Consistency of Trend Break Point Estimator with Underspecified Break Number

Jingjing Yang

Department of Economics, University of Nevada, Reno, 1664 N. Virginia Street, Reno, NV 89557, USA; jingjingy@unr.edu; Tel.: +1-775-784-1112

Abstract: This paper discusses the consistency of trend break point estimators when the number of breaks is underspecified. The consistency of break point estimators in a simple location model with level shifts has been well documented by researchers under various settings, including extensions such as allowing a time trend in the model. Despite the consistency of break point estimators of level shifts, there are few papers on the consistency of trend shift break point estimators in the presence of an underspecified break number. The simulation study and asymptotic analysis in this paper show that the trend shift break point estimator does not converge to the true break points when the break number is underspecified. In the case of two trend shifts, the inconsistency problem worsens if the magnitudes of the breaks are similar and the breaks are either both positive or both negative. The limiting distribution for the trend break point estimator is developed and closely approximates the finite sample performance.

---

1. Introduction

A time series can have multiple breaks. For example, U. S. Treasury bill rates can be observed to have multiple level changes over time, while the Grilli and Yang primary commodity price index shows multiple trend shifts. It is common that the number of breaks is unknown and misspecified. Bai (1995, 1997) [1,2] and Chong (1994, 1995) [3,4] study the consequences of underspecifying the number of break points in linear structural break models. They point out that when the number of breaks in a mean shift model is underspecified, the break point estimator is still consistent for a subset of the true break points. Their discussion covers the mean shift model with and without trend. Bai (1997) [2] shows that the mean break point estimator by sequential estimation is not only consistent but also converges at the same rate as with simultaneous estimation. Bai and Perron (1998) [5] extend the estimation of a single unknown break to multiple unknown breaks under both fixed and shrinking shift magnitudes. Based on the consistency property of the mean shift break point estimator, they propose a sequential procedure for multi-break estimates without estimating the multiple breaks simultaneously. Dynamic programming is introduced by Bai and Perron (2003) [6] to deal with the computational burden in multiple break point estimation. Kejriwal and Perron (2010) [7] extend the work of Perron and Yabu (2009) [8,9] to propose a sequential test of the multiple-trend-shift model robust to persistence in noise.

Although trending components are considered by researchers in the mean shift model, there is little discussion of the consistency of multiple trend shift break point estimators when the number of breaks is underspecified. Consistency analysis is important both for break point estimation and for

structural breaks in the linear regression model. The main motivation of this paper is to address the gap in the literature concerning the consistency of trend shift break point estimators when the break number is underspecified.

The second motivation of this paper is to explore how to approximate the finite sample distributions of the break point estimator for a multiple break model. Specifically, asymptotics of the break point estimator in a trend shift model are provided for the case of an underspecified break number by employing Pitman drifts. The accuracy of the asymptotic approximation to the finite sample distribution is examined. This work follows Yang (2012) [10] who has shown that the finite sample distribution of the single break point estimator is not normal, but depends on the break dates and magnitudes.

In this paper, finite sample simulations are used to illustrate the potential inconsistency of the break point estimator in the trend shift model with an underspecified break number. Then, the limits of the break point estimator under fixed break magnitudes are provided. Both the simulation results and the expression of the limits show that for the trend shift model, the break point estimator can be inconsistent for any of the true break points, while for the mean shift model, the break point estimator converges to one of the true breaks. Then, extending Yang's (2012) work [10] on the single break point estimator, new asymptotics are provided for the break point estimators under local alternatives.

As will be shown in this paper, the mean shift model leads to a consistent break point estimator while the trend shift model does not. Taking first differences of the trend shift model is shown by Yang (2010) [11] to provide a solution to the inconsistency problem. When the break magnitudes are sufficiently large, the first-difference break point estimator has much higher peaks in the density at the true breaks than the levels break point estimator. When the break magnitudes are small, the densities of the two break point estimators depend on the break magnitudes and locations and the strength of the serial correlation. A detailed analysis of the first-difference estimator is omitted in this paper but can be found in Yang (2010) [11] and Yang (2012) [10].

The paper is organized as follows. Section 2 describes the general settings of the mean shift and trend shift models, assumptions, and break point estimators. Section 3 introduces finite sample simulations to demonstrate the consistency properties of different break point estimators. Section 4 derives the expression of the limits of the single break point estimator when the break sizes are fixed and the data sequences have two breaks under I(0) errors. Both mean shift and trend shift break point estimators are discussed. Section 5 establishes the asymptotic distributions of the break point estimators assuming the breaks are Pitman drifts, which approximate the finite sample distributions accurately. Sections 4 and 5 relate the mean shift results to those of Bai (1997) [2]. The last section concludes the paper. Proofs are provided in the Appendix.

2. The Models, Assumptions, and Break Point Estimators

In this section, I define a mean shift and a trend shift model with multiple breaks. For simplicity, I only include the case where a single break model is estimated while the number of breaks is two. The results can be extended to models with more than two breaks.

Let us start with a mean shift model with two breaks:

$$y_t = \mu + \delta_1 DU_t(\lambda_1^c) + \delta_2 DU_t(\lambda_2^c) + u_t, \tag{1}$$

where

$$DU_t(\lambda_i^c) \doteq \begin{cases} 0, & t \leq T_{b,i}^c \\ 1, & t > T_{b,i}^c \end{cases}, \quad i = 1, 2;$$

λ_1^c and λ_2^c are the true break fractions with $T_{b,1}^c = \lambda_1^c T$ and $T_{b,2}^c = \lambda_2^c T$; $T_{b,i}^c$ denotes the time of a break. T is the sample length; δ_1 and δ_2 are the break magnitudes. For convenience of discussion, we define the relative break magnitude ratio $\nu \doteq \delta_2/\delta_1$.

When model (1) is underspecified, the estimated model is given by

$$y_t = \mu + \delta DU_t(\lambda) + u_t, \tag{2}$$

where

$$DU_t(\lambda) \doteq \begin{cases} 0, & t \leq T_b \\ 1, & t > T_b \end{cases} ;$$

λ is the underspecified single break fraction with $T_b = \lambda T$.

For comparison, the trend shift model with two breaks is

$$y_t = \mu + \beta t + \delta_1 DT_t(\lambda_1^c) + \delta_2 DT_t(\lambda_2^c) + u_t, \tag{3}$$

where $DT_t(\lambda_i^c) \doteq (t - T_{b,i}^c) \cdot DU_t(\lambda_i^c), \ i = 1, 2$.

If model (3) is misspecified with only one break, the estimated model is

$$y_t = \mu + \beta t + \delta DT_t(\lambda) + u_t, \tag{4}$$

where $DT_t(\lambda) \doteq (t - T_b) \cdot DU_t(\lambda)$.

It is assumed that the error u_t is I(0), namely

$$u_t = d(L)e_t, \tag{5}$$

where

$$d(L) = \sum_{i=0}^{\infty} d_i L^i, \quad \sum_{i=0}^{\infty} i|d_i| < \infty, \quad d(1)^2 > 0;$$

L is the lag operator; $\{e_t\}$ is a martingale difference sequence with $\sup_t E(e_t^4) < \infty$, $E(e_t|e_{t-1}, e_{t-2}, \cdots) = 0$, and $E(e_t^2|e_{t-1}, e_{t-2}, \cdots) = \sigma_e^2$.

The break point estimators are obtained by minimizing the sum of squared residuals (SSR) over the trimming set $\Lambda \doteq \{\lambda^*, \cdots, 1 - \lambda^*\}$, namely

$$\hat{\lambda}_{MS} = \arg\min_{\lambda \in \Lambda^*} \{SSR_{MS}(\lambda)\},$$

$$\hat{\lambda}_{TS} = \arg\min_{\lambda \in \Lambda^*} \{SSR_{TS}(\lambda)\},$$

where

$$SSR_{MS}(\lambda) \doteq \sum_{t=1}^{T} [y_t - \hat{\mu}_{MS} - \hat{\delta}_{MS} DU_t(\hat{\lambda})]^2, \tag{6}$$

$$SSR_{TS}(\lambda) \doteq \sum_{t=1}^{T} [y_t - \hat{\mu}_{TS} - \hat{\beta}_{TS} t - \hat{\delta}_{TS} DT_t(\hat{\lambda})]^2, \tag{7}$$

with $\hat{\mu}_{MS}$ and $\hat{\delta}_{MS}$ the OLS estimators from model (2) with no restrictions imposed, whereas $\hat{\mu}_{TS}$, $\hat{\beta}_{TS}$, and $\hat{\delta}_{TS}$ are the OLS estimators from model (4) with no restrictions imposed.

3. Illustration of the Inconsistency Problem of the Trend Shift Break Point Estimator

In this section a simple simulation is used to illustrate the consistency/inconsistency of $\hat{\lambda}_{MS}$ and $\hat{\lambda}_{TS}$ in the presence of an underspecified break number. The data are generated based on models (1) and (3) with two breaks, where $T = 100, 250, 500, 1000$, $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$, $\nu = -2, -1, 1, 2$ (we set $\delta_1 = 1$ without loss of generality), and u_t is an *i.i.d.* $N(0,1)$ process. Equations (6) and (7) are used to estimate $\hat{\lambda}_{MS}$ and $\hat{\lambda}_{TS}$ separately in each replication. While trimming is not necessary, to ensure the invertibility of the regression matrix I use 2% trimming, i.e., $\lambda^* = 0.02$. The replications $N = 20{,}000, 10{,}000, 5000, 2500$ are used for $T = 100, 250, 500, 1000$ respectively.

Figure 1a,b plots the histograms of $\hat{\lambda}_{MS}$ with *i.i.d.* errors. In all cases with the increase of T, the distribution of $\hat{\lambda}_{MS}$ has shorter tails and, when $T = 1000$, concentrates at the two break points or one of them depending on the relative break magnitude ratios. Interestingly, when $|\nu| = 1$ and $T = 100$, the density of $\hat{\lambda}_{MS}$ is bimodal, which can be explained by Yang (2012) [10] through the behavior of the mean shift break point estimator, where the break point estimates concentrate around the end points in the no break model.

Figure 1c,d plots the histograms of $\hat{\lambda}_{TS}$. When $\nu = -2$, the density peaks at a point greater than $2/3$. When $\nu = -1$, $\hat{\lambda}_{TS}$ has two equal peaks at $\lambda = 0.2$ and 0.8. When $\nu = 1$, the histogram of $\hat{\lambda}_{TS}$ has only one peak at $\lambda = 0.5$, and with the increase of T the break date estimates are more concentrated. When $\nu = 2$, the histogram of $\hat{\lambda}_{TS}$ peaks at a point between $1/3$ and $2/3$. This shows that when the number of breaks is underspecified, the trend shift break point estimator does not converge to either of the true break points, and that the limit of the break point estimator $\hat{\lambda}_{TS}$ depends on the break magnitudes and locations.

Empirical data also shows that the break point estimators behave differently when the break number is underspecified in mean shift model and trend shift model. Using the US ex-post real interest rate in Figure 2 as an example of mean shifts (the three-month treasury bill rate between the first quarter of 1961 and the third quarter of 1986 deflated by the CPI inflation rate taken from the Citibase data bank), Bai and Perron (1998) [5] detect three mean shifts in years 1965, 1972, and 1980 while a single mean shift point estimator detects one of the real breaks in 1980. Using the extended Grilli and Yang commodity price index as an example of trend shifts (Copper during 1900–2003), Harvey, Leybourne, and Taylor (2009) [12] identify two breaks in 1945 and 1971, while a single trend shift estimator identifies one in 1930, which is not close to the HLT dates.

Both the finite sample histograms and empirical data suggest an interesting pattern: when the break number is underspecified, the mean shift break point estimator converges to a subset of the true break points, while the trend shift counterpart does not converge to either of the true break points and its limit depends on the break dates and magnitudes.
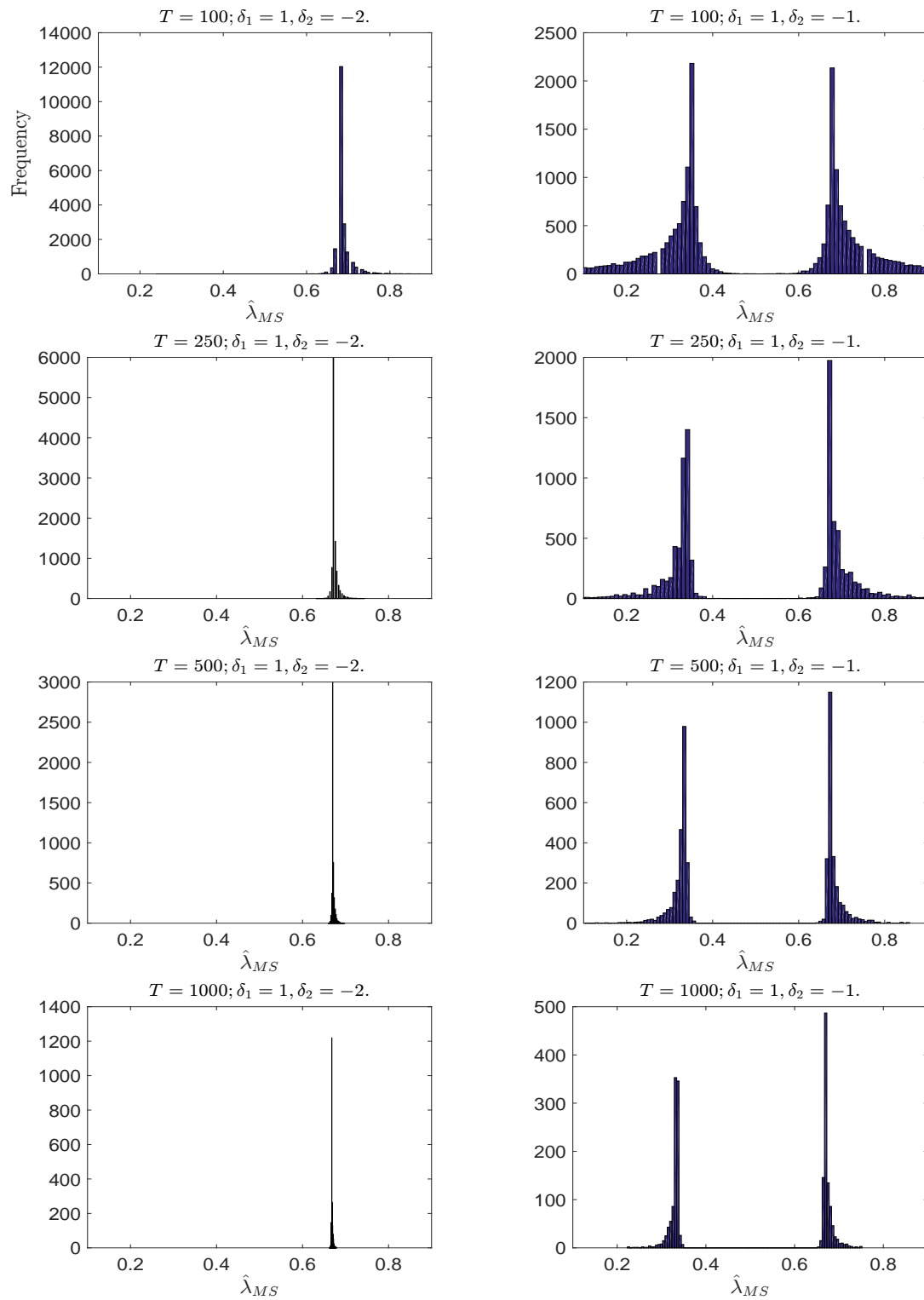
(**a**)

Figure 1. *Cont.*
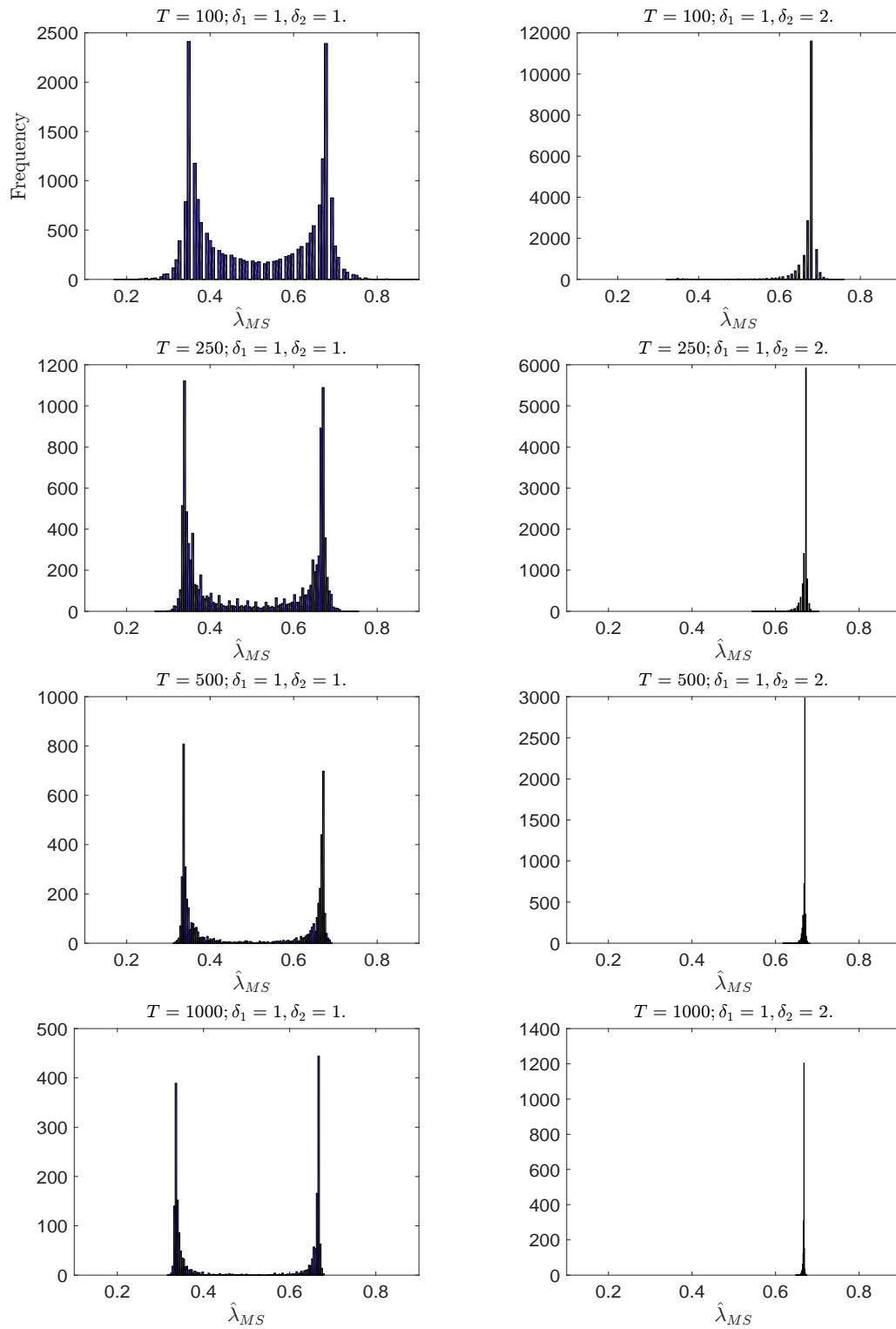
(**b**)

Figure 1. *Cont.*

Figure 1. *Cont.*

(**d**)

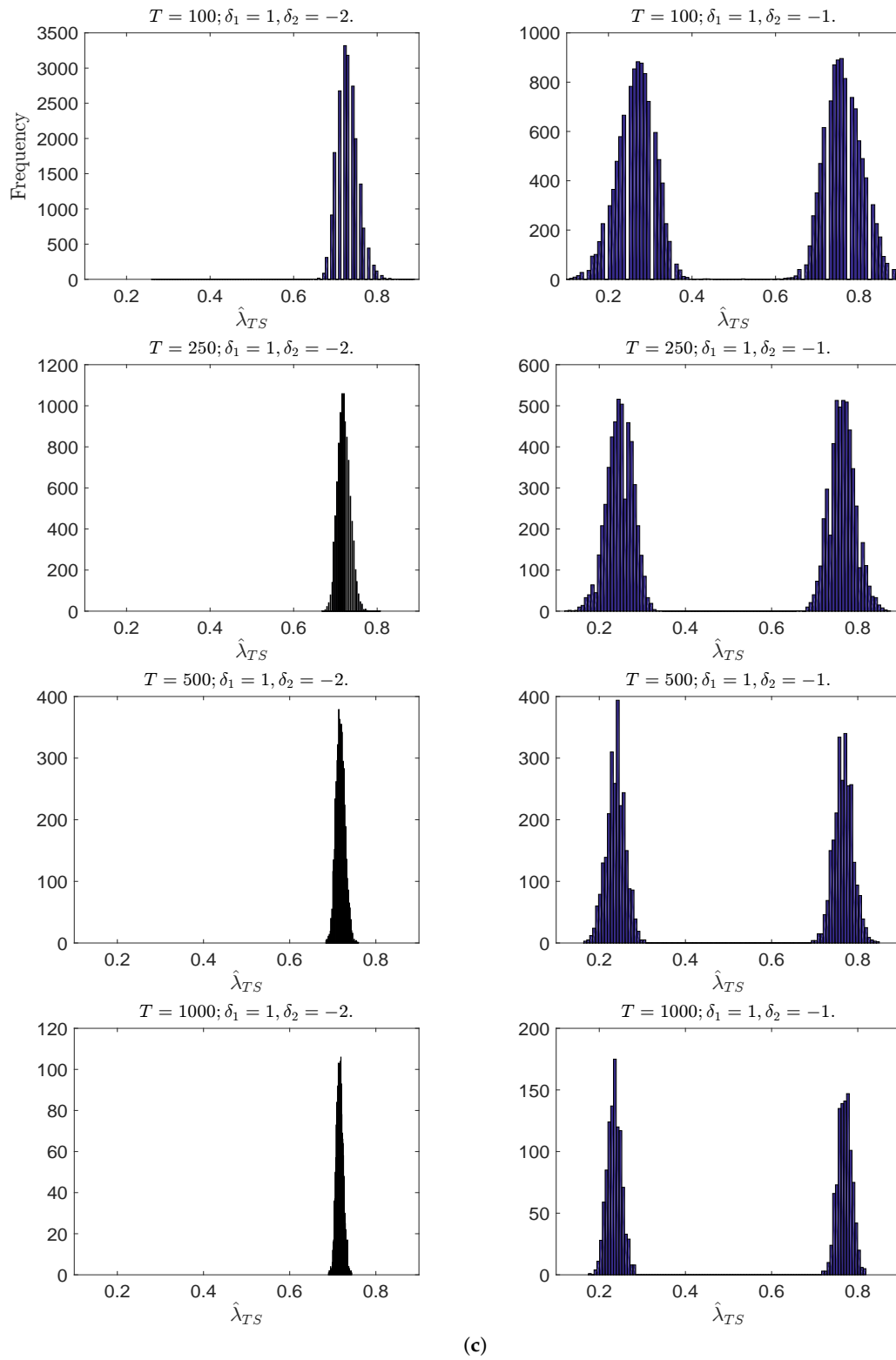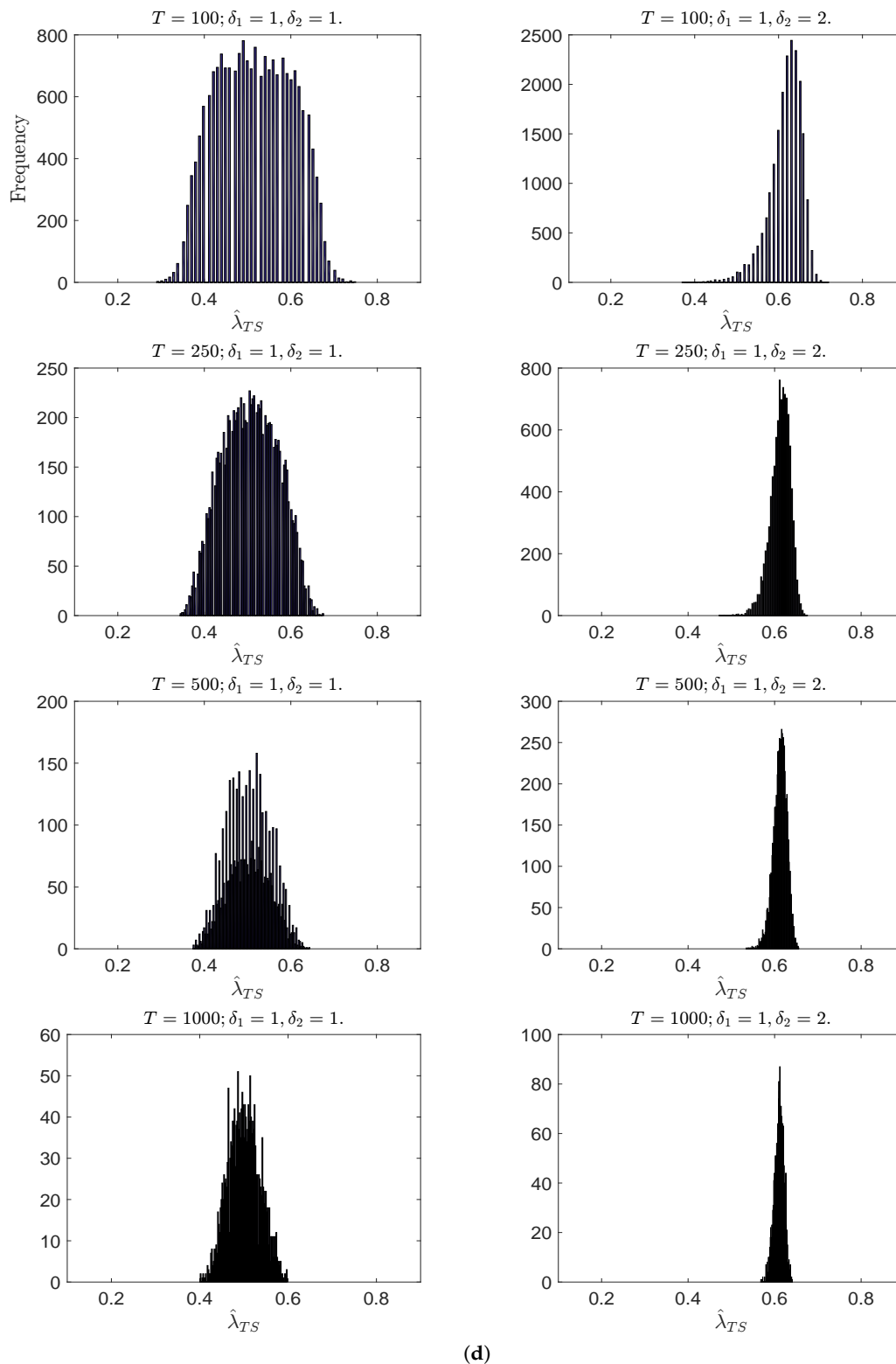**Figure 1.** Histograms of the single break point estimator $\hat{\lambda}_{MS}$ or $\hat{\lambda}_{TS}$ when $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$ and $\delta_1 = 1$ always. From top to bottom on each page: $T = 100, 250, 500, 1000$. (**a**) Histograms of $\hat{\lambda}_{MS}$ when $\nu = -2(\delta_2 = -2), -1(\delta_2 = -1)$; (**b**) Histograms of $\hat{\lambda}_{MS}$ when $\nu = 1(\delta_2 = 1), 2(\delta_2 = 2)$; (**c**) Histograms of $\hat{\lambda}_{TS}$ when $\nu = -2(\delta_2 = -2), -1(\delta_2 = -1)$; (**d**) Histograms of $\hat{\lambda}_{TS}$ when $\nu = 1(\delta_2 = 1), 2(\delta_2 = 2)$.
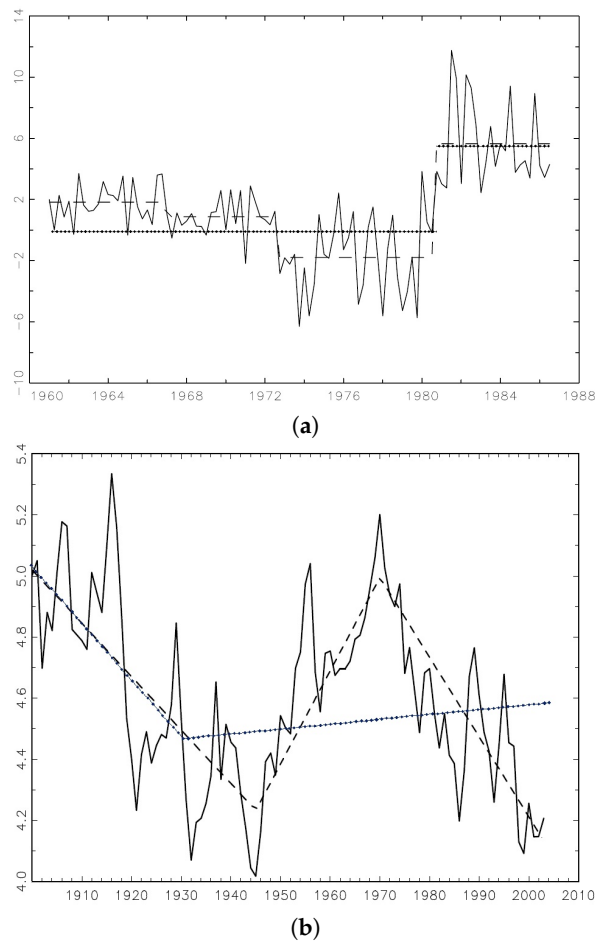
Figure 2. Single break point estimate (dotted line) while multiple mean shifts or trend shifts exist (dashed line). (**a**) US ex-post real interest rate during Q1 1961–Q3 1986; (**b**) Primary commodity price index (Copper) relative to the price of manufacture during 1900–2003.

4. Limits of the Break Point Estimators when the Break Magnitudes are Fixed

Similar to the discussion in Bai (1997) [2] for the mean shift results, the limits of the single trend break point estimator $\hat{\lambda}_{TS}$ are derived in this section when the break sizes are fixed and the data sequences have two trend breaks.

Theorem 1. *Assume there are two break fractions λ_1^c and λ_2^c with fixed break magnitudes in models (1) and (3) while the break number is underspecified as one.*

1. *For the mean shift model (1), under assumption (5) with fixed break magnitudes $\delta_1 = \delta_1^*$ and $\delta_2 = \delta_2^*$, the break point estimator $\hat{\lambda}_{MS}$ converges to one of the true breaks:*

$$(\hat{\lambda}_{MS} - \arg\max_{\lambda \in \Lambda} |G2_{MS}(\lambda, \lambda_1^c) + \nu \cdot G2_{MS}(\lambda, \lambda_2^c)|) = O_p(T^{-1/2}), \tag{8}$$

where $\nu = \delta_2^*/\delta_1^*$ and

$$G2_{MS}(\lambda, \lambda^c) \doteq \frac{\Psi(\lambda, \lambda^c)}{\sqrt{\lambda(1-\lambda)}},$$

$$\Psi(\lambda, \lambda^c) \doteq \begin{cases} (1-\lambda^c)\lambda, & \text{if } \lambda \leq \lambda^c \\ (1-\lambda)\lambda^c, & \text{if } \lambda > \lambda^c \end{cases}.$$

Essentially

$$\left(\hat{\lambda}_{MS} - \lambda_1^c\right) = O_p(T^{-1/2}), \quad if \quad \frac{\lambda_1^c}{\lambda_2^c}\delta_1^2 \geq \frac{(1-\lambda_2^c)}{(1-\lambda_1^c)}\delta_2^2; \quad \left(\hat{\lambda}_{MS} - \lambda_2^c\right) = O_p(T^{-1/2}), \quad otherwise. \quad (9)$$

2. *For the trend shift model (3), under assumption (5) with fixed break magnitudes $\delta_1 = \delta_1^*$ and $\delta_2 = \delta_2^*$, the break point estimator $\hat{\lambda}_{TS}$ has the following limit:*

$$\left(\hat{\lambda}_{TS} - \arg\max_{\lambda\in\Lambda}|G2_{TS}(\lambda,\lambda_1^c) + \nu \cdot G2_{TS}(\lambda,\lambda_2^c)|\right) = O_p(T^{-3/2}), \quad (10)$$

where $\nu = \delta_2^*/\delta_1^*$ and

$$G2_{TS}(\lambda,\lambda^c) \doteq \frac{\int_0^1 F(r,\lambda)F(r,\lambda^c)dr}{\sqrt{\int_0^1 F(r,\lambda)^2 dr}},$$

$$F(r,\lambda) \doteq \begin{cases} \lambda^3 - 2\lambda^2 + \lambda - (2\lambda^3 - 3\lambda^2 + 1)r, & if \ r \leq \lambda \\ \lambda^3 - 2\lambda^2 - (2\lambda^3 - 3\lambda^2)r, & if \ r > \lambda \end{cases}.$$

The limit of $\hat{\lambda}_{MS}$ is either λ_1^c or λ_2^c as shown in Figure 3, which is consistent with the results in Bai (1997) [2] using a different theoretical framework. Not surprisingly, $G2_{MS}(\lambda,\lambda_i^c)$ is maximized at λ_i^c and $\hat{\lambda}_{MS}$ converges to one of the true break points.
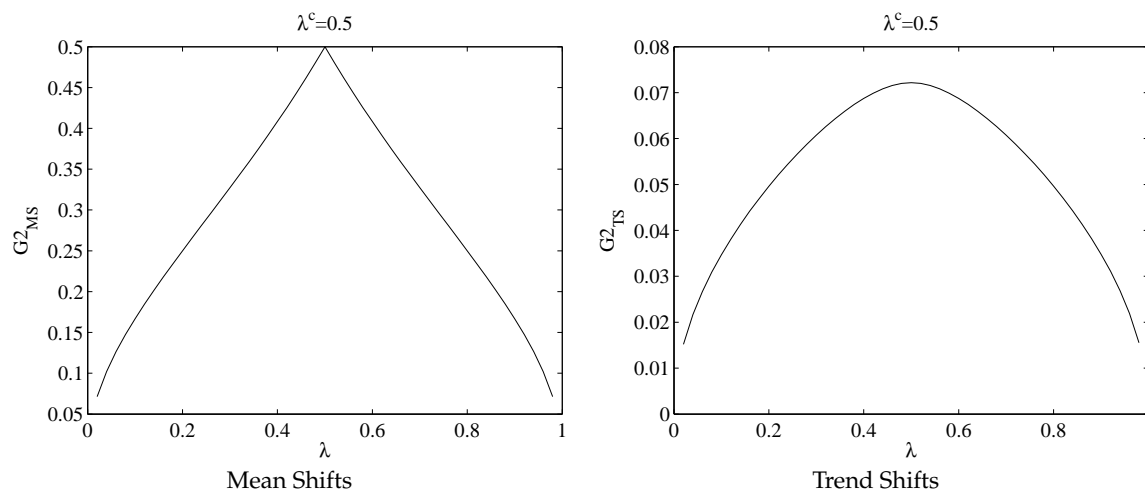

Figure 3. $G2_{MS}(\lambda,\lambda^c)$ and $G2_{TS}(\lambda,\lambda^c)$ with $\lambda^c = 0.5$.

The limit of $\hat{\lambda}_{TS}$ has different patterns. It is still true that $G2_{TS}(\lambda,\lambda_i^c)$ achieves a maximum at $\lambda = \lambda_i^c$ as shown in Figure 3. What makes it different from the mean shift case is when we sum up the two $G2_{TS}$ terms, the function smooths out through the two peaks at each λ_i^c. Hence, when the number of trend breaks is two while assumed to be one, $|G2_{TS}(\lambda,\lambda_1^c) + \nu \cdot G2_{TS}(\lambda,\lambda_2^c)|$ peaks neither at λ_1^c nor at λ_2^c. Figure 4 plots $|G2(\lambda,\lambda_1^c) + \nu \cdot G2(\lambda,\lambda_2^c)|$ with $\nu = \pm 1$ and $\lambda_1^c = 1/4$ and $\lambda_2^c = 3/4$. In both cases $|G2(\lambda,\lambda_1^c) + \nu \cdot G2(\lambda,\lambda_2^c)|$ peaks at neither of the true break points. Certainly, if $|\nu|$ is smaller than 1, $\hat{\lambda}_{TS}$ will be closer to λ_1^c; and if $|\nu|$ is bigger than 1, $\hat{\lambda}_{TS}$ will be closer to λ_2^c. This clearly explains the reason for the inconsistency of the trend shift break point estimator when the break number is underspecified.
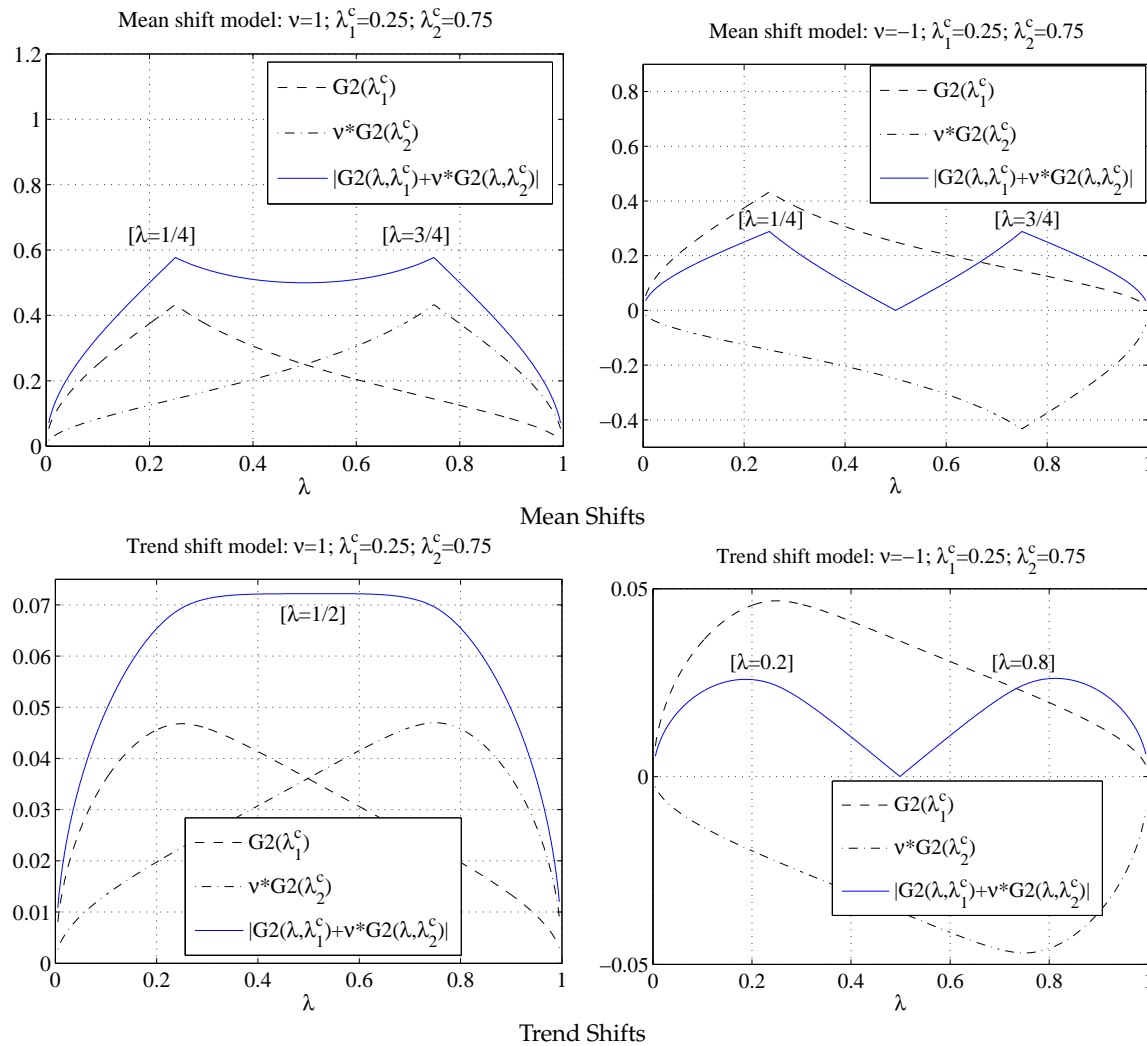
Figure 4. $|G2_{MS}(\lambda, \lambda_1^c) + \nu \cdot G2_{MS}(\lambda, \lambda_2^c)|$ for mean shift model and $|G2_{TS}(\lambda, \lambda_1^c) + \nu \cdot G2_{TS}(\lambda, \lambda_2^c)|$ for trend shift model, where $\nu = 1$ and -1, $\{\lambda_1^c, \lambda_2^c\} = \{1/4, 3/4\}$.

Figure 5 plots the λ's where $|G2_{TS}(\lambda, \lambda_1^c) + \nu \cdot G2_{TS}(\lambda, \lambda_2^c)|$ is maximized along ν when $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$ and $\{1/4, 3/4\}$. When $\nu = 0$, $|G2_{TS}(\lambda, \lambda_1^c) + \nu \cdot G2_{TS}(\lambda, \lambda_2^c)|$ is maximized at λ_1^c. When $|\nu|$ goes to ∞, the limit of the break point estimator will be the true break λ_2^c. Other than these practically uninteresting cases, the limits of $\arg\max\{|G2_{TS}(\lambda, \lambda_1^c) + \nu \cdot G2_{TS}(\lambda, \lambda_2^c)|\}$ will not be the true break points. Take $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$ as an example. When $\nu < -1$, the limiting point is greater than $2/3$. When $-1 < \nu < 0$, the limiting point is less than $1/3$. In both cases, the limiting points are beyond the range of the two true breaks. When $\nu > 0$, the limiting points are between the true breaks. When $\nu = 1$, the limiting point is at $\lambda = 0.5$, the trend shift break point estimator is far away from the true breaks. As ν goes away from 1, the limit of the trend shift break point estimator gets closer to one of the true breaks. The limits tell us the magnitude of the discrepancy between the spurious break and true breaks. Numerically when $|\nu| > 4.3$ or $|\nu| < \frac{1}{4.3}$, the limits of the spurious break point will be between $\pm 2.5\%$ of the true breaks. This threshold can be extended to other cases with different break locations.
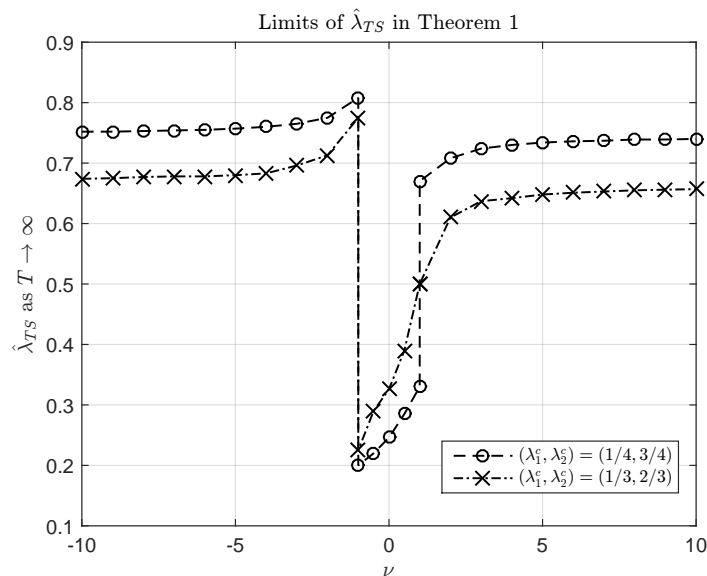
Figure 5. Limits of $\hat{\lambda}_{TS}$ in Theorem 1 when $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$ and $\{1/4, 3/4\}$ along $\nu = -10, \cdots, 10$.

We summarize the findings on the consistency/inconsistency of $\hat{\lambda}_{MS}$ and $\hat{\lambda}_{TS}$ under assumption (5) as follows:

1. For the mean shift model with two breaks, if the break magnitudes are not zero, the single break point estimator $\hat{\lambda}_{MS}$ is consistent for either λ_1 or λ_2:

$$\lim_{T \to \infty} \hat{\lambda}_{MS} \to \lambda_1^c \ \text{or} \ \lambda_2^c.$$

2. For the trend shift model[1] with two breaks, if the break magnitudes are not zero, the single break point estimator $\hat{\lambda}_{TS}$ is inconsistent for either λ_1 or λ_2:

$$\lim_{T \to \infty} \hat{\lambda}_{TS} \not\to \lambda_1^c \ \text{and} \ \lim_{T \to \infty} \hat{\lambda}_{TS} \not\to \lambda_2^c.$$

The limit depends on λ_1^c, λ_2^c, and ν:

$$\lim_{T \to \infty} \hat{\lambda}_{TS} = \arg \max_{\lambda \in \Lambda} |G2_{TS}(\lambda, \lambda_1^c) + \nu \cdot G2_{TS}(\lambda, \lambda_2^c)|.$$

5. Limiting Distributions of $\hat{\lambda}_{MS}$ and $\hat{\lambda}_{TS}$ by Employing Pitman Drifts

As shown in the literature, asymptotic results derived under Pitman drifts often closely approximate the finite sample behavior of the test statistics or estimators involved. In the following, the limiting distributions of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ are developed under Pitman drifts.

Theorem 2. *Assume there are two break points λ_1^c and λ_2^c in the linear model while the break number is underspecified as one.*

---

[1] If DU_t's are included together with DT_t's in model (3), under the condition of fixed break magnitudes, the trend shifts will dominate the mean shifts in the (in)consistency of the break point estimator, following the results in Theorem 1. If $[t \cdot DU_t]$'s are included in model (3), the slope change will force a large level shift. Under this condition, the consistency property of mean shifts will be dominant and the inconsistency problem in break point estimator will not persist anymore.

1.　*For the mean shift model (1), under assumptions (5) and $\delta_1 = T^{-1/2}\delta_1^*$ and $\delta_2 = T^{-1/2}\delta_2^*$, where δ_1^* and δ_2^* are constant scalars, the break point estimator $\hat{\lambda}_{MS}$ has the following limiting distribution:*

$$\hat{\lambda}_{MS} \Rightarrow \arg\max_{\lambda \in \Lambda}\left\{\frac{[(\lambda W(1) - W(\lambda)) + M_1\Psi(\lambda, \lambda_1^c) + M_2\Psi(\lambda, \lambda_2^c)]^2}{\lambda(1-\lambda)}\right\}, \tag{11}$$

where $M_1 \doteq \frac{\delta_1^*}{d(1)}$, $M_2 \doteq \frac{\delta_2^*}{d(1)}$, and

$$\Psi(\lambda, \lambda^c) \doteq \begin{cases} (1-\lambda^c)\lambda, & \text{if } \lambda \leq \lambda^c \\ (1-\lambda)\lambda^c, & \text{if } \lambda > \lambda^c \end{cases}.$$

2.　*For the trend shift model (3), under assumptions (5) and $\delta_1 = T^{-3/2}\delta_1^*$ and $\delta_2 = T^{-3/2}\delta_2^*$, where δ_1^* and δ_2^* are constant scalars, the break point estimator $\hat{\lambda}_{TS}$ has the following limiting distributions:*

$$\hat{\lambda}_{TS} \Rightarrow \quad \arg\max_{\lambda \in \Lambda}\left\{\left[\int_0^1 F(r,\lambda)dW(r) + M_1\int_0^1 F(r,\lambda)F(r,\lambda_1^c)dr + \right.\right.$$
$$\left.\left. M_2\int_0^1 F(r,\lambda)F(r,\lambda_2^c)dr\right]^2 \int_0^1 F(r,\lambda)^2 dr\right\}, \tag{12}$$

where $M_1 \doteq \frac{\delta_1^*}{d(1)} \equiv \frac{\delta_1 T^{3/2}}{d(1)}$, $M_2 \doteq \frac{\delta_2^*}{d(1)} \equiv \frac{\delta_2 T^{3/2}}{d(1)}$, $\nu = M_2/M_1 \equiv \delta_2/\delta_1$, and $F(r,\lambda)$ is defined in Theorem 1.

The asymptotics in Theorem 2 are an extension of work by Yang (2012) [10] from the single-break case to the multiple-break case. To understand the effect of M_1, λ_1^c, M_2, and λ_2^c on the limiting distributions, I decompose the part inside the arg min in Equations (11) and (12) into three parts, where

$$\begin{aligned} G_{MS}(\lambda, \lambda_1^c, \lambda_2^c) &\doteq \quad G1_{MS}(\lambda) + M_1 \cdot G2_{MS}(\lambda, \lambda_1^c) + M_2 \cdot G2_{MS}(\lambda, \lambda_2^c) \\ &\doteq \quad \frac{(\lambda W(1) - W(\lambda))}{\sqrt{\lambda(1-\lambda)}} + M_1 \cdot \frac{\Psi(\lambda, \lambda_1^c)}{\sqrt{\lambda(1-\lambda)}} + M_2 \cdot \frac{\Psi(\lambda, \lambda_2^c)}{\sqrt{\lambda(1-\lambda)}}; \end{aligned} \tag{13}$$

$$\begin{aligned} G_{TS}(\lambda, \lambda^c) &\doteq \quad G1_{TS}(\lambda) + M_1 \cdot G2_{TS}(\lambda, \lambda_1^c) + M_2 \cdot G2_{TS}(\lambda, \lambda_2^c) \\ &\doteq \quad \frac{\int_0^1 F(r,\lambda)dW(r)}{\sqrt{\int_0^1 F(r,\lambda)^2 dr}} + \\ &\qquad M_1 \cdot \frac{\int_0^1 F(r,\lambda)F(r,\lambda_1^c)dr}{\sqrt{\int_0^1 F(r,\lambda)^2 dr}} + M_2 \cdot \frac{\int_0^1 F(r,\lambda)F(r,\lambda_2^c)dr}{\sqrt{\int_0^1 F(r,\lambda)^2 dr}}. \end{aligned} \tag{14}$$

For the asymptotic distribution of $\hat{\lambda}_{MS}$, with the form of $G1_{MS}(\lambda) + M_1 \cdot G2_{MS}(\lambda, \lambda_1^c) + M_2 \cdot G2_{MS}(\lambda, \lambda_2^c)$ in the limiting distributions, Theorem 2 provides a bridge between the asymptotics under the null of no breaks and the asymptotics under local alternatives of up to two breaks.

The asymptotics are continuous at $\{M_1, M_2\} = \{0, 0\}$, i.e., M_1 and M_2 could be as small as possible in the asymptotics. When M_1 and M_2 are small, the random component $G1_{MS}$ dominates G_{MS} and the distribution is close to the case of no breaks. For a small M, $\hat{\lambda}_{TS}$ concentrates more around the middle range exhibiting a bell shape, while $\hat{\lambda}_{MS}$ concentrates more around the boundaries exhibiting a U shape. The detailed explanation is given in Yang (2012) [10]. For a moderate M, the limiting

distribution of $\hat{\lambda}_{MS}$ exhibits a shape of W, resulting from the mixed effects of $G1_{MS}$ and $G2_{MS}$ in the asymptotics. If $T \to \infty$, both M_1 and M_2 increase to ∞,

$$\lim_{T \to \infty} \hat{\lambda}_{MS,TS} \quad \to \quad \arg\max_{\lambda \in \Lambda} |G2_{MS,TS}(\lambda, \lambda_1^c) + \nu \cdot G2_{MS,TS}(\lambda, \lambda_2^c)|.$$

The limiting distributions in Theorem 2 are nonstandard. M_1, M_2, λ_1^c and λ_2^c show up in the approximations, and capture the effects of M's and λ^c's on the asymptotics. Besides other deterministic variables in Theorem 2, the main random variables in the asymptotic distributions are functions of a Wiener process. The Wiener process in the asymptotic distributions was approximated by using standard normal *i.i.d.* random deviates. Integrals were approximated by normalized partial sums of 1000 steps using 10,000 replications.

Figure 6 plots the finite sample distributions of $\hat{\lambda}_{MS}$ and $\hat{\lambda}_{TS}$ with $T = 100$ and asymptotic distributions for $\mu = \beta = 0$, $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$. Errors are *i.i.d.* $N(0, 1)$. The left panels of Figure 6 are for $\hat{\lambda}_{MS}$ and the right panels are for $\hat{\lambda}_{TS}$. From the top to the bottom are the cases of $\{\delta_1, \delta_2\} = \{1, 1\}$, $\{5, 5\}$, $\{1, -1\}$, and $\{5, -5\}$. The pdfs of $\hat{\lambda}_{MS}$ and $\hat{\lambda}_{TS}$ are plotted in separated figures with the same scales to show the performance comparison in the presence of an underspecified break number. Kernel smoothing is used to obtain the pdf based on the simulations. Figure 6 compares the asymptotic limits given by Theorem 2 to finite sample distributions. The two lines in each panel are near-identical, which shows that the asymptotics does a good job of approximating finite sample distributions of the break point estimators.
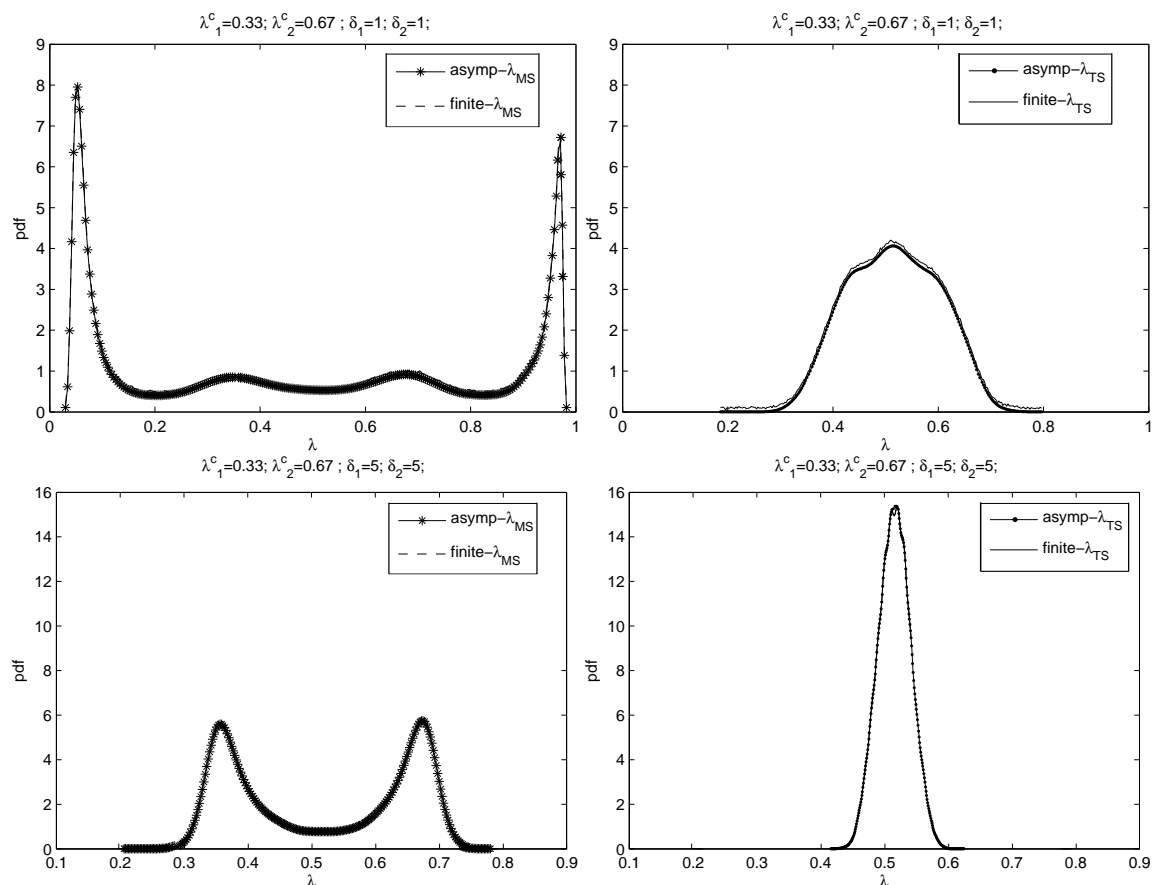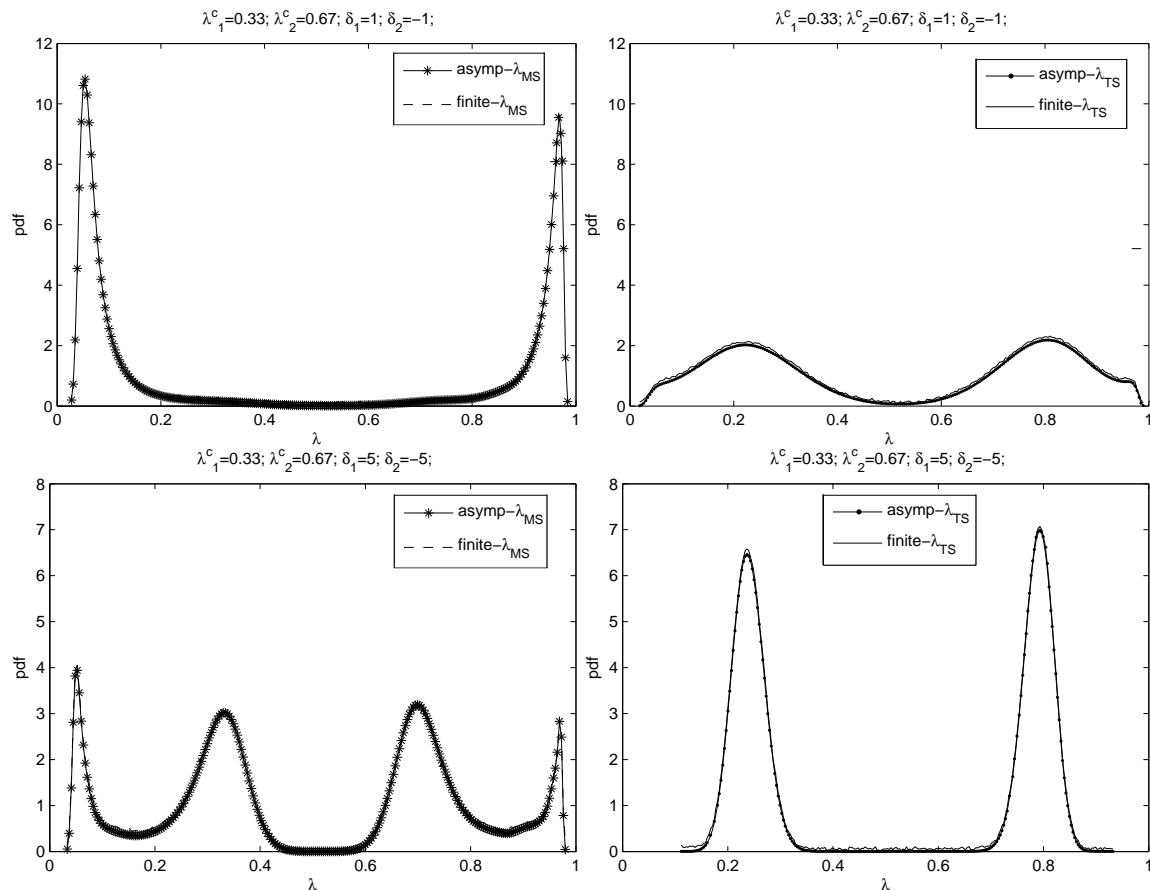

Figure 6. *Cont.*

Figure 6. The asymptotic distributions and the finite sample distributions of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ when $T = 100$, $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$, and u_t is an *i.i.d.* $N(0,1)$ process. The left: the distributions of $\hat{\lambda}_{MS}$; the right: the distributions of $\hat{\lambda}_{TS}$. From the top to the bottom: $\{\delta_1, \delta_2\} = \{1,1\}$, $\{5,5\}$, $\{1,-1\}$, and $\{5,-5\}$.

6. Conclusions

This paper analyzes the consistency of trend shift break point estimators when the number of breaks is underspecifed. The limit of the trend shift break point estimator for fixed break sizes is shown to be dependent on the break magnitudes and locations. In general, the trend shift break point estimator does not consistently estimate one of the true break points. Using the Pitman drift assumption, the limiting distribution of the trend shift break point estimator is shown to closely resemble the finite sample distributions.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A.

Appendix A.1. Proof of Theorem 1

Theorem 1 can be proved simply by following the steps provided in the proof of Theorem 2 below.

Appendix A.2. Proof of Theorem 2

Appendix A.2.1. Asymptotic Distribution of $\hat{\lambda}_{MS}$

Let SSR^0_{MS} be the SSR under the assumption of no breaks. Following Equation (6), we obtain

$$SSR^0_{MS} - SSR_{MS}(\lambda) = [\sum_{t=1}^{T} \tilde{DU}_t(\lambda)^2]\hat{\delta}^2_{MS}.$$

The OLS estimator of δ from (2) is given by

$$\hat{\delta}_{MS} = [\sum_{t=1}^{T} \tilde{DU}_t^2(\lambda)]^{-1} \sum_{t=1}^{T} [\tilde{DU}_t(\lambda)\tilde{y}_t],$$

where $\{\tilde{DU}_t(\lambda)\}$ and $\{\tilde{y}_t\}$ are the residuals from the OLS regressions of $\{DU_t(\lambda)\}$ and $\{y_t\}$ on $[1]'$,

$$\tilde{DU}_t(\lambda) \quad = \quad DU_t(\lambda) - \sum_{t=1}^{T} DU_t/T = DU_t(\lambda) - \bar{DU}(\lambda).$$

When the DGP is given by (1), simple algebra gives

$$
\begin{aligned}
\hat{\delta}_{MS} &= [\sum_{t=1}^{T} \tilde{DU}_t^2(\lambda)]^{-1} \sum_{t=1}^{T} \tilde{DU}_t(\lambda)[\tilde{DU}_t(\lambda_1^c)\delta_1 + \tilde{DU}_t(\lambda_2^c)\delta_2 + u_t] \\
&= [\sum_{t=1}^{T} \tilde{DU}_t^2(\lambda)]^{-1} \sum_{t=1}^{T} \tilde{DU}_t(\lambda)[\tilde{DU}_t(\lambda_1^c)\delta_1 + \tilde{DU}_t(\lambda_2^c)\delta_2] \\
&\quad + [\sum_{t=1}^{T} \tilde{DU}_t^2(\lambda)]^{-1} \sum_{t=1}^{T} \tilde{DU}_t(\lambda)u_t.
\end{aligned}
$$

Multiplying both sides of the above equation by $T^{1/2}$, we have

$$
\begin{aligned}
T^{1/2}\hat{\delta}_{MS} &= [T^{-1}\sum_{t=1}^{T} \tilde{DU}_t^2(\lambda)]^{-1}[T^{-1}\sum_{t=1}^{T} \tilde{DU}_t(\lambda)(\tilde{DU}_t(\lambda_1^c)\delta_1^* + \tilde{DU}_t(\lambda_2^c)\delta_2^*)] + \\
&\quad [T^{-1}\sum_{t=1}^{T} \tilde{DU}_t^2(\lambda)]^{-1}[T^{-1/2}\sum_{t=1}^{T} \tilde{DU}_t(\lambda)u_t].
\end{aligned}
$$

Using

$$[T^{-1}\sum_{t=1}^{T} \tilde{DU}_t^2(\lambda)] \Rightarrow \int_0^1 [I(r > \lambda) - (1-\lambda)]^2 dr = \lambda(1-\lambda),$$

and

$$
\begin{aligned}
[T^{-1}\sum_{t=1}^{T} \tilde{DU}_t(\lambda)\tilde{DU}_t(\lambda^c)] \quad &\Rightarrow \quad \int_0^1 [I(r > \lambda) - (1-\lambda)][I(r > \lambda^c) - (1-\lambda^c)]dr \\
&= \begin{cases} (1-\lambda^c)\lambda, & \text{if } \lambda \le \lambda^c \\ (1-\lambda)\lambda^c, & \text{if } \lambda > \lambda^c \end{cases},
\end{aligned}
$$

and

$$T^{-1/2}[\sum_{t=1}^{T} \tilde{DU}_t(\lambda)u_t] \quad \Rightarrow \quad d(1)\int_0^1 [I(r > \lambda) - (1-\lambda)]dW(r)$$
$$= \quad d(1)[\lambda W(1) - W(\lambda)],$$

we obtain

$$T^{1/2}\hat{\delta}_{MS} \quad \Rightarrow \quad \frac{\delta_1^*}{\lambda(1-\lambda)}\Psi(\lambda, \lambda_1^c) + \frac{\delta_2^*}{\lambda(1-\lambda)}\Psi(\lambda, \lambda_2^c) + \frac{d(1)}{\lambda(1-\lambda)}[\lambda W(1) - W(\lambda)],$$

where

$$\Psi(\lambda, \lambda^c) = \begin{cases} (1-\lambda^c)\lambda, & \text{if } \lambda \le \lambda^c \\ (1-\lambda)\lambda^c, & \text{if } \lambda > \lambda^c \end{cases} .$$

From this result, it immediately follows that

$$SSR_{MS}^0 - SSR_{MS}(\lambda)$$
$$\Rightarrow \quad \frac{1}{\sqrt{\lambda(1-\lambda)}}[d(1)(\lambda W(1) - W(\lambda)) + \delta_1^*\Psi(\lambda, \lambda_1^c) + \delta_2^*\Psi(\lambda, \lambda_2^c)]^2.$$

Applying the CMT theorem gives

$$\hat{\lambda}_{MS} \quad = \quad \arg\max_{\lambda \in \Lambda}\{SSR_{MS}^0 - SSR_{MS}(\lambda))$$
$$\Rightarrow \quad \arg\max_{\lambda \in \Lambda}\{\frac{[(\lambda W(1) - W(\lambda)) + M_1\Psi(\lambda, \lambda_1^c) + M_2\Psi(\lambda, \lambda_2^c)]^2}{\lambda(1-\lambda)}\},$$

where $M_1 = \frac{\delta_1^*}{d(1)}$ and $M_2 = \frac{\delta_2^*}{d(1)}$.

It is straightforward to show that $M_1G2(\lambda, \lambda_1^c) + M_2G2(\lambda, \lambda_2^c)$ is maximized at either λ_1^c or λ_2^c. The first derivative of $G2_{MS}$ w.r.t. λ is given by:

$$G2'_{MS}(\lambda, \lambda^c) = \begin{cases} \frac{(1-\lambda^c)\lambda}{2(1-\lambda)\sqrt{\lambda(1-\lambda)}}, & \text{if } \lambda \le \lambda^c \\ \frac{(1-\lambda)\lambda^c}{2(1-\lambda)\sqrt{\lambda(1-\lambda)}}, & \text{if } \lambda > \lambda^c \end{cases} .$$

Assume $\lambda_1^c < \lambda_2^c$, then it follows that

$$[M_1G2(\lambda, \lambda_1^c) + M_2G2(\lambda, \lambda_2^c)]' = M1\frac{(1-\lambda_1^c)\lambda}{2(1-\lambda)\sqrt{\lambda(1-\lambda)}} + M_2\frac{(1-\lambda_2^c)\lambda}{2(1-\lambda)\sqrt{\lambda(1-\lambda)}},$$
$$\text{when } \lambda \le \lambda_1^c;$$

$$[M_1G2(\lambda, \lambda_1^c) + M_2G2(\lambda, \lambda_2^c)]' = M_1\frac{(1-\lambda)\lambda_1^c}{2(1-\lambda)\sqrt{\lambda(1-\lambda)}} + M_2\frac{(1-\lambda_2^c)\lambda}{2(1-\lambda)\sqrt{\lambda(1-\lambda)}},$$
$$\text{when } \lambda_1^c \le \lambda \le \lambda_2^c;$$

and

$$[M_1G2(\lambda, \lambda_1^c) + M_2G2(\lambda, \lambda_2^c)]' = M_1\frac{(1-\lambda)\lambda_1^c}{2(1-\lambda)\sqrt{\lambda(1-\lambda)}} + M_2\frac{(1-\lambda)\lambda_2^c}{2(1-\lambda)\sqrt{\lambda(1-\lambda)}},$$
$$\text{when } \lambda \ge \lambda_2^c.$$

Through simple algebra, one can show that the peak values of $[M_1 G2(\lambda, \lambda_1^c) + M_2 G2(\lambda, \lambda_2^c)]$ will be obtained at either λ_1^c or λ_2^c.

Appendix A.2.2. Asymptotic Distribution of $\hat{\lambda}_{TS}$

Let SSR_{TS}^0 be the SSR under the assumption of no breaks. From Equation (7), we have the standard result that

$$SSR_{TS}^0 - SSR_{TS}(\lambda) = [\sum_{t=1}^{T} \tilde{DT}_t(\lambda)\tilde{DT}_t(\lambda)]\hat{\delta}_{TS}^2,$$

where

$$\hat{\delta}_{TS} = \sum_{t=1}^{T} \tilde{DT}_t(\lambda)\tilde{DT}_t(\lambda)]^{-1} \sum_{t=1}^{T} \tilde{DT}_t(\lambda)\tilde{y}_t.$$

$\{\tilde{DT}_t(\lambda)\}$ and $\{\tilde{y}_t\}$ are the residuals from the OLS regressions of $\{DT_t(\lambda)\}$ and $\{y_t\}$ on $[1 \ \ t]'$.

When the DGP is given by (3), simple algebra gives

$$T^{3/2}\hat{\delta}_{TS}$$

$$= \left[T^{-1} \sum_{t=1}^{T} T^{-1}\tilde{DT}_t(\lambda)\tilde{DT}_t(\lambda)T^{-1} \right]^{-1} \left[T^{-1} \sum_{t=1}^{T} T^{-1}\tilde{DT}_t(\lambda)\tilde{DT}_t(\lambda_1^c)T^{-1}(T^{3/2}\delta_1) + \right.$$

$$T^{-1} \sum_{t=1}^{T} T^{-1}\tilde{DT}_t(\lambda)\tilde{DT}_t(\lambda_2^c)T^{-1}(T^{3/2}\delta_2) \right] +$$

$$\left[T^{-1} \sum_{t=1}^{T} T^{-1}\tilde{DT}_t(\lambda)\tilde{DT}_t(\lambda)T^{-1} \right]^{-1} \left[T^{-1/2} \sum_{t=1}^{T} T^{-1}\tilde{DT}_t(\lambda)u_t \right].$$

Using

$$T^{-1}\tilde{DT}_t(\lambda) \Rightarrow F(r,\lambda) \doteq (r - \lambda)1(r > \lambda) + (\lambda^3 - 2\lambda^2 + \lambda) - (2\lambda^3 - 3\lambda^2 + 1)r,$$

and

$$T^{-1} \sum_{t=1}^{T} T^{-1}\tilde{DT}_t(\lambda)\tilde{DT}_t(\lambda^c)T^{-1} \Rightarrow \int_0^1 F(r,\lambda)F(r,\lambda^c)dr,$$

and

$$T^{-1/2} \sum_{t=1}^{T} T^{-1}\tilde{DT}_t(\lambda)u_t \Rightarrow d(1) \int_0^1 F(r,\lambda)dW(r),$$

we obtain

$$T^{3/2}\hat{\delta}_{TS}$$

$$\Rightarrow [\int_0^1 F(r,\lambda)^2 dr]^{-1}[\delta_1^* \int_0^1 F(r,\lambda)F(r,\lambda_1^c)dr + \delta_2^* \int_0^1 F(r,\lambda)F(r,\lambda_2^c)dr] +$$

$$[\int_0^1 F(r,\lambda)^2 dr]^{-1}[d(1) \int_0^1 F(r,\lambda)dW(r)]$$

$$= \frac{[\delta_1^* \int_0^1 F(r,\lambda)F(r,\lambda_1^c)dr + \delta_2^* \int_0^1 F(r,\lambda)F(r,\lambda_2^c)dr] + d(1) \int_0^1 F(r,\lambda)dW(r)}{\int_0^1 F(r,\lambda)^2 dr}.$$

From this results, it immediately follows that

$$[SSR_{TS}^0 - SSR_{TS}(\lambda)]$$

$$= [T^{3/2}\hat{\delta}_{TS}]^2 [T^{-1}\sum_{t=1}^{T} T^{-1}\tilde{DT_t}(\lambda)\tilde{DT_t}(\lambda)T^{-1}]$$

$$\Rightarrow \frac{\left[d(1)\int_0^1 F(r,\lambda)dW(r) + \delta_1^*\int_0^1 F(r,\lambda)F(r,\lambda_1^c)dr + \delta_2^*\int_0^1 F(r,\lambda)F(r,\lambda_2^c)dr\right]^2}{\int_0^1 F(r,\lambda)^2 dr}.$$

Furthermore, using the CMT we obtain the limit of the break point estimator as

$$\hat{\lambda}_{TS}$$

$$= \arg\max_{\lambda\in\Lambda}\{SSR_{TS}^0 - SSR_{TS}(\lambda)\}$$

$$\Rightarrow \arg\max_{\lambda\in\Lambda}\Bigg\{\Bigg[\frac{\int_0^1 F(r,\lambda)dW(r)}{\sqrt{\int_0^1 F(r,\lambda)^2 dr}} +$$

$$\frac{M_1\int_0^1 F(r,\lambda)F(r,\lambda_1^c)dr + M_2\int_0^1 F(r,\lambda)F(r,\lambda_2^c)dr}{\sqrt{\int_0^1 F(r,\lambda)^2 dr}}\Bigg]^2\Bigg\},$$

where $M_1 \doteq \frac{\delta^*}{d(1)} \equiv \frac{\delta_1 T^{3/2}}{d(1)}$ and $M_2 \doteq \frac{\delta^*}{d(1)} \equiv \frac{\delta_2 T^{3/2}}{d(1)}$.

Please refer to Yang (2012) [10] for more details about $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$.

References

1. Bai, J.S. Least absolute deviation estimation of a shift. *Econom. Theory* **1995**, *11*, 403–436.
2. Bai, J.S. Estimating multiple breaks one at a time. *Econom. Theory* **1997**, *13*, 315–352.
3. Chong, T. *Consistency of Change-Point Estimators When the Number of Change Points in Structural Change Models is Underspecified*; Working Paper; Department of Economics, University of Rochester: Rochester, NY, USA, 1994.
4. Chong, T. Partial parameter consistency in a misspecified structural change model. *Econ. Lett.* **1995**, *49*, 351–357.
5. Bai, J.S.; Perron, P. Estimating and testing linear models with multiple structural breaks. *Econometrica* **1998**, *66*, 47–78.
6. Bai, J.S.; Perron, P. Computation and analysis of multiple structural change models. *J. Appl. Econom.* **2003**, *18*, 1–22.
7. Kejriwal, M.; Perron, P. A sequential procedure to determine the number of breaks in trend with an integrated or stationary noise component. *J. Time Ser. Anal.* **2010**, *31*, 305–328.
8. Perron, P.; Yabu, T. Estimating deterministic trends with an integrated or stationary noise component. *J. Econom.* **2009**, *151*, 56–69.
9. Perron, P.; Yabu, T. Testing for shifts in trend with an integrated or stationary noise component. *J. Bus. Econ. Stat.* **2009**, *27*, 369–396.
10. Yang, J. Break point estimators for a slope shift: levels versus first differences. *Econom. J.* **2012**, *15*, 154–169.
11. Yang, J. Essays on Estimation and Inference in Models with Deterministic Trends with and without Structural Change. Ph.D. Thesis, Department of Economics, Michigan State University, East Lansing, MI, USA, 2010.
12. Harvey, D.I.; Leybourne, S.J.; Taylor, A.M.R. Simple, robust and powerful tests of the breaking trend hypothesis. *Econom. Theory* **2009**, *25*, 995–1029.