Q}^{(n)}$ in (12). For the adaptive test, we used the Gaussian kernel $k(x) = 1/\sqrt{2\pi}\exp\left(-x^2/2\right)$, with $a_n = 0.9$ and $b_n = 0.01$.

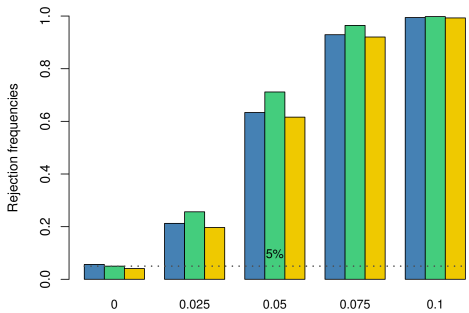
Rejection frequencies are reported in Figure 1, and they strongly support the outstanding overall performance of the adaptive test. Moreover, both the Gaussian and the adaptive test outperform the likelihood ratio test of [17] across all considered error densities g_1. It is also evident that the adaptive test $\hat{Q}^{(n)}$ has significantly greater power than the Gaussian test $Q^{(n)}_{\mathcal{N}}$ under logistic (l), double exponential $(\mathcal{D}e)$, Student (t_3), skew-normal $s\mathcal{N}(10)$, and skew-Student $st_5(10)$ distributions. Additionally, it is noteworthy that the adaptive test $\hat{Q}^{(n)}$ performs equally well as the Gaussian test $Q^{(n)}_{\mathcal{N}}$ under Normal $(\mathcal{N})$ density.



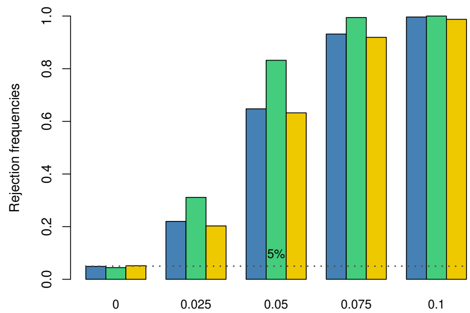
(a) g_1 Gaussian.



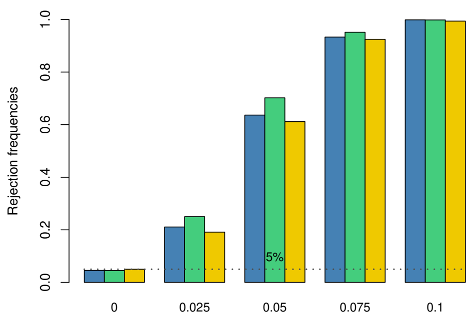
(b) g_1 Logistic.



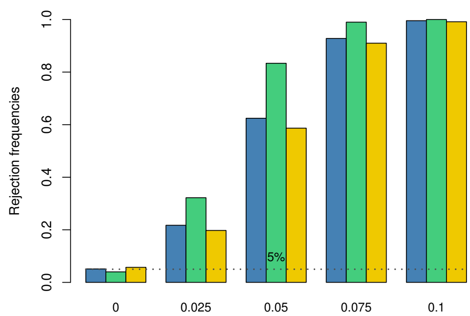
(c) g_1 Double exponential.



(d) g_1 Student-t_3.



(e) g_1 Skew-normal $s\mathcal{N}(10)$.



(f) g_1 Skew-Student $st_5(10)$.

Gaussian test   Adaptive test   Likelihood Ratio test

Figure 1. Rejection frequencies (out of 2500 replications), for $\lambda_2 = 0$ (null hypothesis), 0.025, 0.05, 0.075, 0.1 (alternative hypotheses), with error density g_1 that is Gaussian ($\mathcal{N}$), logistic (l), double exponential ($\mathcal{D}e$), Student (t_3), skew-normal ($s\mathcal{N}(10)$), and skew-Student ($st_5(10)$) of the Gaussian, Adaptive, and Likelihood Ratio tests.

6. Real data analysis

Numerous factors have been presented in the current literature to explain pollution, such as tourism [34], economic growth [19], renewable energy [4], and urbanization [35]. However, previous literature [29] indicates that it is difficult to determine a priori the effect of urbanization on CO_2 emissions. While research on the urbanization–CO_2 linkage is growing, the findings remain inconsistent, which can be attributed to the various econometric techniques used. The current body of literature discussing this issue can be divided into two opposing groups. The first group supports the idea of a linear relationship [23], while the second group supports the idea of a non-linear relationship [31] between these two variables. Moreover, the results of previous studies vary from country to country. For example, [24] reported a negative effect, whereas [2] found a positive effect of urbanization on the environment.

Based on panel data from 15 emerging countries during 1995-2015, [20] discovered a non-linear relationship between CO_2 emissions and urbanization. They found that urbanization contributes relatively strongly to carbon dioxide emissions up to a certain level, and their analysis relied on Hansen's approach. In the present study, we revisit the same dataset, focusing on a shorter period (1995–2004). We apply Gaussian and adaptive tests to this subset, which are considered more powerful than the Hansen test, allowing us to examine threshold effects within a short panel context.

To analyze the potential presence of non-linearity, we utilize the PTR model, as introduced by [17]. Specifically, we examine the following form of the model:

$$y_{it} = \mu_i + \beta_1' x_{it} \mathbf{1}\left(q_{it} \leq \gamma\right) + \beta_2' x_{it} \mathbf{1}\left(q_{it} > \gamma\right) + \varepsilon_{it}, \quad i = 1, \dots, n;\ t = 1, \dots, T,$$

where y_{it} represents the CO_2 emissions in country i during period t, μ_i and ε_{it} respectively represent the country-specific fixed effects and random errors. It is assumed that this error is independently and identically distributed (i.i.d.) with a mean of 0 and a variance of σ^2. In addition, q_{it} denotes the threshold variable for each country in the sample during time t, γ is the threshold value, and $\mathbf{1}(.)$ serves as an indicator for the two regimes with different regression slopes β_1' and β_2'.

To detect the threshold phenomenon, [17] used the following test statistic.

$$F_1 = \frac{S_0 - S_1(\hat{\gamma})}{\hat{\sigma}^2},$$

where S_0 represents the residual sum of squares for errors in the linear model, S_1 represents the residual sum of squares for errors in the panel threshold estimation model, and $\hat{\sigma}^2$ represents the residual variance of the panel threshold estimation. The null hypothesis states that γ is not identified, implying a linear relationship. The alternative hypothesis suggests the presence of at least one threshold.

$$\mathcal{H}_0 : \beta_1 = \beta_2 = \beta \text{ and } \mathcal{H}_1 : \beta_1 \neq \beta_2.$$

However, there are cases where multiple thresholds exist, resulting in three distinct regimes or more. Similar to the F_1 test used for a model with a single threshold, we can assess the significance of the second threshold by applying the likelihood ratio test using the F_2 statistic outlined below.

$$F_2 = \frac{S_1(\hat{\gamma}) - S_2^r(\hat{\gamma})}{\hat{\sigma}^2},$$

where $S_1(\hat{\gamma})$ represents the residual sum of squared errors resulting from estimating the first threshold, $S_2^r(\hat{\gamma})$ and $\hat{\sigma}^2$ are the residual sum of squared errors and the residual variance from the second threshold estimation, respectively.

For any given threshold γ, the slope coefficient can be estimated using Ordinary Least Squares (OLS);

$$\hat{\beta}(\gamma) = \left(X^*(\gamma)' X^*(\gamma)\right)^{-1} X^*(\gamma)' Y^*.$$

The vector of the regression residuals is:

$$\hat{\varepsilon}^*(\gamma) = Y^* - \mathrm{X}^*(\gamma)\hat{\beta}(\gamma),$$

and the sum of the squared errors is:

$$\mathrm{S}_1(\gamma) = \hat{\varepsilon}^*(\gamma)'\hat{\varepsilon}^*(\gamma). \tag{15}$$

[8, 18] suggested using the least squares method to estimate γ. The simplest approach is to minimize the concentrated S_1 in equation (15). Therefore, the least squares estimator of γ is:

$$\hat{\gamma} = \arg\min_{\gamma} S_1(\gamma).$$

When we obtain $\hat{\gamma}$, the slope coefficient estimate is $\hat{\beta} = \hat{\beta}(\hat{\gamma})$. The residual vector is $\hat{\varepsilon}^* = \hat{\varepsilon}^*(\hat{\gamma})$, and the estimator of the residual variance is:

$$\hat{\sigma}^2 = \hat{\sigma}^2(\hat{\gamma}) = \frac{1}{n(T-1)}\hat{\varepsilon}^{*\prime}(\hat{\gamma})\hat{\varepsilon}^*(\hat{\gamma}) = \frac{1}{n(T-1)}\,\mathrm{S}_1(\hat{\gamma}).$$

Our study used a balanced panel dataset that included 15 emerging countries from 1995 to 2004. These countries are Argentina, Bangladesh, Brazil, China, India, Indonesia, Malaysia, Mexico, Pakistan, Philippines, Russian Federation, South Africa, Thailand, Turkey, and Ukraine. The study incorporated the following variables: CO_2 emissions (measured in metric tons per capita), urban population (as a percentage of the total population), GDP per capita (measured in constant 2015 US dollars), renewable energy consumption (as a percentage of total final energy consumption), and trade (as a percentage of GDP). These data series were obtained from the World Bank's World Development Indicators (2024).

The construction of our single threshold model is as follows:

$$\begin{aligned}\ln(co_{2_{it}}) = \mu + \beta_1 \ln(urb_{it})\mathbf{1}\left(\ln(urb_{it}) \le \gamma\right) + \beta_2 \ln(urb_{it})\mathbf{1}\left(\ln(urb_{it}) > \gamma\right)\\ + \beta_3 \ln(gdp_{it}) + \beta_4 \ln(ren_{it}) + \beta_5 \ln(trade_{it}) + \varepsilon_{it}.\end{aligned} \tag{16}$$

To investigate non-linearity, we utilized Hansen's panel threshold approach. After rejecting the null hypothesis, we conducted tests to determine the optimal number of thresholds for our model. To assess the presence of a single threshold, we performed 1000 bootstrap estimations. The results of this analysis are presented in Tables 2 and 3. Table 2 presents the estimates for the single threshold at 3.1608, with a 95% confidence interval ranging from 3.1442 to 3.2223. The anti-log of the threshold value, 3.1608, corresponds to 23.59%, indicating that the threshold level is at 23.59% urbanization. Moving on to Table 3, it displays the significance level of a single threshold. We discovered that the test results for a single threshold are significant at the 5% level, with a bootstrap p-value of 0.0030. Furthermore, the F_1-statistic exceeds the critical value, providing evidence in support of non-linearity and rejecting the notion of a linear relationship between urbanization and carbon emissions.

Table 2. Estimation of a model with a single threshold

95% Confidence Interval (CI)			
Model	Threshold	Lower	Upper
Th-1	3.1608	3.1442	3.2223

Table 3. Test for the single threshold model

Threshold	RSS[a]	MSE[b]	F_1-stat	Probability	Crit 10%	Crit 5%	Crit 1%
Single	0.1918	0.0014	45.0600	0.0030	23.6213	27.5134	33.9653

[a]RSS is the Residual Sum of Squares; [b]MSE is the Mean Squared Error.

We will now analyze the occurrence of double and triple thresholds in the relationship between urbanization and carbon dioxide emissions. The empirical results of the tests for single, double, and triple thresholds are presented in Table 4. This table includes estimates obtained from 1000 bootstrapping iterations, which are used to approximate the presence of double and triple thresholds. The estimates indicate that a single threshold is statistically significant, with a p-value of 0.0030. However, the double and triple thresholds are not statistically significant, with probability values of 0.1870 and 0.3510, respectively. Based on these results, we can conclude that there is only one threshold level in the relationship between urbanization and carbon emissions.

Table 4. Test for multiple threshold models

Threshold	RSS	MSE	F statistic	Probability	Crit 5%
Single	0.1918	0.0014	45.0600	0.0030	27.5134
Double	0.1632	0.0012	24.5000	0.1870	60.4149
Triple	0.1436	0.0010	19.1200	0.3510	73.7865

Prior to discussing the numerical outcomes, the following algorithm outlines the computational steps required to obtain the adaptive test statistic $\hat{Q}^{(n)}$ used in the empirical application.

Algorithm 1 Pseudocode for Computing the Adaptive Test Statistic $\hat{Q}^{(n)}$

Step 1: Data import
Load data and define variables $\ln(co_2)$, $\ln(urb)$, $\ln(gdp)$, $\ln(ren)$, and $\ln(trade)$.
Step 2: Splitting urban variable by threshold
Define the threshold γ and split $\ln(urb)$ into $lnurb^-$ and $lnurb^+$ based on γ.
Step 3: Centering variables
Center all variables and compute $K_1^{(n)}$, $K_2^{(n)}$, and $C_X^{(n)}$.
Step 4: Null model estimation
Based on the centered variables obtained in Step 3, estimate the linear model under H_0 and compute the residuals Z.
Step 5: Kernel-based estimation of density and derivative
Define bandwidth a_n and regularization parameter b_n.
Estimate the density f_1 using a kernel density.
Estimate the derivative f_1' using the derivative of the kernel.
Step 6: Score function and Fisher information
Compute the estimated score $\hat{\phi}_{nT}(Z)$.
Compute the estimated Fisher information $\hat{I}_{nT}$.
Step 7: Test statistic and p-value
Compute the test statistic $\hat{Q}^{(n)}$ defined in equation (12) and the associated p-value.

Now, let's compare the Gaussian test based on $Q_{\mathcal{N}}^{(n)}$, the adaptive test based on $\hat{Q}^{(n)}$, and the Hansen test based on F_1 (refer to Table 5). The observed values of the statistical tests are as follows: $Q_{\mathcal{N}}^{(n)} = 25.0157$ (p-value $=$ 5.6866×10^{-7}), $\hat{Q}^{(n)} = 14.4197$ (p-value $= 0.0002$), and $F_1 = 45.0600$ (p-value $= 0.0030$). These results indicate that all three tests lead to the same conclusion: $\mathcal{H}_0 : \beta_1 = \beta_2 = \beta$ should be rejected at the usual significance levels (α = 5%). Moreover, the findings suggest that the proposed tests (Gaussian and adaptive) are more powerful than the likelihood ratio test.

Table 5. Gaussian, adaptive and likelihood ratio tests

Test	Calculated value	Critical value	p-value
$Q_{\mathcal{N}}^{(n)}$	25.0157	3.8414	5.6866×10^{-7}
$\hat{Q}^{(n)}$	14.4197	3.8414	0.0002
F_1	45.0600	27.5134	0.0030

7. Conclusions

The approach used in this paper allows for testing the classical regression model (without a threshold) against the panel threshold regression model (with large n and small T) for both specified and unspecified error density. Optimal parametric, Gaussian, and adaptive tests are derived based on the LAN property. The adaptive tests exhibit remarkably high ARE values compared to their Gaussian counterparts. Simulation results, based on rejection frequencies, confirm the strong performance of the proposed tests and show that they outperform Hansen's likelihood ratio test. In particular, the adaptive test demonstrates the highest empirical power. A real example illustrating the relationship between urbanization and carbon dioxide emissions confirms the presence of a single threshold effect, rejecting the classical regression model in favor of the threshold specification. The results further highlight the superiority of the Gaussian and adaptive tests over Hansen's likelihood ratio test in detecting threshold effects.

Supplementary materials

For reproducibility and convenience, the R code for the real data application of both the adaptive and Gaussian tests is provided online and can be accessed on GitHub at: https://github.com/DOUNIABOURZIK/Adaptive-and-Gaussian-Tests-Panel-Threshold-Regression.git

Acknowledgement

The authors wish to express their special thanks to the Editor and the Reviewers for their time, careful reading and comments which greatly improved the manuscript.

REFERENCES

1. A. Akharif, M. Fihri, M. Hallin, and A. Mellouk, *Optimal pseudo-Gaussian and rank-based random coefficient detection in multiple regression*, Electronic Journal of Statistics, vol. 14, no. 2, pp. 4207–4243, 2020.
2. U. Al-Mulali, I. Ozturk, and H. H. Lean, *The influence of economic growth, urbanization, trade openness, financial development, and renewable energy on pollution in Europe*, Natural Hazards, vol. 79, no. 1, pp. 621–644, 2015.
3. J. Allal and S. El Melhaoui, *Tests de rangs adaptatifs pour les modèles de régression linéaire avec erreurs ARMA*, Annales des Sciences Mathématiques du Québec, vol. 30, no. 1, pp. 29–54, 2006.
4. M. A. Anwar, S. Nasreen, and A. K. Tiwari, *Forestation, renewable energy and environmental quality: Empirical evidence from Belt and road initiative economies*, Journal of Environmental Management, vol. 291, p. 112684, 2021.
5. A. Azzalini and A. Capitanio, *Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution*, Journal of the Royal Statistical Society, Series B: Statistical Methodology, vol. 65, no. 2, pp. 367–389, 2003.
6. N. Bennala, M. Hallin, and D. Paindaveine, *Pseudo-Gaussian and rank-based optimal tests for random individual effects in large n small T panels*, Journal of Econometrics, vol. 170, no. 1, pp. 50–67, 2012.
7. A. Bouattour, M. Kalai, and K. Helali, *Threshold effects of technology import on industrial employment: a panel smooth transition regression approach*, Journal of Economic Structures, vol. 13, no. 1, 2024.
8. K. S. Chan, *Consistency and limiting distribution of the least squares estimator of a continuous threshold autoregressive model*, The annals of statistics, vol. 21, no. 1, pp. 520–533, 1993.
9. G. Colletaz and C. Hurlin, *Threshold effects in the public capital productivity: An international panel smooth transition approach*, 2006.

10. S. El Melhaoui, J. Allal, and A. Kaaouachi, *Efficient tests for AR(1) dependence against SETAR(2,1,1) models*, Journal of Natural Sciences and Mathematics, vol. 2, no. 1, pp. 1–16, 2008.
11. M. Elsherif, *Nexus between blue economy, renewable energy and environmental sustainability in the MENA region: evidence from panel threshold regression*, International Journal of Energy Economics and Policy, vol. 14, no. 4, pp. 169–185, 2024.
12. M. Fihri, A. Akharif, A. Mellouk, and M. Hallin, *Efficient pseudo-Gaussian and rank-based detection of random regression coefficients*, Journal of Nonparametric Statistics, vol. 32, no. 1, pp. 367–402, 2020.
13. D. Fok, D. van Dijk, and P. Franses, *A multi-level panel STAR model for US manufacturing sectors*, Journal of Applied Econometrics, vol. 20, no. 6, pp. 811–827, 2005.
14. S. Gharbi and H. Othmani, *Threshold effect in the relationship between family ownership and firm performance: A panel smooth transition regression analysis*, Cogent Economics & Finance, vol. 10, no. 1, pp. 2023264, 2022.
15. A. González, T. Teräsvirta, and D. van Dijk, *Panel smooth transition regression models*, Technical report, SSE/EFI Working Paper Series in Economics and Finance, 2005.
16. M. Hallin and B. J. M. Werker, *Semi-parametric efficiency, distribution-freeness and invariance*, Bernoulli, vol. 9, no. 1, pp. 137–165, 2003.
17. B. E. Hansen, *Threshold effects in non-dynamic panels: Estimation, testing, and inference*, Journal of Econometrics, vol. 93, no. 2, pp. 345–368, 1999.
18. B. E. Hansen, *Sample splitting and threshold estimation*, Econometrica, vol. 68, no. 3, pp. 575–603, 2000.
19. R. Khalfaoui, A. K. Tiwari, U. Khalid, and M. Shahbaz, *Nexus between carbon dioxide emissions and economic growth in G7 countries: Fresh insights via wavelet coherence analysis*, Journal of Environmental Planning and Management, vol. 66, no. 12, pp. 31–66, 2021.
20. M. Kocoglu, A. Awan, A. Tunc, and A. Aslan, *The nonlinear links between urbanization and CO_2 in 15 emerging countries: Evidence from unconditional quantile and threshold regression*, Environmental Science and Pollution Research, vol. 29, no. 12, pp. 18177–18188, 2022.
21. L. M. Le Cam, *Asymptotic methods in statistical decision theory*, Springer-Verlag, New York, 1986.
22. L. M. Le Cam and G. L. Yang, *Asymptotics in statistics: Some basic concepts*, 2nd edition, Springer-Verlag, New York, 2000.
23. H. Li, Mu. Hailin, M. Zhang, and N. Li, *Analysis on influence factors of China's CO_2 emissions based on path–STIRPAT model*, Energy Policy, vol. 39, no. 11, pp. 6906–6911, 2011.
24. K. Li and B. Lin, *Impacts of urbanization and industrialization on energy consumption/CO_2 emissions: Does the level of development matter?*, Renewable and Sustainable Energy Reviews, vol. 52, pp. 1107–1122, 2015.
25. A. Lmakri, A. Akharif, and A. Mellouk, *Optimal detection of bilinear dependence in short panels of regression data*, Revista Colombiana de Estadística - Theoretical Statistics, vol. 43, no. 2, pp. 143–171, 2020.
26. G. E. Noether, *On a theorem by Wald and Wolfowitz*, The Annals of Mathematical Statistics, vol. 20, no. 3, pp. 455–458, 1949.
27. S. Regui, A. Akharif, and A. Mellouk, *Locally optimal tests against periodic linear regression in short panels*, Communications in Statistics - Simulation and Computation, vol. 54, no. 6, pp. 2221–2235, 2024.
28. S. Regui, A. Akharif, and A. Mellouk, *Detecting random effects and error correlation in a periodic multiple regression model*, Journal of Statistical Computation and Simulation, pp. 1–44, 2025.
29. M. Salahuddin, M. I. Ali, N. Vink, and J. Gow, *The effects of urbanization and globalization on CO_2 emissions: Evidence from Sub-Saharan Africa countries*, Environmental Science and Pollution Research, vol. 26, no. 3, pp. 2699–2709, 2019.
30. A. Schick, *On efficient estimation in regression models*, The Annals of Statistics, vol. 21, no. 3, pp. 1486–1521, 1993.
31. M. Shahbaz, N. Loganathan, A. T. Muzaffar, K. Ahmed, and M. A. Jabran, *How urbanization affects CO_2 emissions in Malaysia? The application of STIRPAT model*, Renewable and Sustainable Energy Reviews, vol. 57, pp. 83–93, 2016.
32. J. Siddiki and L. R. Bala-Keffi, *Revisiting the relation between financial inclusion and economic growth: a global analysis using panel threshold regression*, Economic Modelling, vol. 135, no. 1, pp. 106707, 2024.
33. A. R. Swensen, *The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend*, Journal of Multivariate Analysis, vol.16 , no. 1, pp. 54–70, 1985.
34. A. K. Tiwari, S. Nasreen, and Z. Iqbal, *Nexus between tourism and environmental pollution in South Asia: A comparative analysis using time-varying and non-parametric techniques*, Current Issues in Tourism, vol. 24, no. 21, pp. 2996–3020, 2020.
35. R. Ulucak, F. Erdogan, and S. H. Bostanci, *A STIRPAT-based investigation on the role of economic growth, urbanization, and energy consumption in shaping a sustainable environment in the Mediterranean region*, Environmental Science and Pollution Research, vol. 28, no. 39, pp. 55290–55301, 2021.
36. A. W. van der Vaart, *Asymptotic statistics*, Cambridge University Press, Cambridge, 1998.
37. A. Wald, *Tests of statistical hypotheses concerning several parameters when the number of observations is large*, Transactions of the American Mathematical Society, vol. 54, no. 3, pp. 426–482, 1943.