ESSAYS ON ESTIMATION AND INFERENCE IN MODELS WITH DETERMINISTIC TRENDS WITH AND WITHOUT STRUCTURAL CHANGE

By

Jingjing Yang

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Economics

2010

ABSTRACT

ESSAYS ON ESTIMATION AND INFERENCE IN MODELS WITH DETERMINISTIC TRENDS WITH AND WITHOUT STRUCTURAL CHANGE

By

Jingjing Yang

Empirical macroeconomists who analyze typical time series, such as GDP, interest rates, stock returns have to worry about the structural change. Possibilities of structural change over time and the properties of structural change parameters are the focus of my dissertation. It includes the choice of break point estimators between using level shift model and trend shift model, break tests robust to I(0)/I(1) errors, and the estimation of break numbers.

I. **Break Point Estimates for a Shift in Trend: Levels versus First Differences**

In the first chapter I analyze the estimation of an unknown break point in a univariate trend shift model under I(1) errors by minimizing the sum of squared residuals. Two break point estimators are considered, one is from the original trend shift model and the other is from its first difference, a mean shift model with I(0) errors. Simulations show a discrepancy between existing asymptotic theories and finite sample distributions of the break point estimators. To achieve a closer approximation, I derive an asymptotic theory for the break point estimators assuming the break magnitude is within a $T^{-1/2}$ neighborhood of zero.

The break point estimator from the trend break model converges to its true value at rate $T^{1/2}$ under I(1) errors, while the break point estimator of the first difference model converges at rate T under I(0) errors. Given this fact, many researchers would think they should use the estimator that converges to the faster rate. However I show that when the break magnitude is small relative to the noise magnitude, the break point estimator from

the trend shift model may have thinner tails and concentrates more on the true break point than that from the first difference transformation. That indicates a preference of the break point estimator from the level model.

II. **Fixed-b Analysis of LM Type Tests for a Shift in Mean**

We analyze lagrange multiplier (LM) tests for a shift in mean of a univariate time series at an unknown date. We consider a class of LM statistics based on nonparametric kernel estimates of the long run variance and we develop a fixed-b asymptotic theory for the statistics. We provide results for the case of $I(0)$ and $I(1)$ errors and use the fixed-b theory to explain finite sample null rejection probabilities and finite sample power of the LM tests.

We show that the choice of bandwidth has a large impact on the size and power of the tests. In particular we find that larger bandwidths lead to non-monotonic power whereas smaller bandwidths give tests with monotonic power. The fixed-b theory suggests that, for a given statistic, kernel and significance level, there exists a "robust" bandwidth such that the fixed-b asymptotic critical value is the same for both $I(0)$ and $I(1)$ errors. In the case of the supremum statistic, the robust bandwidth LM test has good power that is monotonic whereas the power of the mean statistic is non-monotonic.

III. **Consistency of Break Point Estimator under Misspecification of Break Number**

In this chapter, I discuss the inconsistency of sequential trend break point estimators in the presence of underspecification of the number of breaks. The analysis of models with level shifts has been documented by researchers under comprehensive settings 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 point estimators under misspecification. My simulation study and asymptotic analysis show that the trend break point estimators do not converge to the true breaks points under most conditions when the number of estimated breaks is smaller than the true number of breaks. This inconsistency leads to a potential power loss for testing for multiple trend breaks. Taking first difference is proposed to deal with this problem under certain circumstances.

ACKNOWLEDGMENTS

I am very grateful to my advisor, Professor Tim Vogelsang for his continuous support and guidance.

I also wish to express my gratitude to my dissertation committee, Professor Emma Iglesias, Professor Peter Schmidt, and Professor Lijian Yang.

My thanks also go to my parents.

TABLE OF CONTENTS

LIST OF TABLES viii

LIST OF FIGURES x

1 Break Point Estimates for a Shift in Trend: Levels versus First Differences 1
- 1.1 Introduction and Motivation . . . 1
- 1.2 Models, Assumptions, and Two Break Point Estimators . . . 3
- 1.3 Existing analysis of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ . . . 5
- 1.4 Finite Sample Behavior of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ . . . 9
- 1.5 Asymptotic Analysis of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ when δ is Local to 0 at Rate $T^{1/2}$ . . . 12
- 1.6 Break Point Estimators of The Trend Shift Model and its Partial Sum Model 26
- 1.7 Application to One-step Ahead Forecasts . . . 30
- 1.8 Conclusions . . . 37

2 Fixed-b Analysis of LM Type Tests for a Shift in Mean 38
- 2.1 Introduction . . . 38
- 2.2 Model and Assumptions . . . 40
- 2.3 LM Tests for a Shift in Mean . . . 42
- 2.4 Finite Sample Behavior of the LM Tests . . . 43
- 2.5 Fixed-b Asymptotic Analysis of LM Mean Shift Tests . . . 52
- 2.6 Bandwidths That Control Size . . . 60
- 2.7 The $Wald^*$ Statistic of Kejriwal (2009) . . . 67

3 Consistency of the Sequential Trend Break Point Estimator 74
- 3.1 Introduction . . . 74
- 3.2 Model Assumption . . . 76
- 3.3 Existing Analysis and Finite Sample Simulations . . . 78
- 3.4 Break Date Estimator under Multiple Breaks . . . 84
  - 3.4.1 Multiple mean shifts . . . 84
  - 3.4.2 Multiple trend shifts . . . 89
  - 3.4.3 Consistency/Inconsistency conclusion of $\hat{\lambda}_{MS}$ and $\hat{\lambda}_{TS}$ . . . 90
- 3.5 Break point estimators for level and first difference model under multiple breaks . . . 91
- 3.6 Application to Sequential Tests of Multiple Breaks Model . . . 96

APPENDICES **112**
A.1 Extension of the asymptotics in Theorem 1.5.1 to near-I(1) errors 113
A.2 Proof of Theorem 1.5.1 114
A.2.1 Proof of part 1 in Theorem 1.5.1 114
A.2.2 Proof of part 2 in Theorem 1.5.1 117
A.3 Proof that $\arg\max_\lambda G2(\lambda, \lambda^c) = \lambda^c$ 119
A.4 Proof of Corollary 0.1.1 121
A.4.1 Proof of part 1 in Corollary 0.1.1 121
A.4.2 Proof of part 2 in Corollary 0.1.1 121
A.5 Proof of Theorem 1.6.2 122
A.5.1 Proof of part 1 in Theorem 1.6.2 122
A.5.2 Proof of part 2 in Theorem 1.6.2 124
B.1 Proofs and Additional Results of Chapter 2 125
C.1 Proof of Theorem 3.4.4 126
C.2 Proof of Theorem 3.4.5 130
C.3 Analysis of $G2_{TS}$ under two breaks 133
C.4 Proof of Theorem 3.5.6 137
C.4.1 Proof of part 1: $\hat{\lambda}_{TS}$ 137
C.4.2 Proof of part 2: $\hat{\lambda}_{MS}$ 137

BIBLIOGRAPHY **139**

LIST OF TABLES

Table 1.1 Mean squared error (MSE) of one-step ahead forecasts of the trend shift model (1.2.1) under I(1) errors (For one-step forecast $\hat{y}_{t+1}$, the MSE is defined as $(\hat{y}_{t+1} - y_{t+1})^2$.) Simulation settings: $\lambda^c = 0.5$; $\delta = 0, 0.1, 0.2, 0.3, 0.4, 0.5$; u_t is I(1); $T = 101$; and $N = 10,000$. OLS_1^* and OLS_2^* assume that u_t is known to be I(1). 32

Table 1.2 MSE in one-step forecast with the log real per capita GDP series (1870-1996) with trend shift model (1.2.1) and different break point estimators ($\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$), where MSE of one-step forecast are calculated for 1987-1996. 35

Table 1.3 Mean squared error (MSE) of one-step ahead forecasts of the trend shift model (1.2.1) under I(0) errors (For one-step forecast $\hat{y}_{t+1}$, the MSE is defined as $(\hat{y}_{t+1} - y_{t+1})^2$ for point forecasts.) Simulation settings: $\lambda^c = 0.5$; $\delta = 0, 0.01, 0.02, 0.03, 0.04, 0.05$; u_t is I(0); $T = 101$; and $N = 10,000$. OLS_1^* and OLS_2^* assume that u_t is known to be I(0). 36

Table 1.4 MSE in one-step forecast with the log real per capita GDP series (1870-1996) with trend shift model (1.2.1) and different break point estimators ($\hat{\lambda}_{TS}$ and $\hat{\lambda}_{QS}$), where MSE of one-step forecast is calculated for 1987-1996. 36

Table 2.1 Null Rejection Probabilities Using Standard ($b = 0$) $I(0)$ Critical Values, 5% Nominal Level, 15% Trimming, QS Kernel. 45

Table 2.2 Null Rejection Probabilities Using Standard ($b = 0$) $I(0)$ Critical Values, 5% Nominal Level, 15% Trimming, Bartlett Kernel. 46

Table 2.3 Finite Sample Behavior of Data Dependent Bandwidth to Sample Size Ratios, $T = 120$, QS Kernel. 51

Table 2.4 Finite Sample Behavior of Data Dependent Bandwidth to Sample Size Ratios, $T = 120$, Bartlett Kernel 51

Table 2.5 Fixed-b Asymptotic Null Rejection Probabilities Using Standard ($b = 0$) $I(0)$ Critical Values, 5% Nominal Level, 15% Trimming, QS Kernel 61

Table 2.6 Fixed-b Asymptotic Null Rejection Probabilities Using Standard ($b = 0$) $I(0)$ Critical Values, 5% Nominal Level, 15% Trimming, Bartlett Kernel 62

Table 2.7 $I(0)/I(1)$ Robust Bandwidths and Critical Values QS kernel, 15% Trimming. 63

Table 2.8 $I(0)/I(1)$ Robust Bandwidths and Critical Values Bart kernel, 15% Trimming. 66

Table 2.9 Finite Sample Null Rejection Probabilities for Tests Using Size Robust Bandwidths and Fixed-b $I(0)/I(1)$ Critical Values, 5% Nominal Level, 15% Trimming, QS Kernel. 68

Table 2.10 Finite Sample Null Rejection Probabilities for Tests Using Size Robust Bandwidths and Fixed-b $I(0)/I(1)$ Critical Values, 5% Nominal Level, 15% Trimming, Bartlett Kernel. 69

Table 3.1 Sum of densities at the true break λ_1^c and λ_2^c where $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$ under different ρ and $M_1 = M_2$. 95

Table 3.2 Sum of densities at the true break λ_1^c and λ_2^c where $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$ under different ρ and $M_1 = -M_2$. 95

Table 3.3 Sum of densities at the true break λ_1^c and λ_2^c where $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$ under different ρ and $|M_1| \neq |M_2|$, where $M_1 = 50(\delta_1 = 5)$. 95

Table 3.4 Probability of Break Number Selection $\hat{m}$ for Trend Shift Model with 2 breaks: $\{\lambda_1^c, \lambda_2^c\} = \{1/2, 2/3\}, \delta_1 = 1, \theta = 0, T = 120$. 99

Table 3.5 Probability of Break Number Selection $\hat{m}$ for Trend Shift Model with 2 breaks: $\{\lambda_1^c, \lambda_2^c\} = \{1/2, 2/3\}, \delta_1 = 1, \theta = 0.5, T = 120$. 111

LIST OF FIGURES

Figure 1.1 Comparison of the pdf of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ using the asymptotics of Bai(1994) and PZ(2005) with $\lambda^c = 0.5$. x-axis: λ; y-axis: pdf. The left from top to bottom: $M = 1, 3, 5, 7$; the right from top to bottom: $M = 2, 4, 6, 8$. 8

Figure 1.2 Histograms of the break point estimators $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$. $\mu = \beta = 0$; $\lambda^c = 0.5$; $\delta = 0.2, 0.4, 0.6, 0.8$; u_t: I(1) errors; $T = 100$; and $N = 30,000$. 10

Figure 1.3 The asymptotic pdfs of Bai(1994), PZ(2005) and Theorem 1.5.1 with the empirical pdf of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$. $\lambda^c = 0.5$; $M = 2, 4, 6, 8$; u_t: I(1) errors; $T = 100$; and $N = 30,000$. The left: λ_{TS} (solid: Finite sample; dash: PZ(2005); dash-dot: Theorem 1). The right: λ_{MS} (solid: Finite sample; dash: Bai(1994); dash-dot: Theorem 1). . 11

Figure 1.4 Asymptotic pdf of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ by Theorem 1.5.1 at $\lambda^c = 0.5$ and $M = 1, 2, 3, 4, 5, 6, 7, 8$. 15

Figure 1.5 Finite sample histograms and asymptotic pdf of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ in the case of no breaks. (a) Histogram of $\hat{\lambda}_{TS}$ (N=30,000 replications and sample length T=100). (b) Histogram of $\hat{\lambda}_{MS}$ (N=30,000 replications and sample length T=100). (c) Asymptotic pdf of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ under no breaks. 17

Figure 1.6 $G2_{TS}(\lambda, \lambda^c)$ and $G2_{MS}(\lambda, \lambda^c)$ in equation (1.5.16) and (1.5.17) when $\lambda^c = 0.5$. 19

Figure 1.7 The finite sample pdf and the asymptotic pdf by Bai(1994) and PZ(2005) of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ under fixed $M = 2, 4, 6, 8$, $d(1) = 1$, and $T = 100, 200, 500, 1000$. solid: $T = 100$; dash: $T = 200$; dot: $T = 500$; dash-dot: $T = 1000$; '$\cdot$': pdf(Bai); 'o': pdf(Theo. 1). . . . 20

Figure 1.8 The finite sample pdf and theoretical pdf by Bai(1994), PZ(2005), and Theorem 1.5.1 of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ under fixed $\delta = 0.4$, $d(1) = 1$, and $T = 100, 200, 500, 1000$. Solid: finite sample; '$\cdot$': Theorem 1; dash: PZ(the left) or Bai(the right). 21

Figure 1.9 The asymptotic pdfs of Bai(1994), PZ(2005) and Theorem 1.5.1 with the empirical pdfs of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$. $\lambda^c = 0.2$; $M = 2, 4, 6, 8$; u_t: I(1) errors; $T = 100$; and $N = 30,000$. Solid: finite sample; dash-dot: Theorem 1; dash: PZ(left) or Bai(right). 23

Figure 1.10 Asymptotic pdfs of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ by Theorem 1.5.1 for $\lambda^c = 0.2$ and $M = 1, 2, 3, 4, 5, 6, 7, 8$. 24

Figure 1.11 Pdfs of the break point estimators from the "TS-MS" models, under I(1) $u_t's$, $\lambda^c = 0.5$, $T = 100$, $M = 4(\delta = 0.4)$, and different trimmings $\lambda^* = 0.05, 0.1, 0.15, 0.2$. 25

Figure 1.12 Finite and asymptotic pdf by Theorem 1.6.2 and PZ(2005) (only for $\hat{\lambda}_{TS}$) of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{QS}$ ("TS-QS") under I(0) $u_t's$, $\lambda^c = 0.5$, $T = 100$, $d(1) = 1$, and $\delta = 0, 0.02, 0.04, 0.1$. Solid: finite sample; dash-dot: Theorem 2; dash: PZ(left) or Bai(right). 28

Figure 1.13 The real (log) per capita GDP of Italy, Norway, and Sweden, which are of I(1) errors. x-axes: year; y-axes: (log)Per Capita GDP. 33

Figure 1.14 The real (log) per capita GDP of Australia, Canada, Germany, UK, and US, which are of I(0) errors. x-axes: year; y-axes: (log)Per Capita GDP. 34

Figure 2.1 Finite Sample Power, QS kernel, 15% Trimming. 47

Figure 2.2 Finite Sample Power, QS kernel, 15% Trimming. 48

Figure 2.3 Finite Sample Power, Bartlett kernel, 15% Trimming. 49

Figure 2.4 Finite Sample Power, Bartlett kernel, 15% Trimming. 50

Figure 2.5 Finite Sample and Asymptotic Power of $MeanLM$, QS kernel, 15% Trimming. 53

Figure 2.6 Finite Sample and Asymptotic Power of $SupLM$, QS kernel, 15% Trimming. 54

Figure 2.7 Finite and Asymptotic Power of $MeanLM$, Bartlett kernel, 15% Trimming. 55

Figure 2.8 Finite and Asymptotic Power of $SupLM$, Bartlett kernel, 15% Trimming. 56

Figure 2.9 Asymptotic Fixed-b Critical Values, 5% Level, QS kernel, 15% Trimming. 64

Figure 2.10 Asymptotic Fixed-b Critical Values, 5% Level, Bartlett kernel, 15% Trimming. 65

Figure 2.11 Finite Sample Power of Robust Bandwidth Tests, 5% Level, 15% Trimming. 70

Figure 2.12 Finite Sample Power of Robust Bandwidth Tests, 5% Level, 15% Trimming. 71

Figure 3.1 Histogram of single break point estimator $\hat{\lambda}_{MS}$ in two breaks model: $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$. $\delta_1 = 1$ always. From left to right: $\nu = -2(\delta_2 = -2), -1(\delta_2 = -1)$; from top to bottom: $T = 100, 250, 500, 1000$. 80

Figure 3.2 Histogram of single break point estimator $\hat{\lambda}_{MS}$ in two breaks model: $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$. $\delta_1 = 1$ always. From left to right: $\nu = 1(\delta_2 = 1), 2(\delta_2 = 2)$; from top to bottom: $T = 100, 250, 500, 1000$. . 81

Figure 3.3 Histogram of single break point estimator $\hat{\lambda}_{TS}$ in two breaks: $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$. $\delta_1 = 1$ always. The left to right: $\nu = -2(\delta_2 = -2), -1(\delta_2 = -1)$; The top to bottom: $T = 100, 250, 500, 1000$. 82

Figure 3.4 Histogram of single break point estimator $\hat{\lambda}_{TS}$ in two breaks: $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$. $\delta_1 = 1$ always. The left to right: $\nu = 1(\delta_2 = 1), 2(\delta_2 = 2)$; The top to bottom: $T = 100, 250, 500, 1000$. . 83

Figure 3.5 $G2_{MS}(\lambda, \lambda^c)$ under $\lambda^c = 0.5$ for mean shift model 86

Figure 3.6 $|G2_{MS}(\lambda, \lambda_1^c) + \nu \cdot G2_{MS}(\lambda, \lambda_2^c)|$ under different $\nu = 1$ and -1 for mean shift model, where $\{\lambda_1^c, \lambda_2^c\} = \{1/4, 3/4\}$. 87

Figure 3.7 $G2_{TS}(\lambda, \lambda^c)$ under $\lambda^c = 0.5$ for trend shift model 88

Figure 3.8 $|G2_{TS}(\lambda, \lambda_1^c) + \nu \cdot G2_{TS}(\lambda, \lambda_2^c)|$ under $\nu = 1$ and -1 for trend shift model, where $\{\lambda_1^c, \lambda_2^c\} = \{1/4, 3/4\}$. 100

Figure 3.9 Finite sample distribution with the asymptotic distribution of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ at $\nu = -5$. The left to right: $\{\lambda_1^c, \lambda_2^c\} = \{1/4, 3/4\}, \{1/3, 2/3\}$; the top to bottom: $T = 100, 250, 500, 1000$. $\rho = 1$. Solid: finite sample $\hat{\lambda}_{TS}$; dash: finite sample $\hat{\lambda}_{MS}$; dot: asymptotic $\hat{\lambda}_{TS}$; dot-solid: asymptotic $\hat{\lambda}_{MS}$. 101

Figure 3.10 Finite sample distribution with the asymptotic distribution of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ at $\nu = -2$. The left: $\{\lambda_1^c, \lambda_2^c\} = \{1/4, 3/4\}$; the right: $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$; the top to bottom: $T = 100, 250, 500, 1000$. $\rho = 1$. Solid: finite sample $\hat{\lambda}_{TS}$; dash: finite sample $\hat{\lambda}_{MS}$; dot: asymptotic $\hat{\lambda}_{TS}$; dot-solid: asymptotic $\hat{\lambda}_{MS}$. 102

Figure 3.11 Finite sample distribution with the asymptotic distribution of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ at $\nu = -1$. The left: $\{\lambda_1^c, \lambda_2^c\} = \{1/4, 3/4\}$; the right: $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$; the top to bottom: $T = 100, 250, 500, 1000$. $\rho = 1$. Solid: finite sample $\hat{\lambda}_{TS}$; dash: finite sample $\hat{\lambda}_{MS}$; dot: asymptotic $\hat{\lambda}_{TS}$; dot-solid: asymptotic $\hat{\lambda}_{MS}$. 103

Figure 3.12 Finite sample distribution with the asymptotic distribution of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ at $\nu = -0.5$. The left: $\{\lambda_1^c, \lambda_2^c\} = \{1/4, 3/4\}$; the right: $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$; the top to bottom: $T = 100, 250, 500, 1000$. $\rho = 1$. Solid: finite sample $\hat{\lambda}_{TS}$; dash: finite sample $\hat{\lambda}_{MS}$; dot: asymptotic $\hat{\lambda}_{TS}$; dot-solid: asymptotic $\hat{\lambda}_{MS}$. 104

Figure 3.13 Finite sample distribution with the asymptotic distribution of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ at $\nu = 0.5$. The left: $\{\lambda_1^c, \lambda_2^c\} = \{1/4, 3/4\}$; the right: $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$; the top to bottom: $T = 100, 250, 500, 1000$. $\rho = 1$. Solid: finite sample $\hat{\lambda}_{TS}$; dash: finite sample $\hat{\lambda}_{MS}$; dot: asymptotic $\hat{\lambda}_{TS}$; dot-solid: asymptotic $\hat{\lambda}_{MS}$. 105

Figure 3.14 Finite sample distribution with the asymptotic distribution of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ at $\nu = 1$. The left: $\{\lambda_1^c, \lambda_2^c\} = \{1/4, 3/4\}$; the right: $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$; the top to bottom: $T = 100, 250, 500, 1000$. $\rho = 1$. Solid: finite sample $\hat{\lambda}_{TS}$; dash: finite sample $\hat{\lambda}_{MS}$; dot: asymptotic $\hat{\lambda}_{TS}$; dot-solid: asymptotic $\hat{\lambda}_{MS}$. 106

Figure 3.15 Finite sample distribution with the asymptotic distribution of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ at $\nu = 2$. The left: $\{\lambda_1^c, \lambda_2^c\} = \{1/4, 3/4\}$; the right: $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$; the top to bottom: $T = 100, 250, 500, 1000$. $\rho = 1$. Solid: finite sample $\hat{\lambda}_{TS}$; dash: finite sample $\hat{\lambda}_{MS}$; dot: asymptotic $\hat{\lambda}_{TS}$; dot-solid: asymptotic $\hat{\lambda}_{MS}$. 107

Figure 3.16 Finite sample distribution with the asymptotic distribution of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ at $\nu = 5$. The left: $\{\lambda_1^c, \lambda_2^c\} = \{1/4, 3/4\}$; the right: $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$; the top to bottom: $T = 100, 250, 500, 1000$. $\rho = 1$. Solid: finite sample $\hat{\lambda}_{TS}$; dash: finite sample $\hat{\lambda}_{MS}$; dot: asymptotic $\hat{\lambda}_{TS}$; dot-solid: asymptotic $\hat{\lambda}_{MS}$. 108

Figure 3.17 λ to achieve maximal $G2_{TS}(\lambda, \lambda_1^c) + \nu \cdot G2_{TS}(\lambda, \lambda_2^c)$, $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$, $\nu = -10, \cdots, 10$. 109

Figure 3.18 λ to achieve maximal $G2_{TS}(\lambda, \lambda_1^c) + \nu \cdot G2_{TS}(\lambda, \lambda_2^c)$, $\{\lambda_1^c, \lambda_2^c\} = \{1/4, 3/4\}$, $\nu = -10, \cdots, 10$. 110

CHAPTER 1

Break Point Estimates for a Shift in Trend: Levels versus First Differences

1.1 Introduction and Motivation

The break point estimator in the mean shift model or the trend shift model is analyzed extensively by Bai (1994), Bai and Perron (1998)(BP hereafter), Perron and Zhu (2005)(PZ hereafter). The least squares (LS) estimator is considered in these papers, and the break points are estimated by minimizing the sum of squared residuals (SSR). Bai (1994) analyzes the break point estimator of the mean shift model under the assumption that the break magnitude is much greater than $T^{-1/2}$, where T is the sample size. He derives that, for the mean shift model with I(0) errors, the break point estimator converges to the true break at rate T. BP(1998) extend the single unknown break to multiple unknown breaks under both fixed and shrinking shift magnitudes. PZ(2005) analyze the break point estimator of the trend shift model, which allows joint breaks in both the intercept and the trend under both I(0) and I(1) errors. They assume a fixed shift in trend, and show that the break point estimator converges at rate $T^{1/2}$ for the trend shift model under I(1) errors.

The existing literature examines break point estimators of the mean shift model and the

trend shift model separately. I analyze the two estimators from these two models using the same data generating process (DGP). A trend shift model with unit root errors can be transformed into a mean shift model with stationary errors and vice versa. In other words, there are two ways to represent the same DGP: we could start with a trend shift model with I(1) errors and first differencing it to obtain a mean shift model with I(0) errors; or we could start with a level shift model with I(0) error and partial sum it to a trend shift model with I(1) errors.

Based on the convergence rate of the estimators, many researchers would estimate the break point using the first differenced form, which has a faster convergence rate based on Bai and Perron (1998)'s and PZ(2005)'s results. However, it will be shown in this chapter that first differencing is not always better, i.e., the break point estimator from the trend shift model could be preferred though it converges slower. The finite sample results show that the break point estimators have special tail behaviors which are not captured by the existing asymptotic approximations. Therefore I develop a new asymptotic theory to capture the tail behavior. I assume that the break magnitude is within a local $T^{-1/2}$ neighborhood of zero and show the following in this chapter: a) The asymptotics by Bai and Perron (1998) and PZ(2005) indicate a certain range of break magnitudes where the level model break point estimator behaves better than the first difference estimator. However, there is considerable discrepancy between these asymptotics and the finite sample distributions; b) The proposed asymptotics more closely resemble the true distributions of the break point estimators. My results lead to the counter-intuitive result that first differencing is not always the better way to estimate the break point under I(1) errors. In fact the break point estimator from the level model can have thinner tails in the distribution and concentrates more at the true break point when the break magnitude is small relative to noise.

The rest of this chapter is organized as follows. Section 2 describes the model and lays out the assumptions. It defines the break point estimators from the level model and the first difference model. Section 3 summarizes and compares the existing asymptotic

results by Bai and Perron (1998) and PZ(2005). Section 4 provides the finite sample results of the two break estimators, showing the discrepancy between existing asymptotics and finite sample behavior. In Section 5, I develop the new asymptotic theory assuming the break magnitude is local to 0 at rate $T^{-1/2}$ and show that the new theory captures the important finite sample patterns. Extensions to the trend shift model with I(0) errors and its partial sum model are included in Section 6. Section 7 gives an example where using the break point estimator from the trend shift model may reduce one-step ahead forecast errors. Section 8 summarizes the major results of paper.

1.2 Models, Assumptions, and Two Break Point Estimators

For simplicity, I use "TS-MS" to denote a pair of level model (the trend shift model under I(1) errors) and its first difference (the mean shift model under I(0) errors). Let us start with a simple linear trend shift model (TS model):

$$y_t = \mu + \beta t + \delta DT_t(\lambda^c) + u_t, \qquad t = 1, \cdots, T \tag{1.2.1}$$

where δ is the break magnitude, λ^c is the true break point with $T_b^c = \lambda^c T$, and

$$DT_t(\lambda^c) \doteq \begin{cases} 0, & t \le T_b^c \\ t - T_b^c, & t > T_b^c \end{cases}.$$

The error is assumed to be I(1), defined by assumption (A1.a).

$$(A1.a) \qquad u_t = u_{t-1} + \varepsilon_t,$$

where

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

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 $E(e_t^2|e_{t-1}, e_{t-2}, \cdots)$=1.

The first differenced model can be written as:

$$\Delta y_t \;=\; \beta + \delta DU_t(\lambda^c) + \Delta u_t, \qquad t = 2, \cdots, T \tag{1.2.2}$$

where $DU_t(\lambda^c) \doteq 1(t > T_b^c)$; and

$$1(t > T_b^c) \doteq \begin{cases} 0, & t \le T_b^c \\ 1, & t > T_b^c \end{cases} .$$

Because $\{u_t\}$ is I(1), GLS estimates are obtained using this first difference transformation. The error of the first differenced model is I(0), given by

$$\Delta u_t \;=\; \varepsilon_t.$$

Existing asymptotics of the break point estimators depend on assumptions about δ. Typical assumptions of δ in the literature are

$$(A2.a) \qquad \delta = \text{ a constant scalar,}$$

$$(A2.b) \qquad \delta \to 0, \;\; \frac{T^{1/2}\delta}{(\log T)^{1/2}} \to \infty.$$

(A2.a) is the assumption used by PZ(2005). (A2.b) is the assumption used by Bai (1994), where $\delta >> T^{-1/2}$.

Though trimming is not necessary in break point estimation as stated in PZ(2005), it is commonly used in break tests. Consider the grid of possible break dates: $\Lambda^* \doteq [T_{\lambda^*}, T_{\lambda^*} + 1, \cdots, T - T_{\lambda^*}]$. The corresponding grid of the break points is defined as $\Lambda = [\lambda^*, \cdots, 1 - \lambda^*]$, where $\lambda^* \doteq \frac{T_{\lambda^*}}{T}$.

Denote $SSR(\lambda)$ as the sum of squared residuals (SSR) with a single break at $T_b = [\lambda T]$ and SSR^0 as SSR with no break. We further define SSR^0_{TS} and $SSR_{TS}(\lambda)$ as SSR^0 and $SSR(\lambda)$ for the trend shift model (1.2.1), and SSR^0_{MS} and $SSR_{MS}(\lambda)$ as those for the mean shift model (1.2.2).

SSR^0_{TS} and $SSR_{TS}(\lambda)$ are calculated as

$$\begin{aligned} SSR^0_{TS} &\doteq \sum_{t=1}^{T}[y_t - (\tilde{\mu} + \tilde{\beta}t)]^2, \\ SSR_{TS}(\lambda) &\doteq \sum_{t=1}^{T}[y_t - (\hat{\mu} + \hat{\beta}t + \hat{\delta}DT_t(\lambda))]^2, \end{aligned}$$

where $\tilde{\mu}$ and $\tilde{\beta}$ are the OLS estimates of model (1.2.1) with $\delta = 0$ imposed; $\hat{\mu}$, $\hat{\beta}$, and $\hat{\delta}$ are the OLS estimates of model (1.2.1) with no restrictions imposed. SSR^0_{MS} and $SSR_{MS}(\lambda)$ are calculated likewise as:

$$\begin{aligned} SSR^0_{MS} &\doteq \sum_{t=2}^{T}(\Delta y_t - \tilde{\beta})^2 \\ SSR_{MS}(\lambda) &\doteq \sum_{t=2}^{T}[\Delta y_t - (\hat{\beta} + \hat{\delta}DU_t(\lambda))]^2, \end{aligned}$$

where $\tilde{\delta}$ is the OLS estimate of model (1.2.2) with $\delta = 0$ imposed; $\hat{\beta}$ and $\hat{\delta}$ are the OLS estimates of model (1.2.2) with no restrictions imposed.

The break points are estimated by minimizing $SSR_{TS}(\lambda)$ or $SSR_{MS}(\lambda)$ over the set Λ. $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ denote the break point estimator for the level model and its first difference respectively and are defined as

$$\hat{\lambda}_{TS} = \arg\min_{\lambda\in\Lambda} SSR_{TS}(\lambda), \tag{1.2.3}$$

$$\hat{\lambda}_{MS} = \arg\min_{\lambda\in\Lambda} SSR_{MS}(\lambda). \tag{1.2.4}$$

1.3 Existing analysis of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$

Bai and Perron (1998) and PZ(2005) provide limiting distributions of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ under fixed or shrinking break magnitudes. Deng and Perron (2006) extend PZ(2005)'s results from an "unbounded-trend" asymptotic framework to a "bounded-trend" asymptotic framework where the break point estimator is restricted to $[0, 1]$ by normalizing the range of T.

This extension for the trend shift model under I(1) errors is of little use here since the limiting distribution under the "bounded-trend" asymptotic framework is essentially equivalent to that obtained assuming no trend shift. The point of reference in this chapter are the results of Bai and Perron (1998) and PZ(2005) which I now review.

Under assumption $(A1.a)$ and $(A2.a)$, for the level model (1.2.1), PZ(2005) prove that

$$\sqrt{T}(\hat{\lambda}_{TS} - \lambda^c) \quad \xrightarrow{d} \quad N(0, \frac{2d(1)^2}{15\delta^2}). \tag{1.3.5}$$

Under assumption $(A1.a)$ and $(A2.b)$, for the first differenced model (1.2.2), Bai (1994) proves that

$$T(\hat{\lambda}_{MS} - \lambda^c) \quad \xrightarrow{d} \quad \frac{d(1)^2}{\delta^2} \arg\max_r \{W_1(r) - \frac{1}{2}|r|\}, \tag{1.3.6}$$

where $r \in \mathbb{R}$ and $W_1(r)$ is a two-sided Brownian motion on $\mathbb{R}^1$.

Equations (1.3.5) and (1.3.6) show that the break point estimator from the first differenced model converges to zero at speed of T, faster than the $T^{1/2}$ rate in the level model. For a given T, $d(1)$, and δ, if we define $M \doteq \frac{T^{1/2}\delta}{d(1)}$, the limiting distributions by Bai (1994) and PZ(2005) can be approximated in terms of M. We can describe the effect of M on the limiting distributions and compare the performance of the two estimators as M varies.

For a given T, we have the following implications from the existing asymptotic theories.

a) Under assumption $(A1.a)$ and $(A2.a)$, for model (1.2.1),

$$\hat{\lambda}_{TS} - \lambda^c \quad \approx \quad N(0, \frac{2}{15M^2}). \tag{1.3.7}$$

The probability density function of $\hat{\lambda}_{TS}$ is given by

$$h(\hat{\lambda}_{TS}) \quad = \quad \frac{\sqrt{15}M}{\sqrt{4\pi}} \exp(-\frac{15(\hat{\lambda}_{TS} - \lambda^c)^2 M^2}{4}); \tag{1.3.8}$$

b) Under assumption $(A1.a)$ and $(A2.b)$, for model (1.2.2),

$$\hat{\lambda}_{MS} - \lambda^c \quad \approx \quad \frac{1}{M^2} \arg\max_r \{W_1(r) - \frac{1}{2}|r|\}. \tag{1.3.9}$$

It has been shown by Yao (1987) that $\arg\max_r\{W_1(r) - |r|/2\}$ has the distribution function

$$H(x) = 1 + (2\pi)^{-1/2}\sqrt{x}e^{-x/8} - \frac{1}{2}(x+5)\Phi(-\sqrt{x}/2) + \frac{3}{2}e^x\Phi(-3\sqrt{x}/2),$$

for $x > 0$ and $H(x) = 1 - H(-x)$ for $x < 0$, with $\Phi(x)$ the distribution function of a standard normal random variable. The density function can be derived as:

$$\begin{aligned} h(x) &\doteq H(x)' \\ &= \frac{3}{2}e^x\Phi(-\frac{3}{2}\sqrt{x}) - \frac{1}{2}\Phi(-\frac{\sqrt{x}}{2}). \end{aligned}$$

The probability density function of $\hat{\lambda}_{MS}$ is given by

$$\begin{aligned} h(\hat{\lambda}_{MS}) &= \frac{3M^2\exp(M^2(\hat{\lambda}-\lambda^c))}{2}\Phi(-\frac{3M\sqrt{\hat{\lambda}_{MS}-\lambda^c}}{2}) \\ &\quad -\frac{M^2}{2}\Phi(-\frac{M\sqrt{\hat{\lambda}_{MS}-\lambda^c}}{2}). \end{aligned} \tag{1.3.10}$$

Based on equations (1.3.8) and (1.3.10), we can compare the densities of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ with respect to M. In Figure 1.1, the limiting densities are depicted in pairs for $M = 1, 2, \cdots, 8$. We can see that $\hat{\lambda}_{TS}$ mostly ($M > 1$) has thinner tails than $\hat{\lambda}_{MS}$ but the concentrations around λ^c are different. For small values of M, $\hat{\lambda}_{TS}$ is more concentrated around λ^c than $\hat{\lambda}_{MS}$. In this situation, first differencing does not help because $\hat{\lambda}_{MS}$ is dominated by $\hat{\lambda}_{TS}$ in terms of concentration. For large values of M, $\hat{\lambda}_{MS}$ is more concentrated around λ^c. This comparison shows a crossing in the distributions of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ along M. To describe the crossing more accurately, we can define concrete criteria on how concentrated the estimator is around λ^c. For a specific significance level, the critical values (CVs) can describe how tight the estimator is, and we can compare the behaviors under a significance level. Take the 80% significance level as an example. From Figure 1.1 when $\lambda^c = 0.5$, both CVs of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ decrease with the increase of M. However, $\hat{\lambda}_{MS}$ decreases faster than $\hat{\lambda}_{TS}$ but starts with a much bigger value at small M. Therefore, there exists a specific value, M_0, such that for $M \leq M_0$, $\hat{\lambda}_{TS}$ has smaller CVs, i.e., $\hat{\lambda}_{TS}$ has

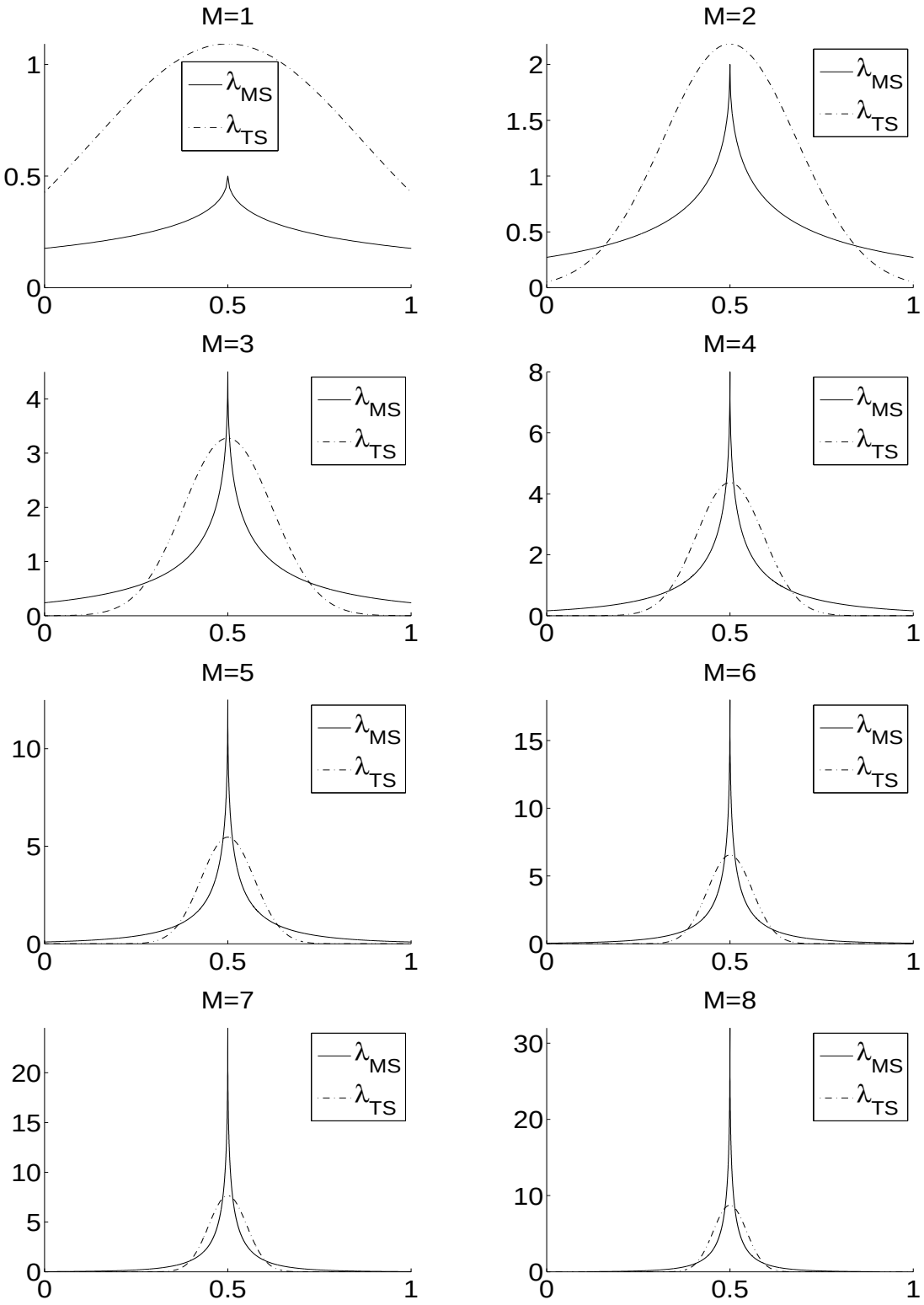

Figure 1.1. Comparison of the pdf of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ using the asymptotics of Bai(1994) and PZ(2005) with $\lambda^c = 0.5$. x-axis: λ; y-axis: pdf. The left from top to bottom: $M = 1, 3, 5, 7$; the right from top to bottom: $M = 2, 4, 6, 8$.

higher densities around λ^c at the significance level of 80%; and for $M \geq M_0$, $\hat{\lambda}_{MS}$ has smaller CVs, and it is tighter around λ^c. Based on the probability density function (pdf) curves in Figure 1.1, we can estimate that for the 80% significance level M_0 is between 7 and 8 for $\lambda^c = 0.5$.

We can also observe that for small values of M, the densities of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ do not collapse to zero when $\hat{\lambda}$ is outside of $[0, 1]$. Because the estimators cannot be outside $[0, 1]$, this implies potential problems of these asymptotic approximations in practice.

1.4 Finite Sample Behavior of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$

In this section, I first use a simple simulation to illustrate the properties of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ in finite samples. I generate data based on model (1.2.1) where u_t is I(1), $d(L) = 1$ and e_t is an iid $N(0, 1)$ process. Set $\mu = \beta = 0$ without loss of generality. Equation (1.2.3) and (1.2.4) are used to estimate $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ in each replication. Trimming is not necessary, however in order to ensure the invertibility of the regression matrix I use 2% trimming, i.e., $\lambda^* = 0.02$. The results are reported for $\lambda^c = 0.5$, $T = 100$, and replications are $N = 30,000$ for all cases.

Figure 1.2 plots the histograms of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ for $\delta = 0.2, 0.4, 0.6, 0.8$. The left are the histograms of $\hat{\lambda}_{TS}$, and the right are the histograms of $\hat{\lambda}_{MS}$. When $\delta = 0.2$, the histogram of $\hat{\lambda}_{TS}$ has one peak at $\lambda = 0.5$ and little mass at $\{0.02\}$ and $\{0.98\}$. It is not close to normal in appearance. More interestingly, the histogram of $\hat{\lambda}_{MS}$ has three peaks around $\{0.02\}$, $\{0.5\}$, and $\{0.98\}$. Compared to $\hat{\lambda}_{TS}$, $\hat{\lambda}_{MS}$ is less concentrated around λ^c for small δ. With an increase of δ, the peaks of the histogram of $\hat{\lambda}_{MS}$ around $\{0.02\}$ and $\{0.98\}$ decrease gradually. For large δ, $\hat{\lambda}_{MS}$ still has fatter tails but concentrates more around λ^c. The comparison of the concentrations matches the asymptotic results in Figure 1.1. However, the fact that there is a large mass on the tails of $\hat{\lambda}_{MS}$ when δ is small is missed by the asymptotic approximation given by (1.3.9).

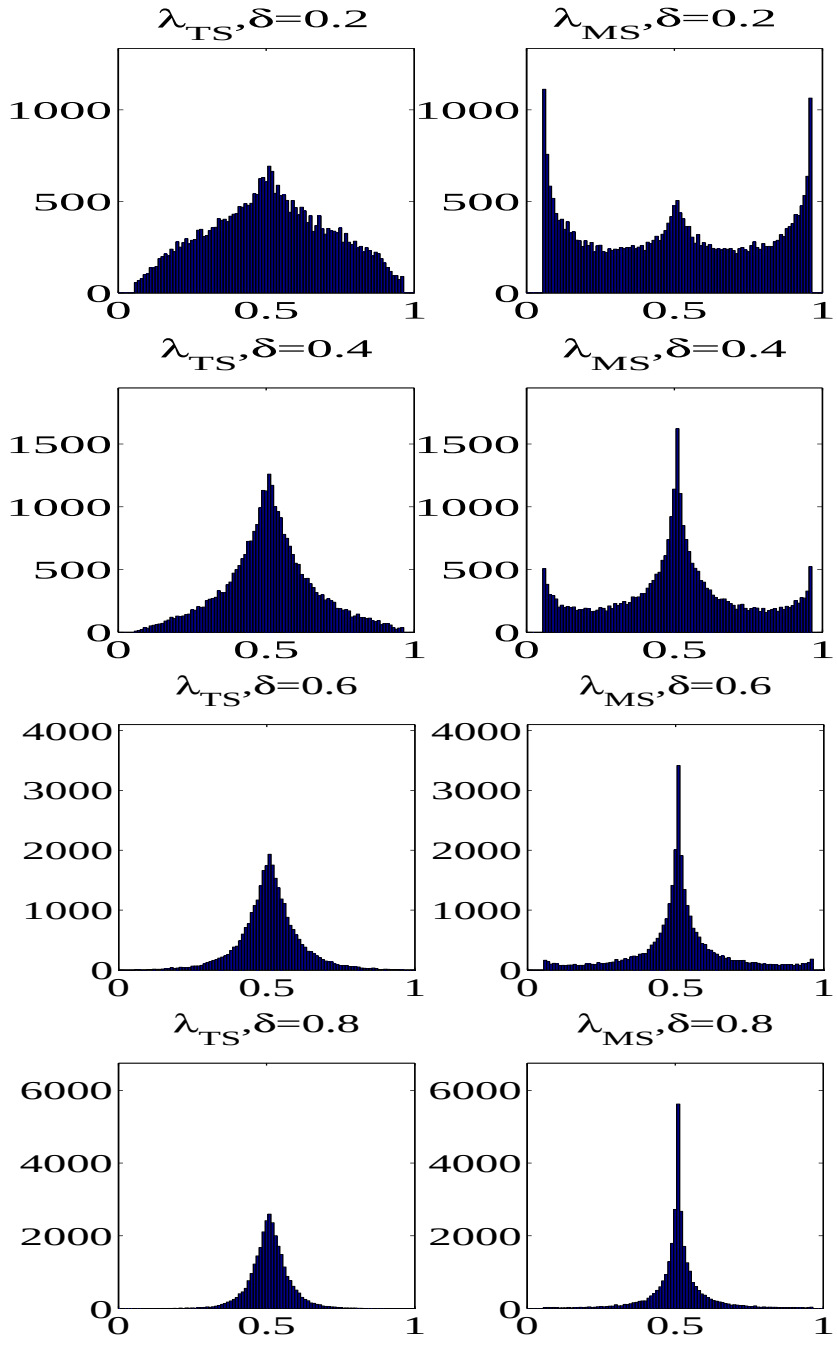

Figure 1.2. Histograms of the break point estimators $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$. $\mu = \beta = 0$; $\lambda^c = 0.5$; $\delta = 0.2, 0.4, 0.6, 0.8$; u_t: I(1) errors; $T = 100$; and $N = 30,000$.

Let us further compare the two asymptotic theories to see how the asymptotic approximations do in practice. Figure 1.3 compares the asymptotic pdfs and the finite sample pdfs. (Each of the plots contains a third density curve corresponding to the new approximation given in Section 5 which will be discussed later.) I obtain the finite sample pdf using a non-parametric kernel density smoothing method [1]. Consistent with the histograms, under small M, $\hat{\lambda}_{MS}$ tends to pile up more on the tails, and less on λ^c. As M grows, $\hat{\lambda}_{MS}$ is more concentrated around λ^c, which is what the existing asymptotics predicts. The existing asymptotics predict the concentration patterns well in finite samples. What these asymptotics do not get right is the tail behavior. Neither of the approximations captures the finite sample tail behavior under small M. PZ(2005)'s result tends to put too little density on the tails, as does Bai (1994)'s density. It becomes less of a concern when M is large. Technically, when M is large, the tails keep going outside $[0, 1]$, but in a practical sense it does not matter.

The previous discussion shows the existing theories by Bai (1994) and PZ(2005) correctly capture the concentration patterns of the two break point estimators, but both of them miss the tail behavior and provide less accurate approximation in finite samples. Because M is a multiplicative factor in equation (1.3.8) and (1.3.10), changes in M cannot be linked to the bimodality or trimodality of the finite sample behavior in Figure 1.2. This suggests that an alternative theoretical explanation for the finite sample patterns is desirable.

[1] For a given set of statistics, $\{X_i\}|i = 1, \cdots, N$, we estimate the pdf $\tilde{f}$ at x by a kernel smooth form, $\tilde{f}(x) = 1/n \cdot \sum K((x - X_i)/h)|i = 1, \cdots, n$, where $K(.)$ is the kernel function and h is the bandwidth. For details see Bowman and Azzalini (1997). In this chapter, I use the standard normal distribution as the kernel function. For the same reason as in PZ(2005), i.e., the optimal data dependant bandwidth may not work well, I choose a simple bandwidth $h = 0.5 * \sigma$ for any error. Simulations show that h does not affect the pdf estimator much.

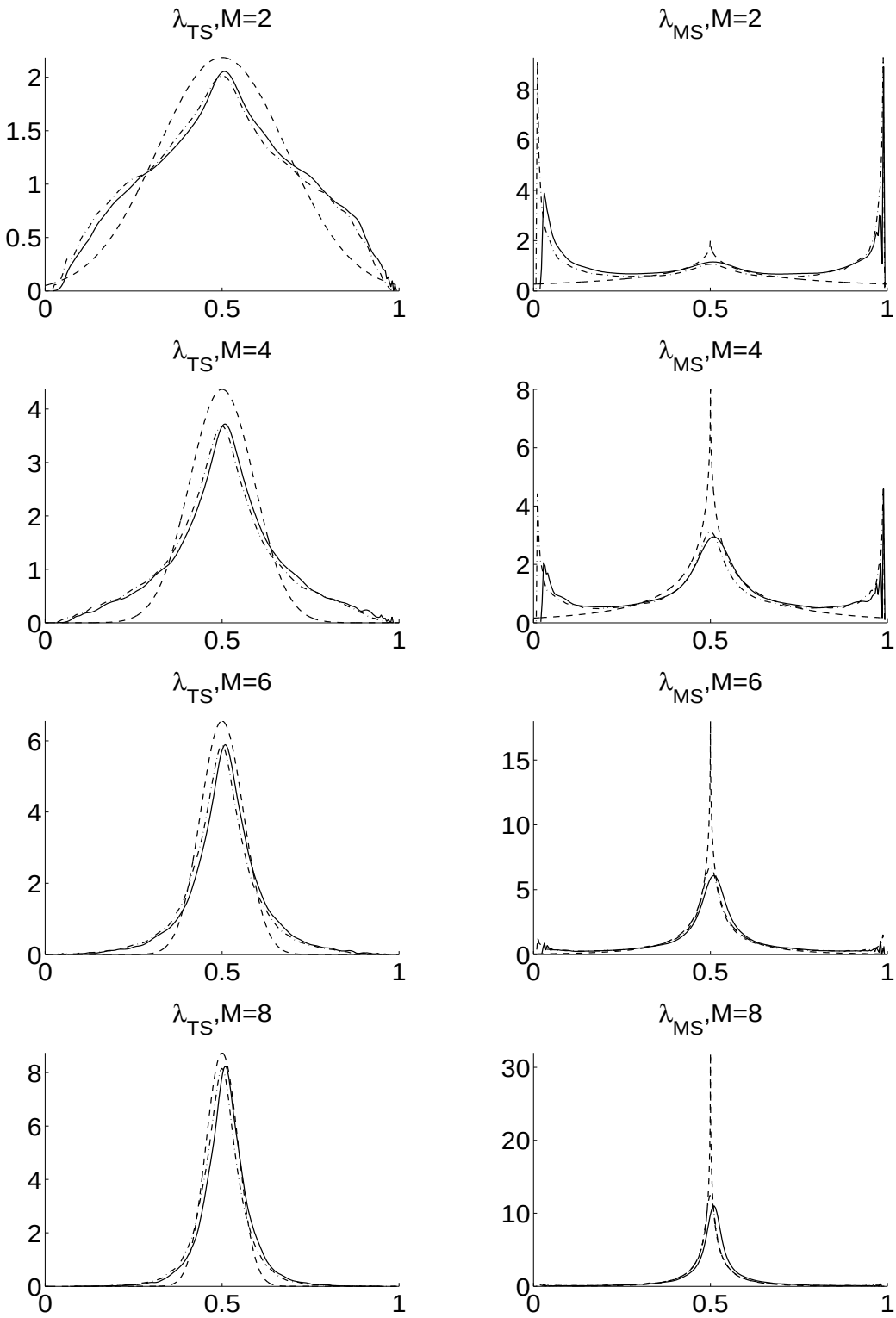

Figure 1.3. The asymptotic pdfs of Bai(1994), PZ(2005) and Theorem 1.5.1 with the empirical pdf of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$. $\lambda^c = 0.5$; $M = 2, 4, 6, 8$; u_t: I(1) errors; $T = 100$; and $N = 30,000$. The left: λ_{TS} (solid: Finite sample; dash: PZ(2005); dash-dot: Theorem 1). The right: λ_{MS} (solid: Finite sample; dash: Bai(1994); dash-dot: Theorem 1).

1.5 Asymptotic Analysis of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ when δ is Local to 0 at Rate $T^{1/2}$

Bai (1994) and PZ(2005) assume that the break magnitude δ is outside a $T^{-1/2}$ neighborhood of zero. An alternative way to develop an asymptotic theory is to assume that δ is within a $T^{-1/2}$ neighborhood using the assumption

$$(A2.c) \qquad \delta = \frac{\delta^*}{T^{1/2}}, \;\; \delta^* = \text{ a constant scalar.} \tag{1.5.11}$$

Next the limiting distributions of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ are derived under assumption (A2.c).

Theorem 1.5.1 *Suppose the regressions of the level model (1.2.1) and its first difference (1.2.2) are estimated using $\lambda \in \Lambda \subseteq (0,1)$ and $T_b^c \doteq \lambda^c T$ is the true break. Under assumption $(A1.a)$ and $(A2.c)$, the break point estimators defined by (1.2.3) and (1.2.4) have the limiting distributions as follows:*

1. For the level model (1.2.1),

$$\hat{\lambda}_{TS} \;\xrightarrow{d}\; \arg\max_{\lambda\in\Lambda}\{\frac{[\int_0^1 F(r,\lambda)W(r)dr + \frac{\delta^*}{d(1)}\int_0^1 F(r,\lambda)F(r,\lambda^c)dr]^2}{\int_0^1 F(r,\lambda)^2 dr}\} \tag{1.5.12}$$

where

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

which implies the approximation

$$\hat{\lambda}_{TS} - \lambda^c \approx \arg\max_{\lambda\in\Lambda}\{\frac{[\int_0^1 F(r,\lambda)W(r)dr + M\int_0^1 F(r,\lambda)F(r,\lambda^c)dr]^2}{\int_0^1 F(r,\lambda)^2 dr} - \lambda^c\} \tag{1.5.13}$$

where $M = \frac{\delta^*}{d(1)} \equiv \frac{\delta T^{1/2}}{d(1)}$.

2. For the first difference model (1.2.2),

$$\hat{\lambda}_{MS} \;\xrightarrow{d}\; \arg\max_{\lambda\in\Lambda}\{\frac{[(\lambda W(1) - W(\lambda)) + \frac{\delta^*}{d(1)}\Psi(\lambda,\lambda^c)]^2}{\lambda(1-\lambda)}\} \tag{1.5.14}$$

where

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

which implies the approximation

$$\hat{\lambda}_{MS} - \lambda^c \approx \arg\max_{\lambda\in\Lambda}\{\frac{[(\lambda W(1) - W(\lambda)) + M\Psi(\lambda,\lambda^c)]^2}{\lambda(1-\lambda)} - \lambda^c\}. \tag{1.5.15}$$

The limits in Theorem 1.5.1 are different from what we have seen before, but just as before, M shows up in the approximations, so we can directly compare them with the existing theory. Figure 1.3 compares the finite sample pdfs with the asymptotic pdfs of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ from equations (1.3.8), (1.3.10), and Theorem 1.5.1 with $\lambda^c = 0.5$, when $M = 2, 4, 6, 8$ (i.e. $\delta = 0.2, 0.4, 0.6, 0.8; T = 100$.) We can see the new asymptotic theory captures the density of $\hat{\lambda}_{TS}$ on the boundary of $[0, 1]$. It also tracks the unusual tail behavior of $\hat{\lambda}_{MS}$, the large mass on the boundary, and predicts the densities of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ in the middle of $[0, 1]$ in finite sample cases as well.

Also I compare $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ using the new asymptotics. Figure 1.4 shows the concentration patterns of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$. Similar to Bai (1994) and PZ(2005), it shows that $\hat{\lambda}_{TS}$ concentrates more around λ^c for a small M. If we consider the high probability in a small area around λ^c as our criterion, we can see $\hat{\lambda}_{TS}$ can be a more precise estimator than $\hat{\lambda}_{MS}$ under small M.

Why does this new asymptotic theory pick up the tail behavior better? To discover the effect of M on the limiting distributions, I decompose the terms inside $\arg\max$ of equations (1.5.12) and (1.5.14) into two parts:

$$\begin{aligned} G_{TS}(\lambda,\lambda^c) &\doteq G1_{TS}(\lambda) + M \cdot G2_{TS}(\lambda,\lambda^c) \\ &\doteq \frac{\int_0^1 F(r,\lambda)W(r)dr}{\sqrt{\int_0^1 F(r,\lambda)^2 dr}} + M \cdot \frac{\int_0^1 F(r,\lambda)F(r,\lambda^c)dr}{\sqrt{\int_0^1 F(r,\lambda)^2 dr}} \end{aligned} \tag{1.5.16}$$

and

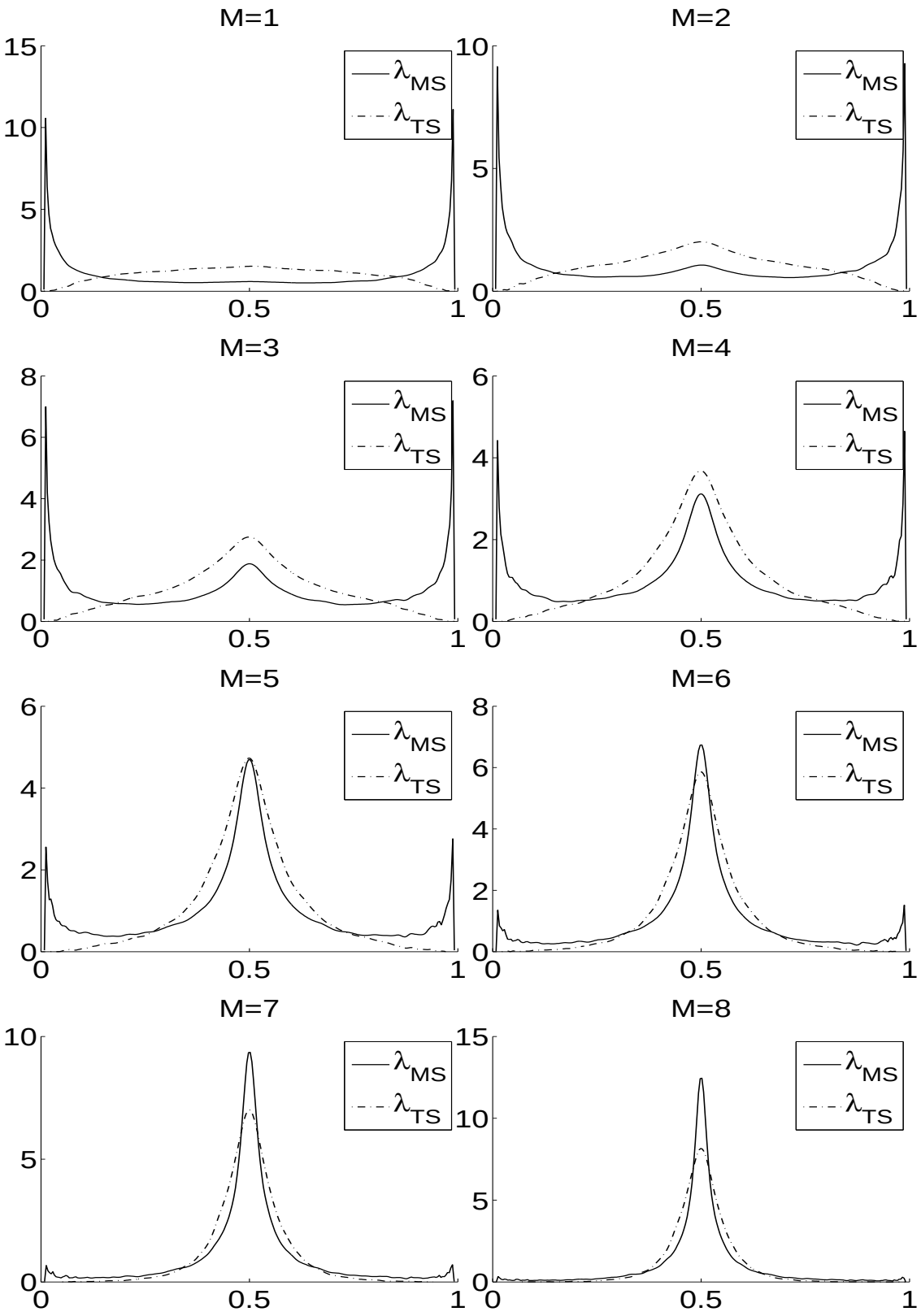

Figure 1.4. Asymptotic pdf of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ by Theorem 1.5.1 at $\lambda^c = 0.5$ and $M = 1, 2, 3, 4, 5, 6, 7, 8$.

$$
\begin{aligned}
G_{MS}(\lambda, \lambda^c) &\doteq G1_{MS}(\lambda) + M \cdot G2_{MS}(\lambda, \lambda^c) \\
&\doteq \frac{(\lambda W(1) - W(\lambda))}{\sqrt{\lambda(1-\lambda)}} + M \cdot \frac{\Psi(\lambda, \lambda^c)}{\sqrt{\lambda(1-\lambda)}} \qquad (1.5.17)
\end{aligned}
$$

For conciseness, denote $G1_{TS}(\lambda, \lambda^c)$ and $G1_{MS}(\lambda, \lambda^c)$ as $G1$, $G2_{TS}(\lambda, \lambda^c)$ and $G2_{MS}(\lambda, \lambda^c)$ as $G2$, and $G_{TS}(\lambda, \lambda^c)$ and $G_{MS}(\lambda, \lambda^c)$ as G.

From the decomposition we can see firstly that the asymptotics in Theorem 1.5.1 is continuous at $M = 0$, i.e., M could be as small as possible in the asymptotics. The existing theories by Bai and PZ need to assume there is a break. If there is no break, their distribution theory breaks down, generating a discontinuity in their asymptotic theory as M converges to zero, while the new approximation is continuous w.r.t. the magnitude of the break.

Figure 1.5 depicts the finite sample histograms and the asymptotic pdfs of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ when $M = 0$. The asymptotic pdfs are obtained by Theorem 1.5.1. Both finite sample histograms and asymptotic pdfs show that when there is no break, the tail behaviors of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ are very different: $\hat{\lambda}_{TS}$ concentrates more in the middle while $\hat{\lambda}_{MS}$ concentrates more around $\{0\}$ and $\{1\}$. The distribution of $\hat{\lambda}_{TS}$ when $M = 0$ goes less often to $\{0\}$ and $\{1\}$ but more to the middle, which is consistent with the results by Nunes, Kuan and Newbold (1995) and Bai (1998) about "spurious breaks"; while the distribution of $\hat{\lambda}_{MS}$ has peaks at $\{0\}$ and $\{1\}$, and is flat in the middle. Theorem 1.5.1 picks up the tails as shown in Figure 1.5, where $\hat{\lambda}_{TS}$ has higher probability in the middle but lower probability on the boundary, while $\hat{\lambda}_{MS}$ has higher probability on the boundary but lower probability in the middle, and both pdfs are flat in the middle range of $[0, 1]$. Although under no break $\hat{\lambda}_{TS}$ is spurious, it forms a major source of the preciseness of $\hat{\lambda}_{TS}$ (in the sense of more concentration in the pdf at certain significance levels) when M is small and λ^c is around 0.5.

With the form of $(G1 + M \cdot G2)$ in the limiting distributions, Theorem 1.5.1 provides a bridge between the $\delta = 0$ asymptotics and the $\delta \neq 0$ asymptotics. When M is small,

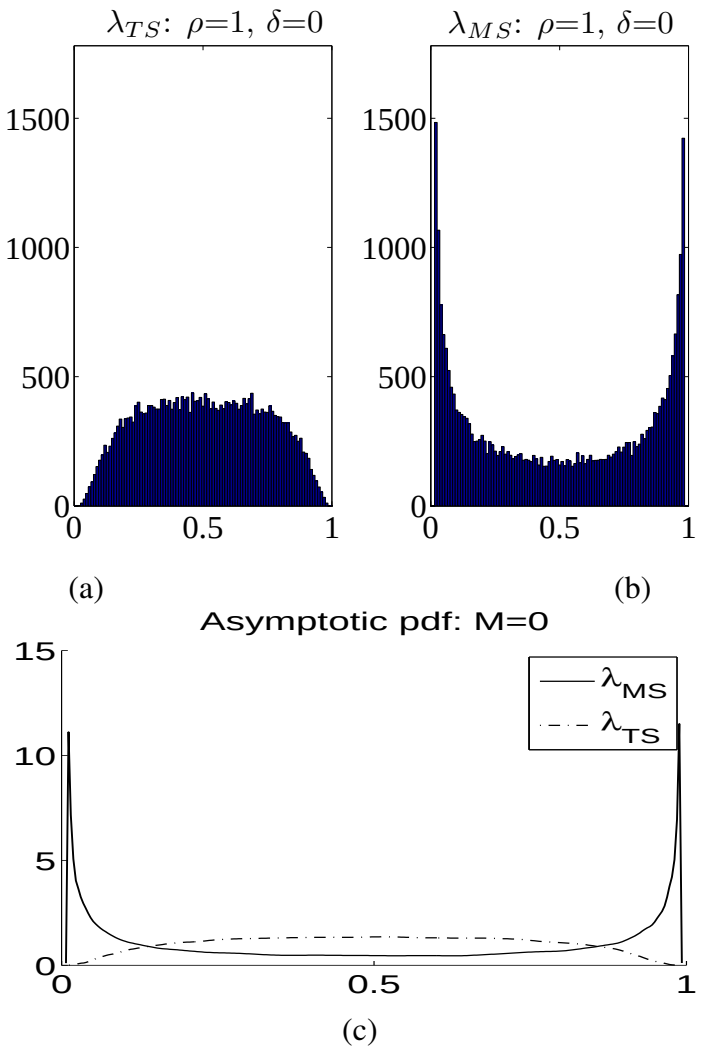

Figure 1.5. Finite sample histograms and asymptotic pdf of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ in the case of no breaks. (a) Histogram of $\hat{\lambda}_{TS}$ (N=30,000 replications and sample length T=100). (b) Histogram of $\hat{\lambda}_{MS}$ (N=30,000 replications and sample length T=100). (c) Asymptotic pdf of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ under no breaks.

the random component $G1$ dominates $G2$ and the distribution tends to have the null tail. When M is significantly different from zero, $G2$ dominates $G1$ and M affects the limiting distribution through $M \cdot G2_{TS}(\lambda, \lambda^c)$ and $M \cdot G2_{MS}(\lambda, \lambda^c)$. Both of the $G2$ parts attain global maxima at the same place, λ^c, as shown in Figure 1.6. (For a detailed proof see Appendix A.3.) If M is big enough, the $G2$ parts are completely dominant in $(G1+M\cdot G2)$, which makes $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ arbitrarily close to λ^c. Therefore, Theorem 1.5.1 explains why as M grows, $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ are consistent to some extent. For a moderately large to small M, the limiting distribution of $\hat{\lambda}_{TS}$ exhibits a shape of "$\frown\!\!\smile$" and $\hat{\lambda}_{MS}$ exhibits a shape of "w", resulting from the mixed effects of $G1$ and $G2$ parts in the asymptotics.

It is useful to see how the asymptotics approximates finite sample behaviors under different sample sizes. Figure 1.7 compares the finite sample distributions with the asymptotic theories by Bai (1994), PZ(2005), and Theorem 1.5.1 under different fixed M but various $T's$ ($M = 2, 4, 6, 8$ and $T = 100, 200, 500, 1000$). We can see that if M is fixed, increasing the sample size does not improve the approximation of the asymptotics by Bai (1994) and PZ(2005) in finite samples. In contrast, the compatibility of the new approximation with the finite sample pdfs at $T = 100, 200, 500, 1000$ shows the approximation of Theorem 1.5.1 is adequate no matter what T is. The reason why increasing the sample size does not improve the approximation of the existing asymptotics in this case is because M is fixed. It is not T per se or δ per se but the relative magnitude of them that drives the shape of finite sample patterns of the break point estimator. This relative effect is picked up by M. For a given value of M, the finite sample behavior of the break point estimator is the same whether T is large and δ is small or T is small and δ is large.

Figure 1.8 looks at fixed δ with T getting bigger. In that case, M is also getting bigger. As is expected that the asymptotic approximations of the existing theories are getting better as M increases, while Theorem 1.5.1 continues to provide a close approximation to finite sample behaviors of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ as shown in this figure.

One additional advantage of the new asymptotics lies in the finite sample approximation

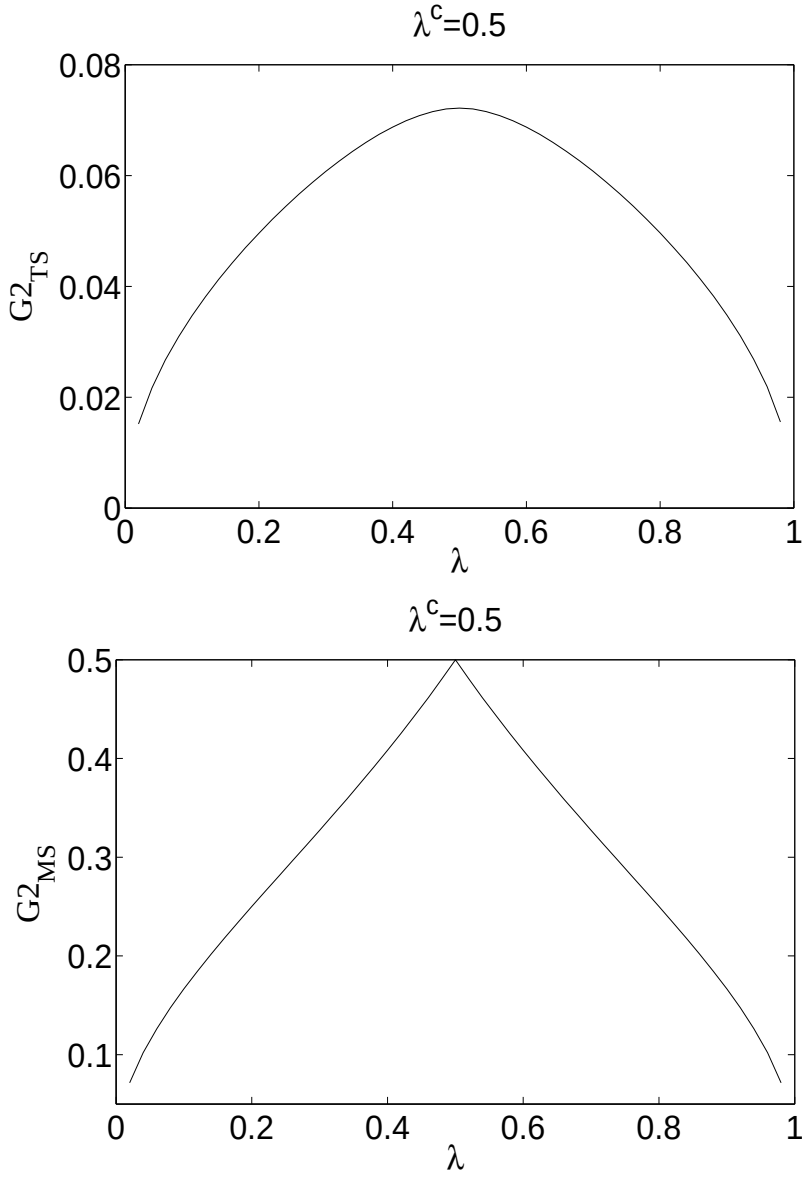

Figure 1.6. $G2_{TS}(\lambda, \lambda^c)$ and $G2_{MS}(\lambda, \lambda^c)$ in equation (1.5.16) and (1.5.17) when $\lambda^c = 0.5$.

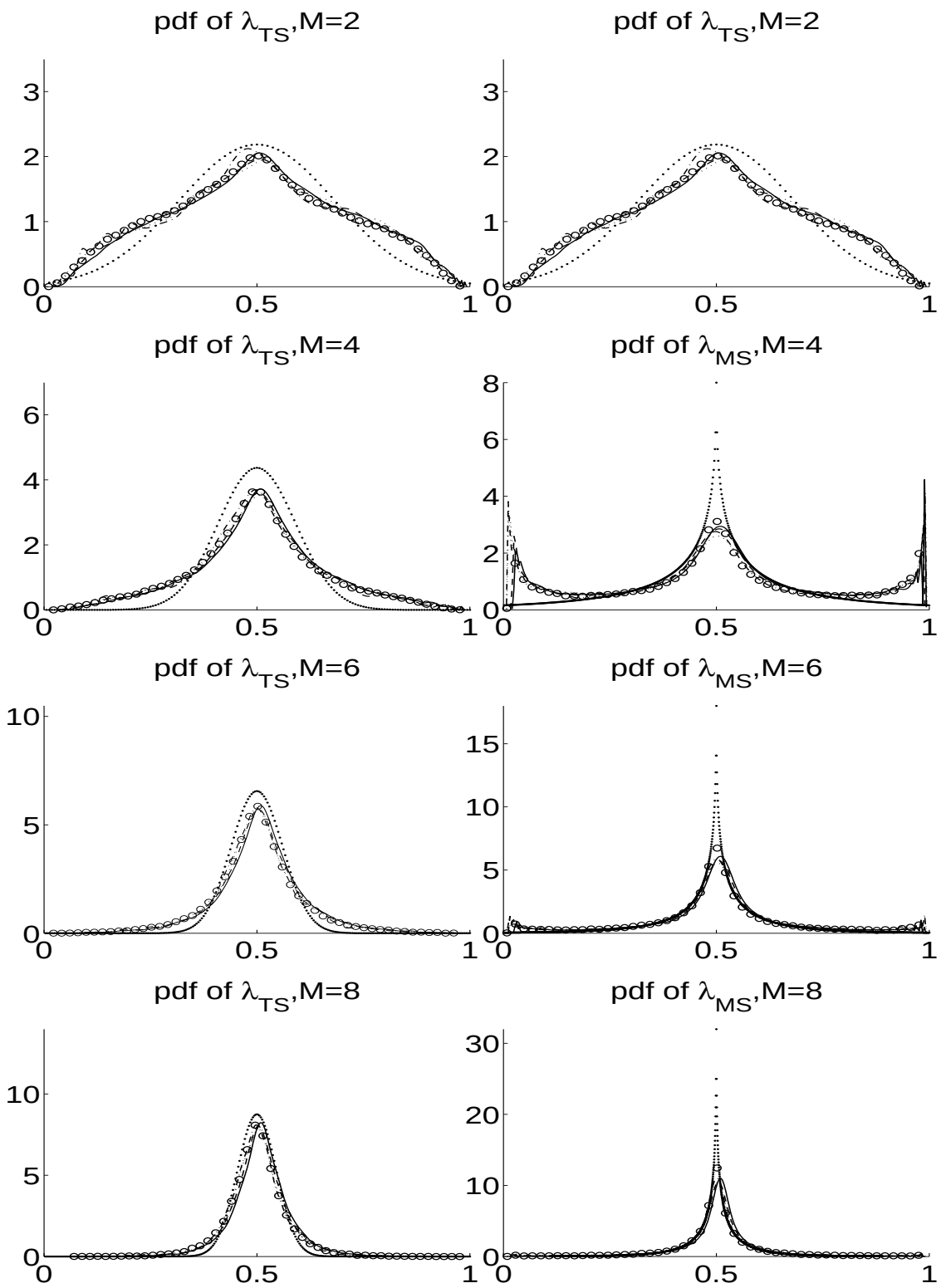

Figure 1.7. The finite sample pdf and the asymptotic pdf by Bai(1994) and PZ(2005) of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ under fixed $M = 2, 4, 6, 8$, $d(1) = 1$, and $T = 100, 200, 500, 1000$. solid: $T = 100$; dash: $T = 200$; dot: $T = 500$; dash-dot: $T = 1000$; '·': pdf(Bai); 'o': pdf(Theo. 1).

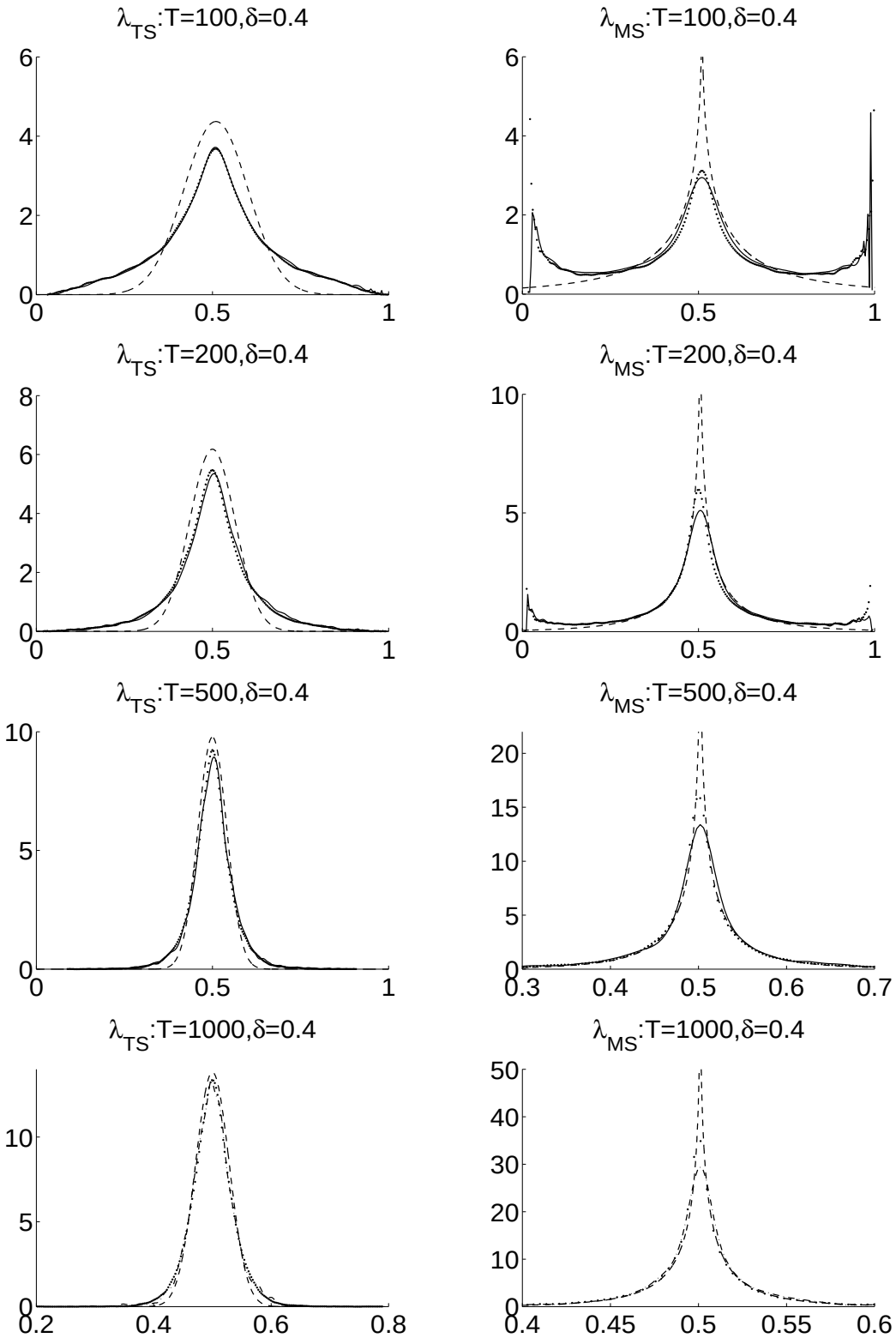

Figure 1.8. The finite sample pdf and theoretical pdf by Bai(1994), PZ(2005), and Theorem 1.5.1 of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ under fixed $\delta = 0.4$, $d(1) = 1$, and $T = 100, 200, 500, 1000$. Solid: finite sample; '$\cdot$': Theorem 1; dash: PZ(the left) or Bai(the right).

under different λ^c. The normalized asymptotics (the limiting distributions of $\hat{\lambda} - \lambda^c$) by Bai (1994) and PZ(2005) are invariant to λ^c. (See equation (1.3.5)and (1.3.6).) In contrast, the finite sample behavior depends on λ^c, and this is captured by Theorem 1.5.1. We can see in equations (1.5.13) and (1.5.15), $G1$, the leading term, does not depend on λ^c, while $G2$ does depend on λ^c and attains its maximum at λ^c; hence the limiting distributions of $\hat{\lambda}_{TS} - \lambda^c$ and $\hat{\lambda}_{MS} - \lambda^c$ are functions of λ^c. Figure 1.9 compares the new and existing aymptotics with finite sample pdfs for $\lambda^c = 0.2$. Compared to the $\lambda^c = 0.5$ case in Figure 1.3, when $\lambda^c \neq 0.5$, the existing asymptotics miss the tail behavior to an even greater extent, while the new asymptotics nails them down.

Figure 1.10 compares $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ at $\lambda^c = 0.2$ using the new asymptotics. Compared to the $\lambda^c = 0.5$ case in Figure 1.4, where the distributions of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ are symmetric around $\lambda = 0.5$, the distributions of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ here are asymmetric around $\lambda = 0.2$. The estimator at $\lambda < 0.2$ has a lower density than that at $\lambda > 0.2$ for $\hat{\lambda}_{TS}$. For $\hat{\lambda}_{MS}$, the behavior on the right and left side are reversed. This is fairly intuitive: when M is small, the $\delta = 0$ asymptotics dominates, which results in this asymmetry in the distributions. We may also notice that $G_{TS}(\lambda, \lambda_1^c) - \lambda_1^c$ is symmetric to $G_{TS}(\lambda, \lambda_2^c) - \lambda_2^c$ around $\lambda = 0$, where $\lambda_2^c \doteq 1 - \lambda_1^c$. This analysis also holds for $\hat{\lambda}_{MS}$.

Trimming is used by most break tests to deal with unknown break dates. Trimming also affects the performance of break point estimators. Theorem 1.5.1 captures the effect of trimming well. The trimming affects the asymptotics through the set Λ in equation (1.5.12) and (1.5.14), which explains big differences in the tail behaviors. Figure 1.11 plots the asymptotic distributions and finite sample distributions with trimming of 0.05, 0.10, 0.15, and 0.20. As expected, with the increase of trimming, the pile up in the tails becomes more pronounced. Theorem 1.5.1 again captures the tail behavior well.

The pattern of concentration around λ^c for $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ changes with different trimming. When trimming=0.05, $\hat{\lambda}_{TS}$ dominates $\hat{\lambda}_{MS}$ in the densities around $\lambda = \lambda^c$. With the increase of trimming, this dominant effect tends to reverse. When trimming=0.2, $\hat{\lambda}_{MS}$ has

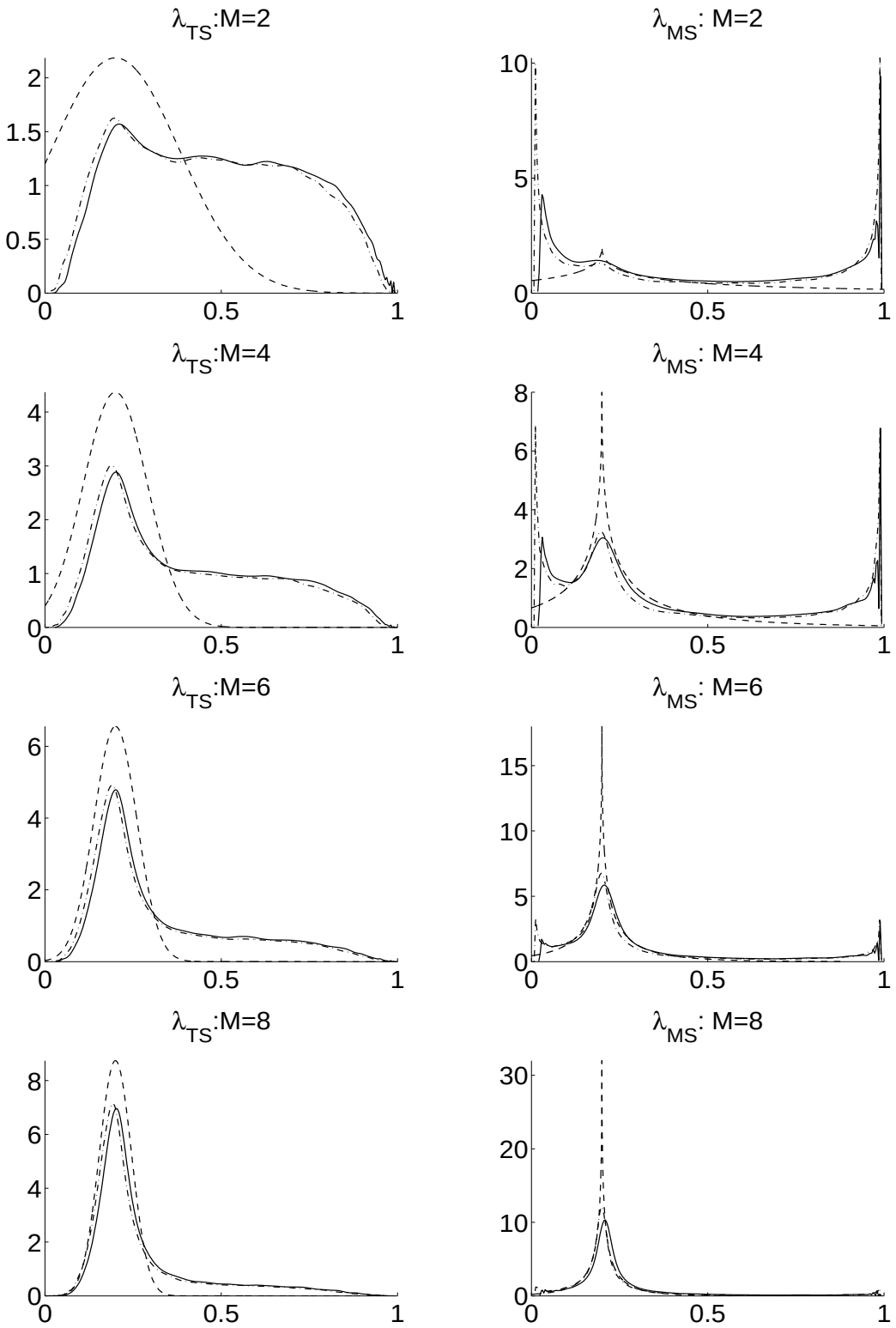

Figure 1.9. The asymptotic pdfs of Bai(1994), PZ(2005) and Theorem 1.5.1 with the empirical pdfs of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$. $\lambda^c = 0.2$; $M = 2, 4, 6, 8$; u_t: I(1) errors; $T = 100$; and $N = 30,000$. Solid: finite sample; dash-dot: Theorem 1; dash: PZ(left) or Bai(right).

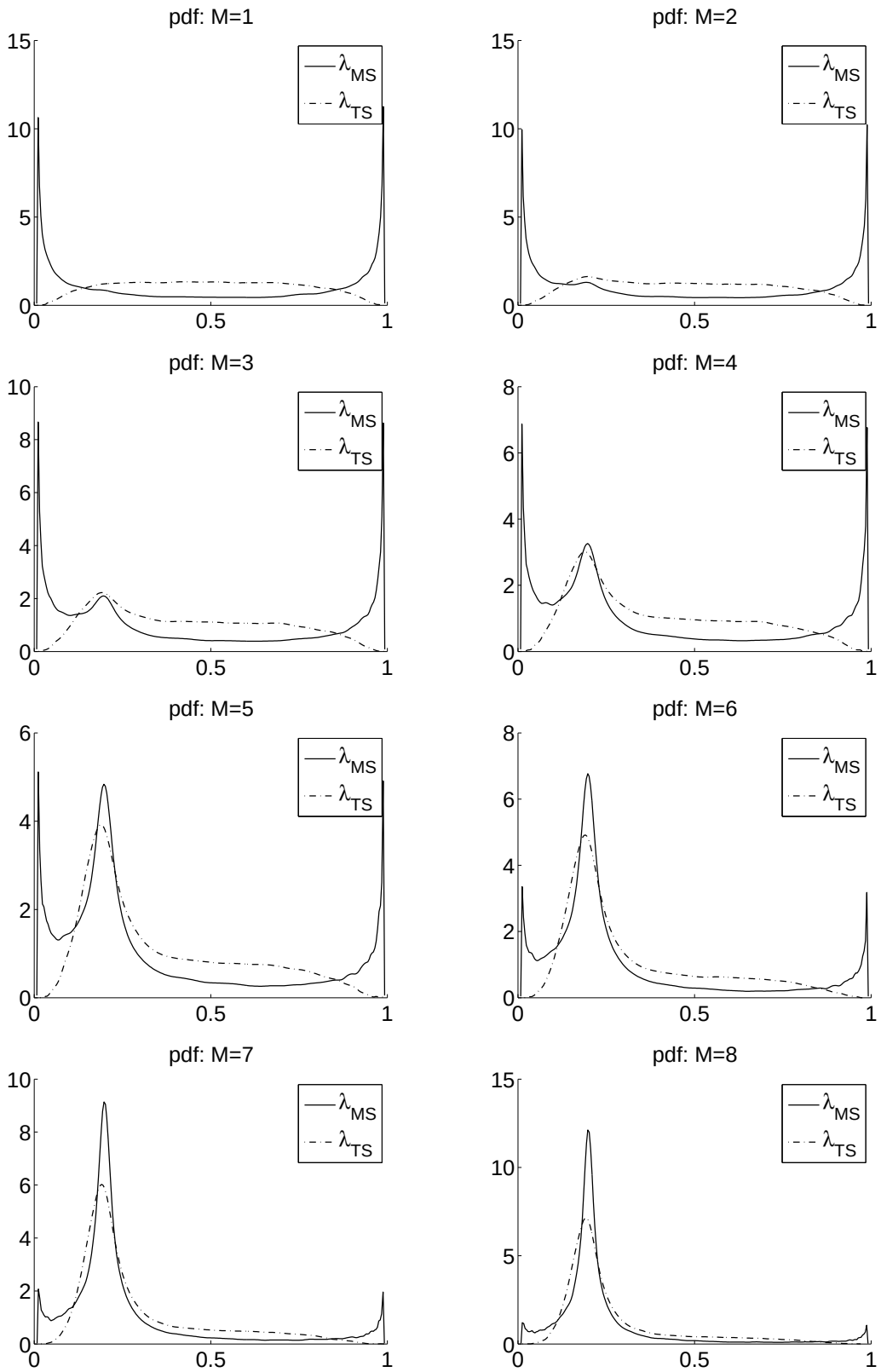

Figure 1.10. Asymptotic pdfs of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ by Theorem 1.5.1 for $\lambda^c = 0.2$ and $M = 1, 2, 3, 4, 5, 6, 7, 8$.

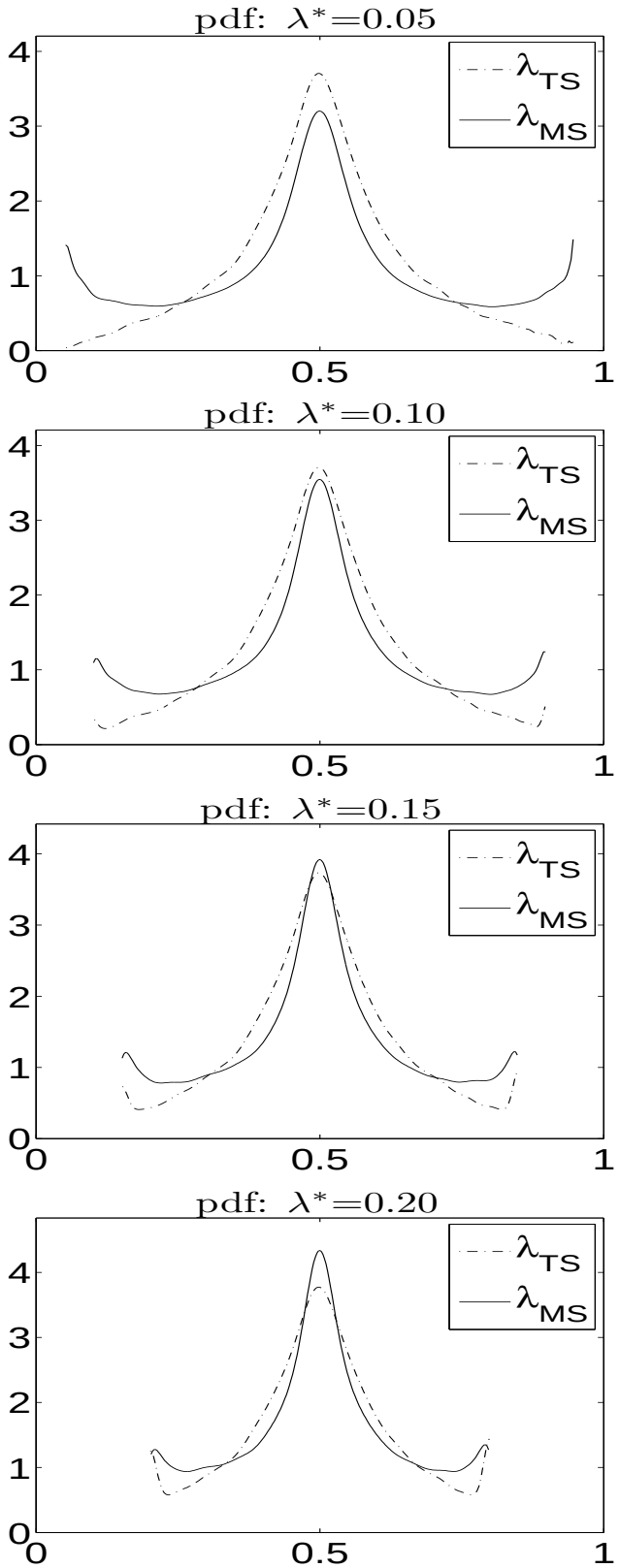

Figure 1.11. Pdfs of the break point estimators from the "TS-MS" models, under I(1) $u_t's$, $\lambda^c = 0.5$, $T = 100$, $M = 4(\delta = 0.4)$, and different trimmings $\lambda^* = 0.05, 0.1, 0.15, 0.2$.

higher density at $\lambda = \lambda^c$. Also the tails change according to different trimmings, especially for $\hat{\lambda}_{TS}$. When trimming=0.05, it has little density in the tails. But when trimming=0.2, $\hat{\lambda}_{TS}$ has considerable mass in the tails while $\hat{\lambda}_{MS}$ does not change that much, which might be the reason why $\hat{\lambda}_{MS}$ becomes dominant.

1.6 Break Point Estimators of The Trend Shift Model and its Partial Sum Model

Given that $\hat{\lambda}_{TS}$ can be more accurate than $\hat{\lambda}_{MS}$ when $u_t \sim I(1)$, it might be possible that when $u_t \sim I(0)$, a more precise estimator of λ^c can be obtained by partial summing the model and inducing a unit root in the error. Similar to the "TS-MS" models, I define a second pair of models: the level trend shift model (TS), and its partial sum, the quadratic shift model (QS). The level model is still the trend shift model (1.2.1). However, the noise assumption is changed from I(1) errors to I(0), defined in assumption $(A1.c)$.

$$(A1.c) \qquad u_t \;=\; e_t, \text{ where } t = 1, \cdots, T,$$

and e_t is defined in (A1.a).

Because of I(0) errors, the break magnitude δ is assumed to be within a $T^{-3/2}$ neighborhood of zero using the assumption

$$(A2.d) \qquad \delta = \frac{\delta^*}{T^{3/2}}, \;\; \delta^* = \text{ a constant scalar.} \tag{1.6.18}$$

Under assumption $(A1.c)$ and $(A2.d)$, for the trend shift model under I(0) errors, PZ(2005) prove that

$$T^{3/2}(\hat{\lambda}_{TS} - \lambda^c) \;\xrightarrow{d}\; N(0, \frac{4}{\lambda^c(1-\lambda^c)}\frac{d(1)^2}{\delta^2}). \tag{1.6.19}$$

If we define $M \doteq \frac{\delta T^{3/2}}{d(1)}$, PZ's result can be rewritten as

$$\hat{\lambda}_{TS} - \lambda^c \approx N(0, \frac{4}{\lambda^c(1-\lambda^c)}\frac{1}{M^2}).$$

Taking the partial sum of model (1.2.1), we can obtain the partial sum model:

$$\begin{aligned} S_t &= \alpha t + \beta\frac{1}{2}(t^2+t) + \delta\sum_{j=1}^{t} DT_j(\lambda^c) + v_t \\ v_t &= v_{t-1} + e_t,\ t = 1, \cdots, T, \end{aligned} \tag{1.6.20}$$

where $S_t \doteq \sum_{j=1}^{t} y_j$, $v_t \doteq \sum_{j=1}^{t} u_j$.

We rewrite the partial sum model as

$$S_t = (\alpha + \frac{\beta}{2})t + \frac{\beta}{2}t^2 + \delta\sum_{j=1}^{t} DT_j(\lambda^c) + \varepsilon_t. \tag{1.6.21}$$

Define $\alpha^* \doteq \alpha + \frac{\beta}{2}$, $\beta^* \doteq \frac{\beta}{2}$, and $DQ_t(\lambda^c) \doteq \sum_{j=1}^{t} DT_j(\lambda^c)$, a quadratic shift. Equation (1.6.21) is expressed as

$$S_t = \alpha^* t + \beta^* t^2 + \delta DQ_t(\lambda^c) + \varepsilon_t. \tag{1.6.22}$$

Define SSR^0_{QS} and $SSR_{QS}(\lambda)$ as SSR^0 and $SSR(\lambda)$ for the quadratic shift model (1.6.22). SSR^0_{QS} and $SSR_{QS}(\lambda)$ are calculated as

$$\begin{aligned} SSR^0_{QS} &\doteq \sum_{t=1}^{T}[S_t - (\tilde{\alpha^*}t + \tilde{\beta^*}t^2)]^2, \\ SSR_{QS}(\lambda) &\doteq \sum_{t=1}^{T}[S_t - (\hat{\alpha^*}t + \hat{\beta^*}t^2 + \hat{\delta}DT_t(\lambda))]^2, \end{aligned}$$

where $\tilde{\alpha^*}$ and $\tilde{\beta^*}$ are the OLS estimates of model (1.6.22) with $\delta = 0$ imposed; $\hat{\alpha^*}$, $\hat{\beta^*}$, and $\hat{\delta}$ are the OLS estimates of model (1.6.22) with no restrictions imposed.

The break point estimator $\hat{\lambda}_{QS}$ is obtained by minimizing the $SSR_{QS}(\lambda)$:

$$\hat{\lambda}_{QS} = \arg\min_{\lambda\in\Lambda} SSR_{QS}(\lambda). \tag{1.6.23}$$

Similar to what we see before in the distribution of $\hat{\lambda}_{TS}$ from the trend shift model under I(1) errors, there is also a discrepancy between their asymptotic theory and the finite

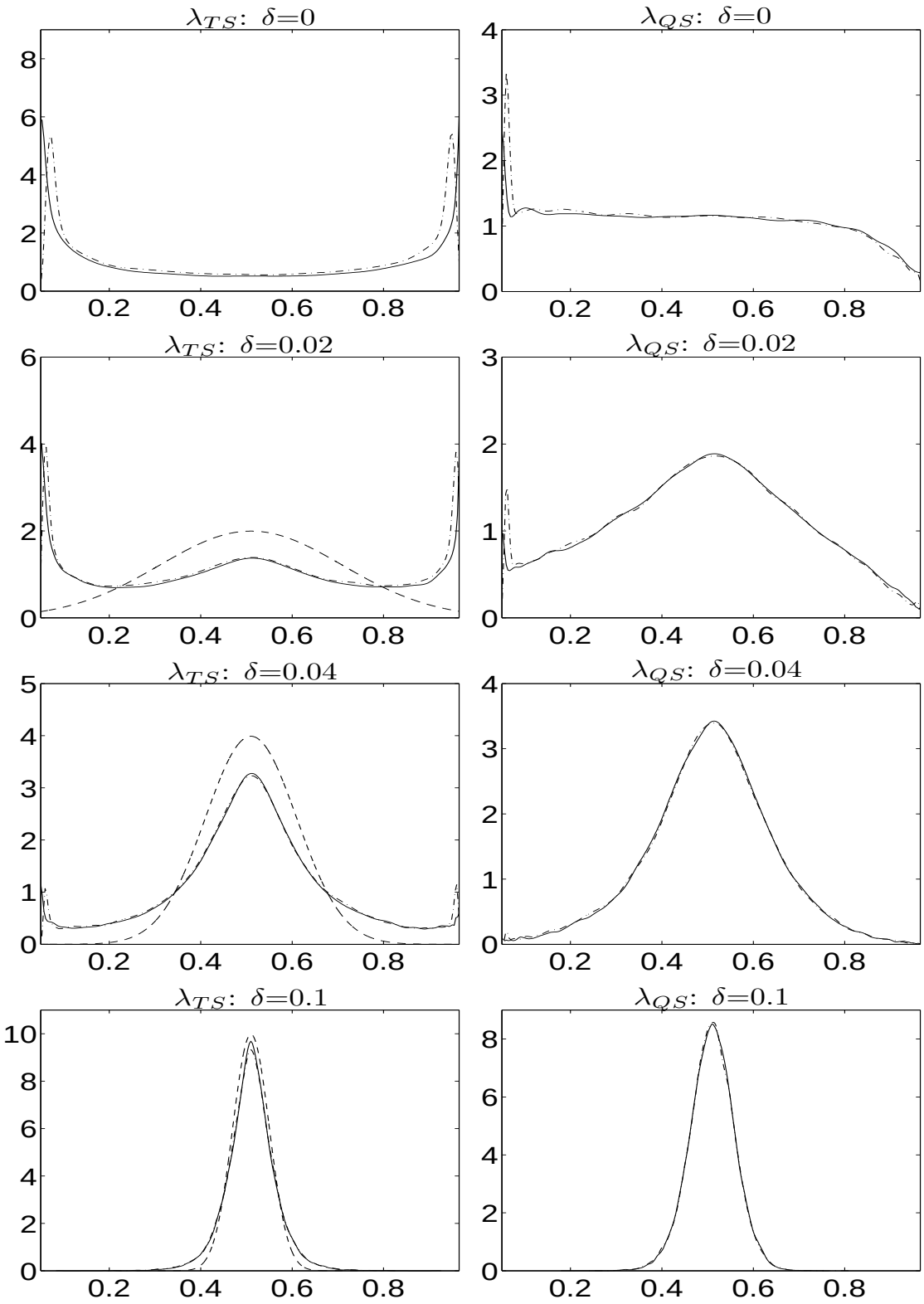

Figure 1.12. Finite and asymptotic pdf by Theorem 1.6.2 and PZ(2005) (only for $\hat{\lambda}_{TS}$) of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{QS}$ ("TS-QS") under I(0) $u_t's$, $\lambda^c = 0.5$, $T = 100$, $d(1) = 1$, and $\delta = 0, 0.02, 0.04, 0.1$. Solid: finite sample; dash-dot: Theorem 2; dash: PZ(left) or Bai(right).

sample behaviors of this estimator. On the left column of Figure 1.12 are the finite sample histograms ($T = 100$) of $\hat{\lambda}_{TS}$ (The histograms on the right column are for the break point estimator from the partial sum model which will be analyzed in Theorem 1.6.2). The finite sample histograms exhibit not only non-normal distributions but also the complicated tail behaviors of the break point estimator. The limiting distribution for $\hat{\lambda}_{QS}$ is not available in the literatures. Using the same approach as Theorem 1.5.1, I derive new asymptotic limits of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{QS}$ from the trend shift model under I(0) errors and its partial sum model. The proof is similar to that of Theorem 1.5.1. (See Appendix A.5)

Theorem 1.6.2 *Suppose the regressions of the level model (1.2.1) and its partial sum (1.6.20) are estimated using $\lambda \in \Lambda \subseteq (0,1)$ and $T_b^c \doteq \lambda^c T$ is the true break. Under assumption $(A1.c)$ and $(A2.d)$, the break point estimators by minimizing the $SSR(\lambda)$ have the limiting distributions as follows.*

1. For the level model (1.2.1) under I(0) errors (A1.c),

$$\hat{\lambda}_{TS} \Rightarrow \arg\max_{\lambda\in\Lambda}\{\frac{[\int_0^1 F(r,\lambda)dW(r) + \frac{\delta^*}{d(1)}\int_0^1 F(r,\lambda)F(r,\lambda^c)dr]^2}{\int_0^1 F(r,\lambda)^2 dr}\} \tag{1.6.24}$$

where

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

which implies the approximation

$$\hat{\lambda}_{TS} - \lambda^c \approx \arg\max_{\lambda\in\Lambda}\{\frac{[\int_0^1 F(r,\lambda)dW(r) + M\int_0^1 F(r,\lambda)F(r,\lambda^c)dr]^2}{\int_0^1 F(r,\lambda)^2 dr} - \lambda^c\}, \tag{1.6.25}$$

where $M = \frac{\delta^*}{d(1)} \equiv \frac{\delta T^{3/2}}{d(1)}$.

2. For the partial sum model (1.6.20),

$$\hat{\lambda}_{QS} \Rightarrow \arg\max_{\lambda\in\Lambda}\{\frac{[\int_0^1 Q(r,\lambda)W(r)dr + \frac{\delta^*}{d(1)}\int_0^1 Q(r,\lambda)Q(r,\lambda^c)dr]^2}{\int_0^1 Q(r,\lambda)^2 dr}\} \tag{1.6.26}$$

where

$$Q(r,\lambda) \doteq \frac{(r-\lambda)^2}{2}1(r>\lambda) - (-\lambda + 2\lambda^2 - 2\lambda^4 + \lambda^5)r - (\frac{1}{2} - \frac{5\lambda^2}{3} + \frac{5\lambda^4}{2} - \frac{4\lambda^5}{3})r^2.$$

which implies the approximation

$$\hat{\lambda}_{QS} - \lambda^c \approx \arg\max_{\lambda\in\Lambda}\{\frac{[\int_0^1 Q(r,\lambda)W(r)dr + M\int_0^1 Q(r,\lambda)Q(r,\lambda^c)dr]^2}{\int_0^1 Q(r,\lambda)^2 dr} - \lambda^c\}. \tag{1.6.27}$$

Similar to Theorem 1.5.1, the distribution can be decomposed into two parts: one is determined by the null asymptotic distribution and the other is determined by the break magnitude. Also, the asymptotic distributions are continuous in the break magnitude. For the same reason, both $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{QS}$ are consistent to some extent when M is big. Figure 1.12 shows similar patterns of a comparison between the break point estimators from the "TS-QS" models and those from "TS-MS" . Generally with small δ^*, the break point estimator from the partial sum model has thinner tails than from the level model. This means that for small values of the break magnitude, it would be better to use the partial sum model to obtain the break point estimator.

1.7 Application to One-step Ahead Forecasts

The previous analysis shows $\hat{\lambda}_{TS}$ could be more precise than $\hat{\lambda}_{MS}$, which means choosing $\hat{\lambda}_{TS}$ might be sensible in applications that use break point estimates, such as modeling, tests, and forecasts. In this section, we will see if the thinner tails of $\hat{\lambda}_{TS}$ can result in better focasts compared to $\hat{\lambda}_{MS}$. Ng and Vogelsang (2002)(NV hereafter) discuss the forecasting of the dynamic time series in the presence of the deterministic components, where the MSE of the forecast is considered as a criterion to evaluate different modeling approaches. Two approaches, OLS_1 and OLS_2, defined in NV(2002), are used for the trend shift model in this chapter. The definition of these approaches are described as follows.

Model (1.2.1) is the focus, which is the same model used by NV(2002). Let m_t denote the deterministic part of y_t and $m_t \doteq \mu + \beta t + \delta DT_t(\lambda^c)$. The error is defined as

$$u_t \quad = \quad \alpha u_{t-1} + \varepsilon_t. \tag{1.7.28}$$

where $\varepsilon_t \sim i.i.d.\ N(0,1)$. The assumption on u_t is extended to cover both I(1) and I(0) cases by allowing $|\alpha| \leq 1$.

1. OLS_1

The data generating process (DGP) can be written as

$$y_t \quad = \quad d_0 + d_1 t + d_2 DT_t(\lambda^c) + \alpha y_{t-1} + \varepsilon_t. \tag{1.7.29}$$

The feasible one-step forecast:

$$y_{t+1|T} \quad = \quad \hat{d}_0 + \hat{d}_1 t + \hat{d}_2 DT_t(\lambda^c) + \hat{\alpha} y_{t-1} + \varepsilon_t. \tag{1.7.30}$$

The OLS_1 approach first applies OLS to equation (1.7.29) to obtain $\hat{d}_0$, $\hat{d}_1$, $\hat{d}_2$, $\hat{\alpha}$, and $\hat{u}_t$; then the estimated parameters are used in equation (1.7.30) to obtain the $y_{T+1|T}$.

2. OLS_2

The feasible one-step forecast is given by

$$\hat{y}_{T+1|T} \quad = \quad \hat{m}_{T+1} + \hat{\alpha}(y_T - \hat{m}_T). \tag{1.7.31}$$

The OLS_2 approach first applies OLS to equation (1.2.1) to obtain $\hat{u}_t$, $\hat{\mu}$, $\hat{\beta}$, and $\hat{\delta}$; then applies OLS to equation (1.7.28) with $\hat{u}_t$ to obtain $\hat{\alpha}$; finally, $\hat{y}_{T+1|T}$ is obtained based on equation (1.7.31).

The closer $\hat{\lambda}$ is to the true break λ^c, the closer $DT_t(\hat{\lambda})$ is to $DT_t(\lambda^c)$, and hence the less model misspecification is a concern. A smaller mismatch between the estimated model and the true model leads to a smaller MSE of forecast. I compare the MSE of one-step forecasts by OLS_1 and OLS_2 to illustrate the effect of different accuracies of the break point estimators $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ on the forecasts.

Table 1.1. Mean squared error (MSE) of one-step ahead forecasts of the trend shift model (1.2.1) under I(1) errors (For one-step forecast $\hat{y}_{t+1}$, the MSE is defined as $(\hat{y}_{t+1}-y_{t+1})^2$.) Simulation settings: $\lambda^c = 0.5$; $\delta = 0, 0.1, 0.2, 0.3, 0.4, 0.5$; u_t is I(1); $T = 101$; and $N = 10,000$. OLS_1^* and OLS_2^* assume that u_t is known to be I(1).

	OLS_1		OLS_2		OLS_1^*		OLS_2^*	
δ	$\hat{\lambda}_{TS}$	$\hat{\lambda}_{MS}$	$\hat{\lambda}_{TS}$	$\hat{\lambda}_{MS}$	$\hat{\lambda}_{TS}$	$\hat{\lambda}_{MS}$	$\hat{\lambda}_{TS}$	$\hat{\lambda}_{MS}$
0	1.214	1.300	1.240	1.267	1.107	1.323	1.049	1.090
0.1	1.208	1.301	1.230	1.269	1.109	1.317	1.046	1.089
0.2	1.185	1.286	1.209	1.290	1.093	1.282	1.033	1.090
0.3	1.174	1.258	1.195	1.302	1.091	1.237	1.025	1.087
0.4	1.159	1.231	1.181	1.301	1.083	1.197	1.016	1.076
0.5	1.148	1.196	1.167	1.291	1.078	1.162	1.010	1.063

First, I provide simulation results with the setting: $\lambda^c = 0.5$; $\delta = 0, 0.1, 0.2, 0.3, 0.4, 0.5$; u_t is I(1); $T = 101$; and $N = 10,000$. Table 1.1 gives the OLS_1 and OLS_2 MSE of one-step forecasts with $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ under different δ. The MSE of one-step forecasts using $\hat{\lambda}_{TS}$ are smaller than those using $\hat{\lambda}_{MS}$ with both OLS_1 and OLS_2 methods for all $\delta's$ in this example. This happens because $\hat{\lambda}_{TS}$ concentrates more around λ^c and has thinner tails than $\hat{\lambda}_{MS}$, which leads to less misspecification in $DT_t(\hat{\lambda})$. With the same break point estimator, the MSE by OLS_2 are mostly bigger than OLS_1, which is consistent to the conclusion in NV(2002).

Next, I describe an empirical illustration of forecast errors using $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$. Similar to PZ(2005), I estimate the break points and calculate one-step forecast of the annual (log) real per capita GDP series between 1870 and 1996. All the data are taken from the Groningen Growth and Development Centre [2] (See Figure 1.13 and 1.14). World War I and II along with other factors may affect the location of possible breaks in different ways on different countries. Sweden seems to have a break in around 1920, while Italy more likely has the break in around 1945. To choose the series with errors that are likely I(1), I follow PZ(2005) and use the series of Italy, Norway and Sweden (See Figure 1.13). I use the trend

[2] http://www.ggdc.net/maddison/Historical_Statistics.

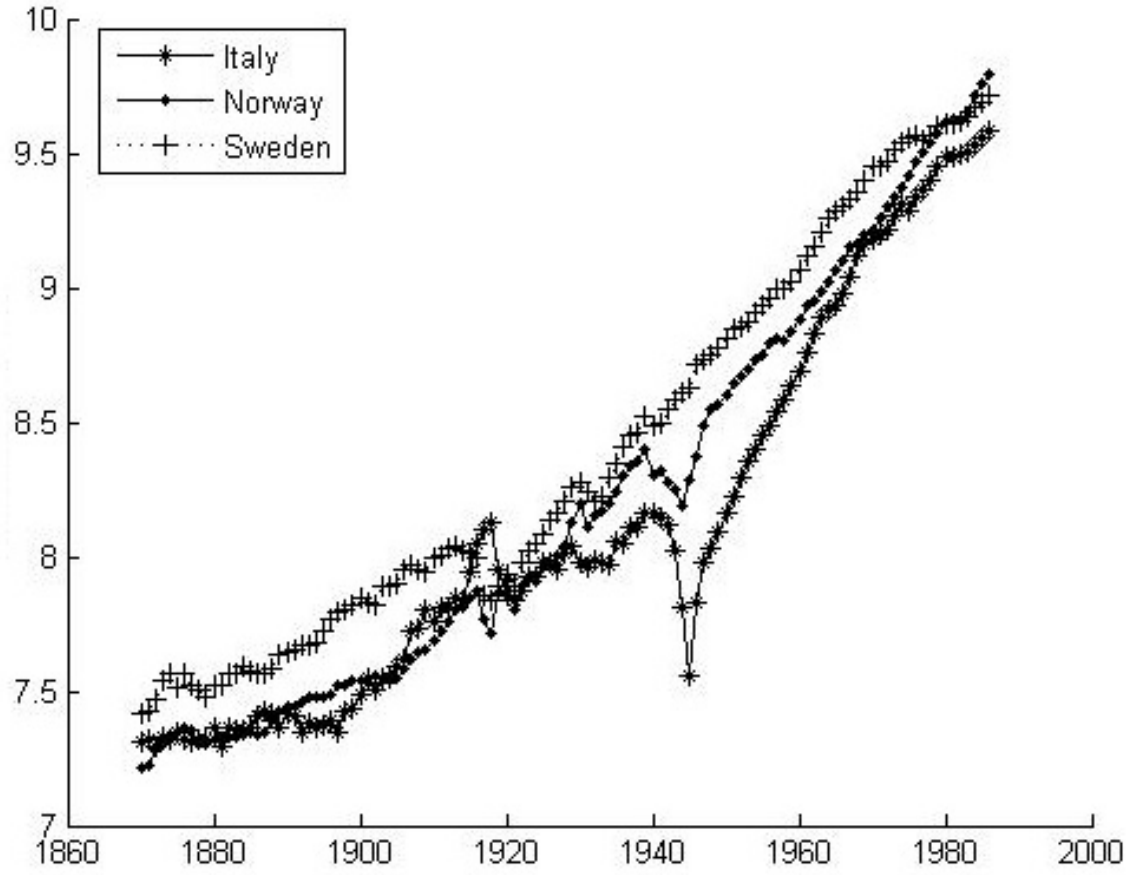

Figure 1.13. The real (log) per capita GDP of Italy, Norway, and Sweden, which are of I(1) errors. x-axes: year; y-axes: (log)Per Capita GDP.

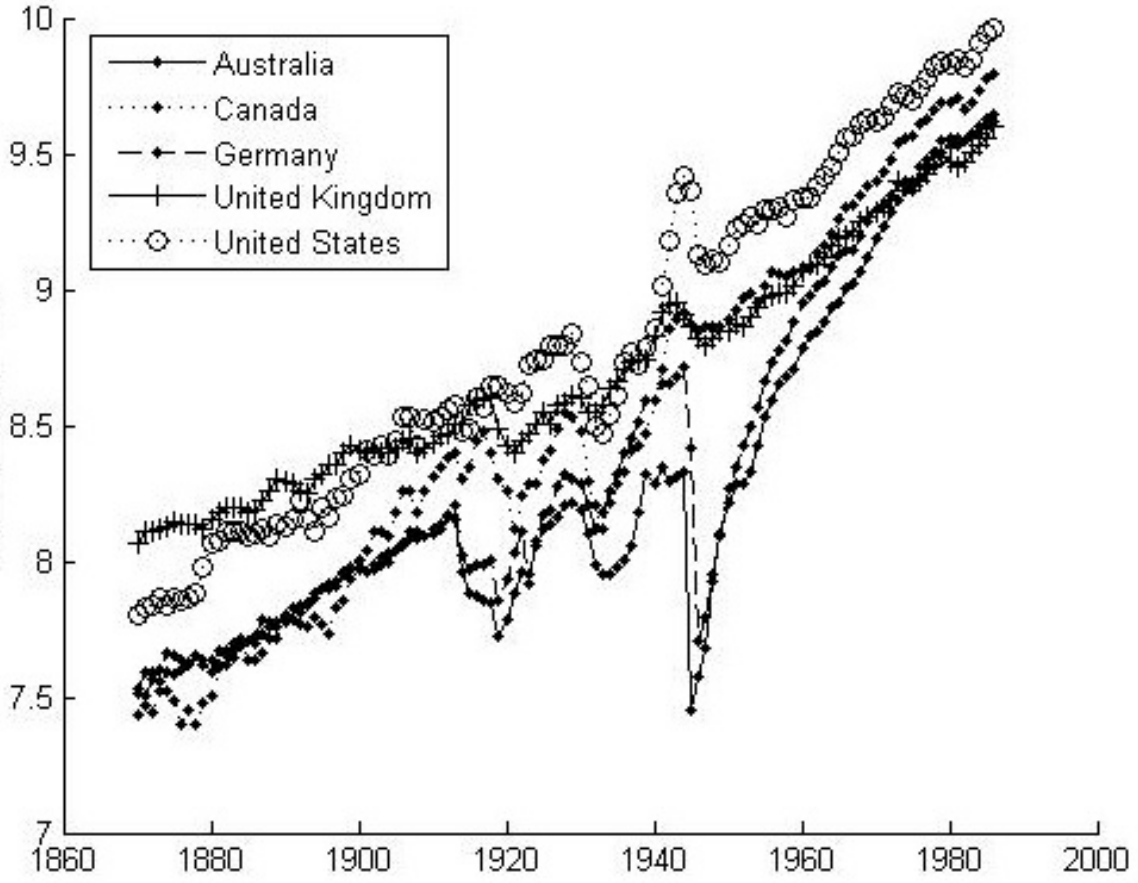

Figure 1.14. The real (log) per capita GDP of Australia, Canada, Germany, UK, and US, which are of I(0) errors. x-axes: year; y-axes: (log)Per Capita GDP.

Table 1.2. MSE in one-step forecast with the log real per capita GDP series (1870-1996) with trend shift model (1.2.1) and different break point estimators ($\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$), where MSE of one-step forecast are calculated for 1987-1996.

	OLS_1		OLS_2	
Country	$\hat{\lambda}_{TS}$	$\hat{\lambda}_{MS}$	$\hat{\lambda}_{TS}$	$\hat{\lambda}_{MS}$
Italy	0.0016	0.0017	0.0018	0.0015
Norway	0.0003	0.0006	0.0002	0.0008
Sweden	0.0014	0.0015	0.0016	0.0014

shift model (1.2.1) for parameter estimation. The one-step forecast is only applied on the data from year 1987 to 1996, based on estimated model using data of the whole period prior to each forecasted year. Consider the one-step forecast in 1990 as an example. It is based on the model estimated using the data from 1970 to 1989, and the MSE error is computed as $(\hat{y}_{1990} - y_{1990})^2$. The overall error is the average MSE over these years. Table 1.2 lists the average MSE errors of forecast for the real per capita GDP from 1987 to 1996 by OLS_1 and OLS_2. For OLS_1, the MSE using $\hat{\lambda}_{TS}$ is smaller than that using $\hat{\lambda}_{MS}$ for all three countries. For OLS_2, it is hard to conclude which break point estimator leads to a smaller MSE.

Since some series have I(0) errors around the trend and Theorem 1.6.2 reveals the advantage of $\hat{\lambda}_{QS}$ over $\hat{\lambda}_{TS}$ under small break magnitude, it would be interesting to look at both the Monte Carlo simulations and the empirical data to see how $\hat{\lambda}_{QS}$ and $\hat{\lambda}_{TS}$ behave in finite samples. Table 1.3 lists the MSE of one-step forecast by OLS_1 and OLS_2 with $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{QS}$ under different δ. The setting is: $\lambda^c = 0.5$; $\delta = 0, 0.01, 0.02, 0.03, 0.04, 0.05$; u_t is I(0); $T = 101$; and $N = 10,000$. The MSE of one-step forecast using $\hat{\lambda}_{QS}$ are smaller than those using $\hat{\lambda}_{TS}$ by both OLS_1 and OLS_2 for all $\delta' s$, which is what we expected.

Next $\hat{\lambda}_{QS}$ is used to see whether its more concentration around λ^c can lead to smaller MSE in one-step forecast than $\hat{\lambda}_{TS}$ for the GDP series. To choose the (log) GDP series with I(0) errors, I follow PZ(2005) and choose Australia, Germany, United Kingdom, and United States. Similar to the previous application, I choose the data from 1870 to 1996.

Table 1.3. Mean squared error (MSE) of one-step ahead forecasts of the trend shift model (1.2.1) under I(0) errors (For one-step forecast $\hat{y}_{t+1}$, the MSE is defined as $(\hat{y}_{t+1} - y_{t+1})^2$ for point forecasts.) Simulation settings: $\lambda^c = 0.5$; $\delta = 0, 0.01, 0.02, 0.03, 0.04, 0.05$; u_t is I(0); $T = 101$; and $N = 10,000$. OLS_1^* and OLS_2^* assume that u_t is known to be I(0).

	OLS_1		OLS_2		OLS_1^*		OLS_2^*	
δ	$\hat{\lambda}_{TS}$	$\hat{\lambda}_{QS}$	$\hat{\lambda}_{TS}$	$\hat{\lambda}_{QS}$	$\hat{\lambda}_{TS}$	$\hat{\lambda}_{QS}$	$\hat{\lambda}_{TS}$	$\hat{\lambda}_{QS}$
0	1.454	1.158	1.453	1.155	2.539	2.204	2.124	2.031
0.01	1.417	1.152	1.416	1.149	2.502	2.202	2.114	2.030
0.02	1.336	1.134	1.335	1.132	2.400	2.197	2.091	2.026
0.03	1.261	1.110	1.260	1.108	2.303	2.188	2.068	2.024
0.04	1.191	1.098	1.190	1.096	2.238	2.179	2.049	2.022
0.05	1.150	1.094	1.149	1.092	2.204	2.173	2.037	2.022

Table 1.4. MSE in one-step forecast with the log real per capita GDP series (1870-1996) with trend shift model (1.2.1) and different break point estimators ($\hat{\lambda}_{TS}$ and $\hat{\lambda}_{QS}$), where MSE of one-step forecast is calculated for 1987-1996.

	OLS_1		OLS_2	
Country	$\hat{\lambda}_{TS}$	$\hat{\lambda}_{QS}$	$\hat{\lambda}_{TS}$	$\hat{\lambda}_{QS}$
Australia	0.0035	0.0007	0.0033	0.0021
Germany	0.0028	0.0012	0.0031	0.0015
United Kingdom	0.0004	0.0005	0.0004	0.0009
United States	0.0003	0.0002	0.0002	0.0003

Figure 1.14 shows the raw data of the (log) real per capita GDP of these countries. I use the same forecast methods on this data set. The one-step forecasts are provided from 1987 to 1996, and the overall error is the average MSE over 10 years. Table 1.4 lists the average MSE of one-step forecast for the real per capita GDP from 1987 to 1996 when $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{QS}$ are used. We can see that, for United Kingdom and United States, the MSE of one-step forecast is similar when $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{QS}$ are applied under both OLS_1 and OLS_2 forecasts. For Australia and Germany, MSE of one-step forecast with $\hat{\lambda}_{QS}$ is considerably lower.

1.8 Conclusions

In this chapter, I derive a new asymptotic theory for two break point estimators: one ($\hat{\lambda}_{TS}$) is from the trend shift model and the other ($\hat{\lambda}_{MS}$) is from the first difference, the mean shift model. Existing theories do not fully capture the finite sample behaviors, especially the tail behavior of the finite sample distributions with small break magnitude. This discrepancy is stronger when the true break is not in the middle of the sample. To better approximate the finite sample distributions, a new asymptotic theory is developed under the assumption that the break magnitude is within a small neighborhood of zero. The new asymptotic theory captures the finite sample behaviors of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$, especially the tails in the densities. Under the same break magnitude, $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ are compared in precision using the new asymptotics. Both theoretical analysis and simulations reveal that, under small break magnitude, $\hat{\lambda}_{TS}$ concentrates more around the true break. Using $\hat{\lambda}_{TS}$ instead of $\hat{\lambda}_{MS}$ can decrease the MSE in one-step ahead forecasts.

There are other potentially interesting topics accompanying the comparison of the break point estimators using the new approximation. A possible improvement of break tests could be achieved if we choose the break estimator properly according to break magnitude in a data dependent way. Also, this limiting distribution analysis of the single break estimators would help the research on the multiple break point estimates, e.g. the break point estimates in the presence of under-specification of the break numbers.

CHAPTER 2

Fixed-b Analysis of LM Type Tests for a Shift in Mean

2.1 Introduction

In this chapter we provide a theoretical analysis of lagrange multiplier (LM) tests for a shift in the mean of a univariate time series at an unknown date. We consider a class of LM statistics based on nonparametric kernel estimators of the long run variance. The main theoretical contribution of this part is to develop a fixed-b asymptotic theory for the long run variance estimator. The fixed-b limit of the LM statistics depends on the kernel and bandwidth needed to implement the long run variance estimator and the fixed-b limit also depends on the magnitude of the mean shift under the alterative. This allows us to theoretically capture to impact of the choice of bandwidth on both the size and power of the tests. In particular we show that the bandwidth plays an important role on determining whether the tests exhibit non-monotonic power (power that is not necessarily increasing as the magnitude of the mean shift increases). Small bandwidths lead to tests with monotonic power whereas large bandwidths lead to non-monotonic power.

We derive fixed-b results for both the case of stationary $I(0)$ errors and nearly integrated

$I(1)$ errors. We obtain an unexpected and very useful finding. There exist bandwidths such that, for a given significance level, the critical values of the LM statistics are the same for both $I(0)$ and $I(1)$. Use of these "robust" bandwidths and the associated fixed-b critical values provides tests that are asymptotically robust to whether the errors are $I(0)$ or $I(1)$. Such a simple way of obtaining robustness to the strength of serial correlation in the errors should appeal to empirical researchers. Our robust LM tests complement the $I(0)/I(1)$ robust tests in literature that have been developed for Wald-type tests. See Vogelsang (1997), Vogelsang (1998) and Sayginsoy and Vogelsang (2010). While the various $I(0)/I(1)$ robust tests are asymptotically valid whether the errors are $I(0)$ or $I(1)$, in finite samples the Wald-type tests tend to over-reject when there is a negative moving average component and an autoregressive root near one in the errors. In contrast, the robust LM tests do not over-reject in this case although they do tend to over-reject when there is a negative moving average component but no autoregressive component is present. These complementary finite sample properties of robust $Wald$ and LM tests could be exploited to provide more robust inference overall.

The approach and analysis in this chapter is related to some recent papers in the econometrics literature on LM tests for a shift in mean. The possibility of non-monotonic power of LM tests for a shift in mean where documented by Vogelsang (1999). The reason that power can be non-monotonic is simple. LM statistics use long run variance estimators based on residuals from the model estimated under the null hypothesis of no mean shift. Therefore, when there is a shift in mean, the long run variance estimator is not invariant to the magnitude of the mean shift. A large shift in mean can cause the denominator of an LM statistic to be large and this can cause power to be low. While Vogelsang (1999) pin-pointed the long run variance estimator as the source of non-monotonic power, he did not examine the role played by the choice of bandwidth. A recent paper by Kejriwal (2009) proposed the use of a hybrid long run variance estimator that can restore monotonic power to LM tests. The hybrid estimator blends components of long run variance estimators based on

null and alternative residuals. We show in this chapter that there is a direct link between the statistics proposed by Kejriwal (2009) and the LM statistics based on a specific bandwidth choice. Our theory shows that an explanation for the monotonic power of the statistics proposed by Kejriwal (2009) is the use of a data dependent bandwidth based on alternative residuals.

The results in this chapter on non-monotonic power add to a small but growing literature on non-monotonic power of tests for a shift mean. This literature was started by Perron (1991) where simulation results were given for some well known tests for a shift in mean. Vogelsang (1999) provided some theoretical explanations for non-monotonic power for a large group of statistics. Other papers have given results for specific statistics with contributions by Vogelsang (1997), Crainiceanu and Vogelsang (2007), Deng and Perron (2008), Juhl and Xiao (2009) and Kejriwal (2009). Our research parallels the work by Crainiceanu and Vogelsang (2007) in establishing a direct link between the bandwidth and non-monotonic power. Crainiceanu and Vogelsang (2007) showed that the bandwidth choice of the long run variance estimator is directly linked to non-monotonic power of CUSUM and related tests for a shift in mean.

The remainder of the chapter is organized as follows. In the next section we describe the model and lay out the assumptions. In Section 3 we define the statistics and their finite sample properties are illustrated Section 4. In Section 5 we develop the fixed-b asymptotic theory for the LM tests and show that the fixed-b theory explains the important finite sample patterns. In Section 6 we compute the $I(0)/I(1)$ "robust" bandwidths and examine their finite sample performance when used with fixed-b critical values. For the most part, the bandwidths effectively control the over-rejection problem caused by strong serial correlation and in some cases retain good power. Section 7 establishes a direct relationship between the hybrid Wald statistics proposed by Kejriwal (2009) and the LM statistics. The proof of the main theoretical result of this chapter is given in an appendix.

2.2 Model and Assumptions

Consider a simple mean shift model

$$y_t = \mu + \delta DU_t(T_b) + u_t, \quad t = 1, 2, ..., T, \tag{2.2.1}$$

where $DU_t(T_b) = 1(t > T_b)$ and $1(\cdot)$ is the indicator function. We denote the true break data as T_b^0, and following standard practice in the structural change literature, we assume that the break point, $\lambda_0 = T_b^0/T$, remains fixed as the sample size increases. Throughout this chapter we assume that λ_0 in unknown. Following Canjels and Watson (1997) and Bunzel and Vogelsang (2005) among others we assume the error term is given by

$$u_t = \rho u_{t-1} + \varepsilon_t, \, t = 1, \cdots, T \tag{2.2.2}$$

$$\varepsilon_t = d(L)e_t, \quad d(L) = \sum_{i=0}^{\infty} d_i L^i, \quad \sum_{i=0}^{\infty} i|d_i| < \infty, \quad d(1)^2 > 0 \tag{2.2.3}$$

where 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 $E(e_t^2|e_{t-1}, e_{t-2}, \cdots) = 1$. When $|\rho| < 1$, the errors are $I(0)$ and when $\rho = 1 - c/T$, where c is a constant the errors are nearly $I(1)$. The pure unit root error case is given when $c = 0$.

Under assumptions (2.2.2) and (2.2.3), some standard results are (see, for example, Phillips (1987)):

$$T^{-1/2} \sum_{t=1}^{[rT]} u_t \Rightarrow \sigma W(r) \text{ if } |\rho| < 1,$$

$$T^{-1/2} u_{[rT]} \Rightarrow d(1) V_c(r) \text{ if } \rho = 1 - \frac{c}{T},$$

where $\sigma^2 = d(1)^2/(1-\rho)^2$, $W(r)$ is a standard Wiener process, $V_c(r) = \int_0^r \exp\{-c(r-s)\} dW(s)$, $[rT]$ is the integer part of rT where $r \in [0, 1]$ and $\Rightarrow$ denotes weak convergence.

The parameter σ^2 needs to be estimated in order to test the null hypothesis that the mean of y_t is stable. Here we focus on the class of nonparametric spectral density estimators given by,

$$\tilde{\sigma}^2(m) = \tilde{\gamma}_0 + 2 \sum_{j=1}^{T-1} k(j/m) \tilde{\gamma}_j, \; \tilde{\gamma}_j = T^{-1} \sum_{t=j+1}^{T} \tilde{u}_t \tilde{u}_{t-j};$$

where $\tilde{u}_t = y_t - \bar{y}$ are the OLS residuals from regression (2.2.1) with $\delta = 0$ (no shift in mean) imposed on the model. As usual, $k(x)$ is the kernel function and m is the bandwidth (or truncation lag for kernels that truncate). A kernel is labelled type 1 if $k(x)$ is twice continuously differentiable everywhere, and as a type 2 kernel if $k(x)$ is continuous, twice continuously differentiable everywhere except at $|x| = 1$ and $k(x) = 0$ for $|x| \geq 1$. For type 2 kernels define the derivative from the left at $x = 1$ as $k'_{-}(1) = \lim_{h \to 0} [(k(1) - k(1-h))/h]$.

2.3 LM Tests for a Shift in Mean

We focus on testing the null hypothesis that there is no shift in mean:

$$H_0 : \delta = 0,$$

against the alternative

$$H_1 : \delta \neq 0.$$

For a given break date, T_b define the LM test as

$$LM(T_b, m) = \frac{SSR_0 - SSR(T_b)}{\tilde{\sigma}^2(m)}$$

where $SSR_0 = \sum_{t=1}^{T} \tilde{u}_t^2$ is the sum of squared residuals under the null hypothesis and $SSR(T_b)$ is the sum of squared residuals from the regression

$$y_t = \mu + \delta DU_t(T_b) + u_t. \tag{2.3.4}$$

Because we treat the break date as unknown, we follow Andrews (1993) and Andrews and Ploberger (1994) and consider supremum and mean tests of the form

$$MeanLM_m = T^{-1} \sum_{T_b \in \Lambda^*} LM(T_b, m), \qquad SupLM_m = \sup_{T_b \in \Lambda^*} LM(T_b, m)$$

where $\Lambda^* = \{T_b^*, T_b^* + 1, \cdots, T - T_b^*\}$ is the set of possible break dates. The parameter $\lambda^* = T_b^*/T$ is held fixed as T increases and λ^* determines the amount of trimming used in computing the statistics. Note that because T_b only shows up through the $-SSR(T_b)$ component of $LM(T_b, m)$, it follow that $SupLM_m = LM(\widehat{T}_b, m)$ where $\widehat{T}_b = \arg\min_{T_b \in \Lambda^*} SSR(T_b)$.

2.4 Finite Sample Behavior of the LM Tests

In this section we use a simple simulation design to illustrate the impact of the bandwidth choice on the performance of the Mean and Sup LM statistics. We generate data according model (2.2.2) for the case where $d(L) = 1$ and e_t is an iid $N(0, 1)$ process, i.e. a Gaussian $AR(1)$ model. We set $\mu = 0$ without loss of generality. We focus exclusively on the quadratic spectral (QS) kernel and we consider two data dependent bandwidth rules for m based on Andrews (1991) using the AR(1) plug-in method. In the first case we use the null OLS residuals, $\widetilde{u}_t$ when computing the AR(1) estimate needed for the bandwidth formula. In the second case we compute the AR(1) estimate using the alternative OLS residuals

$$\widehat{u}_t(\widehat{T}_b) = y_t - \widehat{\mu} - \widehat{\delta} DU_t(\widehat{T}_b)$$

where $\widehat{\mu}$, $\widehat{\delta}$ are the OLS estimates from regression (2.3.4) using the estimated break date $\widehat{T}_b = \arg\min_{T_b \in \Lambda^*} SSR(T_b)$ We denote the respective bandwidths by $\widetilde{m}$ and $\widehat{m}$. We report results for the sample size $T = 120$ and we use $5,000$ replications in all cases. We use 15% trimming, i.e. $\lambda^* = 0.15$. We compute rejection probabilities using critical values taken from the I(0) asymptotic distribution of the statistics using results in Andrews (1993) and Andrews and Ploberger (1994) which require consistency of $\tilde{\sigma}^2(m)$. Results are reported for the nominal level of 0.05.

Empirical null rejections are reported in Table 2.1 and 2.2. The first column gives the values of ρ used in the simulations. Columns two through four give results for the LM tests using the null bandwidth, $\widetilde{m}$ and whereas columns five through seven give results using the

alternative bandwidth, $\widehat{m}$. For each bandwidth we report the average, across replications, of the bandwidth relative to the sample size: $\widetilde{b} = \widetilde{m}/T$ and $\widehat{b} = \widehat{m}/T$. Some interesting patterns appear in the table. When the data is iid ($\rho = 0$), both data dependent bandwidths are small and the tests have rejections not far from the nominal level. As ρ increases and serial correlation becomes stronger, both bandwidths increase although $\widetilde{b}$ increases faster than $\widehat{b}$; in fact $\widetilde{b}$ is quite large when $\rho = 1$. This makes sense because $\widetilde{b}$ is based on an AR(1) estimate that has less downward bias and is hence closer to one when ρ is close to one and this inflates the bandwidth. For the $MeanLM$ statistics, we see that rejections become larger than 0.05 as ρ becomes closer to one with severe over-rejections when $\rho = 1$. This is the usual over-rejection problem cased by a unit root in the error. The pattern for the $SupLM$ statistics are different. As ρ increases, both $SupLM$ statistics tend to under-reject and when $\rho = 1$, $SupLM_{\widehat{m}}$ substantially under-rejects whereas $SupLM_{\widetilde{m}}$ slightly over-rejects. It is surprising that $SupLM_{\widehat{m}}$ under-rejects in the unit root case. Clearly there is a complicated relationship between the values of ρ, the bandwidths and whether a test tends to over-reject or under-reject.

Figures 2.1, 2.2, 2.3 and 2.4 depict finite sample power for the case where the break occurs in the middle of the sample ($\lambda_0 = 0.5$). The four panels correspond to the values of $\rho = 0.0.0.5, 0.9, 1.0$. For the cases where $\rho < 1$, we see that power is non-monotonic when the data dependent bandwidth is computed under the null. This finding of non-monotonic power for LM-type tests was also documented by Vogelsang (1999) and Kejriwal (2009). Interestingly, if the data dependent bandwidth is based on alternative residuals, we see that the LM tests have monotonic power. This suggests the bandwidth has an important effect on finite sample power functions. Table 2.3 and 2.4 and reports the average bandwidth ratios (across replications) for the four values of ρ and a grid of values for the mean shift magnitude δ. Notice that as δ increases, $\widetilde{b}$ steadily grows and can become very large when the mean shift is large. This is not surprising given the well known result of Perron (1990). Because the AR(1) parameter is being estimated using null residuals, the estimated AR(1)

Table 2.1. Null Rejection Probabilities Using Standard ($b = 0$) $I(0)$ Critical Values, 5% Nominal Level, 15% Trimming, QS Kernel.

Panel A: $T = 120$

ρ	$ave(\widetilde{b})$	$MeanLM_{\tilde{m}}$	$SupLM_{\tilde{m}}$	$ave(\widehat{b})$	$MeanLM_{\hat{m}}$	$SupLM_{\hat{m}}$
0	0.012	0.022	0.013	0.012	0.052	0.04
0.5	0.048	0.026	0.006	0.044	0.069	0.038
0.7	0.082	0.018	0.001	0.072	0.077	0.028
0.9	0.199	0.002	0.002	0.142	0.096	0.012
1	0.589	0.219	0.041	0.214	0.265	0.016

Panel B: Fixed-b Asymptotic Rejections

	b	$MeanLM$	$SupLM$
$I(0)$	0.02	0.049	0.044
$I(1), c = 60$	0.04	0.092	0.047
	0.06	0.065	0.016
	0.08	0.051	0.006
$I(1), c = 36$	0.06	0.096	0.025
	0.08	0.069	0.006
	0.10	0.051	0.002
$I(1), c = 12$	0.12	0.093	0.000
	0.14	0.063	0.000
	0.16	0.041	0.001
	0.18	0.023	0.001
	0.20	0.014	0.003
	0.22	0.008	0.005
$I(1), c = 0$	0.18	0.281	0.000
	0.20	0.182	0.000
	0.22	0.077	0.000
	0.58	0.059	0.087
	0.60	0.071	0.098

Table 2.2. Null Rejection Probabilities Using Standard ($b = 0$) $I(0)$ Critical Values, 5% Nominal Level, 15% Trimming, Bartlett Kernel.

Panel A: $T = 120$

ρ	$ave(\widetilde{b})$	$MeanLM_{\tilde{m}}$	$SupLM_{\tilde{m}}$	$ave(\widehat{b})$	$MeanLM_{\hat{m}}$	$SupLM_{\hat{m}}$
0	0.012	0.051	0.037	0.013	0.056	0.042
0.5	0.055	0.079	0.042	0.050	0.090	0.058
0.7	0.088	0.085	0.023	0.078	0.104	0.051
0.9	0.185	0.066	0.002	0.140	0.162	0.035
1	0.501	0.191	0.012	0.197	0.501	0.042

Panel B: Fixed-b Asymptotic Rejections

	b	$MeanLM$	$SupLM$
$I(0)$	0.02	0.048	0.041
$I(1), c = 60$	0.04	0.139	0.106
	0.06	0.099	0.047
	0.08	0.077	0.023
$I(1), c = 36$	0.06	0.149	0.079
	0.08	0.110	0.033
	0.10	0.090	0.012
$I(1), c = 12$	0.12	0.175	0.004
	0.14	0.141	0.001
	0.16	0.111	0.000
	0.18	0.087	0.001
	0.20	0.069	0.001
	0.22	0.050	0.001
$I(1), c = 0$	0.18	0.478	0.000
	0.20	0.434	0.000
	0.22	0.382	0.000
	0.50	0.016	0.020
	0.52	0.019	0.026

Table 2.3. Finite Sample Behavior of Data Dependent Bandwidth to Sample Size Ratios, $T = 120$, QS Kernel.

	$\rho = 0$		$\rho = 0.5$		$\rho = 0.7$		$\rho = 0.9$		$\rho = 1$	
δ	$ave(\widetilde{b})$	$ave(\widehat{b})$	$ave(\widetilde{b})$	$ave(\widehat{b})$	$ave(\widetilde{b})$	$ave(\widehat{b})$	$ave(\widetilde{b})$	$ave(\widehat{b})$	$ave(\widetilde{b})$	$ave(\widehat{b})$
0	0.012	0.012	0.048	0.044	0.082	0.072	0.199	0.142	0.589	0.214
0.5	0.014	0.012	0.051	0.044	0.085	0.072	0.201	0.142	0.589	0.214
1	0.023	0.012	0.059	0.045	0.092	0.072	0.207	0.142	0.590	0.214
1.5	0.035	0.012	0.072	0.045	0.104	0.072	0.218	0.141	0.592	0.214
2	0.049	0.012	0.088	0.045	0.120	0.073	0.231	0.141	0.594	0.214
3	0.082	0.012	0.128	0.046	0.162	0.074	0.268	0.142	0.599	0.214
4	0.120	0.012	0.176	0.046	0.212	0.075	0.314	0.143	0.607	0.216
5	0.162	0.012	0.228	0.047	0.266	0.076	0.364	0.147	0.615	0.220
6	0.207	0.012	0.282	0.047	0.322	0.077	0.416	0.152	0.625	0.227
7	0.252	0.012	0.336	0.047	0.377	0.078	0.467	0.156	0.636	0.235
8	0.297	0.012	0.388	0.047	0.429	0.078	0.515	0.161	0.647	0.245
9	0.342	0.012	0.437	0.047	0.479	0.078	0.559	0.165	0.660	0.256
10	0.385	0.012	0.484	0.047	0.524	0.078	0.600	0.167	0.673	0.268
25	0.771	0.012	0.834	0.047	0.853	0.078	0.870	0.171	0.821	0.334
50	0.921	0.012	0.942	0.047	0.947	0.078	0.947	0.171	0.905	0.334
100	0.969	0.012	0.975	0.047	0.976	0.078	0.974	0.171	0.950	0.334

parameter approaches one as δ increases and this inflates the bandwidth. In contrast, $\widehat{b}$ changes very little as δ increases when ρ is small and $\widehat{b}$ increases much more slowly as δ increases when ρ is close to one. This reflects the fact that $\widehat{b}$ is based on the alternative residuals which are nearly invariant to δ when $\widehat{T}_b$ is close to T_b^0. It appears that large bandwidths are leading to tests with non-monotonic power.

To see the link between the bandwidth and monotonic power more clearly, we simulated finite sample for the case of $T = 120$ and $\rho = 0.7$ using both data dependent bandwidths and several fixed values of b. Results for the $MeanLM$ and $SupLM$ statistics are given, respectively, in the top panels of Figures 2.5, 2.6, 2.7 and 2.8. For $MeanLM$ we see that power is monotonic for $b = 0.02, 0.1$. For $b = 0.18$ power is lower but is still monotonic. By just increasing b to 0.2, power suddenly becomes non-monotonic. This is the equivalent of changing m from 22 to 24 given the sample size of 120. As b increases further power completely collapses. Similar patterns hold for $SupLM$ although the change from monotonic power to non-monotonic power happens more quickly as b increases. Because

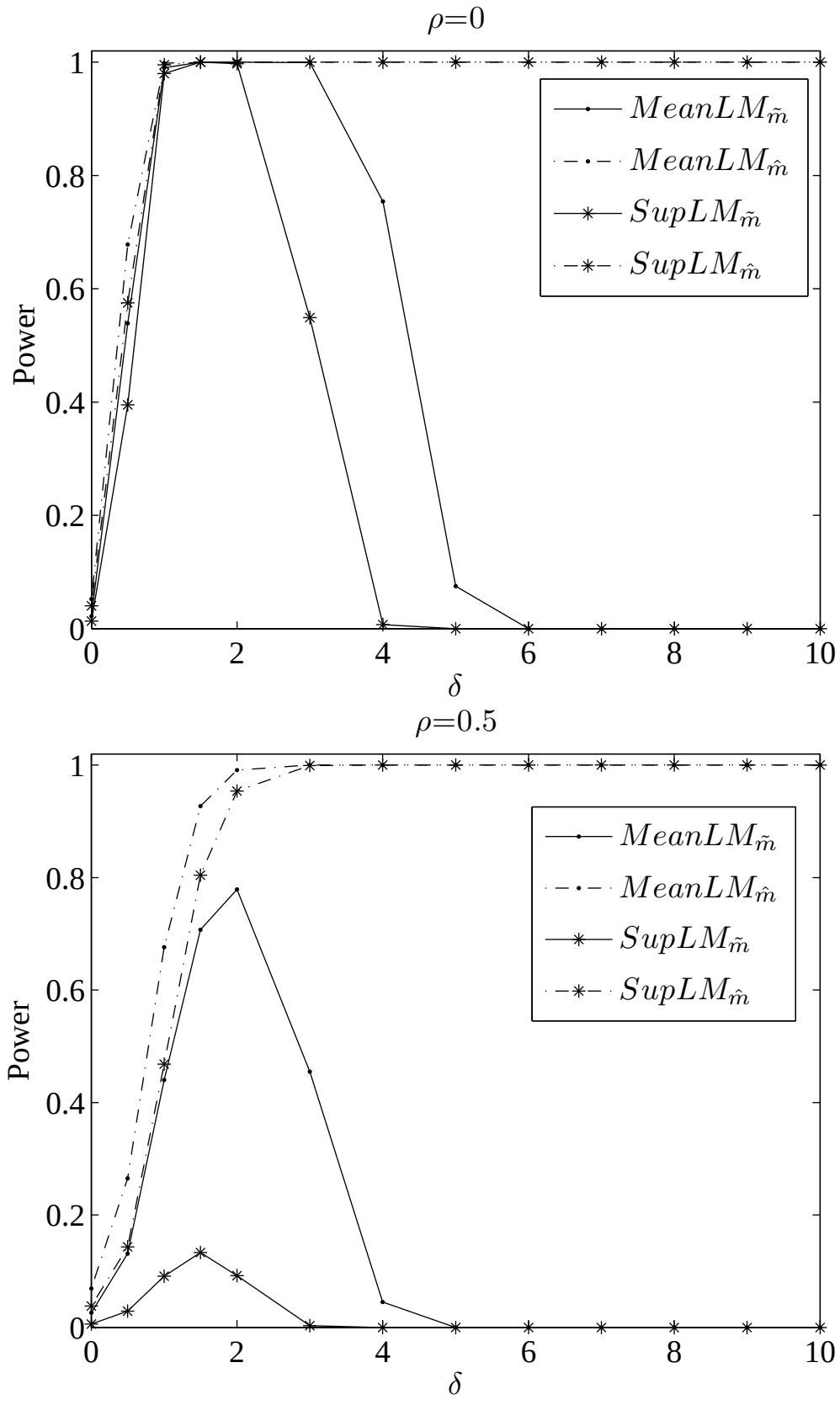

Figure 2.1. Finite Sample Power, QS kernel, 15% Trimming.

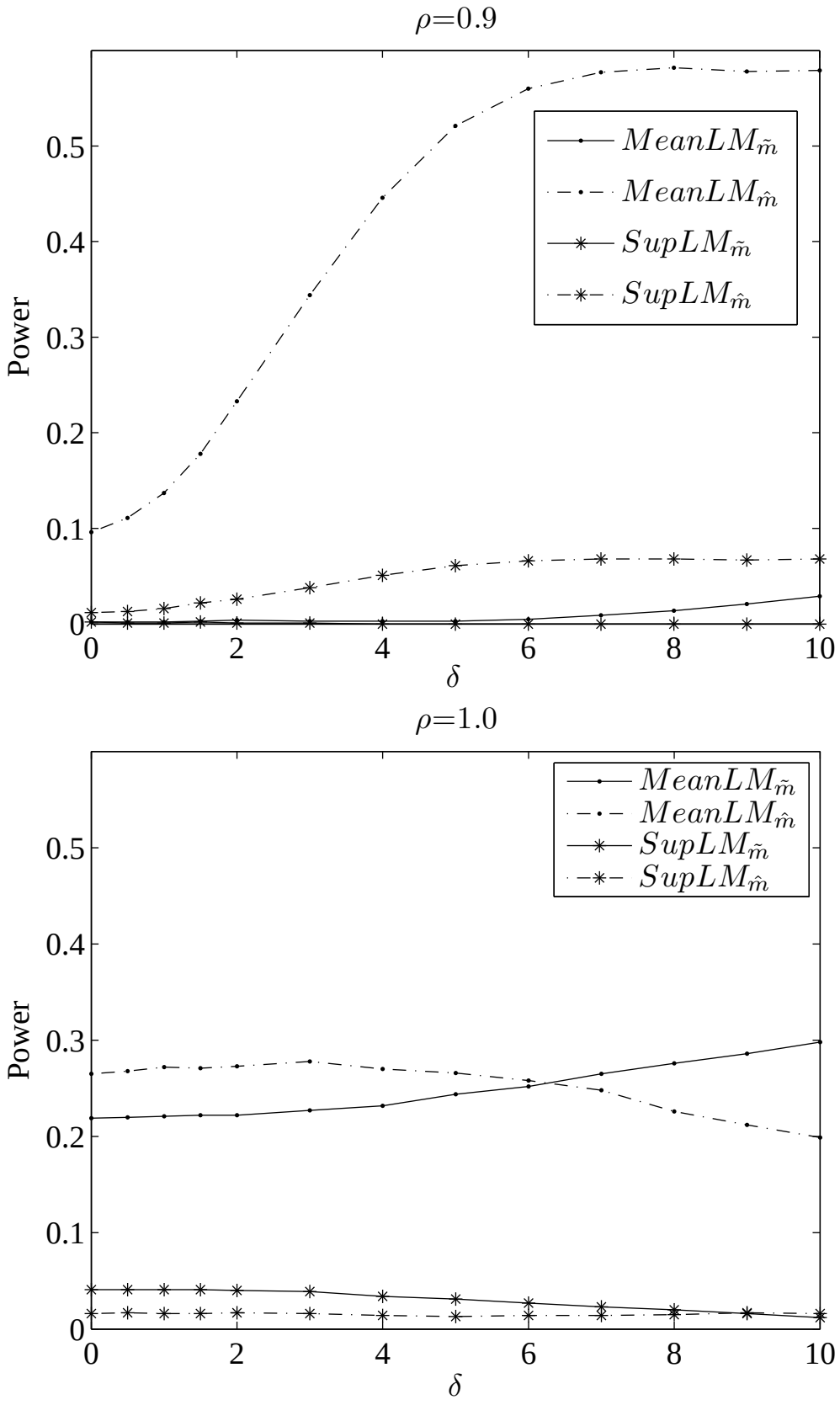

Figure 2.2. Finite Sample Power, QS kernel, 15% Trimming.

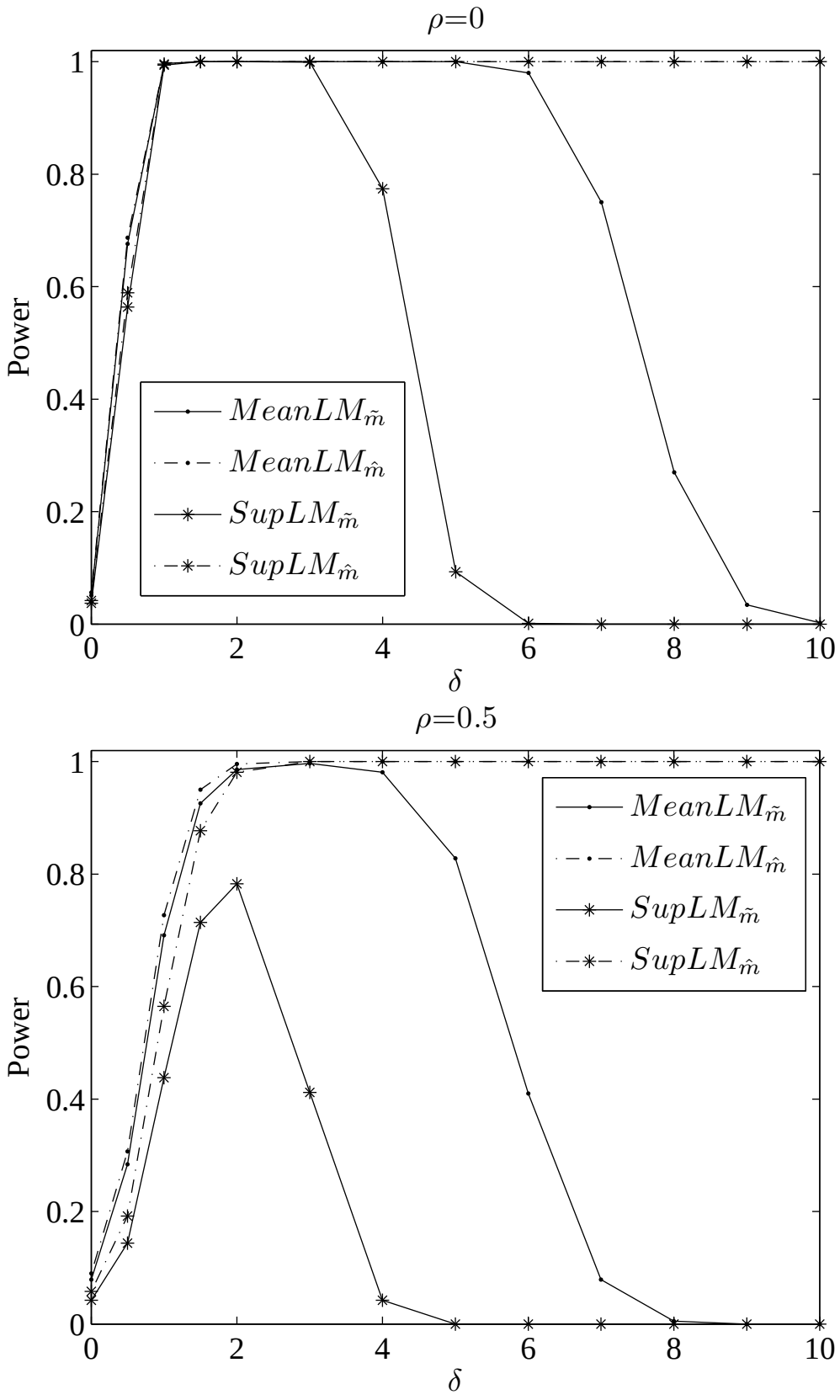

Figure 2.3. Finite Sample Power, Bartlett kernel, 15% Trimming.

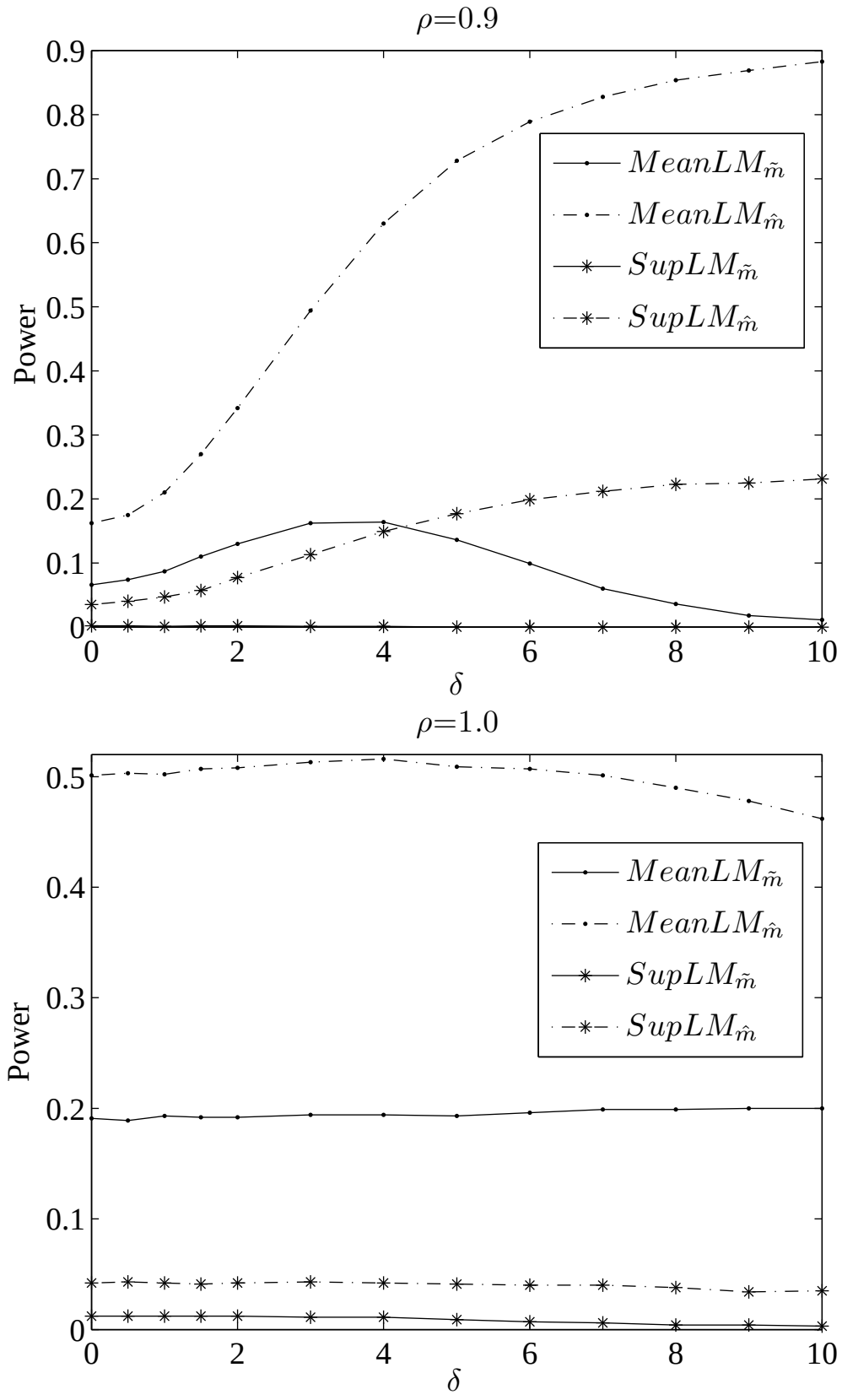

Figure 2.4. Finite Sample Power, Bartlett kernel, 15% Trimming.

Table 2.4. Finite Sample Behavior of Data Dependent Bandwidth to Sample Size Ratios, $T = 120$, Bartlett Kernel

	$\rho = 0$		$\rho = 0.5$		$\rho = 0.7$		$\rho = 0.9$		$\rho = 1$	
δ	$ave(\widetilde{b})$	$ave(\widehat{b})$	$ave(\widetilde{b})$	$ave(\widehat{b})$	$ave(\widetilde{b})$	$ave(\widehat{b})$	$ave(\widetilde{b})$	$ave(\widehat{b})$	$ave(\widetilde{b})$	$ave(\widehat{b})$
0	0.012	0.013	0.055	0.050	0.088	0.078	0.185	0.140	0.501	0.197
0.5	0.014	0.013	0.058	0.051	0.090	0.078	0.187	0.140	0.502	0.197
1	0.025	0.013	0.066	0.051	0.097	0.079	0.192	0.140	0.503	0.197
1.5	0.041	0.013	0.078	0.051	0.108	0.079	0.200	0.140	0.504	0.197
2	0.056	0.012	0.093	0.052	0.122	0.080	0.210	0.140	0.505	0.197
3	0.088	0.012	0.129	0.052	0.157	0.081	0.238	0.140	0.510	0.197
4	0.122	0.012	0.169	0.053	0.196	0.082	0.271	0.141	0.515	0.199
5	0.158	0.012	0.210	0.053	0.238	0.083	0.307	0.144	0.522	0.202
6	0.193	0.012	0.250	0.053	0.279	0.083	0.344	0.148	0.530	0.207
7	0.228	0.012	0.290	0.053	0.318	0.084	0.379	0.152	0.539	0.213
8	0.262	0.012	0.327	0.053	0.355	0.084	0.412	0.155	0.548	0.220
9	0.294	0.012	0.361	0.053	0.389	0.084	0.443	0.159	0.558	0.229
10	0.324	0.012	0.393	0.053	0.420	0.084	0.470	0.161	0.568	0.238
25	0.579	0.012	0.619	0.053	0.631	0.084	0.649	0.164	0.670	0.288
50	0.673	0.012	0.686	0.053	0.690	0.084	0.695	0.164	0.707	0.288
100	0.702	0.012	0.706	0.053	0.707	0.084	0.709	0.164	0.712	0.288

the alternative data dependent bandwidth, $\widehat{m}$, leads to relatively small values of b, the LM tests using $\widehat{m}$ are, on average, using bandwidths in the monotonic power range. In contrast, because $\widetilde{m}$ increases as δ increases, LM tests based on $\widetilde{m}$ are using bandwidths in the non-monotonic power range when δ is large. In other words, as δ increases, the $LM_{\widetilde{m}}$ tests jump from "low b" power curves to "high b" power curves and this results in non-monotonic power.

In summary, the finite sample simulations show that patterns in null rejection probabilities and power depend on the bandwidth and this relationship in turn depends on the value of ρ. This suggests that a theoretical explanation for the finite sample patterns requires an asymptotic theory that depends on the bandwidth and the strength of the serial correlation. The natural candidate is fixed-b asymptotic theory used in conjunction with nearly I(1) asymptotics which we explore in the next section.

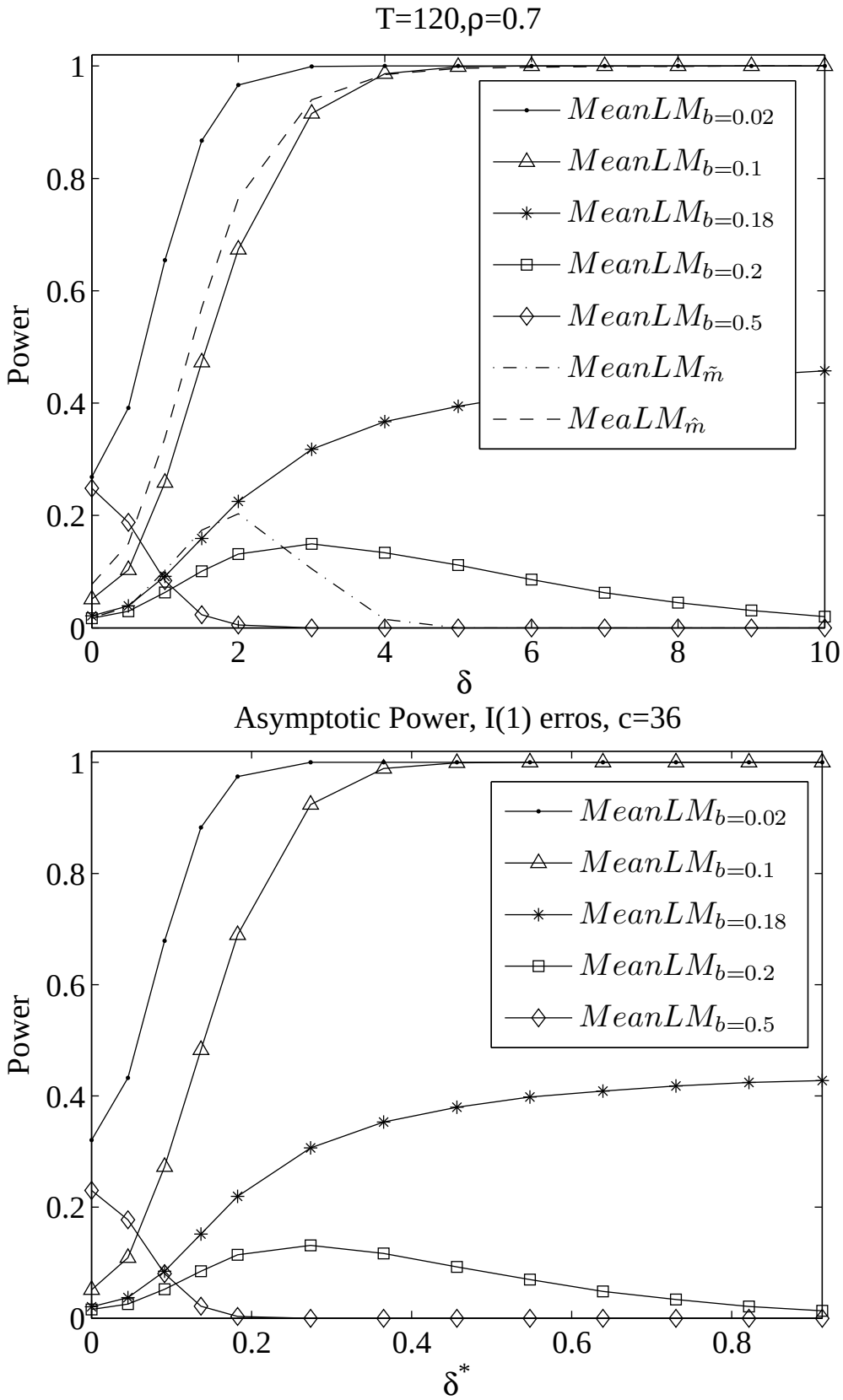

Figure 2.5. Finite Sample and Asymptotic Power of $MeanLM$, QS kernel, 15% Trimming.

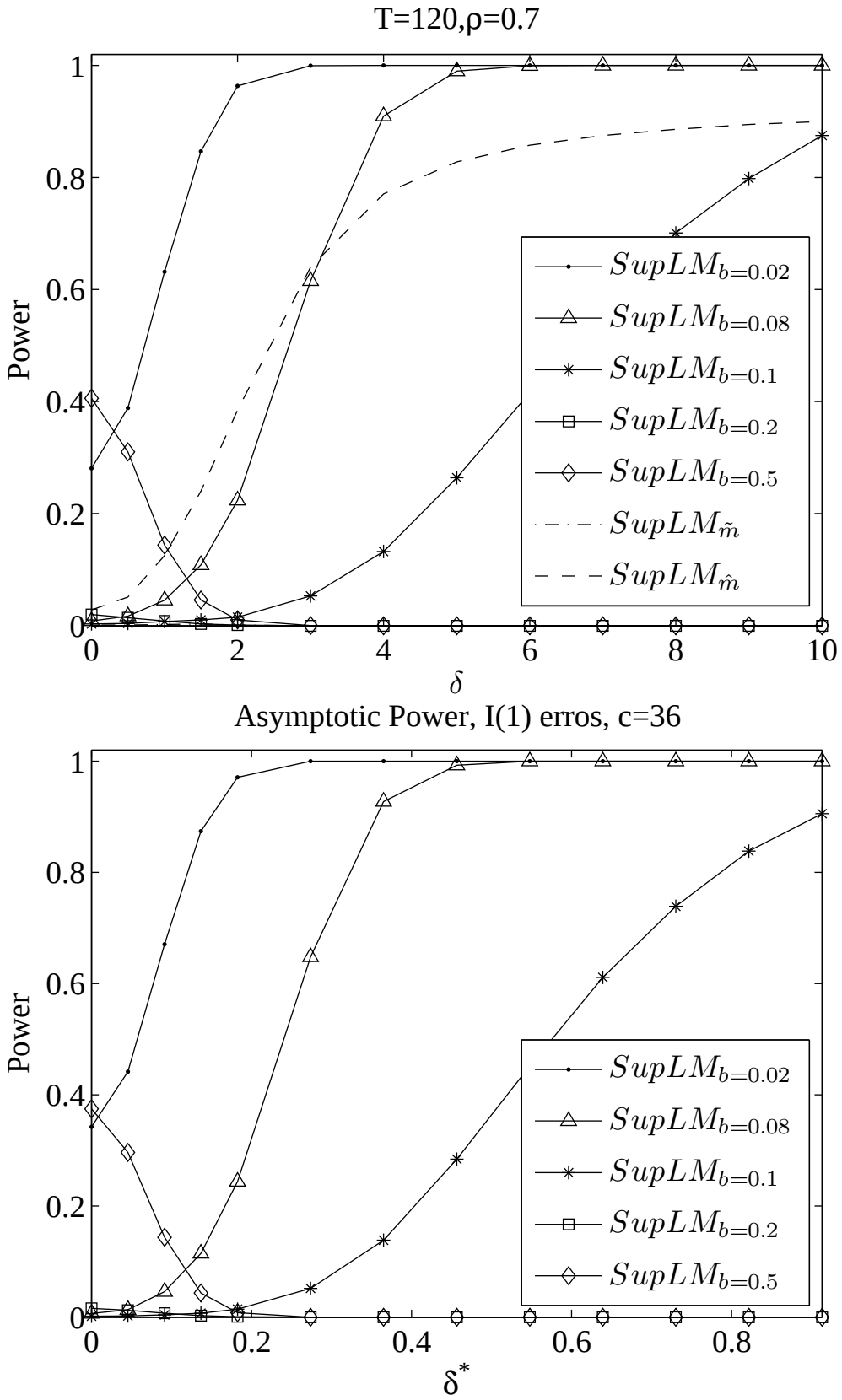

Figure 2.6. Finite Sample and Asymptotic Power of $SupLM$, QS kernel, 15% Trimming.

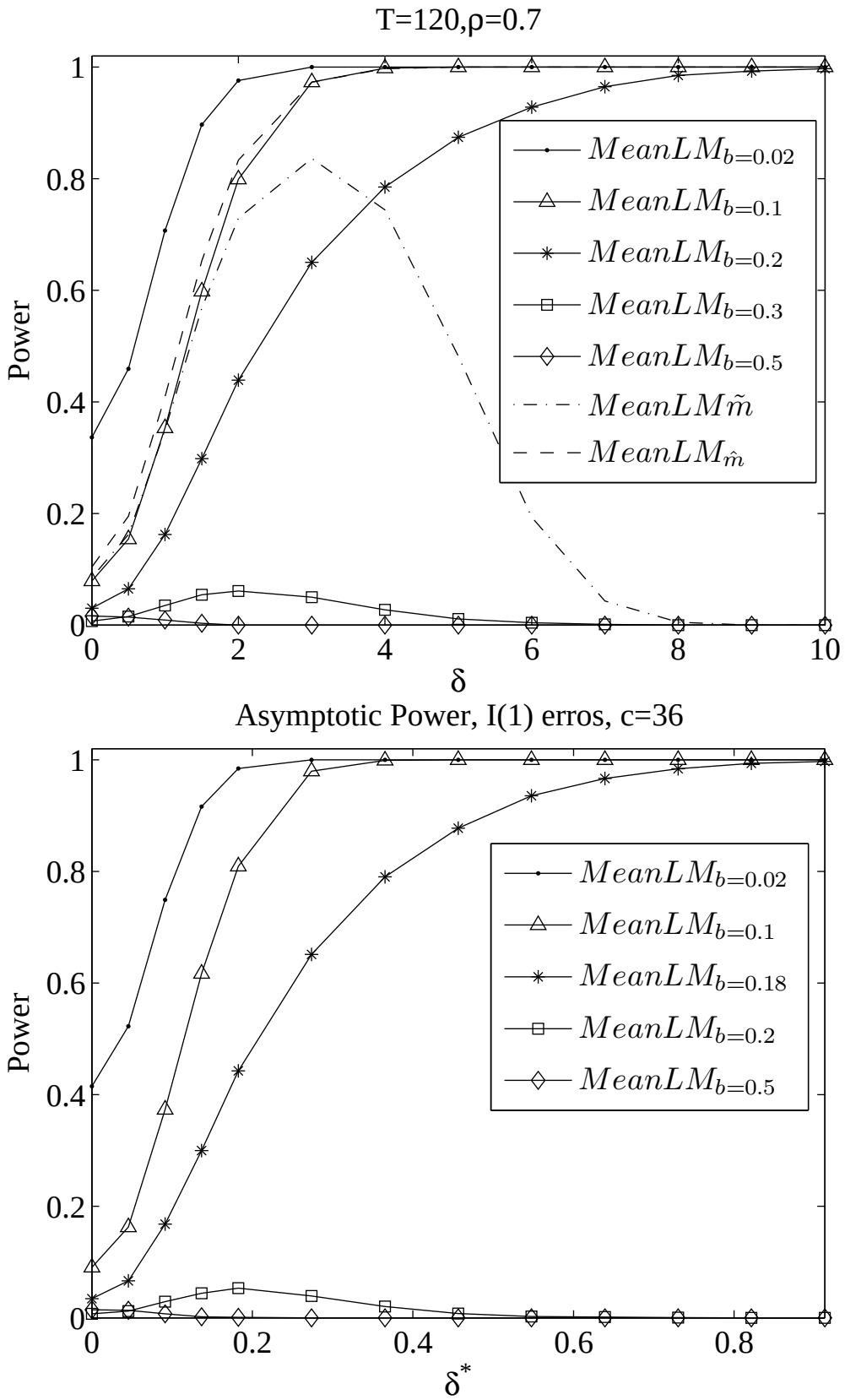

Figure 2.7. Finite and Asymptotic Power of $MeanLM$, Bartlett kernel, 15% Trimming.

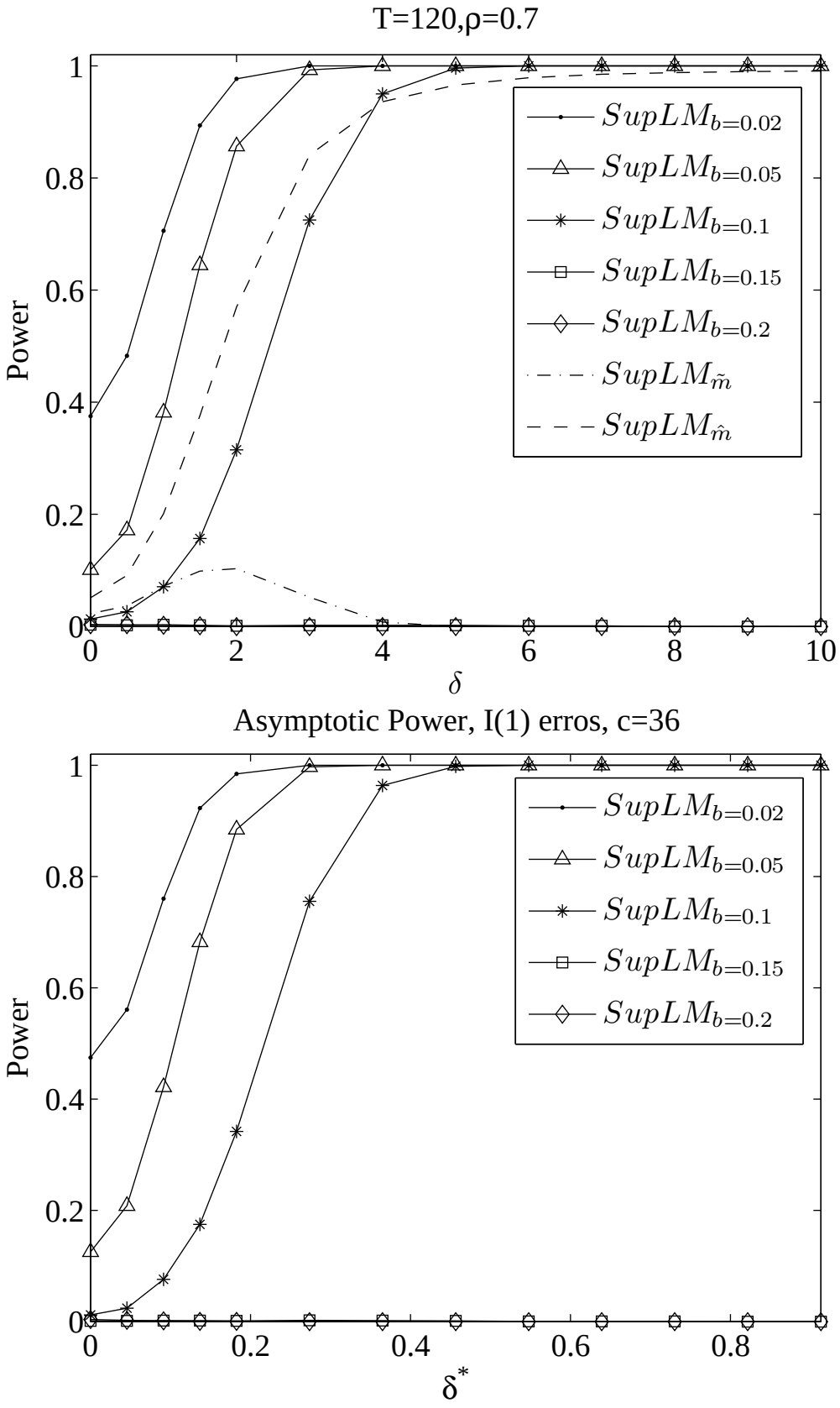

Figure 2.8. Finite and Asymptotic Power of $SupLM$, Bartlett kernel, 15% Trimming.

2.5 Fixed-b Asymptotic Analysis of LM Mean Shift Tests

In this section, we provide fixed-b asymptotic results for the LM tests. These results complement the fixed-b results derived by Sayginsoy and Vogelsang (2010) for the case of nonparametric HAC $Wald$ statistics for testing for a shift in mean. We derive results under the local alternative

$$H_A : \delta = \delta_0 g(T)$$

where $g(T) = T^{-1/2}$ if $|\rho| < 1$ and $g(T) = T^{1/2}$ if $\rho = 1 - c/T$. Because the numerator of LM, $SSR_0 - SSR(T_b)$, is identical to the $Wald$ statistic, its limit follows directly from the results of Sayginsoy and Vogelsang (2010) (Theorems 1 & 2). Our theoretical contribution is obtaining the fixed-b limit of $\tilde{\sigma}^2(m)$ under H_A. Obviously, results for the null distribution of the LM tests follow by setting $\delta_0 = 0$. The following theorem gives the limiting distribution of LM under the local alternative.

Theorem 2.5.3 *Suppose the true model is given by (2.2.1) with break date $T_b^0 = \lambda_0 T$. Suppose the LM statistic is computed using model (2.2.1) using the break date $T_b = \lambda T$. Let $m = bT$, where $b \in (0, 1]$ is fixed as T increases. Under the local alternative H_A, as $T \to \infty$,*

$$LM(T_b, m) \Rightarrow \frac{\lambda(1-\lambda)[P_i(\lambda) + \Psi(\lambda, \lambda_0)\delta^*]^2}{\Phi_i(b, \delta_0)}, \;\; i = 0 \text{ if } \{u_t\} \text{ is I(0), } i = 1 \text{ if } \{u_t\} \text{ is I(1)}$$

where

$$P_0(\lambda) = \frac{1}{\lambda(1-\lambda)} \int_0^1 [\mathbf{1}(r > \lambda) - (1-\lambda)] dW(r),$$
$$P_1(\lambda) = \frac{1}{\lambda(1-\lambda)} \int_0^1 [\mathbf{1}(r > \lambda) - (1-\lambda)] V_c(r) dr,$$

$$\Psi(\lambda, \lambda_0) = \begin{cases} \frac{1-\lambda_0}{\lambda(1-\lambda)^2}, & \text{if } \lambda \leq \lambda_0, \\ \frac{\lambda_0}{\lambda^2(1-\lambda)}, & \text{if } \lambda > \lambda_0. \end{cases}$$

$$\delta^* = \delta_0/\sigma, \;\; \text{if } \{u_t\} \text{ is I(0)}; \;\; \delta_0/d(1), \;\; \text{if } \{u_t\} \text{ is I(1)}.$$

$$Q_0(r) = \frac{\delta_0}{\sigma}[(r-\lambda_0)\mathbf{1}(r>\lambda_0) - r(1-\lambda_0)] + W(r) - rW(1),$$

if $\{u_t\}$ *is I(0),*

$$Q_1(r) = \frac{\delta_0}{d(1)}[(r-\lambda_0)\mathbf{1}(r>\lambda_0) - r(1-\lambda_0)] + \int_0^r V_c(s)ds - r\int_0^1 V_c(s)ds,$$

if $\{u_t\}$ *is I(1).*

$$\Phi_i(b,\delta_0) =$$

$$\begin{cases} \int_0^1\int_0^1 -\frac{1}{b^2}k''(\frac{r-s}{b})Q_i(r)Q_i(s)drds, \\ \qquad\qquad \textit{if } k(.) \textit{ is of type 1}; \\ \int\int_{|r-s|<b} -\frac{1}{b^2}k''((r-s)/(b))Q_i(r)Q_i(s)drds + \frac{2}{b}k'_-(1)\int_0^{1-b} Q_i(r+b)Q_i(r)dr, \\ \qquad\qquad \textit{if } k(.) \textit{ is of type 2}; \\ \frac{2}{b}\int_0^1 Q_i(r)^2dr - \frac{2}{b}\int_0^{1-b} Q_i(r+b)Q_i(r)dr, \\ \qquad\qquad \textit{if } k(.) \textit{ is Bartlett}. \end{cases}$$

$$MeanLM_m \Rightarrow \int_{\lambda^*}^{1-\lambda^*} \frac{\lambda(1-\lambda)[P_i(\lambda) + \Psi(\lambda,\lambda_0)\delta^*]^2}{\Phi_i(b,\delta_0)}d\lambda,$$

$$SupLM_m \Rightarrow \sup_{\lambda\in[\lambda^*,1-\lambda^*]} \frac{\lambda(1-\lambda)[P_i(\lambda) + \Psi(\lambda,\lambda_0)\delta^*]^2}{\Phi_i(b,\delta_0)}.$$

The proof of Theorem 1 is given in the Appendix. The limits given by Theorem 1 depend on the kernel, bandwidth and magnitude of the mean shift. In particular, the denominator of LM depends on δ^*. This contrasts with standard asymptotic theory where $\tilde{\sigma}^2(m)$ would remain consistent under the local alternative H_A and the denominator of LM would not depend on δ^* (nor the kernel and bandwidth). Notice that as δ^* increases, the numerator of LM becomes larger. However, the denominator also increases and this can cause power to fall. Whether or not power can be non-monotonic is difficult to discern from the limiting random variables given that the relative impact of an increase of δ^* on the numerator and denominator depends on the kernel and bandwidth.

Because of the complicated forms of the fixed-b limits, we cannot theoretically determine the impact of kernel or bandwidth choice on null rejection probabilities when using standard ($b=0$) $I(0)$ critical values nor can we determine when power will be monotonic or non-monotonic. Because we can easily simulate the fixed-b asymptotic limits, we can

easily compute null rejection probabilities and power for a given kernel and bandwidth. We can simulate the limits for the case where the errors are $I(0)$ and, for cases where strong serial correlation is present in the data, can simulate the limits using the $I(1)$ results with c chosen to reflect the strength of the serial correlation.

Table 2.1 and 2.2 (Panel B) report simulated asymptotic null rejection probabilities using standard ($b = 0$) $I(0)$ 5% critical values. We set $\delta_0 = 0$. We simulated the fixed-b limits using $10,000$ replications and the Wiener processes in the limits were approximated using scaled partial sums of 1,000 iid $N(0,1)$ random deviates. We report results for the QS kernel and a selection of bandwidths, b, in neighborhoods around the average finite sample data dependent bandwidths. The first row of Panel B reports rejections for the $I(0)$ case with $b = 0.02$. Focusing on the QS kernel in Table 2.1 we see that rejections are very similar to what was obtained in finite samples with $\rho = 0$ where the data dependent bandwidths were small. The next three rows give results for $I(1)$ errors with $c = 60$ which corresponds to $\rho = 0.5$ in a sample with $T = 120$. The tendency of the $MeanLM$ test to slightly over-reject is captured by the asymptotics as is the tendency of the $SupLM$ test to under-reject. As we go farther down the table, c decreases and we see that asymptotic rejections for $MeanLM$ tend to fall but then increase dramatically when $c = 0$ and b is below the finite sample average of the data dependent bandwidths. Rejections for $SupLM$ remain very small as c increases although when $c = 0$, $SupLM$ tends to over-reject when b is close to 0.6 but tends to substantially under-reject when b is close to 0.2. Overall, the fixed-b asymptotic rejections capture all of the salient patterns in the finite sample rejections. One thing apparent from Panel B of Table 2.1 and 2.2 is that null rejections are very sensitive to the value of b.

To further explore the impact of the bandwidth on null rejections, we report results for a full grid of bandwidths in Table 2.5 and 2.6. Keeping in mind that these are asymptotic fixed-b rejections when using standard ($b = 0$) $I(0)$ critical values, several interesting patterns emerge. When the errors are $I(0)$, increasing the bandwidth tends to cause

over-rejections and over-rejections can be very large when b is large. When the errors are $I(1)$, the patterns are different. When small bandwidths are used, there can be large over-rejections especially when $c = 0$. As the bandwidth increases, there is tendency to under-reject especially with the $SupLM$ statistic. It is striking how little b has to increase for $SupLM$ to switch from extreme over-rejections to extreme under-rejections. As b becomes large, both $MeanLM$ and $SupLM$ again tend to severely over-reject regardless of the value of c. The upshot of Table 2.5 and 2.6 is that fixed-b asymptotics correctly predicts that the finite sample null behavior of both statistics will be very sensitive to the bandwidth and the strength of the serial correlation.

We now turn to asymptotic power simulations. We focus on the case of $c = 36$ which corresponds to the finite sample case of $T = 120$ and $\rho = 0.7$ as was depicted in the upper panels of Figures 2.5, 2.6, 2.7 and 2.8. We simulated asymptotic power using standard ($b = 0$) $I(0)$ 5% critical values for the same grid of b values used in the finite sample power simulations. The results are reported in the lower panels of Figures 2.5, 2.6, 2.7 and 2.8. We used a grid for δ^* that maps into the values of δ used in the $T = 120$ case. We see that the fixed-b asymptotic rejections capture the salient features of the finite sample power results. As b increases power goes from monotonic to non-monotonic. Fixed-b asymptotic theory successfully captures the impact of b on power and correctly predicts when non-monotonic power will occur.

2.6 Bandwidths That Control Size

A close inspection of the rejections reported in Table 2.5 and 2.6 suggests there are bandwidths such that the critical values of a given statistic are the same in the $I(0)$ and $I(1, c = 0)$ cases. For example, rejections for $MeanLM$ decrease as b increases in both the $I(0)$ and $I(1, c = 0)$ cases, but the decrease is faster in the $I(1, c = 0)$ case. There appears to be a value of b between 0.22 and 0.24 where rejections in the two cases are the

Table 2.5. Fixed-b Asymptotic Null Rejection Probabilities Using Standard ($b = 0$) $I(0)$ Critical Values, 5% Nominal Level, 15% Trimming, QS Kernel

	$MeanLM$					$SupLM$				
		$I(1)$					$I(1)$			
b	$I(0)$	$c = 60$	$c = 36$	$c = 12$	$c = 0$	$I(0)$	$c = 60$	$c = 36$	$c = 12$	$c = 0$
0.020	0.049	0.185	0.306	0.647	0.939	0.044	0.177	0.321	0.681	0.946
0.040	0.047	0.092	0.151	0.394	0.825	0.039	0.047	0.090	0.333	0.801
0.060	0.045	0.065	0.096	0.268	0.721	0.035	0.016	0.025	0.122	0.616
0.080	0.042	0.051	0.069	0.188	0.646	0.030	0.006	0.006	0.017	0.254
0.100	0.038	0.042	0.051	0.133	0.576	0.032	0.004	0.002	0.001	0.001
0.120	0.038	0.037	0.042	0.093	0.507	0.040	0.003	0.001	0.000	0.000
0.140	0.034	0.031	0.033	0.063	0.440	0.053	0.005	0.002	0.000	0.000
0.160	0.032	0.026	0.027	0.041	0.368	0.069	0.011	0.005	0.001	0.000
0.180	0.031	0.022	0.020	0.023	0.281	0.091	0.017	0.009	0.001	0.000
0.200	0.029	0.018	0.016	0.014	0.182	0.121	0.029	0.016	0.003	0.000
0.220	0.029	0.016	0.013	0.008	0.077	0.150	0.046	0.026	0.005	0.000
0.240	0.031	0.016	0.011	0.004	0.012	0.184	0.066	0.042	0.010	0.001
0.260	0.032	0.019	0.012	0.004	0.001	0.223	0.090	0.059	0.016	0.002
0.280	0.038	0.023	0.017	0.004	0.001	0.258	0.113	0.082	0.026	0.003
0.300	0.053	0.034	0.025	0.006	0.001	0.297	0.144	0.107	0.038	0.004
0.320	0.074	0.051	0.039	0.013	0.001	0.332	0.177	0.136	0.053	0.007
0.340	0.099	0.074	0.060	0.027	0.003	0.370	0.210	0.166	0.072	0.011
0.360	0.126	0.098	0.084	0.043	0.008	0.402	0.237	0.192	0.091	0.017
0.380	0.151	0.123	0.105	0.060	0.014	0.431	0.266	0.219	0.112	0.023
0.400	0.174	0.147	0.129	0.075	0.020	0.458	0.293	0.245	0.133	0.028
0.420	0.199	0.169	0.150	0.091	0.024	0.484	0.322	0.271	0.153	0.033
0.440	0.218	0.189	0.171	0.106	0.028	0.510	0.347	0.296	0.170	0.040
0.460	0.241	0.206	0.188	0.120	0.030	0.538	0.374	0.323	0.189	0.045
0.480	0.266	0.228	0.207	0.137	0.033	0.568	0.402	0.350	0.207	0.051
0.500	0.292	0.252	0.231	0.151	0.036	0.598	0.428	0.376	0.230	0.056
0.520	0.318	0.278	0.257	0.170	0.041	0.626	0.455	0.403	0.252	0.063
0.540	0.345	0.307	0.284	0.193	0.045	0.658	0.488	0.430	0.272	0.070
0.560	0.380	0.338	0.314	0.217	0.052	0.686	0.516	0.460	0.297	0.078
0.580	0.417	0.376	0.347	0.245	0.059	0.712	0.546	0.490	0.324	0.087
0.600	0.456	0.414	0.388	0.280	0.071	0.734	0.575	0.520	0.351	0.098
0.620	0.499	0.457	0.430	0.323	0.085	0.757	0.605	0.549	0.379	0.108
0.640	0.546	0.507	0.480	0.374	0.102	0.780	0.635	0.580	0.412	0.121
0.660	0.599	0.557	0.534	0.428	0.129	0.798	0.662	0.609	0.443	0.139
0.680	0.652	0.614	0.589	0.490	0.164	0.817	0.689	0.637	0.472	0.157
0.700	0.710	0.672	0.652	0.553	0.209	0.834	0.713	0.666	0.507	0.174
0.720	0.762	0.734	0.716	0.622	0.258	0.851	0.736	0.692	0.536	0.194
0.740	0.818	0.794	0.779	0.694	0.315	0.866	0.754	0.713	0.563	0.216
0.760	0.870	0.849	0.839	0.773	0.380	0.880	0.773	0.736	0.590	0.235
0.780	0.916	0.903	0.896	0.857	0.472	0.894	0.793	0.755	0.615	0.255
0.800	0.954	0.946	0.945	0.930	0.594	0.903	0.811	0.773	0.638	0.274
0.820	0.976	0.973	0.976	0.977	0.812	0.914	0.828	0.791	0.659	0.294
0.840	0.988	0.987	0.989	0.993	0.979	0.923	0.843	0.808	0.680	0.311
0.860	0.995	0.994	0.995	0.997	0.996	0.932	0.857	0.824	0.702	0.326
0.880	0.998	0.997	0.998	0.999	0.999	0.943	0.872	0.840	0.723	0.341
0.900	0.999	0.999	0.999	0.999	1.000	0.951	0.884	0.854	0.744	0.358
0.920	1.000	0.999	0.999	0.999	1.000	0.959	0.897	0.870	0.764	0.375
0.940	1.000	0.999	1.000	1.000	1.000	0.965	0.908	0.880	0.786	0.392
1.000	1.000	1.000	1.000	1.000	1.000	0.982	0.939	0.918	0.837	0.447

Table 2.6. Fixed-b Asymptotic Null Rejection Probabilities Using Standard ($b = 0$) $I(0)$ Critical Values, 5% Nominal Level, 15% Trimming, Bartlett Kernel

	$MeanLM$					$SupLM$				
		$I(1)$					$I(1)$			
b	$I(0)$	$c = 60$	$c = 36$	$c = 12$	$c = 0$	$I(0)$	$c = 60$	$c = 36$	$c = 12$	$c = 0$
0.020	0.048	0.269	0.415	0.754	0.966	0.041	0.301	0.475	0.805	0.975
0.040	0.044	0.139	0.222	0.502	0.883	0.037	0.106	0.192	0.491	0.876
0.060	0.042	0.099	0.149	0.367	0.793	0.030	0.047	0.079	0.282	0.757
0.080	0.040	0.077	0.110	0.281	0.724	0.024	0.023	0.033	0.131	0.613
0.100	0.036	0.065	0.090	0.224	0.668	0.018	0.009	0.012	0.038	0.350
0.120	0.033	0.052	0.073	0.175	0.618	0.014	0.005	0.004	0.004	0.024
0.140	0.029	0.045	0.059	0.141	0.572	0.013	0.002	0.002	0.001	0.001
0.160	0.026	0.039	0.049	0.111	0.525	0.013	0.002	0.001	0.000	0.000
0.180	0.023	0.034	0.042	0.087	0.478	0.016	0.003	0.002	0.001	0.000
0.200	0.019	0.028	0.034	0.069	0.434	0.021	0.004	0.003	0.001	0.000
0.220	0.015	0.022	0.027	0.050	0.382	0.027	0.006	0.004	0.001	0.000
0.240	0.010	0.017	0.021	0.034	0.329	0.035	0.009	0.006	0.002	0.001
0.260	0.008	0.011	0.014	0.024	0.267	0.045	0.012	0.009	0.003	0.001
0.280	0.005	0.009	0.010	0.016	0.199	0.056	0.017	0.013	0.004	0.001
0.300	0.003	0.006	0.007	0.010	0.134	0.074	0.025	0.019	0.007	0.001
0.320	0.002	0.004	0.005	0.005	0.076	0.089	0.032	0.026	0.010	0.002
0.340	0.001	0.002	0.002	0.003	0.039	0.107	0.041	0.032	0.013	0.003
0.360	0.001	0.001	0.001	0.001	0.017	0.128	0.051	0.039	0.017	0.003
0.380	0.001	0.001	0.001	0.000	0.008	0.153	0.060	0.049	0.022	0.004
0.400	0.000	0.000	0.000	0.001	0.004	0.177	0.074	0.058	0.027	0.006
0.420	0.000	0.000	0.001	0.002	0.005	0.202	0.090	0.072	0.034	0.008
0.440	0.001	0.002	0.003	0.006	0.008	0.227	0.104	0.086	0.043	0.010
0.460	0.001	0.004	0.007	0.010	0.010	0.255	0.122	0.102	0.053	0.013
0.480	0.003	0.007	0.010	0.015	0.012	0.279	0.137	0.115	0.063	0.016
0.500	0.006	0.011	0.014	0.019	0.016	0.302	0.153	0.131	0.074	0.020
0.520	0.008	0.015	0.018	0.022	0.019	0.326	0.169	0.145	0.084	0.026
0.540	0.012	0.018	0.022	0.027	0.021	0.352	0.186	0.159	0.093	0.030
0.560	0.015	0.022	0.026	0.030	0.022	0.376	0.203	0.171	0.103	0.033
0.580	0.017	0.025	0.030	0.032	0.022	0.406	0.220	0.184	0.112	0.036
0.600	0.019	0.027	0.033	0.033	0.022	0.434	0.238	0.201	0.123	0.039
0.620	0.021	0.030	0.033	0.034	0.021	0.460	0.260	0.218	0.133	0.041
0.640	0.022	0.030	0.033	0.034	0.020	0.489	0.279	0.237	0.142	0.044
0.660	0.022	0.030	0.033	0.033	0.020	0.518	0.301	0.258	0.151	0.047
0.680	0.021	0.029	0.033	0.032	0.018	0.550	0.325	0.277	0.163	0.051
0.700	0.021	0.030	0.032	0.031	0.017	0.581	0.346	0.298	0.176	0.055
0.720	0.019	0.029	0.030	0.029	0.016	0.612	0.370	0.319	0.192	0.060
0.740	0.018	0.027	0.028	0.028	0.014	0.642	0.399	0.344	0.209	0.066
0.760	0.019	0.025	0.028	0.027	0.011	0.669	0.429	0.369	0.227	0.072
0.780	0.020	0.028	0.031	0.030	0.010	0.695	0.459	0.397	0.247	0.080
0.800	0.027	0.039	0.043	0.042	0.012	0.718	0.484	0.423	0.268	0.089
0.820	0.040	0.060	0.070	0.069	0.019	0.741	0.511	0.451	0.292	0.097
0.840	0.069	0.101	0.116	0.124	0.038	0.762	0.538	0.476	0.316	0.109
0.860	0.117	0.165	0.186	0.208	0.087	0.782	0.568	0.505	0.341	0.118
0.880	0.187	0.254	0.284	0.341	0.238	0.805	0.596	0.536	0.366	0.127
0.900	0.288	0.364	0.404	0.507	0.620	0.825	0.621	0.562	0.390	0.138
0.920	0.391	0.479	0.528	0.646	0.816	0.842	0.645	0.584	0.412	0.150
0.940	0.489	0.583	0.634	0.754	0.889	0.859	0.669	0.610	0.437	0.162
1.000	0.701	0.771	0.812	0.898	0.960	0.897	0.735	0.677	0.509	0.199

same. A similar phenomenon occurs for the $SupLM$ statistic for b between 0.08 and 0.1. This suggests that, for a given significance level, there could be a bandwidth such that the $I(0)$ and $I(1, c = 0)$ critical values are the same. This is a promising situation for practice because use of these bandwidths in conjunction with the fixed-b critical value would result in a test that is robust to the strength of the serial correlation in the error.

To show that these robust bandwidths indeed exist, we plot in Figure 2.9 and 2.10 the 5% fixed-b asymptotic critical values for a grid of values of b using the QS kernel with 15% trimming. The plot includes $I(0)$ critical values and $I(1)$ critical values for $c = 0, 12, 36, 60$. There clearly exists a value of b for each statistic where the $I(0)$ and $I(1, c = 0)$ critical value curves cross and for other values of c the critical value curves are no higher. These bandwidths deliver tests with asymptotic size equal to 5% and are appropriately labeled "robust bandwidths". In unreported plots we found similar patterns at other significance levels.

Table 2.7 and 2.8 provides robust bandwidths for each statistic for a range of significance levels. The corresponding asymptotic fixed-b critical value (which is the same for $I(0)$ and $I(1)$ errors) is also reported.

To assess the performance of the robust bandwidths and associated fixed-b approximation in practice, we simulated null rejection probabilities and finite sample power using the same simulation design as used in Section 4. Null rejection probabilities are reported in Table 2.9 and 2.10. While reporting results for $AR(1)$ errors as before, we also include results where an $MA(1)$ component is added to the error:

$$u_t = \rho u_{t-1} + e_t + \theta e_{t-1},$$

$$u_0 = e_0 = 0.$$

We report results for sample sizes $T = 60, 120, 240$ and we continue to focus on the QS kernel and use 15% trimming. For the most part, null rejection probabilities are at or below 0.05. The exceptions occur when θ is negative and ρ is close to zero. When $\rho = 0$ and $\theta = -0.8$ large over-rejections can occur. This happens because the data is close to being

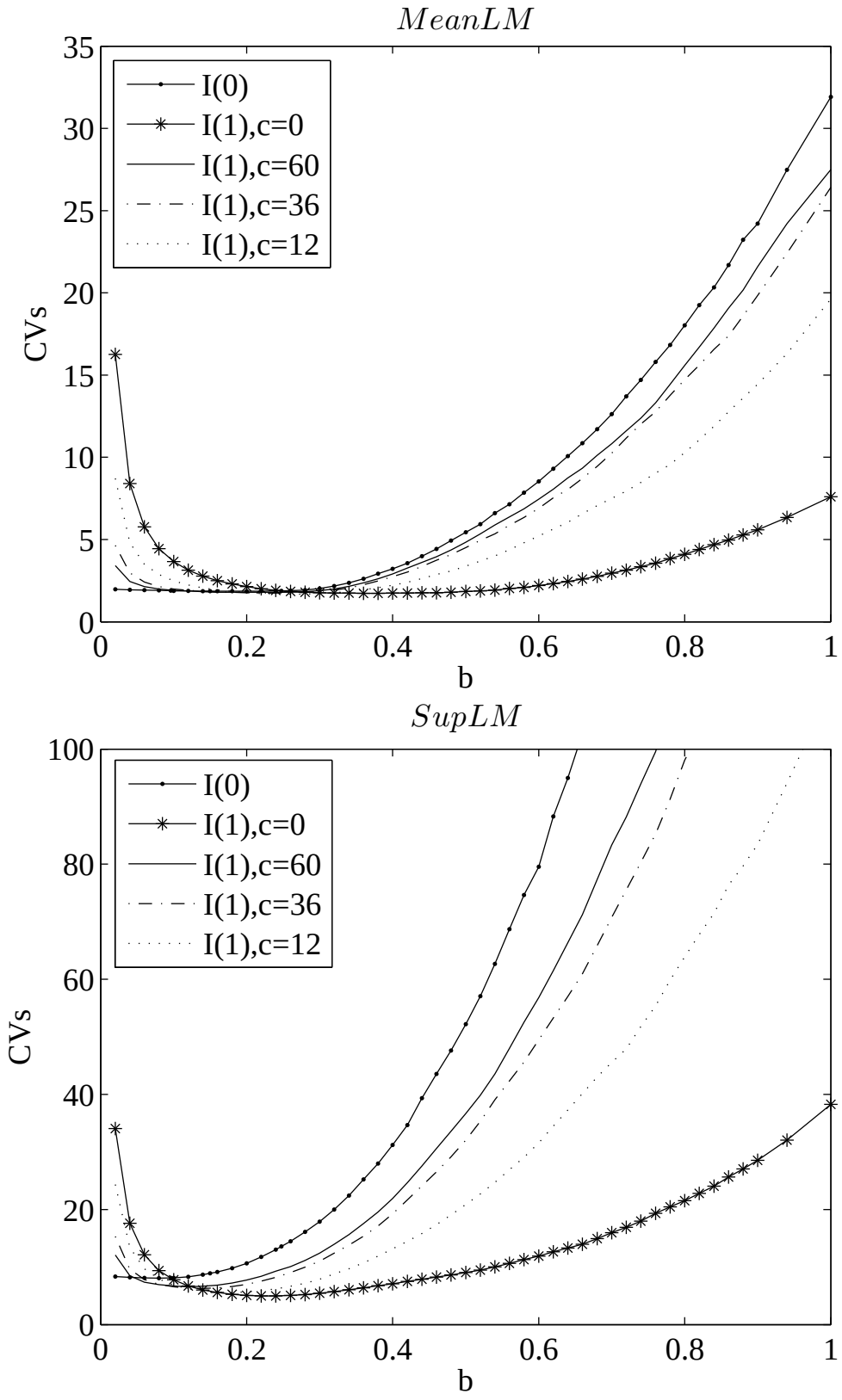

Figure 2.9. Asymptotic Fixed-b Critical Values, 5% Level, QS kernel, 15% Trimming.

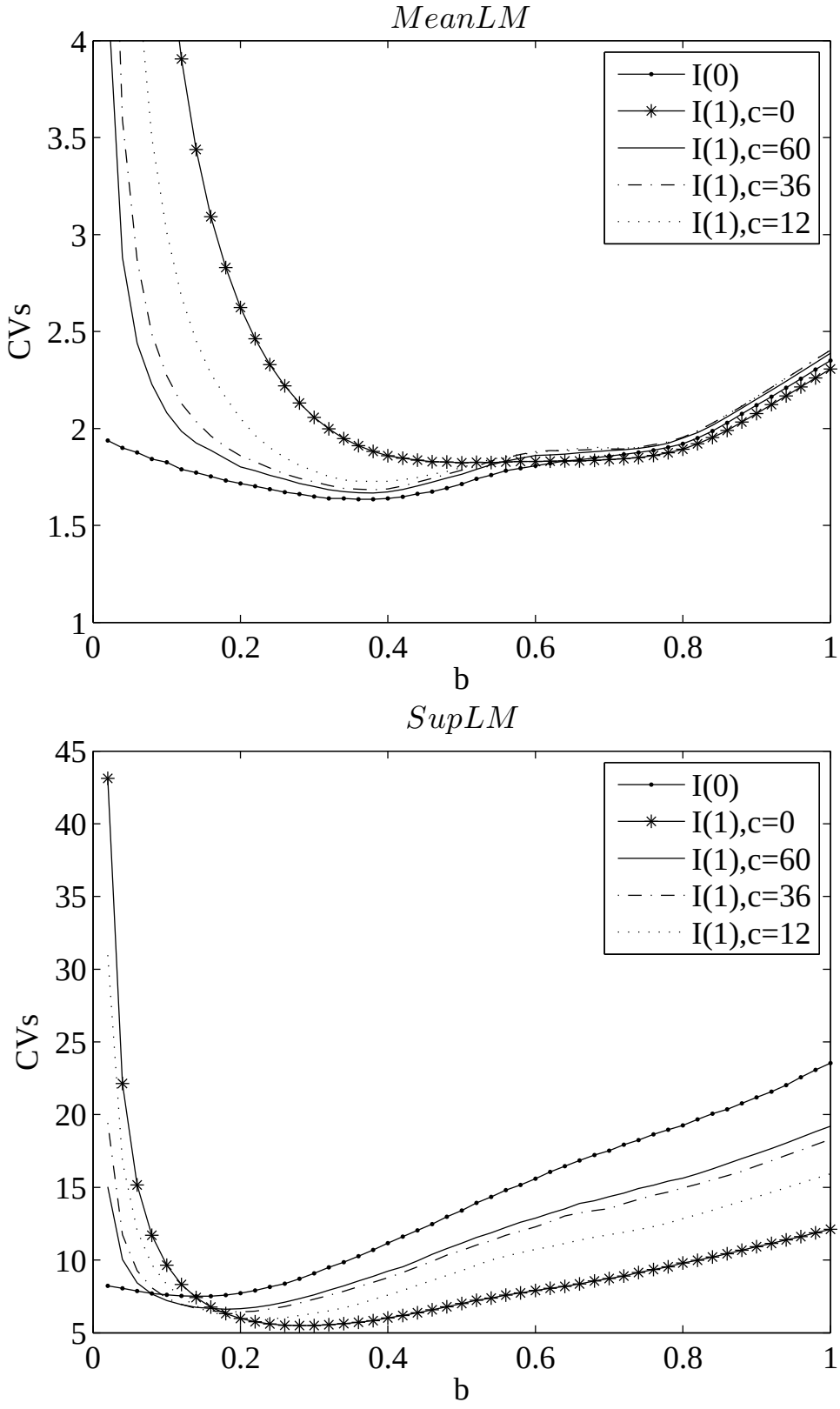

Figure 2.10. Asymptotic Fixed-b Critical Values, 5% Level, Bartlett kernel, 15% Trimming.

Table 2.7. $I(0)/I(1)$ Robust Bandwidths and Critical Values QS kernel, 15% Trimming.

	$MeanLM$		$SupLM$	
level	b	cv	b	cv
20%	0.306	1.581	0.118	6.331
19%	0.303	1.593	0.117	6.417
18%	0.300	1.608	0.116	6.493
17%	0.297	1.622	0.115	6.576
16%	0.295	1.637	0.113	6.649
15%	0.293	1.649	0.112	6.742
14%	0.290	1.665	0.111	6.841
13%	0.287	1.678	0.109	6.932
12%	0.283	1.696	0.108	7.032
11%	0.280	1.712	0.107	7.125
10%	0.276	1.730	0.105	7.235
9%	0.272	1.749	0.104	7.347
8%	0.268	1.772	0.102	7.488
7%	0.263	1.796	0.101	7.642
6%	0.258	1.825	0.099	7.801
5%	0.250	1.864	0.096	8.018
4%	0.240	1.918	0.093	8.281
3%	0.234	1.967	0.091	8.562
2%	0.225	2.031	0.088	8.907
1%	0.212	2.136	0.081	9.659

Table 2.8. $I(0)/I(1)$ Robust Bandwidths and Critical Values Bart kernel, 15% Trimming.

	$MeanLM$		$SupLM$	
level	b	cv	b	cv
20%	0.905	2.015	0.166	6.102
19%	0.915	2.041	0.165	6.163
18%	0.919	2.055	0.163	6.224
17%	0.918	2.058	0.162	6.290
16%	0.887	1.995	0.160	6.358
15%	0.850	1.924	0.159	6.428
14%	0.822	1.878	0.157	6.493
13%	0.809	1.862	0.156	6.569
12%	0.794	1.844	0.154	6.648
11%	0.781	1.833	0.152	6.736
10%	0.741	1.797	0.150	6.847
9%	0.700	1.777	0.148	6.950
8%	0.686	1.782	0.145	7.065
7%	0.664	1.791	0.144	7.175
6%	0.662	1.811	0.141	7.332
5%	0.640	1.832	0.138	7.499
4%	0.629	1.867	0.134	7.705
3%	0.626	1.908	0.130	7.953
2%	0.625	1.974	0.126	8.274
1%	0.688	2.051	0.120	8.773

over-differenced in which case the asymptotic theory breaks down. It is interesting to note that the portion of the nuisance parameter space where over-rejections occur is different than what was found by Sayginsoy and Vogelsang (2010) for robust versions of Wald-type statistics. Simulations in Sayginsoy and Vogelsang (2010) show that robust Wald statistics tend to over-reject when θ is negative but ρ is close to 1. This means that robust versions of LM and Wald tests for a shift in mean can be used in a complementary way if an empirical researcher is concerned about over-rejections caused by a negative moving average component in the error.

Because we have shown that the choice of bandwidth has a substantial impact on the power of LM tests, we need to explore the power of the LM tests when using the robust bandwidths. Figures 2.11 and 2.12 power plots for the tests based on the robust bandwidths. We see that for the $MeanLM_{0.248}$ test, the robust bandwidth is large enough to make $MeanLM_{0.248}$ have non-monotonic power which is unfortunate. On the other hand, the robust bandwidth for $SupLM$ is relatively small and we see that $SupLM_{0.096}$ has good power that is monotonic. This is true for both asymptotic and finite sample power.

In practice when the QS kernel is used, we recommend that the $SupLM_{0.096}$ statistic be used because it is robust to strong serial correlation under the null hypothesis of no shift in mean and it has good power to detect a shift in mean. If a negative moving average component is a concern, $SupLM_{0.096}$ can be used in conjunction with one of the robust Wald-type tests of Sayginsoy and Vogelsang (2010) which have complementary over-rejection problems in the presence of a negative moving average component.

2.7 The $Wald^*$ Statistic of Kejriwal (2009)

In this section we establish a link between the LM tests analyzed in this chapter and a class of "hybrid" tests recently proposed by Kejriwal (2009). We show that the hybrid tests of Kejriwal (2009) are transformations of the LM statistics. Therefore the fixed-b asymptotic

Table 2.9. Finite Sample Null Rejection Probabilities for Tests Using Size Robust Bandwidths and Fixed-b $I(0)/I(1)$ Critical Values, 5% Nominal Level, 15% Trimming, QS Kernel.

		$T = 60$		$T = 120$		$T = 240$	
θ	ρ	$MeanLM$ $b = 0.25$	$SupLM$ $b = 0.096$	$MeanLM$ $b = 0.25$	$SupLM$ $b = 0.096$	$MeanLM$ $b = 0.25$	$SupLM$ $b = 0.096$
0.8	0	0.045	0.012	0.045	0.021	0.046	0.031
	0.5	0.030	0.006	0.036	0.012	0.040	0.019
	0.7	0.019	0.005	0.028	0.007	0.034	0.013
	0.9	0.011	0.010	0.013	0.005	0.019	0.006
	1	0.072	0.051	0.060	0.059	0.055	0.054
0.4	0	0.046	0.014	0.046	0.024	0.047	0.033
	0.5	0.031	0.007	0.036	0.013	0.040	0.020
	0.7	0.020	0.005	0.028	0.007	0.033	0.013
	0.9	0.012	0.011	0.013	0.006	0.019	0.006
	1	0.073	0.051	0.059	0.059	0.055	0.054
0	0	0.055	0.023	0.051	0.034	0.050	0.041
	0.5	0.034	0.010	0.038	0.014	0.042	0.022
	0.7	0.022	0.006	0.028	0.009	0.035	0.014
	0.9	0.013	0.013	0.014	0.006	0.019	0.006
	1	0.073	0.055	0.059	0.060	0.055	0.054
-0.4	0	0.103	0.071	0.076	0.086	0.060	0.084
	0.5	0.046	0.018	0.045	0.026	0.046	0.032
	0.7	0.029	0.012	0.032	0.014	0.038	0.019
	0.9	0.015	0.019	0.014	0.009	0.020	0.007
	1	0.073	0.067	0.060	0.065	0.054	0.057
-0.8	0	0.524	0.308	0.433	0.574	0.285	0.614
	0.5	0.187	0.049	0.157	0.151	0.112	0.195
	0.7	0.088	0.024	0.083	0.055	0.069	0.078
	0.9	0.035	0.032	0.026	0.028	0.027	0.020
	1	0.073	0.105	0.062	0.106	0.057	0.083
-1	0	0.843	0.401	0.929	0.892	0.947	0.998
	0.5	0.451	0.050	0.743	0.302	0.884	0.823
	0.7	0.226	0.016	0.475	0.089	0.753	0.418
	0.9	0.068	0.010	0.120	0.016	0.271	0.039
	1	0.055	0.023	0.051	0.034	0.050	0.041

Table 2.10. Finite Sample Null Rejection Probabilities for Tests Using Size Robust Bandwidths and Fixed-b $I(0)/I(1)$ Critical Values, 5% Nominal Level, 15% Trimming, Bartlett Kernel.

		$T = 60$		$T = 120$		$T = 240$	
θ	ρ	$MeanLM$ $b = 0.25$	$SupLM$ $b = 0.096$	$MeanLM$ $b = 0.25$	$SupLM$ $b = 0.096$	$MeanLM$ $b = 0.25$	$SupLM$ $b = 0.096$
0.8	0	0.076	0.016	0.064	0.026	0.053	0.033
	0.5	0.088	0.013	0.070	0.019	0.056	0.026
	0.7	0.090	0.013	0.075	0.016	0.061	0.021
	0.9	0.070	0.018	0.077	0.012	0.074	0.014
	1	0.066	0.041	0.057	0.053	0.052	0.050
0.4	0	0.076	0.019	0.063	0.027	0.052	0.035
	0.5	0.087	0.013	0.070	0.020	0.055	0.027
	0.7	0.089	0.013	0.075	0.016	0.061	0.022
	0.9	0.070	0.019	0.077	0.013	0.074	0.014
	1	0.067	0.041	0.057	0.053	0.052	0.050
0	0	0.070	0.021	0.059	0.030	0.050	0.039
	0.5	0.085	0.014	0.068	0.022	0.055	0.028
	0.7	0.089	0.014	0.075	0.018	0.060	0.023
	0.9	0.071	0.020	0.078	0.013	0.074	0.015
	1	0.068	0.045	0.058	0.054	0.053	0.051
-0.4	0	0.046	0.022	0.043	0.039	0.041	0.051
	0.5	0.074	0.020	0.064	0.029	0.053	0.035
	0.7	0.087	0.018	0.073	0.022	0.057	0.027
	0.9	0.073	0.025	0.076	0.017	0.074	0.017
	1	0.066	0.056	0.058	0.060	0.052	0.054
-0.8	0	0.004	0.009	0.004	0.017	0.006	0.040
	0.5	0.032	0.012	0.024	0.028	0.025	0.054
	0.7	0.057	0.016	0.044	0.031	0.038	0.046
	0.9	0.071	0.035	0.073	0.035	0.066	0.030
	1	0.067	0.099	0.058	0.099	0.053	0.075
-1	0	0.000	0.005	0.000	0.010	0.000	0.009
	0.5	0.007	0.005	0.000	0.008	0.000	0.011
	0.7	0.022	0.004	0.003	0.007	0.000	0.012
	0.9	0.060	0.008	0.033	0.009	0.009	0.008
	1	0.070	0.021	0.059	0.030	0.050	0.039

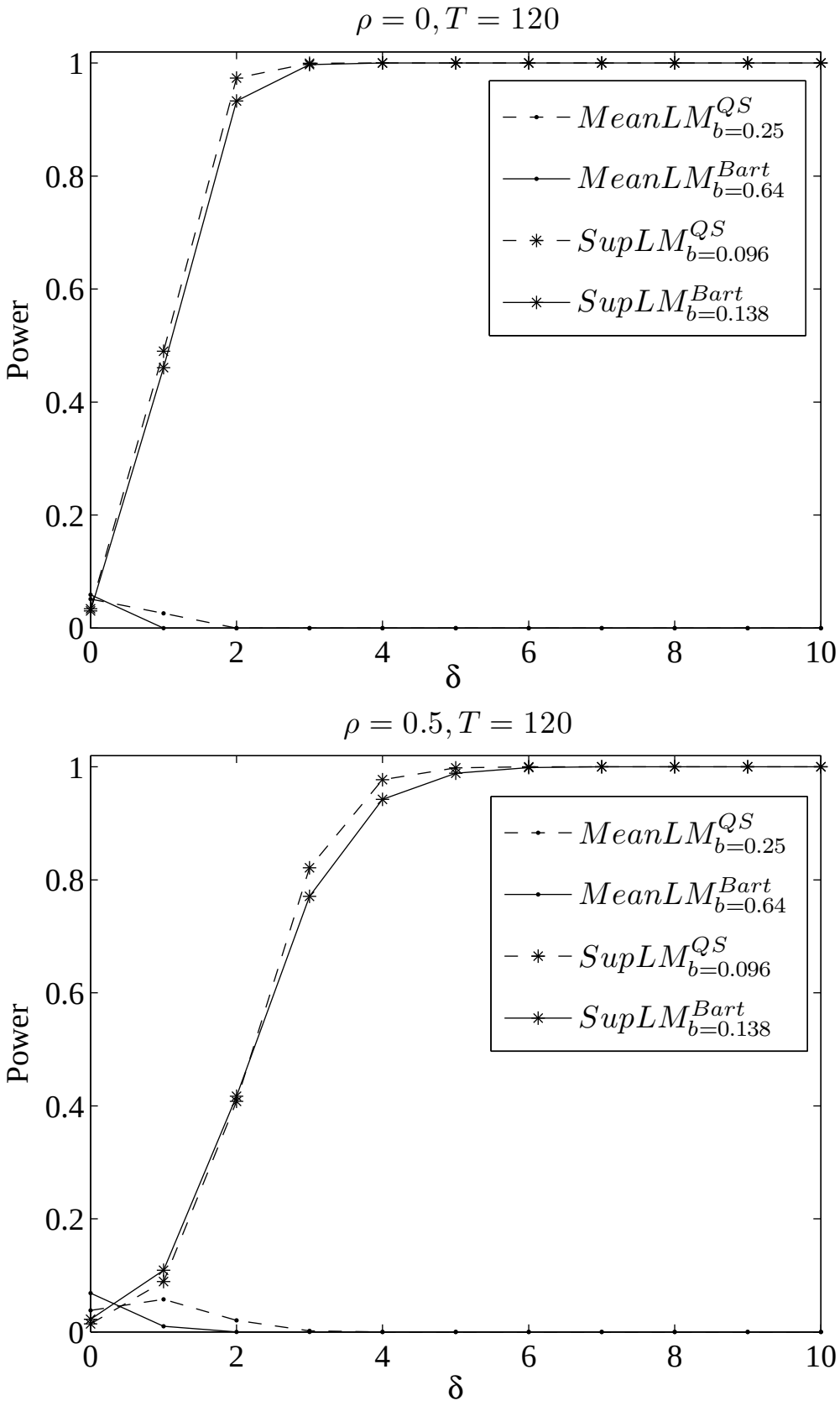

Figure 2.11. Finite Sample Power of Robust Bandwidth Tests, 5% Level, 15% Trimming.

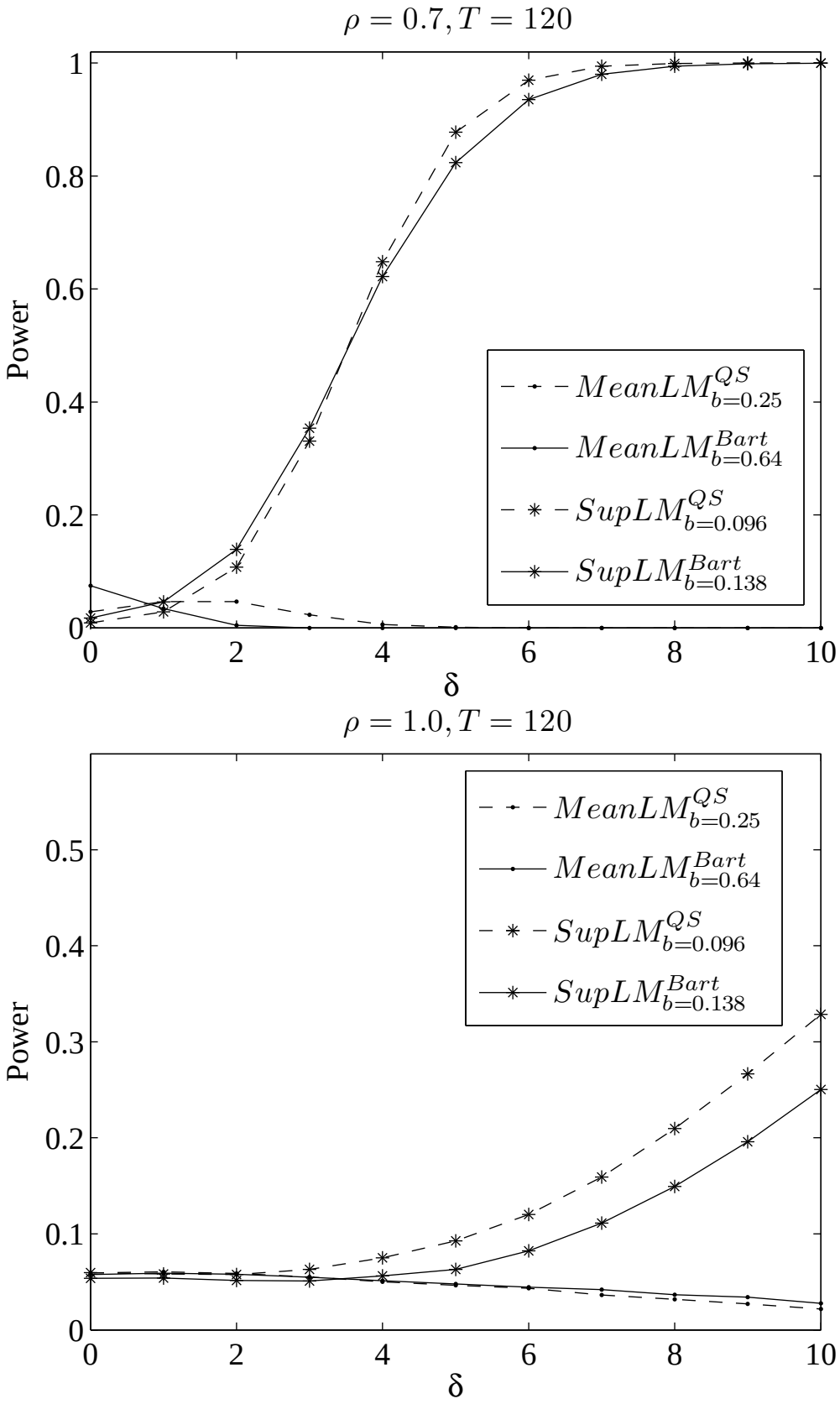

Figure 2.12. Finite Sample Power of Robust Bandwidth Tests, 5% Level, 15% Trimming.

theory we developed for the LM tests applies directly to the tests of Kejriwal (2009). Our results provide an alternative theoretical explanation for the monotonic power properties of the hybrid tests that complement the theoretical explanations developed by Kejriwal (2009).

The basic idea used by Kejriwal (2009) was to construct a test statistic that blends aspects of LM and $Wald$ statistics in a way to maintain the relatively better size properties of LM tests but to have the relatively better power (monotonic) properties of $Wald$ tests. The statistic proposed by Kejriwal (2009) is defined as

$$Wald^{*}(T_b, m) = \frac{SSR_0 - SSR(T_b)}{\widetilde{\sigma}^{*2}(m)}$$

where $\widetilde{\sigma}^{*2}(m)$ is a "hybrid" HAC estimator given by

$$\widetilde{\sigma}^{*2}(m) = \hat{\gamma}_0 + 2\sum_{j=1}^{T-1} k(j/m)\tilde{\gamma}_j, \;\; \hat{\gamma}_0 = T^{-1}\sum_{t=1}^{T} \widehat{u}_t(\widehat{T}_b)^2,$$

and $\tilde{\gamma}_j$ is defined as before using null residuals. Notice that $\widetilde{\sigma}^{*2}(m)$ can be constructed by taking $\widetilde{\sigma}^{2}(m)$ and replacing $\tilde{\gamma}_0$ with $\hat{\gamma}_0$. This estimator is a hybrid because it combines parts of HAC estimators based on null and alternative residuals.

By using some straightforward algebra it follows that

$$\widetilde{\sigma}^{*2}(m) = T^{-1}\sum_{t=1}^{T} \widehat{u}_t(\widehat{T}_b)^2 + T^{-1}\sum_{t=1}^{T} \tilde{u}_t^2 + 2T^{-1}\sum_{j=1}^{T-1} k(j/m) \sum_{t=j+1}^{T} \tilde{u}_t\tilde{u}_{t-j} - T^{-1}\sum_{t=1}^{T} \tilde{u}_t^2$$

$$= T^{-1}SSR(\widehat{T}_b) + \tilde{\sigma}^2(m) - T^{-1}SSR_0 = \tilde{\sigma}^2(m) - T^{-1}(SSR_0 - SSR(\widehat{T}_b))$$

$$= \tilde{\sigma}^2(m)\left[1 - T^{-1}LM(\widehat{T}_b, m)\right] = \tilde{\sigma}^2(m)\left[1 - T^{-1}SupLM_m\right].$$

Plugging this formula for $\widetilde{\sigma}^{*2}(m)$ into $Wald^{*}$ gives

$$Wald^{*}(T_b, m) = \frac{(SSR_0 - SSR(T_b))}{\tilde{\sigma}^2(m)\left[1 - T^{-1}SupLM_m\right]} = \frac{LM(T_b, m)}{1 - T^{-1}SupLM_m}.$$

It immediately follows that

$$MeanW_m^{*} = \frac{MeanLM_m}{1 - T^{-1}SupLM_m}, \quad SupW_m^{*} = \frac{SupLM_m}{1 - T^{-1}SupLM_m},$$

and we see that the $Wald^*$ statistics are transformations of the $MeanLM$ and $SupLM$ statistics. Notice that $MeanW_m^*$ and $MeanLM_m$ are asymptotically equivalent to first order under fixed-b asymptotics. Hence they have the same asymptotic power. The finding for $SupW_m^*$ is stronger because $SupW_m^*$ is a monotonic transformation of $SupLM_m$ and the two statistics give the exact same test. Hence $SupW_m^*$ and $SupLM_m$ have the same power functions (finite and asymptotic).

Because Kejriwal (2009) recommends using the alternative data dependent bandwidth, $\widehat{m}$, the theoretical explanations we provide for the $LM_{\widehat{m}}$ statistics directly apply to the $Wald_{\widehat{m}}^*$ statistics. In particular, our fixed-b theoretical results show that the reason $MeanW_{\widehat{m}}^*$ and $SupW_{\widehat{m}}^*$ have monotonic power is because they are equivalent to LM tests based on the bandwidth $\widehat{m}$ which tends to be relatively small. If the null data dependent bandwidth, $\widetilde{m}$, were used for the $Wald^*$ statistics, they would have non-monotonic power just like the LM tests.

CHAPTER 3

Consistency of the Sequential Trend Break Point Estimator

3.1 Introduction

A time series can have multiple breaks, and it is common that the break number is unknown and we misspecify it. Chong (1994), Chong (1995), and Bai (1995) analyzed the consistency of change-point estimators when the number of change points is under-specified. They pointed out that even when the number of breaks for the mean shift model is underspecified, the break date estimators are still consistent to the subset of the true break points. Although the trending components are considered by some researchers in mean shifts model, the existing literatures have seldom detailed discussions on the consistency of the misspecified multiple trend shifts estimator. The consistency analysis is important to both break point estimates and the break hypothesis tests. This is one main motivation of this chapter: there should be more concerns on the consistency of the break point estimators for the multiple trend shifts model.

The second motivation of this chapter is to explore how to better approximate the finite sample distributions for multiple breaks model. The existing literatures on multiple break

point estimators only provide consistency analysis on multiple mean shifts model but no distributions, which would be a big plus to understand how multiple breaks estimator performs. Aiming at the consistency analysis, I provide the asymptotics of the multiple trend breaks estimator under misspecified break number using Pitman shifts. This work follows Yang (2010) which has shown that the finite sample distribution of the single break point estimator is not normal but depends on both the break magnitude and the break date. The asymptotics of multiple break points estimator would be more complicated. Also there is a need to verify its approximation to the finite samples.

In the following, I will first use finite sample simulation to show the possible inconsistency of the break point estimator in the multiple trend shifts model with under-specified break number. Then a new asymptotics will be provided for the break point estimator under local alternative, which is an extension of Yang (2010)'s work on the single break point estimator. It proves that for the trend shift model, the break point estimator can be inconsistent with any of the true break points, while for the mean shift model, the break point estimator converges to one of the true breaks. Furthermore, first-differencing the multiple trend shift model is suggested to eliminate the inconsistency problem. Applying first difference to the sequential process of the multiple trend break estimation may improve the power of the trend shift test.

The chapter is organized as follows. The second section includes the definition of model, assumptions, and the break point estimators. The third section reviews the existing analysis on the break point estimators. Finite sample simulations are introduced to demonstrate consistency properties of different break point estimators. The fourth section provides the asymptotic distributions of the single break point estimators when the data sequences have two breaks under I(0) errors. Both mean shift and trend shift break point estimators are discussed. The fifth section provides the trend shift break point estimator and the first difference estimator under I(1) errors. The asymptotic results discover how the inconsistency can be partially solved by first differencing. The performance in the pdf of two estimators

are compared. The sixth section provides the application of first difference break point estimator to the sequential trend break test proposed by KP(2010).

3.2 Model Assumption

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 the models with more than two breaks.

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

$$y_t = \mu + \delta_1 DU_t(\lambda_1^c) + \delta_2 DU_t(\lambda_2^c) + u_t. \tag{3.2.1}$$

In model (3.2.1) λ_1^c and λ_2^c are the true break points with $T_1^c = T\lambda_1^c$ and $T_2^c = T\lambda_2^c$. To be convenient in discussion, we define $\nu \doteq \delta_2/\delta_1$. The underspecified model (3.2.1) is denoted by

$$y_t = \mu + \delta DU_t(\lambda^c) + u_t, \tag{3.2.2}$$

where $DU_t(\lambda^c) \doteq I(t > T_b^c)$, $T_b^c = T\lambda^c$, T is the sample length, and

$$I(t > T_b^c) \doteq \begin{cases} 0, & t \leq T_b^c \\ 1, & t > T_b^c \end{cases}.$$

As a comparison, the trend shift model with two breaks is

$$y_t = \mu + \beta t + \delta_1 DT_t(\lambda_1^c) + \delta_2 DT_t(\lambda_2^c) + u_t, \tag{3.2.3}$$

where δ_1 and δ_2 are the break magnitudes, λ_1^c and λ_2^c are the true break points. And if the model (3.2.3) is misspecified with only 1 break, it is denoted as

$$y_t = \mu + \beta t + \delta DT_t(\lambda^c) + u_t, \tag{3.2.4}$$

where

$$DT_t(\lambda^c) \doteq \begin{cases} 0, & t \leq T_b^c \\ t - T_b^c, & t > T_b^c \end{cases}.$$

I(0) errors are defined by assumption (A1.a).

$$(C1.a) \qquad u_t = \rho u_{t-1} + \varepsilon_t,$$

where $|\rho| < 1$ and

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

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 $E(e_t^2|e_{t-1}, e_{t-2}, \cdots) = 1$.

And I(1) errors are defined by assumption (C1.b)

$$(C1.b) \qquad u_t = \rho u_{t-1} + \varepsilon_t,$$

where $\rho = 1 - \frac{c}{T}$, $c \geq 0$ is a constant scalar.

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

$$\begin{aligned} \hat{\lambda}_{MS} &= \arg\min_{\lambda \in \Lambda^*} \{SSR_{MS}(\lambda)\}, \\ \hat{\lambda}_{TS} &= \arg\min_{\lambda \in \Lambda^*} \{SSR_{TS}(\lambda)\}, \end{aligned}$$

where

$$SSR_{MS}(\lambda) \doteq \sum_{t=1}^{T} [y_t - \hat{\mu}_{MS} - \hat{\delta}_{MS} DU_t(\hat{\lambda})]^2, \tag{3.2.5}$$

$$SSR_{TS}(\lambda) \doteq \sum_{t=1}^{T} [y_t - \hat{\mu}_{TS} - \hat{\beta}_{TS} t - \hat{\delta}_{TS} DT_t(\hat{\lambda})]^2. \tag{3.2.6}$$

$\hat{\beta}_{MS}$ and $\hat{\delta}_{MS}$ are the OLS estimates of Model (3.2.2) with no restrictions imposed. $\hat{\mu}_{TS}$, $\hat{\beta}_{TS}$, and $\hat{\delta}_{TS}$ are the OLS estimates of Model (3.2.4) with no restrictions imposed.

3.3 Existing Analysis and Finite Sample Simulations

Chong (1994), Chong (1995), and Bai (1995) studied the consequences of under-specifying the number of change points in structural change models. A general case with a single break in the intercept is estimated when the data sequence has two breaks. Their discussion covers the mean shift model with or without trending. Bai and Perron (1998)(BP hereafter) extended the estimate of single unknown break to multiple unknown breaks under both fixed and shrinking shift magnitudes. They concluded that the break point estimator still converges to one of the true breaks for the mean shifts model. Based on this argument, they proposed a sequential procedure for multi-break estimates without estimating the multiple breaks simultaneously. Dynamic programming was introduced by Bai and Perron (2003) to deal with the computational burden in multiple break point estimation. Kejriwal and Perron (2010)(KP hereafter) extended the sequential tests to the multiple trend shifts model to be robust to the persistence in the noise.

In the following, I first use a simple simulation to illustrate the properties of $\hat{\lambda}_{MS}$ and $\hat{\lambda}_{TS}$ in finite samples in the presence of under-specification of break number. I generate data based on model (3.2.1) and (3.2.3) with two breaks, where $T = 100, 250, 500, 1000$, $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$, $\nu = -2, -1, 1, 2$ ($\delta_1 = 1$), $d(L) = 1$ and ε_t is an iid $N(0,1)$ process. And set $\delta_1 = 1$ without loss of generality. Equation (3.2.5) and (3.2.6) are used to estimate $\hat{\lambda}_{MS}$ and $\hat{\lambda}_{TS}$ separately in each replication. Trimming is not necessary, however in order to ensure the invertibility of the regression matrix, I use 2% trimming, i.e., $\lambda^* = 0.02$. The replications $N = 20000, 10000, 5000, 2500$ for $T = 100, 250, 500, 1000$ cases.

Figure 3.1 and 3.2 plot the histograms of $\hat{\lambda}_{MS}$ when $\mu = -2, -1, 1, 2$, $T = 100, 250, 500, 1000$ and error u_t is i.i.d. $N(0,1)$. Interestingly, when $|\nu| = 1$ and $T = 100$, the histogram of $\hat{\lambda}_{MS}$ has a wide spread over the area around both ends or the middle $\hat{\lambda} = 0.5$. This can be explained by Yang (2010) through the behavior of the mean shift

break point estimator, where the break point estimates concentrate around the ends of the gridding area in no break case. With the increase of T, the tails of $\hat{\lambda}_{MS}$ decrease gradually. When $T = 1000$, the histogram of $\hat{\lambda}_{TS}$ concentrates at one true break $\lambda_c = 1/3$ or $\lambda_c = 2/3$. For $\mu = 1$, there are two peaks in histogram because δ_1 and δ_2 has the same effects on the break point estimates.

Figure 3.3 and 3.4 plot the histograms of $\hat{\lambda}_{TS}$ when $\nu = -2, -1, 1, 2(\delta_1 = 1)$, $T = 100, 250, 500, 1000$. For $\nu = -2$, the only peak in histogram of $\hat{\lambda}_{TS}$ is at $\lambda > 2/3$. When $\nu = -1$, $\hat{\lambda}_{TS}$ has two equivalent peaks in histogram at around $\lambda = 0.2$ and 0.8. When $\nu = 1$, the histogram of $\hat{\lambda}_{TS}$ roughly has only one peak in the histograms for the stationary cases, at around $\lambda = 0.5$. When $T = 100$, the break point estimates are less concentrated around $\lambda = 0.5$. With the increase of T, the pattern of only one peak in the histograms still holds, and the break estimates are more concentrated around $\lambda = 0.5$. When $\nu = 2$, the histogram of $\hat{\lambda}_{TS}$ concentrates at $1/3 < \lambda < 2/3$. In all these cases, the break date estimates are mostly at 0.5. When $\nu = 2$ and $T = 1000$, the histograms concentrates on the points other than the true breaks. It shows that when the misspecification of break number exists, the break point estimator for the trend shift model does not converge to either of the true breaks. And if λ_1^c and λ_2^c are different, the concentration of the break point estimators varies, i.e., the limits of the break point estimator $\hat{\lambda}_{TS}$ depend on both break magnitudes and break dates. How these parameters matters will be analyzed later.

The finite sample histograms show two interesting points: a) In the case of underspecification of the break number, the mean shift break 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. The limits of the break point estimators depend on the break dates and the break magnitudes. b) When break magnitude increases, the break point estimators have

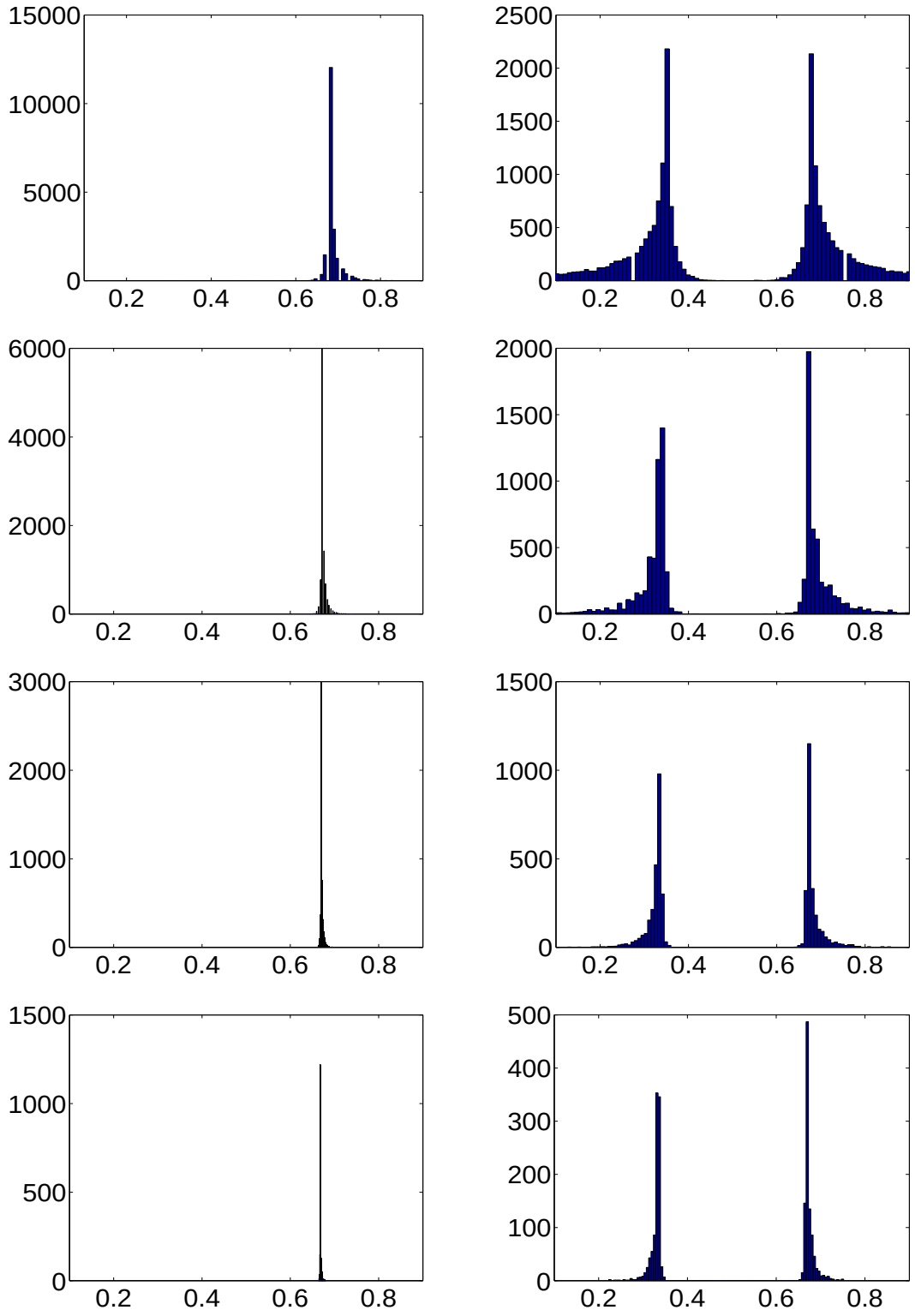

Figure 3.1. Histogram of single break point estimator $\hat{\lambda}_{MS}$ in two breaks model: $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$. $\delta_1 = 1$ always. From left to right: $\nu = -2(\delta_2 = -2), -1(\delta_2 = -1)$; from top to bottom: $T = 100, 250, 500, 1000$.

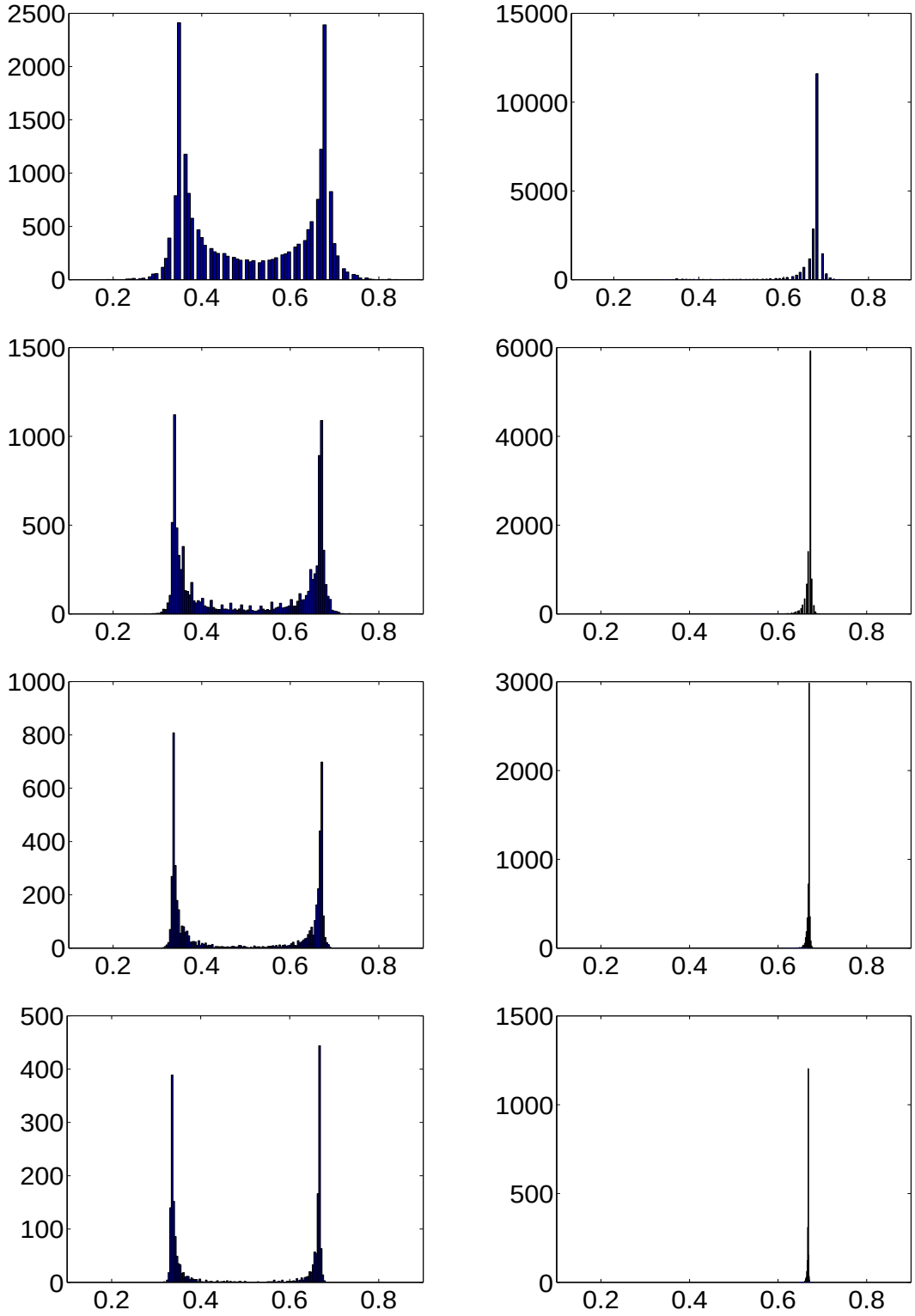

Figure 3.2. Histogram of single break point estimator $\hat{\lambda}_{MS}$ in two breaks model: $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$. $\delta_1 = 1$ always. From left to right: $\nu = 1(\delta_2 = 1), 2(\delta_2 = 2)$; from top to bottom: $T = 100, 250, 500, 1000$.

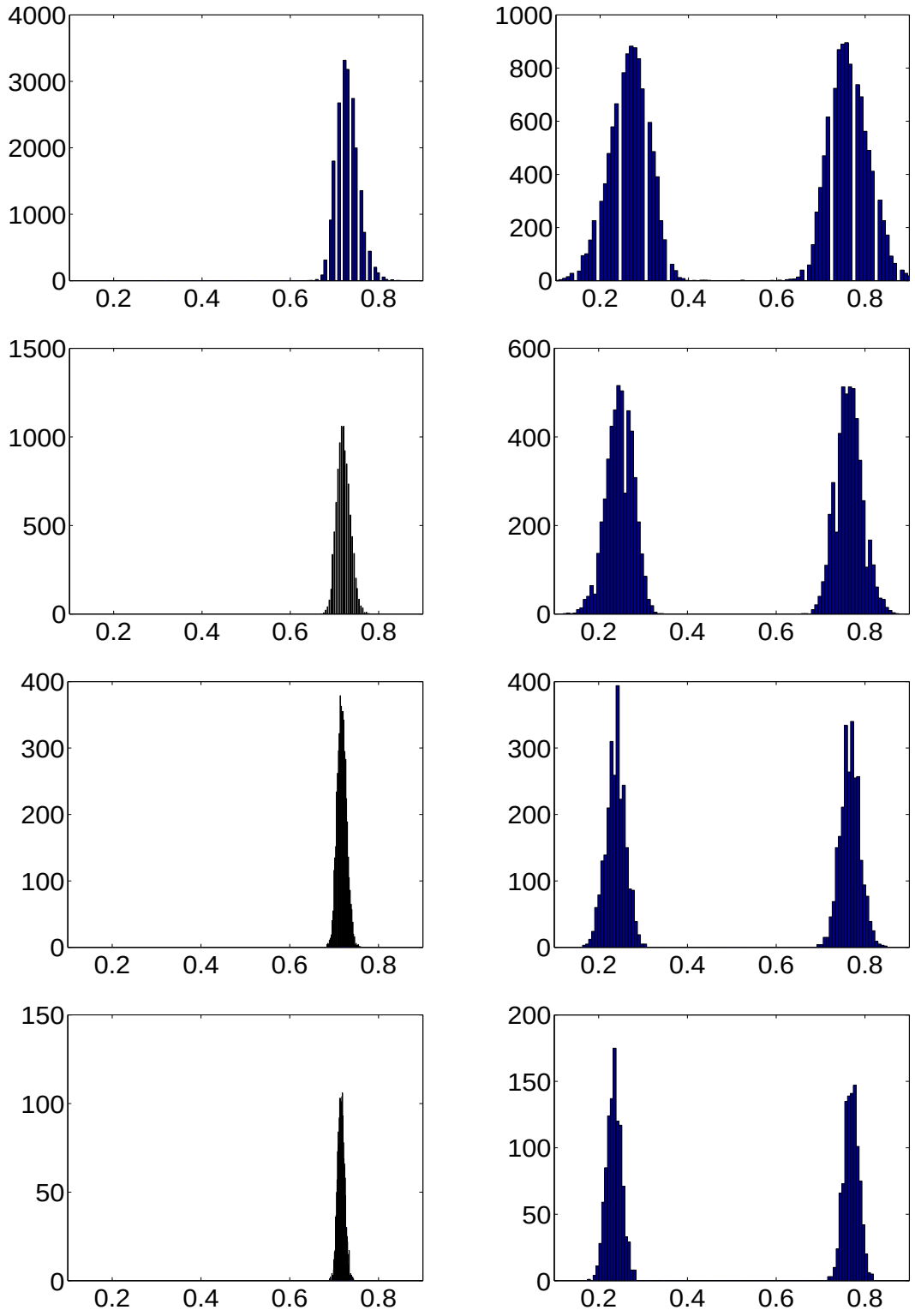

Figure 3.3. Histogram of single break point estimator $\hat{\lambda}_{TS}$ in two breaks: $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$. $\delta_1 = 1$ always. The left to right: $\nu = -2(\delta_2 = -2), -1(\delta_2 = -1)$; The top to bottom: $T = 100, 250, 500, 1000$.

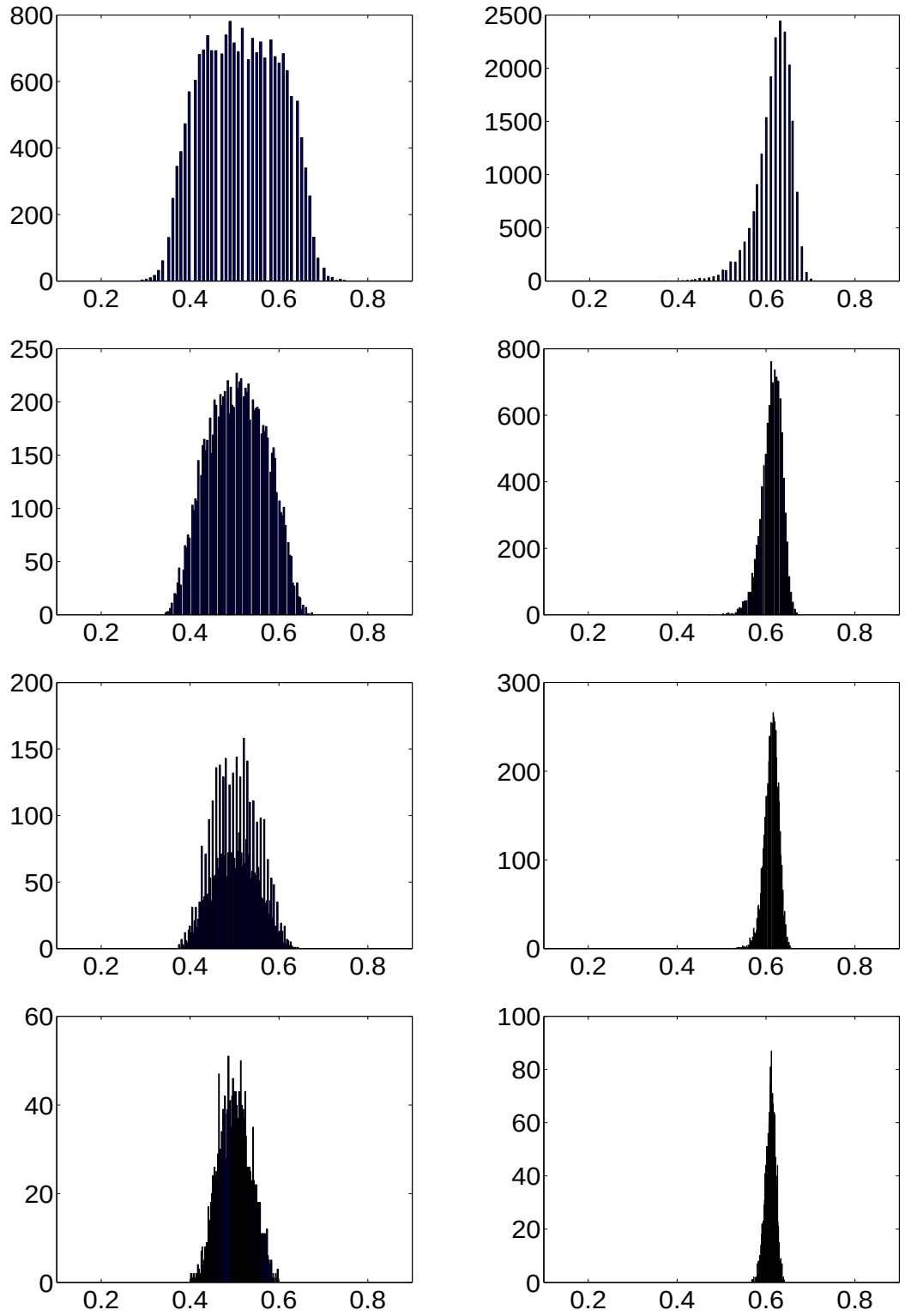

Figure 3.4. Histogram of single break point estimator $\hat{\lambda}_{TS}$ in two breaks: $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$. $\delta_1 = 1$ always. The left to right: $\nu = 1(\delta_2 = 1), 2(\delta_2 = 2)$; The top to bottom: $T = 100, 250, 500, 1000$.

complicated distributions which have not been explained by current results.

3.4 Break Date Estimator under Multiple Breaks

I can assume that break magnitude δ_1 and δ_2 are within a $T^{-1/2}$ neighborhood:

$$(C2.a) \qquad \delta_1 = \frac{\delta_1^*}{T^{1/2}}, \ \ \delta_2 = \frac{\delta_2^*}{T^{1/2}}, \ \text{ where } \ \delta_1^*, \delta_2^* = \text{ constant scalars.} \quad (3.4.7)$$

Also I can assume that break magnitude δ_1 and δ_2 are within a $T^{-3/2}$ neighborhood:

$$(C2.b) \qquad \delta_1 = \frac{\delta_1^*}{T^{3/2}}, \ \ \delta_2 = \frac{\delta_2^*}{T^{3/2}}, \ \text{ where } \ \delta_1^*, \delta_2^* = \text{ constant scalars.} \quad (3.4.8)$$

The limiting distributions of $\hat{\lambda}_{MS}$ and $\hat{\lambda}_{TS}$ with under-specified break number are derived under assumption $(C2.a)$ or $(C2.b)$ for different models and errors.

3.4.1 Multiple mean shifts

Theorem 3.4.4 *Assume the mean shift model has two break points, λ_1^c and λ_2^c, as in (3.2.1). When the break number is underspecified as one and the assumptions (C1.a) and (C2.a) hold such that $\delta_1 = T^{-1/2}\delta_1^*$ and $\delta_2 = T^{-1/2}\delta_2^*$, where δ_1^* and δ_2^* are constant, the break point estimator $\hat{\lambda}_{MS}$ has the limiting distributions as follows:*

$$\hat{\lambda}_{MS} \ \xrightarrow{d} \ \arg\max_{\lambda\in\Lambda}\{\frac{[(\lambda W(1) - W(\lambda)) + \frac{\delta_1^*}{d(1)}\Psi(\lambda,\lambda_1^c) + \frac{\delta_2^*}{d(1)}\Psi(\lambda,\lambda_2^c)]^2}{\lambda(1-\lambda)}\} \quad (3.4.9)$$

where

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

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

$$\hat{\lambda}_{MS} \approx \arg\max_{\lambda\in\Lambda}\{\frac{[(\lambda W(1) - W(\lambda)) + M_1\Psi(\lambda,\lambda_1^c) + M_2\Psi(\lambda,\lambda_2^c)]^2}{\lambda(1-\lambda)}\}. \quad (3.4.10)$$

To discover the effect of M_1, λ_1^c, M_2, and λ_2^c on the limiting distributions, I decompose the terms inside $\arg\max$ in equation (3.4.10) into three parts:

$$\begin{aligned} & G_{MS}(\lambda, \lambda_1^c, \lambda_2^c) \\ \doteq\ & \frac{\lambda W(1) - W(\lambda) + M_1 \Psi(\lambda, \lambda_1^c) + M_2 \Psi(\lambda, \lambda_2^c)}{\sqrt{\lambda(1-\lambda)}} \qquad (3.4.11) \\ \doteq\ & G1_{MS}(\lambda) + M_1 \cdot G2_{MS}(\lambda, \lambda_1^c) + M_2 \cdot G2_{MS}(\lambda, \lambda_2^c) \\ \doteq\ & \frac{(\lambda W(1) - W(\lambda))}{\sqrt{\lambda(1-\lambda)}} + M_1 \cdot \frac{\Psi(\lambda, \lambda_1^c)}{\sqrt{\lambda(1-\lambda)}} + M_2 \cdot \frac{\Psi(\lambda, \lambda_2^c)}{\sqrt{\lambda(1-\lambda)}} \qquad (3.4.12) \end{aligned}$$

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 3.4.4 provides a bridge between the $\delta = 0$ asymptotics and the $\delta \neq 0$ 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 break.

Theorem 3.4.4 also explains why as M grows, $\hat{\lambda}_{MS}$ are closer to the true breaks. With the increase of T, where M_1 and M_2 increase, $G2_{MS}$ parts will be dominant in $G1_{MS} + M_1 \cdot G2_{MS}(\lambda_1^c) + M_2 \cdot G2_{MS}(\lambda_2^c)$. 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 $G2_{MS}$ parts in the asymptotics. If $T \to \infty$, both M_1 and M_2 increase to ∞,

$$\begin{aligned} \lim_{T\to\infty} \hat{\lambda}_{MS} &= \lim_{T\to\infty} \arg\max_{\lambda\in\Lambda} [G1_{MS}(\lambda) + M_1 \cdot G2_{MS}(\lambda, \lambda_1^c) + M_2 \cdot G2_{MS}(\lambda, \lambda_2^c)]^2 \\ &= \lim_{T\to\infty} \arg\max_{\lambda\in\Lambda} [G1_{MS}(\lambda)/M_1 + G2_{MS}(\lambda, \lambda_1^c) + \nu \cdot G2_{MS}(\lambda, \lambda_2^c)]^2 \\ &\to \arg\max_{\lambda\in\Lambda} |G2_{MS}(\lambda, \lambda_1^c) + \nu \cdot G2_{MS}(\lambda, \lambda_2^c)|. \end{aligned}$$

where $\nu \doteq M_2/M_1 = \delta_2/\delta_1$. Therefore, the limit of $\hat{\lambda}_{MS}$ is determined by $|G2_{MS}(\lambda, \lambda_1^c) + \nu \cdot G2_{MS}(\lambda, \lambda_2^c)|$, which attains global maximum at either λ_1^c or λ_2^c as shown in Figure 3.5 (The proof is straightforward and omitted here). Hence, $\hat{\lambda}_{MS}$ converges to λ_1^c or λ_2^c as $T \to \infty$, that is the break point estimator $\hat{\lambda}_{MS}$ is consistent to one of the true breaks. Our results agree the existing literatures in consistency analysis, and moreover the distribution is derived here.

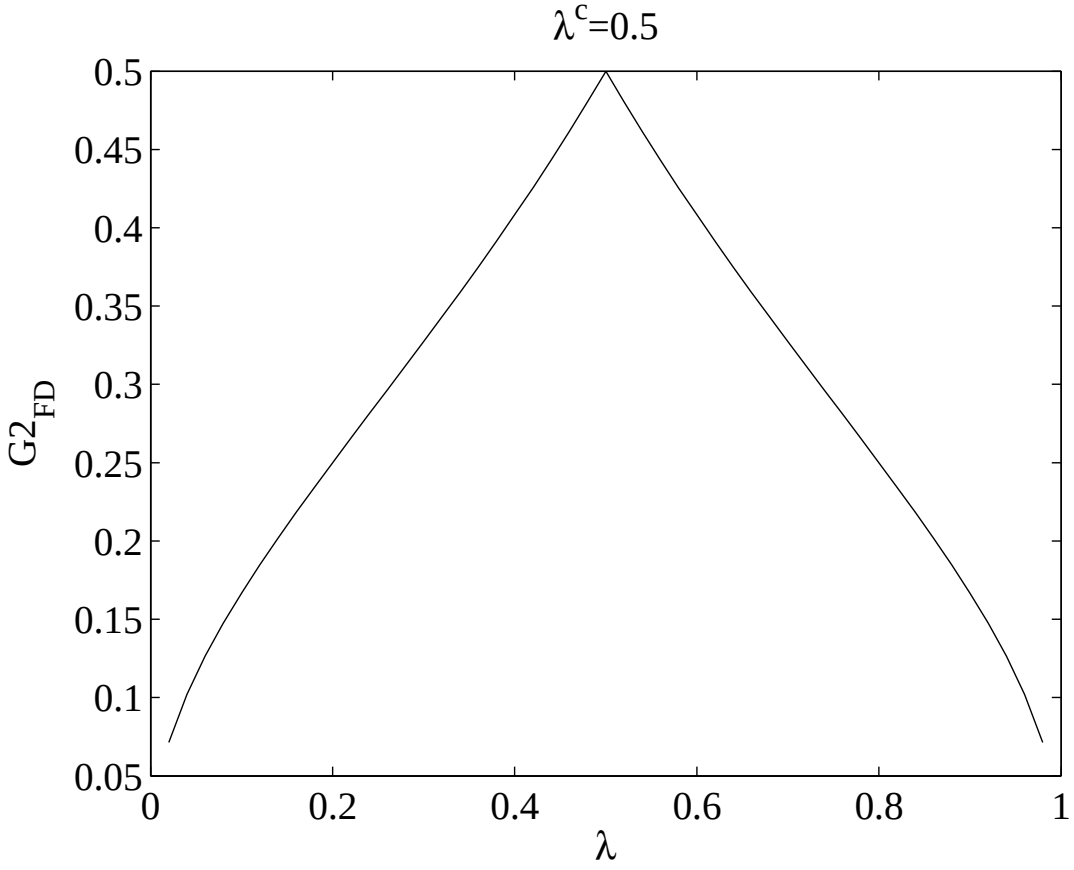

Figure 3.5. $G2_{MS}(\lambda, \lambda^c)$ under $\lambda^c = 0.5$ for mean shift model

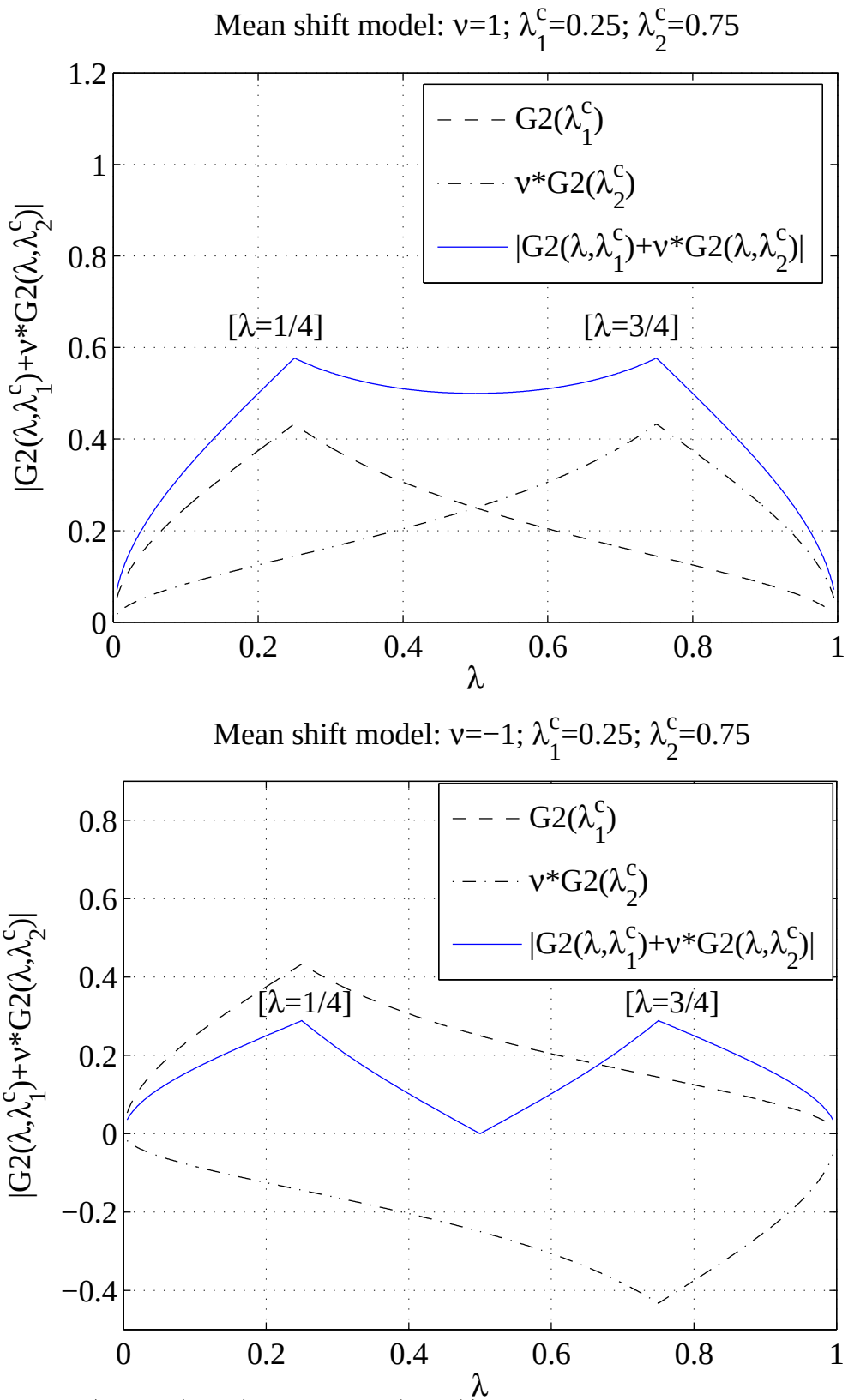

Figure 3.6. $|G2_{MS}(\lambda, \lambda_1^c) + \nu \cdot G2_{MS}(\lambda, \lambda_2^c)|$ under different $\nu = 1$ and -1 for mean shift model, where $\{\lambda_1^c, \lambda_2^c\} = \{1/4, 3/4\}$.

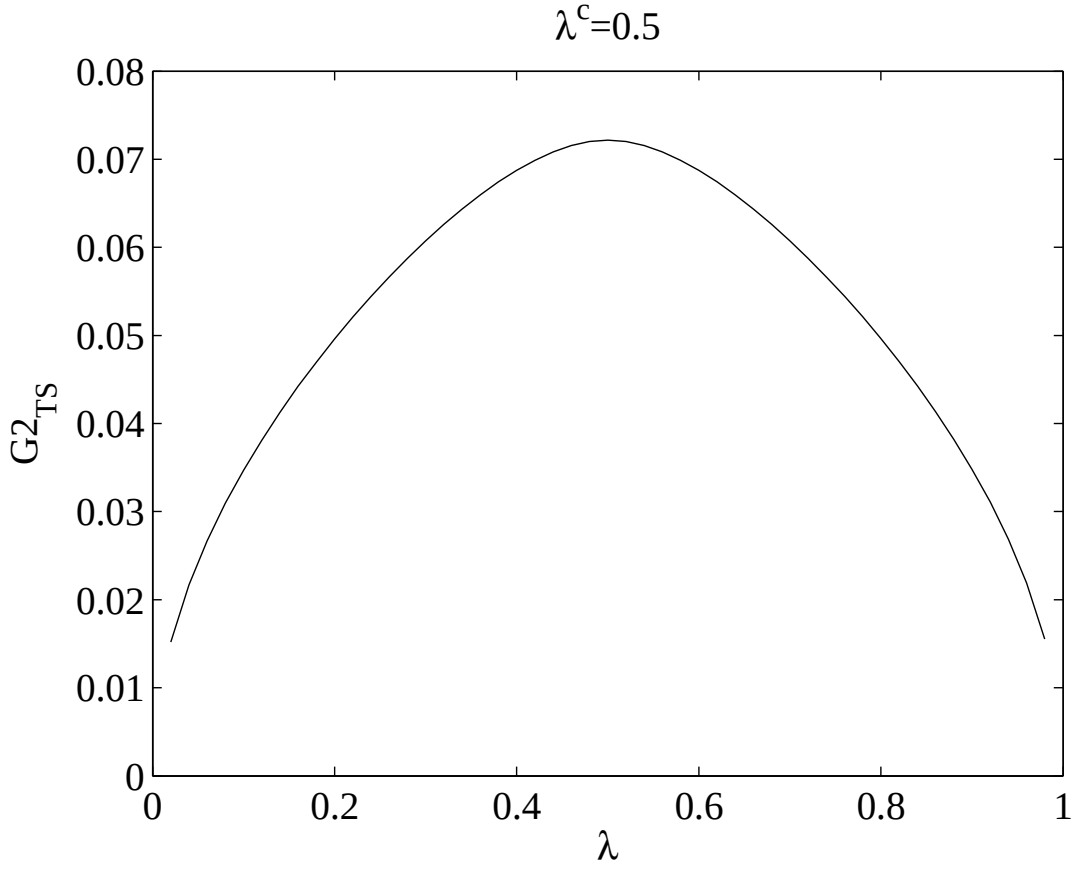

Figure 3.7. $G2_{TS}(\lambda, \lambda^c)$ under $\lambda^c = 0.5$ for trend shift model

3.4.2 Multiple trend shifts

When the break number is underspecified, we estimate one break point by model (3.2.4), while the true model is (3.2.3). In the following I explore the consistency of break points by deriving the asymptotics under local alternative. Theorem 3.4.5 gives the limiting distribution of $\hat{\lambda}_{TS}$ under assumption (C1.a) and (C2.b) and shows that it is inconsistent to any of the true break points.

Theorem 3.4.5 *Assume the trend shift model has two break points, λ_1^c and λ_2^c, as in (3.2.3). When the break number is underspecified as one and assumption (C1.a) and (C2.b) hold such that $\delta_1 = T^{-3/2}\delta_1^*$ and $\delta_2 = T^{-3/2}\delta_2^*$, the break point estimator $\hat{\lambda}_{TS}$ has the limiting distributions as follows:*

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

$$\arg\max_{\lambda\in\Lambda}\{[\int_0^1 F(r,\lambda)dW(r) + \frac{\delta_1^*}{d(1)}\int_0^1 F(r,\lambda)F(r,\lambda_1^c)dr + \frac{\delta_2^*}{d(1)}\int_0^1 F(r,\lambda)F(r,\lambda_2^c)dr]^2 / \int_0^1 F(r,\lambda)^2 dr\} \tag{3.4.13}$$

where

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

If we define $M_1 = \frac{\delta_1^*}{d(1)} \equiv \frac{\delta_1 T^{3/2}}{d(1)}$ and $M_2 = \frac{\delta_2^*}{d(1)} \equiv \frac{\delta_2 T^{3/2}}{d(1)}$, there is

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

$$\arg\max_{\lambda\in\Lambda}\{[\int_0^1 F(r,\lambda)dW(r) + M_1\int_0^1 F(r,\lambda)F(r,\lambda_1^c)dr + M_2\int_0^1 F(r,\lambda)F(r,\lambda_2^c)dr]^2 \int_0^1 F(r,\lambda)^2 dr\}. \tag{3.4.14}$$

For the asymptotic distribution of $\hat{\lambda}_{TS}$, we can also decompose the part inside $\arg\min$ into $G1_{TS}$ and $G2_{TS}$ part, where

$$\begin{aligned} & G_{TS}(\lambda, \lambda^c) \\ \doteq\ & G1_{TS}(\lambda) + M_1 \cdot G2_{TS}(\lambda, \lambda_1^c) + M_2 \cdot G2_{TS}(\lambda, \lambda_2^c) \\ \doteq\ & \frac{\int_0^1 F(r,\lambda) dW(r)}{\sqrt{\int_0^1 F(r,\lambda)^2 dr}} + \\ & M_1 \cdot \frac{\int_0^1 F(r,\lambda) F(r,\lambda_1^c) dr}{\sqrt{\int_0^1 F(r,\lambda)^2 dr}} + M_2 \cdot \frac{\int_0^1 F(r,\lambda) F(r,\lambda_2^c) dr}{\sqrt{\int_0^1 F(r,\lambda)^2 dr}} \end{aligned} \tag{3.4.15}$$

Similar to previous discussion on mean shifts, if M_1 and M_2 are small, G_{TS} is dominated by that of $G1_{TS}$. The asymptotic distribution will be close to the distribution when there is no break. Later on $G2_{TS}$ part starts to dominate. If $T \to \infty$ ($M_1, M_2 \to \infty$),

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

And it is still true that $G2(\lambda, \lambda_i^c)$ achieves maximum at $\lambda = \lambda_i^c$ as shown in Figure 3.7. What makes it different from the mean shifts case is: when we stack one part of $G2$ to the other, the function smooths out through the two peaks at the λ_i^c's. Hence when the number of trend breaks is two while assumed to be one, $|G2_{TS}(\lambda, \lambda_1^c) + \nu \cdot G2_{TS}(\lambda, \lambda_2^c)|$ achieves maximum neither at λ_1^c nor at λ_2^c. Figure 3.8 plots $|G2(\lambda, \lambda_1^c) + \nu \cdot G2(\lambda, \lambda_2^c)|$ under different ν when $\lambda_1^c = 1/3$ and $\lambda_2^c = 2/3$, which shows in both cases $|G2(\lambda, \lambda_1^c) + \nu \cdot G2(\lambda, \lambda_2^c)|$ reaches peak at neither of the true break points. Certainly, if $|\nu|$ is smaller than 1, $\hat{\lambda}_{TS}$ will be closer to λ_1^c; if $|\nu|$ is bigger than 1, $\hat{\lambda}_{TS}$ will be closer to λ_2^c. This indicates the inconsistency of the trend shifts estimator when the break number is underspecified.

3.4.3 Consistency/Inconsistency conclusion of $\hat{\lambda}_{MS}$ and $\hat{\lambda}_{TS}$

In this section, we summarize the previous analysis on the consistency/inconsistency of $\hat{\lambda}_{MS}$ and $\hat{\lambda}_{TS}$ under assumption (C1.a) and (C2.a):

1. For mean shift model with 2 breaks, if the break magnitude $\delta_1 \neq 0$ and $\delta_2 \neq 0$, the single break point estimator $\hat{\lambda}_{MS}$ is consistent to either λ_1 or λ_2:

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

2. For trend shift model with two breaks, if the break magnitude $\delta_1 \neq 0$ and $\delta_2 \neq 0$, the single break point estimator $\hat{\lambda}_{TS}$ is inconsistent to either λ_1 or λ_2 :

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

The limit depends on λ_1^c, λ_2^c, and ν, and will be

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

3.5 Break point estimators for level and first difference model under multiple breaks

In this section, I first difference the trend shift model with multiple breaks to solve the inconsistency problem under near-I(1) errors. We choose near-I(1) errors because it can provide a good approximation to finite sample case with different persistence in the errors. In the case of near-I(1) errors, I assume that δ_1 and δ_2 are within a $T^{-1/2}$ neighborhood as in (C2.a): $\delta_1 = \frac{\delta_1^*}{T^{1/2}}$ and $\delta_2 = \frac{\delta_2^*}{T^{1/2}}$.

As we can see in the previous section, the limit of the break point estimator with under-specified break number depends on the model, that is, the mean shift model leads to the consistent break point estimator while the trend shift model does not. What follows is that if we take the first difference on the trend shift model, we might solve the inconsistency problem.

Let us start with the trend shift model with two breaks:

$$y_t \ = \ \mu + \beta t + \delta_1 DT_t(\lambda_1^c) + \delta_2 DT_t(\lambda_2^c) + u_t. \tag{3.5.16}$$

First Difference it, we get:

$$\Delta y_t \quad = \quad \beta + \delta_1 DU_t(\lambda_1^c) + \delta_2 DU_t(\lambda_2^c) + \Delta u_t. \tag{3.5.17}$$

The asymptotics are different for $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ during this procedure. I extend the results in Theorem 3.4.4 and 3.4.5 for I(0) errors to near-I(1) errors to describe the performance of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ under different λ_1^c, δ_1, λ_1^c, and δ_2.

Theorem 3.5.6 *Assume there are two breaks in the trend shift model (3.5.16). Suppose the regressions in the level model (3.5.16) and its first difference (3.5.17) are estimated assuming a single break $\lambda \in \Lambda \subseteq (0,1)$ where λ_1^c and λ_2^c are the true breaks. Under the assumption $(C1.b)$ and $(C2.a)$, the break point estimators by minimizing the $SSR(\lambda)$ have the limiting distributions as follows.*

1. For the level model (3.2.3),

$$\hat{\lambda}_{TS} \Rightarrow$$
$$\arg\max_{\lambda\in\Lambda}\{\frac{[\int_0^1 F(r,\lambda)V_c(r)dr + M_1\int_0^1 F(r,\lambda)F(r,\lambda_1^c)dr + M_2\int_0^1 F(r,\lambda)F(r,\lambda_2^c)dr]^2}{\int_0^1 F(r,\lambda)^2 dr}\}$$

where $V_c(r) \doteq \int_0^r \exp(-c(r-s))dW(s)$, $M \doteq \frac{\delta^*}{d(1)}$ and $F(r,\lambda)$ is defined in Theorem 3.4.5.

2. For the first difference (3.2.1),

$$\hat{\lambda}_{MS} \Rightarrow$$
$$\arg\max_{\lambda\in\Lambda}\{[\frac{(\lambda W(1) - W(\lambda) - c\int_0^1 (1(r>\lambda)-(1-\lambda))V_c(r)dr)}{\sqrt{\lambda(1-\lambda)}} + \frac{M_1(\Psi(\lambda,\lambda_1^c) + M_2(\Psi(\lambda,\lambda_2^c)}{\sqrt{\lambda(1-\lambda)}}]^2\}$$

where $M \doteq \frac{\delta^*}{d(1)}$ and $\Psi(\lambda,\lambda^c)$ is defined in Theorem 3.4.4.

The asymptotics in Theorem 3.5.6 are the extension of the work by Yang (2010) from single break case to multiple breaks case. Compared to Theorem 3.4.5 and 3.4.4, $G1_{TS}$ and $G1_{MS}$ are different in Theorem 3.5.6:

$$G1_{TS} \doteq \frac{\int_0^1 F(r,\lambda)V_c(r)dr}{\sqrt{\int_0^1 F(r,\lambda)^2 dr}},$$

$$G1_{MS} \doteq \frac{(\lambda W(1) - W(\lambda) - c\int_0^1 (1(r>\lambda) - (1-\lambda))V_c(r)dr)}{\sqrt{\lambda(1-\lambda)}}.$$

Though $G1$ is different, when $T \to \infty$, $G1$ will not show in the limit equation, therefore the limits of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ are the same as in Theorem 3.4.5 and 3.4.4. Hence $\hat{\lambda}_{TS}$ on the level model is inconsistent, while $\hat{\lambda}_{MS}$ on the first difference model converges to either λ_1^c or λ_2^c. That is the first difference break point estimator $\hat{\lambda}_{MS}$ solves the inconsistency problem of $\hat{\lambda}_{TS}$.

Figure 3.9, 3.10, 3.11, 3.12, 3.13, 3.14, 3.15 and 3.16 plot the finite sample distribution and asymptotic distribution of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ for $\nu = -5, -2, -1, -0.5, 0.5, 1, 2, 5$ under $T = 100, 250, 500, 1000$, $\rho = 1$, $\mu = \beta = 0$, and $\{\lambda_1^c, \lambda_2^c\} = \{1/4, 3/4\}$ or $\{1/3, 2/3\}$. For all figures, without loss of generality, I assume $\delta_1 = 1$ and is fixed. In these Figures, the pdfs of λ_{TS} and λ_{MS} are plotted in the same panel to compare the performance in presence of under-specified break numbers. We use kernel smoothing to obtain the pdf based on the simulations [3] .

Figures 3.9 to 3.16 show the evolution of the distributions along the increase of ν. For $\nu < -1$ and $\nu > 1$, the break point estimator tends to be closer to λ_2^c; For $|\nu| < 1$, the break point estimator will be closer to λ_1^c. With the increase of T, we see that $\hat{\lambda}_{TS}$ converges to some points which are not the true ones. Therefore, from the point of the consistency to the true break points, we prefer $\hat{\lambda}_{MS}$. For $\nu > 0$ as in Figure 3.9 to 3.12, $\hat{\lambda}_{TS}$ would converge to point between λ_1^c and λ_2^c, while for $\nu < 0$ in Figure 3.13 to 3.16, $\hat{\lambda}_{TS}$ would converge to point in $[0, \lambda_1^c]$ or $[\lambda_2^c, 1]$.

[3] In this chapter, I use the standard normal distribution as the kernel function. For the same reason as in PZ(2005), that is, the optimal data dependant bandwidth may not work well, I choose a simple bandwidth equals to 0.5σ for any error, where σ is the STD of the data sequence. Simulations show that h does not affect the pdf estimator much.

For either of the two estimators, whether the pdf has one peak or two peaks and the location of the peak(s) depend on several factors: the space between the breaks, the relative magnitude, the signs of two breaks, and the persistence of the errors as well. For example, when break magnitudes are the same, $\hat{\lambda}_{TS}$ tends to have one peak in pdf if the distance in between is small, i.e. $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$. While for $\{\lambda_1^c, \lambda_2^c\} = \{1/4, 3/4\}$, $\hat{\lambda}_{TS}$ has two peaks in pdf. The effect of persistence on the shapes of density seems non-monotonic.

To compare the performance of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ quantitatively, we use the sum of pdf values at the true break points as a criterion to compare the precision. We consider 3 cases: $\nu = 1$, $\nu = -1$, and $|\nu| \neq 1$. Table 3.1, 3.2, and 3.3 list the sum of pdf values at the true break point. We can see that $\hat{\lambda}_{TS}$ stands at an advantage when the break magnitudes are small, disregarding the sign of the breaks and how strong the persistence in the errors is. One exception is when we are in stationary case and the difference between the magnitudes and the magnitudes themselves are small. In that case, we may prefer $\hat{\lambda}_{MS}$. This can be explained clearly by both finite sample pdf and the theoretical limiting distributions. When the break magnitude is small, the null hypothesis plays the major role in the distribution. $\hat{\lambda}_{TS}$ has high pdf around the true break, and $\hat{\lambda}_{MS}$ has most mass around 0 and 1. Therefore when break magnitude is small, we prefer $\hat{\lambda}_{TS}$. However, persistence may not be such a dominant factor as break magnitudes. With the increase of break magnitude, this condition changes. The pdf of $\hat{\lambda}_{MS}$ gradually gets denser at the true breaks, while that of $\hat{\lambda}_{TS}$ gradually becomes smaller. When the break magnitude becomes big enough, $\hat{\lambda}_{MS}$ has much higher pdf at the true breaks. And eventually the inconsistency problem for $\hat{\lambda}_{TS}$ drives it to converge to the points other than the true break points. Then even in a stationary case, we still want to "over-difference" the trend shift model and get a higher density at the true values.

In Figure 3.9 to 3.16, the asymptotic distributions are plotted together with the finite sample distributions via simulations. We can see that the asymptotic distribution approximates well to the finite sample distributions. When the pdf is calculated, the bandwidth

Table 3.1. Sum of densities at the true break λ_1^c and λ_2^c where $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$ under different ρ and $M_1 = M_2$.

	$\rho = 0$	$\rho = 0$	$\rho = 0.5$	$\rho = 0.5$	$\rho = 1$	$\rho = 1$
$M_1 = M_2$	$\hat{\lambda}_{TS}$	$\hat{\lambda}_{MS}$	$\hat{\lambda}_{TS}$	$\hat{\lambda}_{MS}$	$\hat{\lambda}_{TS}$	$\hat{\lambda}_{MS}$
10	0.11	0.71	1.00	1.01	1.42	0.87
20	0.00	0.97	0.71	2.51	1.50	1.27
50	0.00	6.02	0.00	5.13	1.52	3.42
100	0.00	8.17	0.00	6.71	2.23	6.11
150	0.00	11.31	0.00	9.92	2.76	8.97

Table 3.2. Sum of densities at the true break λ_1^c and λ_2^c where $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$ under different ρ and $M_1 = -M_2$.

	$\rho = 0$	$\rho = 0$	$\rho = 0.5$	$\rho = 0.5$	$\rho = 1$	$\rho = 1$
$M_1 = -M_2$	$\hat{\lambda}_{TS}$	$\hat{\lambda}_{MS}$	$\hat{\lambda}_{TS}$	$\hat{\lambda}_{MS}$	$\hat{\lambda}_{TS}$	$\hat{\lambda}_{MS}$
10	1.89	0.07	1.73	0.12	1.71	0.86
20	1.31	0.48	1.47	0.81	1.51	1.03
50	0.17	2.11	0.97	2.51	1.77	2.01
100	0.02	6.31	0.00	5.97	1.51	4.13
150	0.00	8.12	0.00	7.46	1.43	6.47

used in the kernel smooth plays an important role. To provide a comparison under the same base, we should choose the same bandwidth in the pdf calculation. For different T, under the same M_1 and M_2, our results also show that the approximation of Theorem 3.5.6 to the finite sample distribution is pretty good. The data will be provided upon request.

Figure 3.17 and Figure 3.18 plot the λ to achieve the maximal value of $|G2_{TS}(\lambda, \lambda_1^c) + \nu \cdot G2_{TS}(\lambda, \lambda_2^c)|$ along ν at $(\lambda_1^c, \lambda_2^c) = (1/3, 2/3)$ and $(1/4, 3/4)$, which shows where the

Table 3.3. Sum of densities at the true break λ_1^c and λ_2^c where $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$ under different ρ and $|M_1| \neq |M_2|$, where $M_1 = 50 (\delta_1 = 5)$.

	$\rho = 0$	$\rho = 0$	$\rho = 0.5$	$\rho = 0.5$	$\rho = 1$	$\rho = 1$
M_2	$\hat{\lambda}_{TS}$	$\hat{\lambda}_{MS}$	$\hat{\lambda}_{TS}$	$\hat{\lambda}_{MS}$	$\hat{\lambda}_{TS}$	$\hat{\lambda}_{MS}$
10	4.97	19.16	6.31	13.17	4.27	4.53
20	2.35	20.37	4.12	14.76	3.97	4.46
40	0.00	5.76	0.00	5.07	1.51	3.36
100	0.00	0.00	0.00	0.00	0.31	0.96
150	0.00	0.00	0.00	0.00	0.00	0.00

limits of $\hat{\lambda}_{TS}$ would be when $T \to \infty$. When $\nu = 0$, the λ to achieve maximum will be λ_1^c, one of the true break points. Other than that, all the limits would not be at the true break point. With the increase of ν from -10 to 10, there are two kinks in the limit at $|\nu| = 1$. And when $|\nu|$ goes to ∞, the limit of the break point estimator will be the true break λ_2^c. 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 point is between the two true break points. When $\nu > 0$, the limiting point will be between the true break points. And when $\nu = 1$, the limiting point is at $\lambda = 0.5$. This can be extended to be a general pattern.

3.6 Application to Sequential Tests of Multiple Breaks Model

The previous analysis shows that $\hat{\lambda}_{MS}$ can deal with the inconsistency of $\hat{\lambda}_{TS}$ in the presence of under-specification of break numbers, which means using $\hat{\lambda}_{MS}$ might be better than $\hat{\lambda}_{TS}$ in some applications with unknown break number. BP(1998) and BP(2003) proposed the sequential process to test mean shift hypothesis and locate the break points step by step, where the consistency of the break point estimator is critical. PY(2008) develop a test for an unknown break point in a univariate trend break model where the noise component can be either stationary or integrated. A bias-corrected estimate of the serial correlation parameter is used and a super efficient estimate of ρ is applied to choose the test for $I(0)$ or $I(1)$ errors. KP(2010) applied the sequential test in BP(2003) to a multiple trend break model and made it robust to I(0) and I(1) errors using the method in PY(2008).

A general model is defined in these papers as follows.

$$y_t = x_t'\Psi + u_t \tag{3.6.18}$$

$$u_t = \rho u_{t-1} + \varepsilon_t. \tag{3.6.19}$$

For a linear trend shift model, $x_t \doteq (1, t, DT_t)'$, $\Psi \doteq (\mu, \beta, \delta)'$. Testing hypothesis H_0 : $R\Psi = \gamma$, PY(2008)'s test is defined as

$$W_{FS} = (R\hat{\Psi} - \gamma)'[s^2 R(X'X)^{-1}R](R\hat{\Psi} - \gamma),$$

where $R = (0, 0, 1)$, $\gamma = 0$, $X = (x_1, (1-\hat{\rho})x_2, \cdots, (1-\hat{\rho})x_T)'$,$s^2 = T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t^2$, and $\hat{\varepsilon}_t$ are the residuals associated with the feasible GLS regression.

KP(2010) described the sequential way as follows: first, we obtain the estimates of the break dates $T_1, \cdots, T_l$ as global minimizers of the sum of squared residuals from the model with l breaks estimated by OLS:

$$(\hat{T}_1, \cdots, \hat{T}_l) = \arg \min_{(T_1, \cdots, T_l)} SSR(T_1, \cdots, T_l).$$

This can be achieved using the dynamic programming algorithm proposed by BP(2003). Second, we test for the presence of an additional break in each of the $(l+1)$ segments partitioned by $(\hat{T}_1, \cdots, \hat{T}_l)$. The test statistics test the null hypothesis of, say, l breaks, versus the alternative hypothesis of $(l+1)$ breaks is proposed. In practice, l starts from 0. Break tests in each segment follows PY(2009) in details.

The inconsistency problem of the trend break point estimator with under-specified break number is not considered by KP(2010) and other related literatures. Theorem 3.5.6 proves the inconsistency of $\hat{\lambda}_{TS}$, which could hurt the sequential test. Hence instead of using $\hat{\lambda}_{TS}$ as in KP(2010) that does not take into account the inconsistency problem, I use $\hat{\lambda}_{MS}$ as break point estimates to show how solving the inconsistency problem can improve the power of the test. To do so, we should make sure this does not change the null behavior of the KP(2010) test. KP(2010)'s result are based on the assumption that the break point estimators are consistent to the true breaks. Hence the null distribution of KP(2010) test does not depend on the break point estimators, i.e. consistency of the break point estimators ensures the break point estimators do not asymptotically show up in the null distribution. From this point, we can use $\hat{\lambda}_{MS}$ instead of $\hat{\lambda}_{TS}$ to solve the inconsistency problem.

I consider the similar variety of data generating processes as in KP(2010), especially the setting of errors. m denotes the selected break number.

Simulation DGP can be described as follows for a two-break case ($m = 2$),

$$y_t = \delta_1 DT_t(\lambda_1^c) + \delta_2 DT_t(\lambda_2^c) + u_t, \tag{3.6.20}$$

$$u_t = \rho u_{t-1} + e_t + \theta e_{t-1}. \tag{3.6.21}$$

In Table 3.4 and 3.5, we set $\delta_1 = 1$, $\lambda_1^c = 1/3$, and $\lambda_2^c = 2/3$, $\theta = 0$ and 0.5 respectively. The true break number $m = 2$. The results here show that, for big δ or when $\rho = 1, 0.9, 0.8$, $\hat{\lambda}_{MS}$ is always preferred in that the power is raised by introducing $\hat{\lambda}_{MS}$ and that the possibility to over-estimate ($\hat{m} \geq 3$) is lowered using $\hat{\lambda}_{MS}$. When $\rho = 0.5$ and break magnitude is small, the improvement from using $\hat{\lambda}_{MS}$ is not as much as the other cases. This is because when ρ and δ are both small, there is a comprehensive effect from the over-difference and consistency problem. The effect from over-difference outweights the value of consistency. The power comparison between choosing $\hat{\lambda}_{MS}$ and $\hat{\lambda}_{TS}$ at $T = 240$ (which will be provided upon request) is similar to when $T = 120$. The asymptotic result predicts the finite sample distribution well, that is, the finite sample performance is only a matter of $M_1, M_2, \lambda_1^c, \lambda_2^c$, and does not depend on T much. When $\theta = 0.5$, the power improve from $\hat{\lambda}_{MS}$ is less than $\theta = 0$.

Table 3.4. Probability of Break Number Selection $\hat{m}$ for Trend Shift Model with 2 breaks: $\{\lambda_1^c, \lambda_2^c\} = \{1/2, 2/3\}, \delta_1 = 1, \theta = 0, T = 120$.

		$\hat{m}$ using $\hat{\lambda}_{TS}$				$\hat{m}$ using $\hat{\lambda}_{MS}$			
ρ	δ_2	0.00	1.00	2.00	$\geq$ 3.00	0.00	1.00	2.00	$\geq$ 3.00
1.00	0.50	0.00	0.58	0.30	0.12	0.00	0.63	0.31	0.06
	0.60	0.00	0.43	0.43	0.14	0.00	0.47	0.48	0.05
	0.70	0.00	0.33	0.52	0.15	0.00	0.38	0.54	0.08
	0.80	0.00	0.23	0.60	0.17	0.00	0.27	0.64	0.09
	0.90	0.00	0.16	0.67	0.17	0.00	0.18	0.70	0.12
	1.00	0.00	0.10	0.72	0.18	0.00	0.16	0.78	0.06
0.90	0.50	0.00	0.63	0.30	0.07	0.00	0.66	0.33	0.01
	0.60	0.00	0.44	0.46	0.10	0.00	0.49	0.48	0.03
	0.70	0.00	0.27	0.58	0.15	0.00	0.28	0.60	0.12
	0.80	0.00	0.16	0.71	0.13	0.00	0.18	0.76	0.06
	0.90	0.00	0.07	0.79	0.14	0.00	0.10	0.81	0.09
	1.00	0.00	0.05	0.80	0.15	0.00	0.09	0.83	0.08
0.80	0.50	0.00	0.58	0.37	0.05	0.00	0.61	0.38	0.01
	0.60	0.00	0.40	0.52	0.08	0.00	0.42	0.53	0.05
	0.70	0.00	0.23	0.65	0.12	0.00	0.26	0.68	0.06
	0.80	0.00	0.10	0.79	0.11	0.00	0.13	0.82	0.05
	0.90	0.00	0.05	0.84	0.11	0.00	0.08	0.88	0.04
	1.00	0.00	0.02	0.86	0.12	0.00	0.03	0.86	0.11
0.50	0.50	0.00	0.03	0.88	0.09	0.00	0.06	0.87	0.07
	0.60	0.00	0.02	0.87	0.11	0.00	0.03	0.85	0.12
	0.70	0.00	0.01	0.88	0.11	0.00	0.03	0.86	0.11
	0.80	0.00	0.01	0.90	0.09	0.00	0.02	0.88	0.10
	0.90	0.00	0.00	0.90	0.10	0.00	0.01	0.91	0.08
	1.00	0.00	0.00	0.90	0.10	0.00	0.00	0.93	0.07

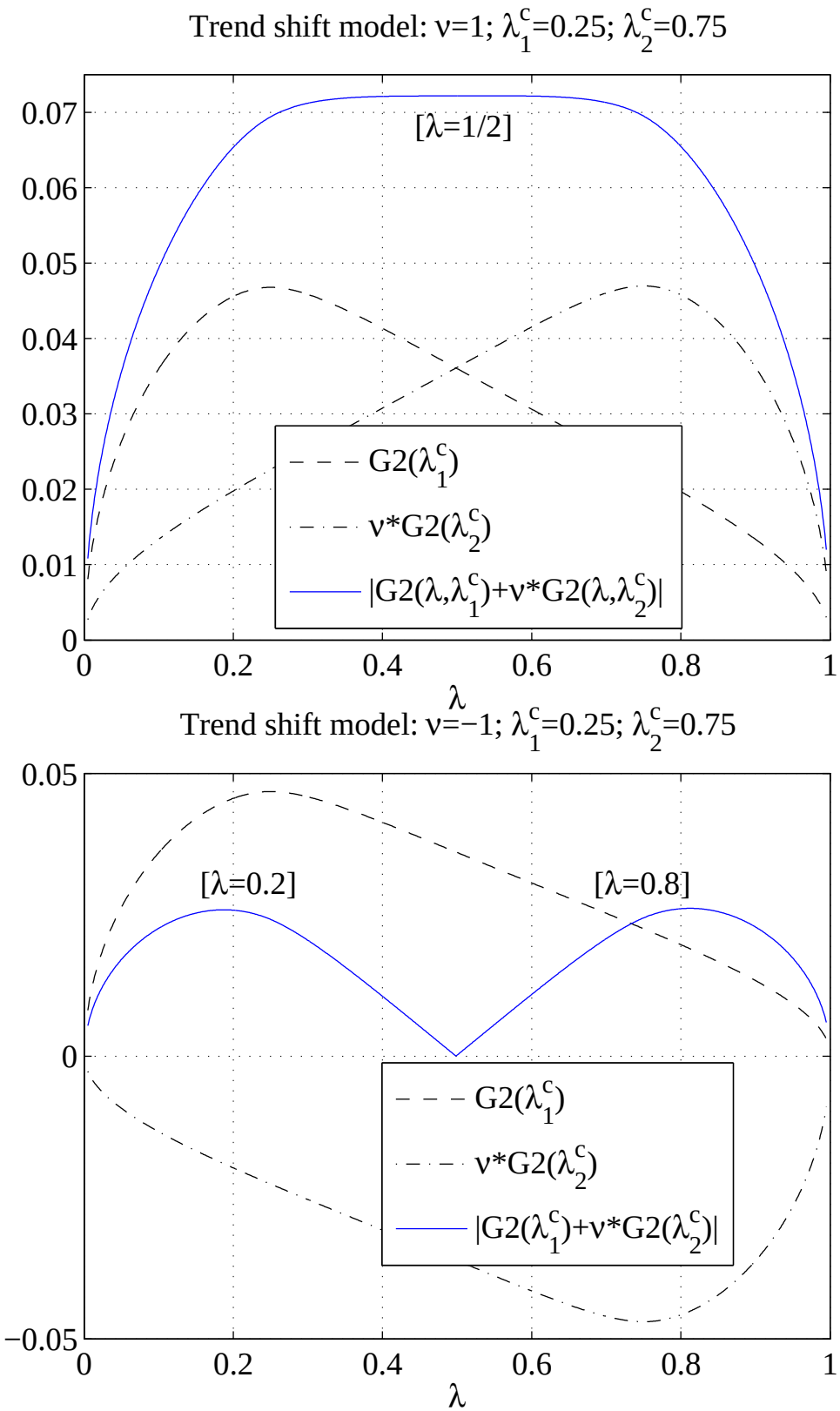

Figure 3.8. $|G2_{TS}(\lambda, \lambda_1^c) + \nu \cdot G2_{TS}(\lambda, \lambda_2^c)|$ under $\nu = 1$ and -1 for trend shift model, where $\{\lambda_1^c, \lambda_2^c\} = \{1/4, 3/4\}$.

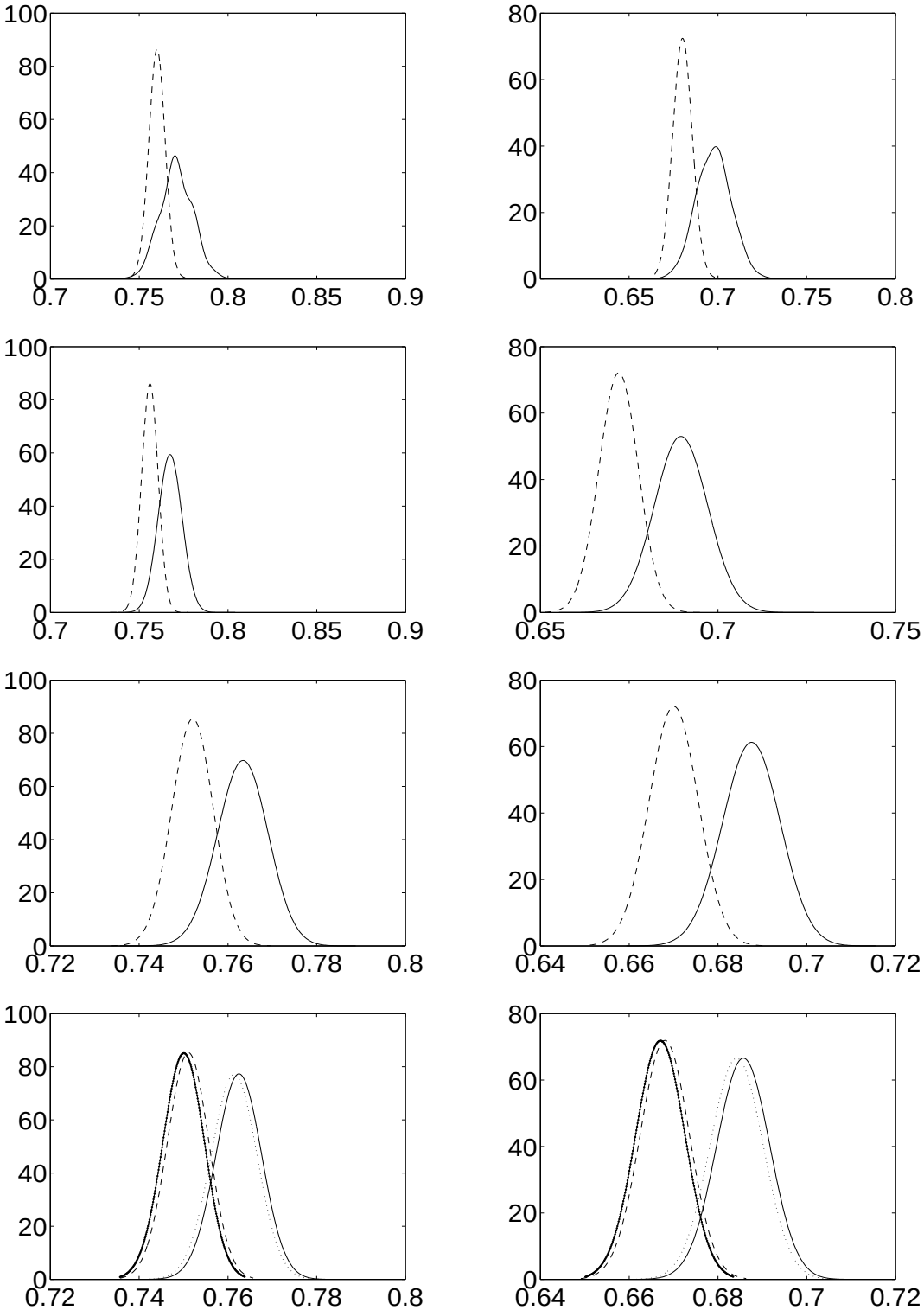

Figure 3.9. Finite sample distribution with the asymptotic distribution of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ at $\nu = -5$. The left to right: $\{\lambda_1^c, \lambda_2^c\} = \{1/4, 3/4\}, \{1/3, 2/3\}$; the top to bottom: $T = 100, 250, 500, 1000$. $\rho = 1$. Solid: finite sample $\hat{\lambda}_{TS}$; dash: finite sample $\hat{\lambda}_{MS}$; dot: asymptotic $\hat{\lambda}_{TS}$; dot-solid: asymptotic $\hat{\lambda}_{MS}$.

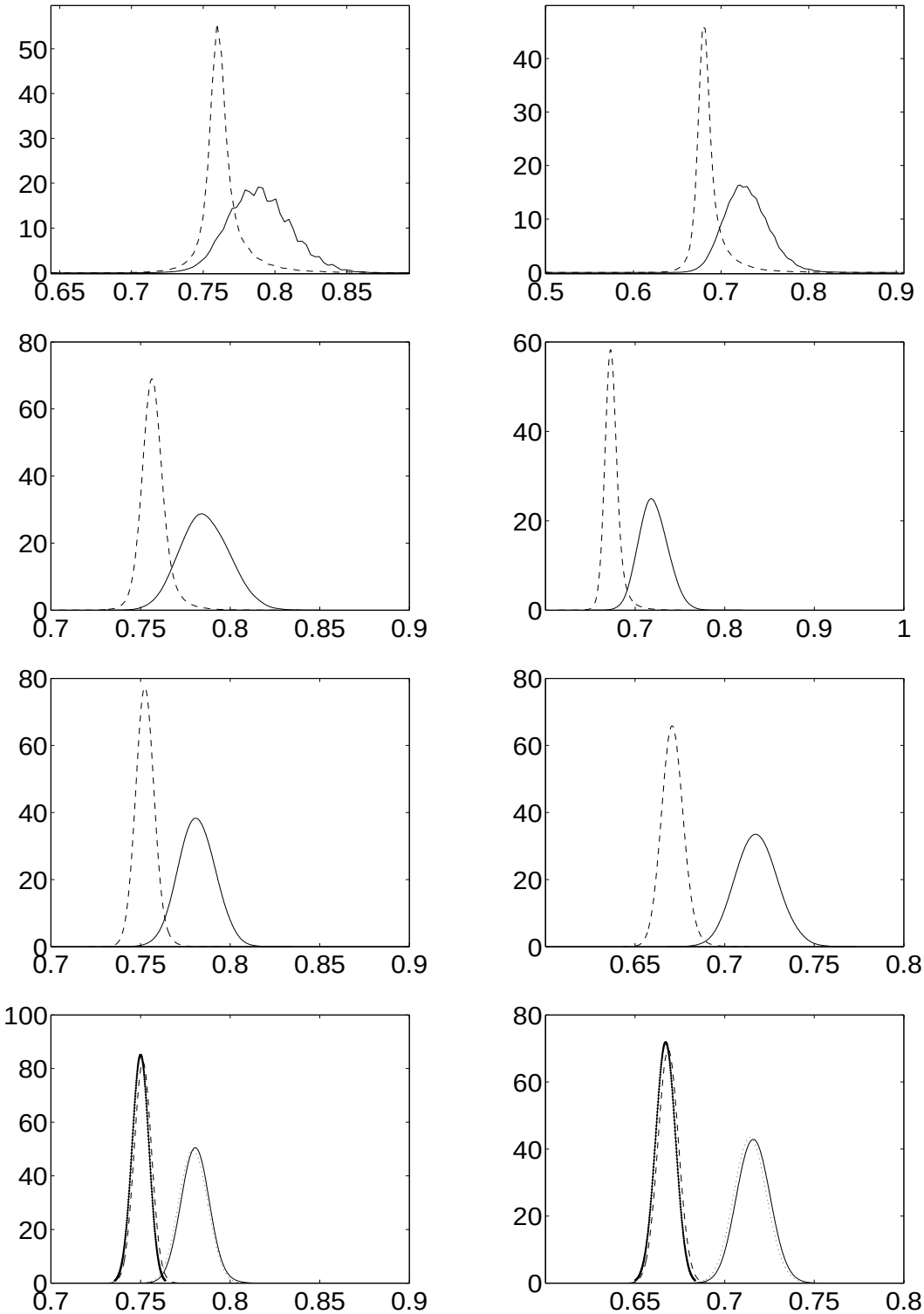

Figure 3.10. Finite sample distribution with the asymptotic distribution of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ at $\nu = -2$. The left: $\{\lambda_1^c, \lambda_2^c\} = \{1/4, 3/4\}$; the right: $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$; the top to bottom: $T = 100, 250, 500, 1000$. $\rho = 1$. Solid: finite sample $\hat{\lambda}_{TS}$; dash: finite sample $\hat{\lambda}_{MS}$; dot: asymptotic $\hat{\lambda}_{TS}$; dot-solid: asymptotic $\hat{\lambda}_{MS}$.

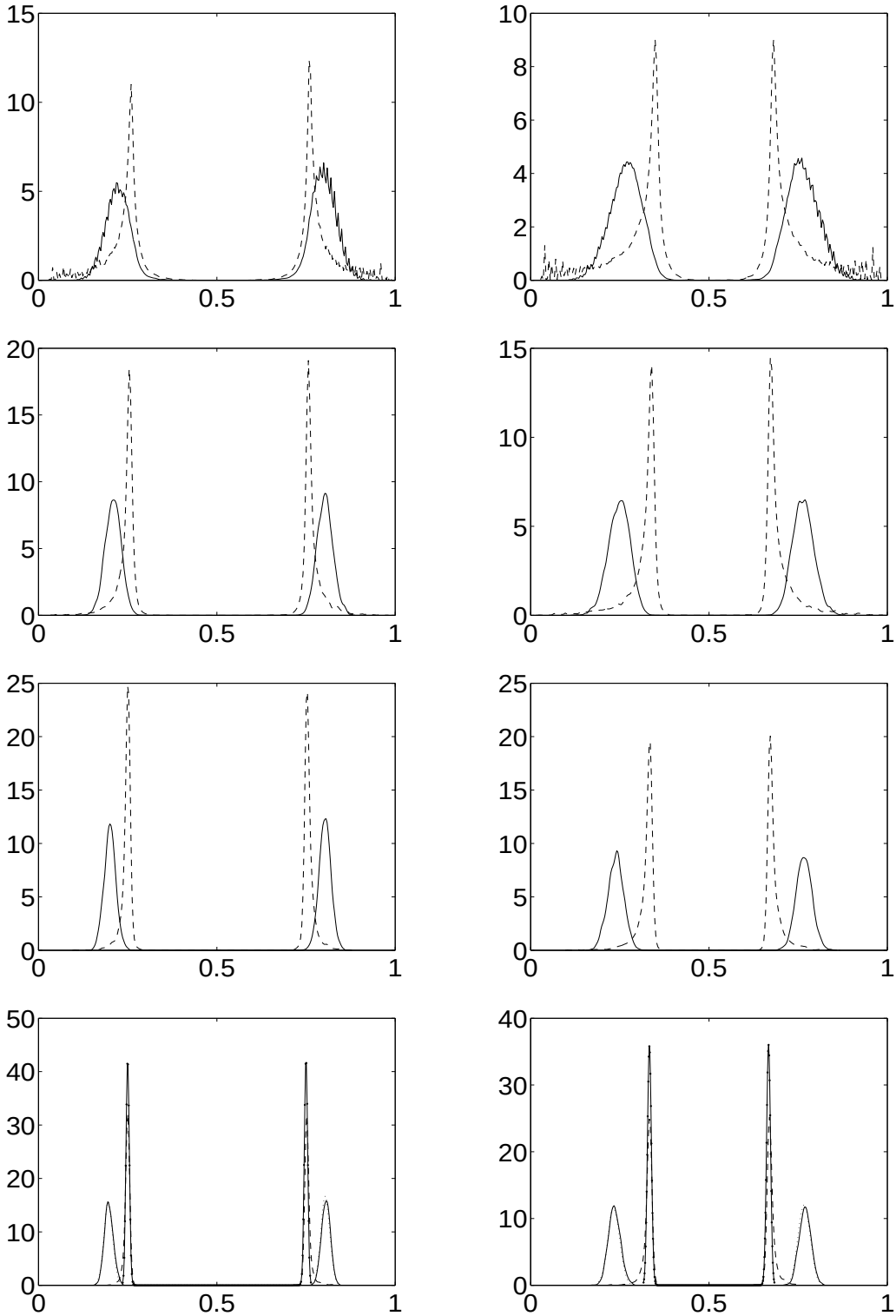

Figure 3.11. Finite sample distribution with the asymptotic distribution of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ at $\nu = -1$. The left: $\{\lambda_1^c, \lambda_2^c\} = \{1/4, 3/4\}$; the right: $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$; the top to bottom: $T = 100, 250, 500, 1000$. $\rho = 1$. Solid: finite sample $\hat{\lambda}_{TS}$; dash: finite sample $\hat{\lambda}_{MS}$; dot: asymptotic $\hat{\lambda}_{TS}$; dot-solid: asymptotic $\hat{\lambda}_{MS}$.

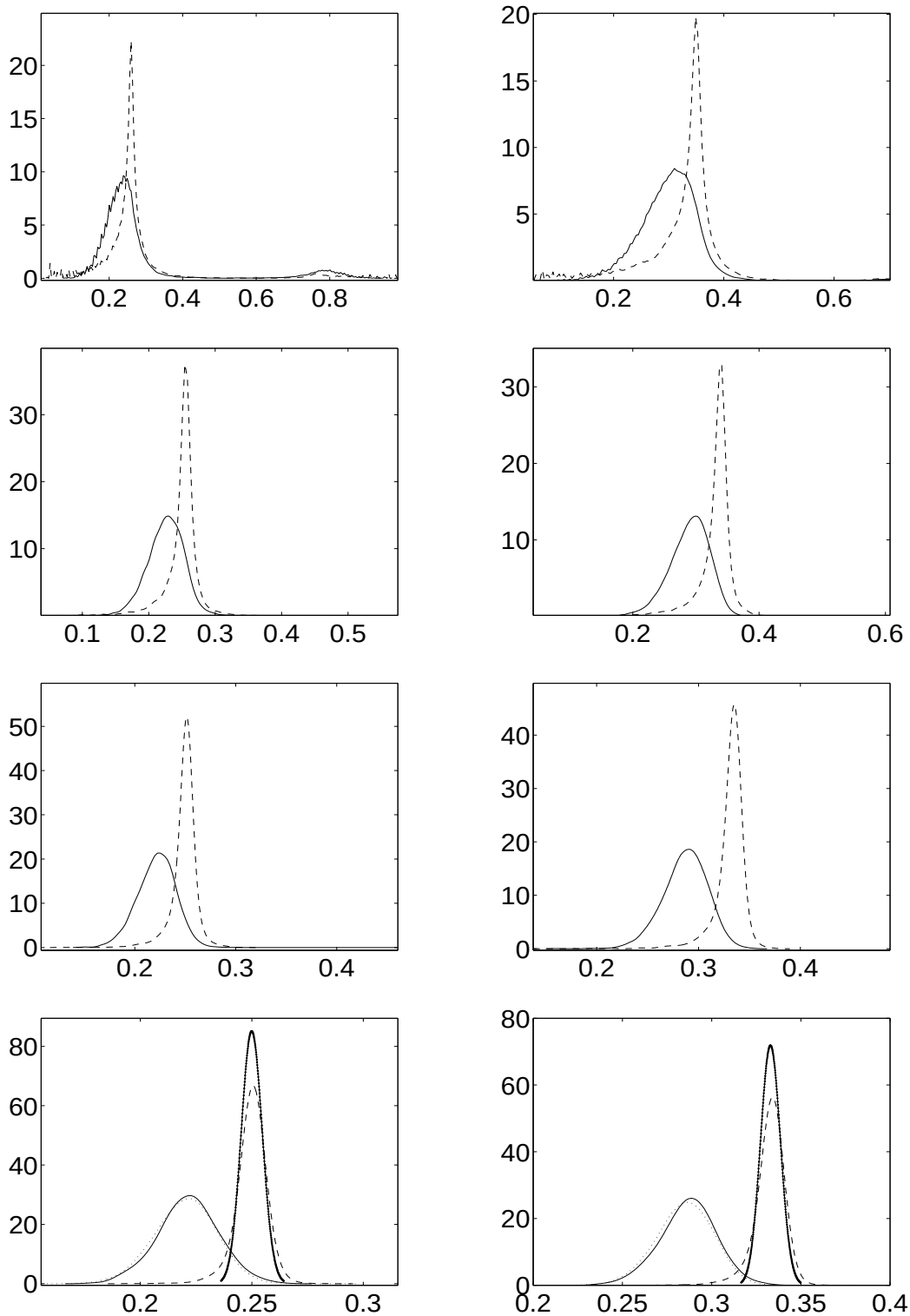

Figure 3.12. Finite sample distribution with the asymptotic distribution of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ at $\nu = -0.5$. The left: $\{\lambda_1^c, \lambda_2^c\} = \{1/4, 3/4\}$; the right: $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$; the top to bottom: $T = 100, 250, 500, 1000$. $\rho = 1$. Solid: finite sample $\hat{\lambda}_{TS}$; dash: finite sample $\hat{\lambda}_{MS}$; dot: asymptotic $\hat{\lambda}_{TS}$; dot-solid: asymptotic $\hat{\lambda}_{MS}$.

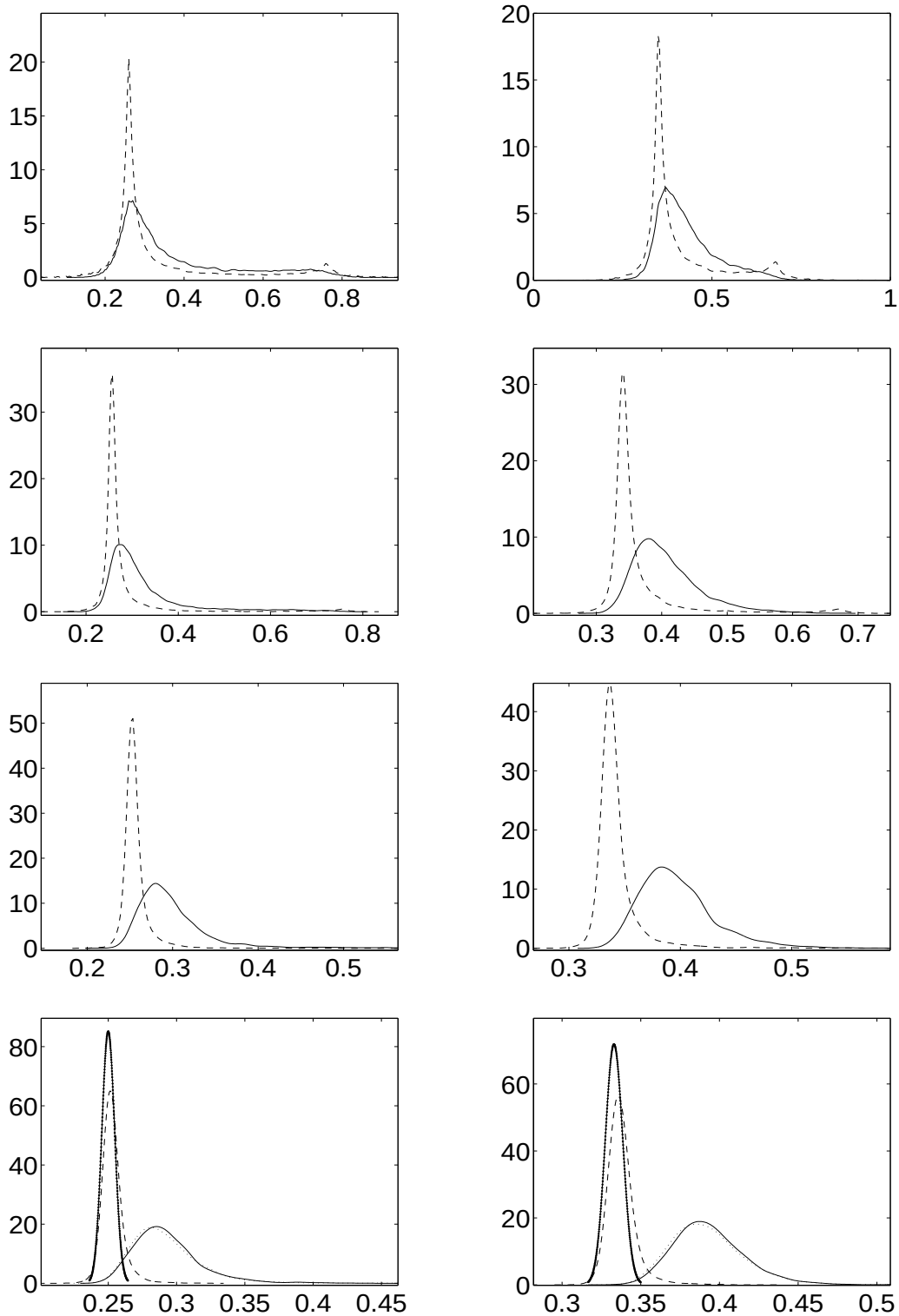

Figure 3.13. Finite sample distribution with the asymptotic distribution of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ at $\nu = 0.5$. The left: $\{\lambda_1^c, \lambda_2^c\} = \{1/4, 3/4\}$; the right: $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$; the top to bottom: $T = 100, 250, 500, 1000$. $\rho = 1$. Solid: finite sample $\hat{\lambda}_{TS}$; dash: finite sample $\hat{\lambda}_{MS}$; dot: asymptotic $\hat{\lambda}_{TS}$; dot-solid: asymptotic $\hat{\lambda}_{MS}$.

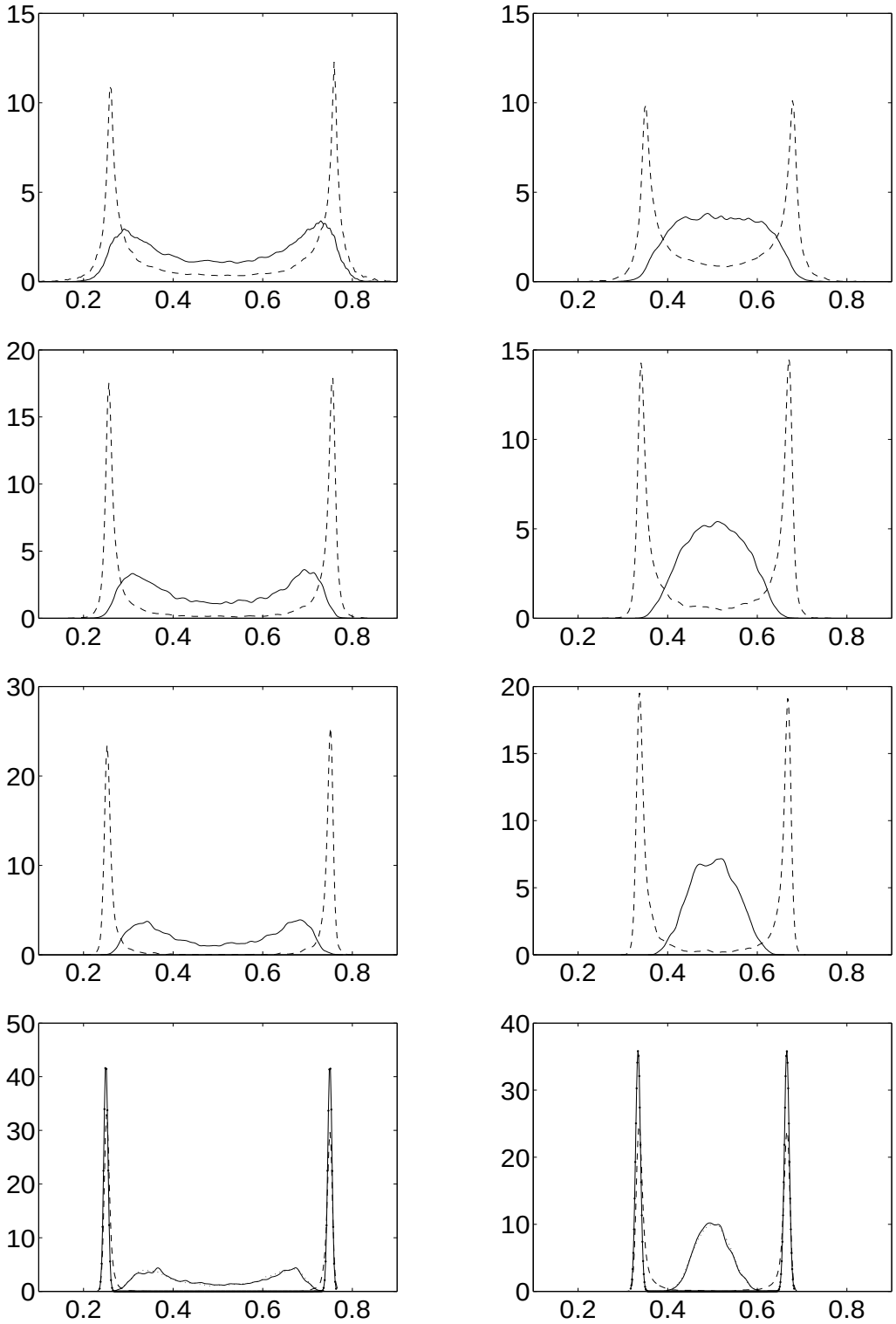

Figure 3.14. Finite sample distribution with the asymptotic distribution of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ at $\nu = 1$. The left: $\{\lambda_1^c, \lambda_2^c\} = \{1/4, 3/4\}$; the right: $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$; the top to bottom: $T = 100, 250, 500, 1000$. $\rho = 1$. Solid: finite sample $\hat{\lambda}_{TS}$; dash: finite sample $\hat{\lambda}_{MS}$; dot: asymptotic $\hat{\lambda}_{TS}$; dot-solid: asymptotic $\hat{\lambda}_{MS}$.

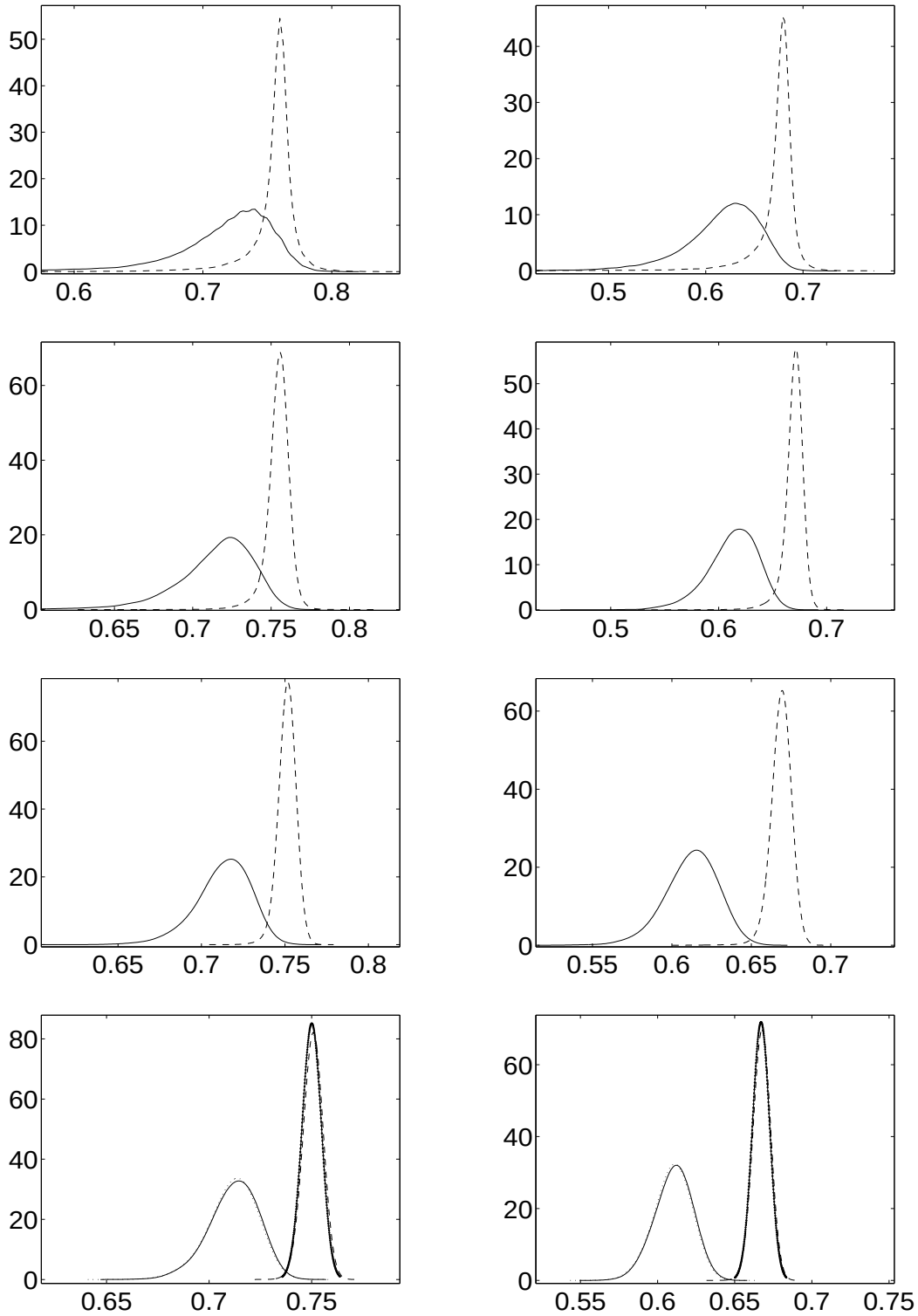

Figure 3.15. Finite sample distribution with the asymptotic distribution of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ at $\nu = 2$. The left: $\{\lambda_1^c, \lambda_2^c\} = \{1/4, 3/4\}$; the right: $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$; the top to bottom: $T = 100, 250, 500, 1000$. $\rho = 1$. Solid: finite sample $\hat{\lambda}_{TS}$; dash: finite sample $\hat{\lambda}_{MS}$; dot: asymptotic $\hat{\lambda}_{TS}$; dot-solid: asymptotic $\hat{\lambda}_{MS}$.

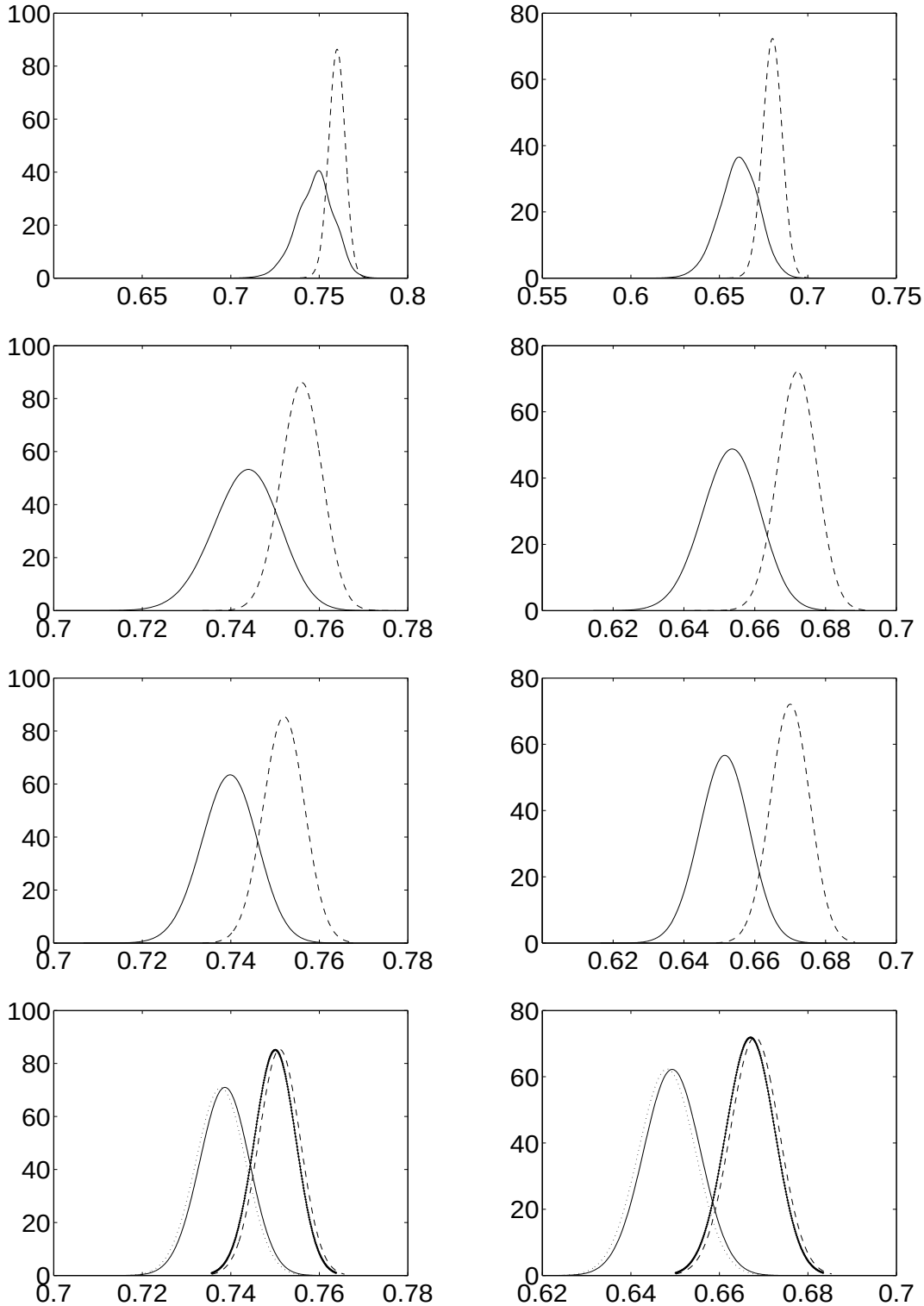

Figure 3.16. Finite sample distribution with the asymptotic distribution of $\hat{\lambda}_{TS}$ and $\hat{\lambda}_{MS}$ at $\nu = 5$. The left: $\{\lambda_1^c, \lambda_2^c\} = \{1/4, 3/4\}$; the right: $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$; the top to bottom: $T = 100, 250, 500, 1000$. $\rho = 1$. Solid: finite sample $\hat{\lambda}_{TS}$; dash: finite sample $\hat{\lambda}_{MS}$; dot: asymptotic $\hat{\lambda}_{TS}$; dot-solid: asymptotic $\hat{\lambda}_{MS}$.

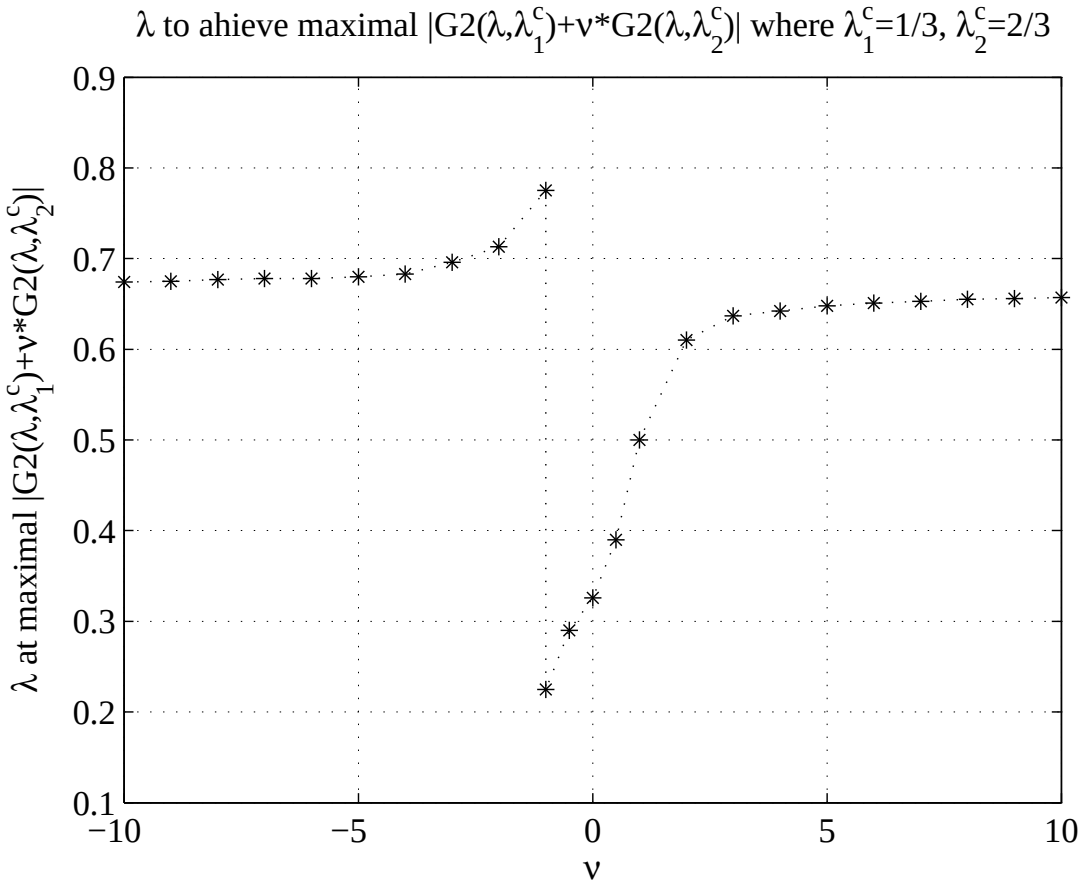

Figure 3.17. λ to achieve maximal $G2_{TS}(\lambda, \lambda_1^c)+\nu \cdot G2_{TS}(\lambda, \lambda_2^c)$, $\{\lambda_1^c, \lambda_2^c\} = \{1/3, 2/3\}$, $\nu = -10, \cdots, 10$.

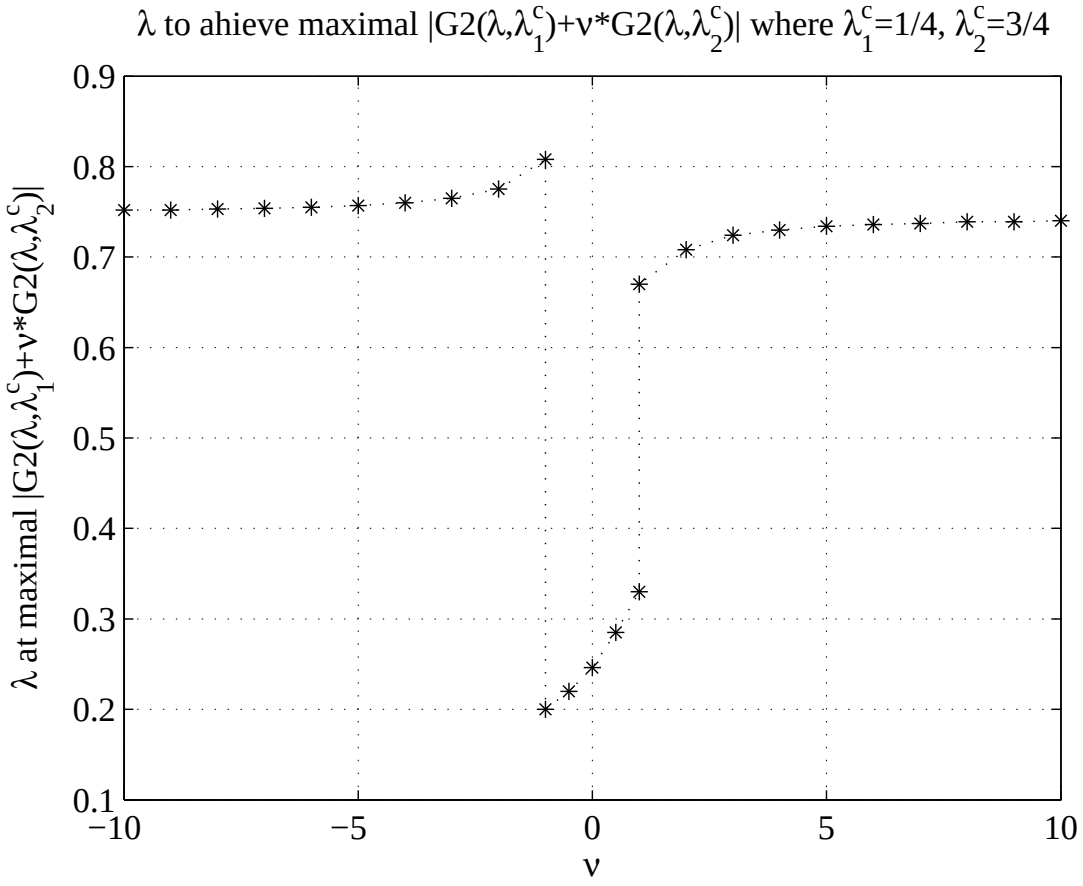

Figure 3.18. λ to achieve maximal $G2_{TS}(\lambda, \lambda_1^c)+\nu\cdot G2_{TS}(\lambda, \lambda_2^c)$, $\{\lambda_1^c, \lambda_2^c\} = \{1/4, 3/4\}$, $\nu = -10, \cdots, 10$.

Table 3.5. Probability of Break Number Selection $\hat{m}$ for Trend Shift Model with 2 breaks: $\{\lambda_1^c, \lambda_2^c\} = \{1/2, 2/3\}, \delta_1 = 1, \theta = 0.5, T = 120$.

		$\hat{m}$ using $\hat{\lambda}_{TS}$				$\hat{m}$ using $\hat{\lambda}_{MS}$			
ρ	δ_2	0.00	1.00	2.00	$\geq$3.00	0.00	1.00	2.00	$\geq$3.00
1.00	0.50	0.00	0.59	0.24	0.17	0.00	0.61	0.26	0.13
	0.60	0.00	0.50	0.30	0.20	0.00	0.50	0.32	0.18
	0.70	0.00	0.43	0.35	0.22	0.00	0.45	0.35	0.20
	0.80	0.00	0.35	0.43	0.22	0.00	0.33	0.42	0.25
	0.90	0.00	0.28	0.47	0.25	0.00	0.27	0.48	0.25
	1.00	0.00	0.25	0.50	0.25	0.00	0.26	0.51	0.23
0.90	0.50	0.00	0.61	0.28	0.11	0.00	0.62	0.27	0.11
	0.60	0.00	0.51	0.35	0.14	0.00	0.52	0.36	0.12
	0.70	0.00	0.39	0.45	0.16	0.00	0.39	0.47	0.14
	0.80	0.00	0.29	0.53	0.18	0.00	0.31	0.56	0.13
	0.90	0.00	0.19	0.62	0.19	0.00	0.20	0.66	0.12
	1.00	0.00	0.14	0.69	0.17	0.00	0.16	0.73	0.11
0.80	0.50	0.00	0.55	0.36	0.09	0.00	0.54	0.38	0.08
	0.60	0.00	0.45	0.45	0.10	0.00	0.44	0.47	0.09
	0.70	0.00	0.26	0.62	0.12	0.00	0.27	0.65	0.08
	0.80	0.00	0.18	0.69	0.13	0.00	0.19	0.72	0.09
	0.90	0.00	0.09	0.78	0.13	0.00	0.06	0.81	0.13
	1.00	0.00	0.06	0.80	0.14	0.00	0.04	0.84	0.12
0.50	0.50	0.00	0.55	0.36	0.09	0.00	0.53	0.36	0.12
	0.60	0.00	0.45	0.45	0.10	0.00	0.42	0.42	0.16
	0.70	0.00	0.26	0.62	0.12	0.00	0.25	0.60	0.15
	0.80	0.00	0.18	0.69	0.13	0.00	0.18	0.71	0.11
	0.90	0.00	0.09	0.78	0.13	0.00	0.10	0.82	0.08
	1.00	0.00	0.06	0.80	0.14	0.00	0.07	0.83	0.10

APPENDICES

A.1 Extension of the asymptotics in Theorem 1.5.1 to near-I(1) errors

The results in Theorem 1.5.1 are easily extended from pure I(1) errors to near-I(1) errors. Assume

$$(A1.b) \qquad u_t \;=\; \rho u_{t-1} + \varepsilon_t, \text{ where } t = 2, \cdots, T, 0$$

where $\rho = 1 - \frac{c}{T}$, $c \geq 0$ is a constant scalar. The asymptotics of the break point estimator is derived under near-I(1) errors to show that the limiting distributions depend on c.

Corollary 0.1.1 *Suppose the regressions in the level model (1.2.1) and its first difference (1.2.2) are estimated by using $\lambda \in \Lambda \subseteq (0,1)$ and $T_b^c \doteq \lambda^c T$ is the true break. Under the assumption $(A1.b)$ and $(A2.c)$, the break point estimators by minimizing the $SSR(\lambda)$ have the limiting distributions as follows.*

1. For the level model (1.2.1),

$$\hat{\lambda}_{TS} \Rightarrow \arg\max_{\lambda\in\Lambda}\{\frac{[\int_0^1 F(r,\lambda)V_c(r)dr + M\int_0^1 F(r,\lambda)F(r,\lambda^c)dr]^2}{\int_0^1 F(r,\lambda)^2 dr}\}$$

where $V_c(r) \doteq \int_0^1 \exp(-c(r-s))dW(s)$, $M \doteq \frac{\delta^*}{d(1)}$ and $F(r,\lambda)$ is defined in Theorem 1.5.1.

2. For the first difference (1.2.2),

$$\hat{\lambda}_{MS} \Rightarrow$$
$$\arg\max_{\lambda\in\Lambda}\{\frac{[(\lambda W(1) - W(\lambda) - c\int_0^1 (1(r>\lambda) - (1-\lambda))V_c(r)dr) + M(\Psi(\lambda,\lambda^c)]^2}{\lambda(1-\lambda)}\}$$

where $M \doteq \frac{\delta^*}{d(1)}$ and $\Psi(\lambda,\lambda^c)$ is defined in Theorem 1.5.1.

A.2 Proof of Theorem 1.5.1

A.2.1 Proof of part 1 in Theorem 1.5.1

The break point estimator $\hat{\lambda}_{TS}$ is obtained by minimizing the $SSR_{TS}(\lambda)$ (See (1.2.3)). Because SSR^0 does not depend on λ, we can equivalently define $\hat{\lambda}_{TS}$ as

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

where SSR^0_{TS} denotes the SSR under the assumption of no beaks. Using the Frisch and Waugh (1933) Theorem,

$$\hat{\delta} = [\sum_{t=1}^{T} \tilde{DT}_t(\lambda)\tilde{DT}_t(\lambda)]^{-1}\sum_{t=1}^{T}\tilde{DT}_t(\lambda)\tilde{y}_t, \tag{0.2.22}$$

where $\{\tilde{DT}_t(\lambda)\}$ and $\{\tilde{y}_t\}$ are the residuals from the OLS regressions of $\{DT_t(\lambda)\}$ and $\{y_t\}$ on $[1\ \ t]'$. There is a standard result (See Sayginsoy and Vogelsang (2010)) that

$$SSR^0_{TS} - SSR_{TS}(\lambda) = [\sum_{t=1}^{T}\tilde{DT}_t(\lambda)\tilde{DT}_t(\lambda)]\hat{\delta}^2. \tag{0.2.23}$$

Consider $T^{-1}\tilde{DT}_t(\lambda)$. Simple algebra gives $T^{-1}\tilde{DT}_t(\lambda)$

$$\begin{aligned}
&= T^{-1}DT_t(\lambda) - \sum_{t=1}^{T} T^{-1}DT_t(\lambda)[1\ \ t]\begin{bmatrix}1 & 0\\ 0 & T^{-1}\end{bmatrix}\times \\
&\quad \left\{\sum_{t=1}^{T}\begin{bmatrix}1 & 0\\ 0 & T^{-1}\end{bmatrix}\begin{bmatrix}1\\ t\end{bmatrix}\begin{bmatrix}1 & t\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & T^{-1}\end{bmatrix}\right\}^{-1}\begin{bmatrix}1 & 0\\ 0 & T^{-1}\end{bmatrix}\begin{bmatrix}1\\ t\end{bmatrix} \\
&\Rightarrow\ (r-\lambda)1(r>\lambda) - \\
&\quad \int_0^1 (r-\lambda)1(r>\lambda)[1\ \ r]dr[\int_0^1\begin{bmatrix}1\\ r\end{bmatrix}\begin{bmatrix}1 & r\end{bmatrix}dr]^{-1}\begin{bmatrix}1\\ r\end{bmatrix} \\
&= (r-\lambda)1(r>\lambda) - [\int_0^1 (r-\lambda)1(r>\lambda)dr\ \ \int_0^1 (r-\lambda)1(r>\lambda)rdr] \\
&\quad \begin{bmatrix}4 & -6\\ -6 & 12\end{bmatrix}\begin{bmatrix}1\\ r\end{bmatrix}.
\end{aligned}$$

Because

$$\begin{aligned}\int_0^1 (r-\lambda)1(r>\lambda)dr &= \frac{\lambda^2}{2}-\lambda+\frac{1}{2},\\ \int_0^1 (r-\lambda)1(r>\lambda)rdr &= \frac{\lambda^3}{6}-\frac{\lambda}{2}+\frac{1}{3},\end{aligned}$$

we have

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

For simplicity, we define

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

Because $\{u_t\}$ is I(1),

$$T^{-1/2}u_{[rT]} \Rightarrow d(1)W(r),$$

where $W(r)$ is the standard Wiener process. Well known results give

$$T^{-3/2}\sum_{t=1}^{T} T^{-1}\tilde{DT}_t(\lambda)u_t \Rightarrow d(1)\int_0^1 \tilde{F}(r,\lambda)W(r)dr.$$

Scaling (0.2.23) by T^{-2} gives

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

From the previous results, it follows that

$$\begin{aligned}
T^{1/2}\hat{\delta} &= [T^{-1}\sum_{t=1}^{T} T^{-1}\tilde{DT}_t(\lambda)\tilde{DT}_t(\lambda)T^{-1}]^{-1} \\
&\quad [T^{-1}\sum_{t=1}^{T} T^{-1}\tilde{DT}_t(\lambda)\tilde{DT}_t(\lambda^c)T^{-1}(T^{1/2}\delta)] + \\
&\quad [T^{-1}\sum_{t=1}^{T} T^{-1}\tilde{DT}_t(\lambda)\tilde{DT}_t(\lambda)T^{-1}]^{-1}[T^{-3/2}\sum_{t=1}^{T} T^{-1}\tilde{DT}_t(\lambda)u_t] \\
\Rightarrow\ & [\int_0^1 F(r,\lambda)^2dr]^{-1}[\delta^*\int_0^1 F(r,\lambda)F(r,\lambda^c)dr] \\
&\quad +[\int_0^1 F(r,\lambda)^2dr]^{-1}[d(1)\int_0^1 F(r,\lambda)W(r)dr] \\
=\ & \frac{d(1)\int_0^1 F(r,\lambda)W(r)dr + \delta^*\int_0^1 F(r,\lambda)F(r,\lambda^c)dr}{\int_0^1 F(r,\lambda)^2dr},
\end{aligned}$$

and

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

which gives

$$T^{-2}[SSR^0_{TS} - SSR_{TS}(\lambda)] \ \Rightarrow\ \frac{[d(1)\int_0^1 F(r,\lambda)W(r)dr + \delta^*\int_0^1 F(r,\lambda)F(r,\lambda^c)dr]^2}{\int_0^1 F(r,\lambda)^2dr}.$$

Furthermore using the continuous mapping theorem (CMT), we obtain the limit of the break point estimator as

$$\begin{aligned}
\hat{\lambda}_{TS} &= \arg\max_{\lambda\in\Lambda}\{SSR^0_{TS} - SSR_{TS}(\lambda)) \\
&= \arg\max_{\lambda\in\Lambda}\{T^{-2}[SSR^0_{TS} - SSR_{TS}(\lambda)]) \\
&\Rightarrow \arg\max_{\lambda\in\Lambda}\{\frac{[d(1)\int_0^1 F(r,\lambda)W(r)dr + \delta^*\int_0^1 F(r,\lambda)F(r,\lambda^c)dr]^2}{\int_0^1 F(r,\lambda)^2dr}\} \\
&= \arg\max_{\lambda\in\Lambda}\{\frac{[\int_0^1 F(r,\lambda)W(r)dr + M\int_0^1 F(r,\lambda)F(r,\lambda^c)dr]^2}{\int_0^1 F(r,\lambda)^2dr}\},
\end{aligned}$$

where $M \doteq \frac{\delta^*}{d(1)}$.

A.2.2 Proof of part 2 in Theorem 1.5.1

Using similar arguments as the level model, we have

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

Under the assumptions of Model (1.2.2), the OLS estimate of δ is given by

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

where

$$\begin{aligned}
\tilde{DU}_t(\lambda) &= DU_t(\lambda) - \sum_{t=1}^{T} DU_t/T = DU_t(\lambda) - \bar{DU}(\lambda), \\
\tilde{\Delta y}_t &= \Delta y_t - \sum_{t=1}^{T} \Delta y_t/T = \Delta y_t - \bar{\Delta y}.
\end{aligned}$$

Simple algebra gives

$$\begin{aligned}\hat{\delta} &= [\sum_{t=1}^{T} \tilde{DU}_t^2(\lambda)]^{-1} \sum_{t=1}^{T} \tilde{DU}_t(\lambda)[\tilde{DU}_t(\lambda^c)\delta + \Delta u_t] \\ &= [\sum_{t=1}^{T} \tilde{DU}_t^2(\lambda)]^{-1} \sum_{t=1}^{T} \tilde{DU}_t(\lambda)\tilde{DU}_t(\lambda^c)\delta + [\sum_{t=1}^{T} \tilde{DU}_t^2(\lambda)]^{-1} \sum_{t=1}^{T} \tilde{DU}_t(\lambda)\Delta u_t.\end{aligned}$$

Because $\Delta u_t = \varepsilon_t$,

$$\begin{aligned}T^{1/2}\hat{\delta} &= [T^{-1}\sum_{t=1}^{T} \tilde{DU}_t^2(\lambda)]^{-1}[T^{-1}\sum_{t=1}^{T} \tilde{DU}_t(\lambda)\tilde{DU}_t(\lambda^c)\delta^*] + \\ &\quad [T^{-1}\sum_{t=1}^{T} \tilde{DU}_t^2(\lambda)]^{-1}[T^{-1/2}\sum_{t=1}^{T} \tilde{DU}_t(\lambda)\varepsilon_t];\end{aligned}$$

also because

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

and

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

and

$$\begin{aligned}T^{-1/2}[\sum_{t=1}^{T} \tilde{DU}_t(\lambda)\varepsilon_t] &\Rightarrow d(1)\int_0^1 [I(r > \lambda) - (1-\lambda)]dW(r) \\ &= d(1)[\lambda W(1) - W(\lambda)];\end{aligned}$$

we obtain

$$T^{1/2}\hat{\delta} \Rightarrow \frac{\delta^*}{\lambda(1-\lambda)}\Phi(\lambda, \lambda^c) + \frac{d(1)}{\lambda(1-\lambda)}[\lambda W(1) - W(\lambda)],$$

where

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

Using this result, it immediately follows that

$$\begin{aligned} SSR^0_{MS}-SSR_{MS}(\lambda) &= [\sum_{t=1}^{T}\tilde{DU}_t(\lambda)^2]\hat{\delta}^2 \\ &= [T^{-1}\sum_{t=1}^{T}\tilde{DU}_t(\lambda)^2][T^{1/2}\hat{\delta}]^2 \\ &\Rightarrow \frac{1}{\sqrt{\lambda(1-\lambda)}}[d(1)(\lambda W(1)-W(\lambda))+\delta^*\Psi(\lambda,\lambda^c)]^2. \end{aligned}$$

Applying the CMT theorem gives

$$\begin{aligned} \hat{\lambda}_{TS} &= \arg\max_{\lambda\in\Lambda}\{SSR^0_{MS}-SSR_{MS}(\lambda)) \\ &\Rightarrow \arg\max_{\lambda\in\Lambda}\{\frac{[d(1)(\lambda W(1)-W(\lambda))+\delta^*\Psi(\lambda,\lambda^c)]^2}{\lambda(1-\lambda)}\} \\ &= \arg\max_{\lambda\in\Lambda}\{\frac{[(\lambda W(1)-W(\lambda))+M\Psi(\lambda,\lambda^c)]^2}{\lambda(1-\lambda)}\} \end{aligned}$$

where $M=\frac{\delta^*}{d(1)}$.

A.3 Proof that $\arg\max_\lambda G2(\lambda,\lambda^c)=\lambda^c$

a) First I derive $G2_{TS}(\lambda,\lambda^c)$, which is a function of λ and λ^c. Then I prove that it always achieves the global maximum at $\lambda=\lambda^c$ in $[0,1]$. This result suggests that $\arg\max\{G1_{TS}(\lambda)+G2_{TS}(\lambda,\lambda^c)\}$ converges to λ^c as M increases.

Take $\lambda\leq\lambda^c$, we have

$$\begin{aligned} G2_{TS}(\lambda,\lambda^c) &= \frac{(1-\lambda)^2(1-\lambda^c)^2(\lambda+\lambda^c+2\lambda\lambda^c+2)/6}{\sqrt{(1-\lambda^3)(1-\lambda^c)^3/3}} \\ &= \frac{(1-\lambda^c)^2}{2\sqrt{3}}\frac{(\lambda+\lambda^c+2\lambda\lambda^c+2)}{\sqrt{\lambda^2+\lambda+1}}. \end{aligned}$$

Taking the derivative of $G2_{TS}$ with respect to λ gives

$$G2'_{TS}(\lambda) = \frac{\sqrt{3}(1-\lambda^c)^2}{4}\frac{(\lambda^c-\lambda)}{(\lambda^2+\lambda+1)^{3/2}}.$$

We can see that $G2'_{TS}(\lambda) \geq 0$ when $\lambda \leq \lambda^c$, which proves that the maximum value of $G2_{TS}$ is obtained at $\lambda = \lambda^c$ for $\lambda \leq \lambda^c$.

Now take $\lambda \geq \lambda^c$, giving

$$G2_{TS}(\lambda, \lambda^c) = \frac{\sqrt{3}(\lambda^c)^2}{4}\frac{[(6-3\lambda^c)+(2\lambda^c-3)\lambda]}{\sqrt{\lambda^2-3\lambda+3}}.$$

The derivative of $G2_{TS}(\lambda, \lambda^c)$ with respect to λ is

$$G2'_{TS}(\lambda) = \frac{\sqrt{3}(\lambda^c)^2}{4}\frac{(\lambda^c-\lambda)}{(\lambda^2-3\lambda+3)^{3/2}}.$$

The fact that $G2'_{TS}(\lambda) \leq 0$ shows that the maximum value of $G2_{TS}$ is obtained at $\lambda = \lambda^c$ when $\lambda \geq \lambda^c$.

b) I calculate $G2_{MS}(\lambda, \lambda^c)$ and show that it achieves the global maximum at $\lambda = \lambda^c$ in $[0, 1]$. The conclusion is similar to that in part a): $\arg\max\{G1_{MS}(\lambda) + G2_{MS}(\lambda, \lambda^c)\}$ converges to λ^c as M increases.

When $\lambda \leq \lambda^c$

$$\begin{aligned} G2_{MS}(\lambda, \lambda^c) &= \frac{(1-\lambda^c)\lambda}{\sqrt{\lambda(1-\lambda)}} \\ &= \frac{(1-\lambda^c)\sqrt{\lambda}}{\sqrt{(1-\lambda)}} \\ &\leq \sqrt{(1-\lambda^c)\lambda^c} \end{aligned}$$

hence $G2_{MS}(\lambda, \lambda^c)$ attains maximum at $\lambda = \lambda^c$ for $\lambda \leq \lambda^c$. It can be proved similarly for $\lambda \geq \lambda^c$.

Combining a) and b) we obtain that $\arg\max_\lambda G2_{TS}(\lambda, \lambda^c) = \lambda^c$ and $\arg\max_\lambda G2_{MS}(\lambda, \lambda^c) = \lambda^c$.

A.4 Proof of Corollary 0.1.1

A.4.1 Proof of part 1 in Corollary 0.1.1

Under assumption (A1.b): $u_t = \rho u_{t-1} + \varepsilon_t$, where $\rho \doteq 1 - \frac{c}{T}$,

$$T^{-3/2} \sum_{t=1}^{T} T^{-1} \tilde{DT}_t(\lambda) u_t \Rightarrow d(1) \int_0^1 F(r, \lambda) V_c(r) dr,$$

where $V_c(r) = \int_0^1 \exp(-c(r-s)) dW(s)$.

A.4.2 Proof of part 2 in Corollary 0.1.1

Because $u_t = (1 - \frac{c}{T}) u_{t-1} + \varepsilon_t$, it follows that $\Delta u_t = -\frac{c}{T} u_{t-1} + \varepsilon_t$. This gives

$$\begin{aligned}(T-1)^{-1/2} \sum_{t=2}^{T} \tilde{DU}_t(\lambda) \Delta u_t &= (T-1)^{-1/2} \sum_{t=2}^{T} \tilde{DU}_t(\lambda) \varepsilon_t \\ &\quad -(T-1)^{-1/2} (T)^{-1} c \sum_{t=2}^{T} \tilde{DU}_t(\lambda) u_{t-1},\end{aligned}$$

where

$$(T-1)^{-1/2} \sum_{t=2}^{T} \tilde{DU}_t(\lambda) \varepsilon_t \Rightarrow d(1)[\lambda W(1) - W(\lambda)],$$

and

$$(T-1)^{-1/2} (T)^{-1} c \sum_{t=2}^{T} \tilde{DU}_t(\lambda) u_{t-1} \Rightarrow d(1) c \int_0^1 (1(r > \lambda) - (1-\lambda)) V_c(r) dr.$$

The rest of the proof is straightforward and follows the proof of Theorem 1.5.1.

A.5 Proof of Theorem 1.6.2

A.5.1 Proof of part 1 in Theorem 1.6.2

The break point estimator $\hat{\lambda}_{TS}$ is obtained by minimizing the $SSR_{TS}(\lambda)$ (See (1.2.3)). Because SSR^0 does not depend on λ, we can equivalently define $\hat{\lambda}_{TS}$ as

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

where SSR^0_{TS} denotes the SSR under the assumption of no beaks. Using the Frisch and Waugh (1933) Theorem,

$$\hat{\delta} = [\sum_{t=1}^{T} \tilde{DT}_t(\lambda)\tilde{DT}_t(\lambda)]^{-1}\sum_{t=1}^{T}\tilde{DT}_t(\lambda)\tilde{y}_t, \tag{0.5.24}$$

where $\{\tilde{DT}_t(\lambda)\}$ and $\{\tilde{y}_t\}$ are the residuals from the OLS regressions of $\{DT_t(\lambda)\}$ and $\{y_t\}$ on $[1\ \ t]'$. There is a standard result (See Sayginsoy and Vogelsang (2010)) that

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

Consider $T^{-1}\tilde{DT}_t(\lambda)$, simple algebra in the proof of Theorem 1.5.1 gives

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

Because $\{u_t\}$ is I(0),

$$T^{-1/2}\sum_{t=1}^{[rT]} u_t \Rightarrow d(1)W(r),$$

where $W(r)$ is the standard Wiener process. Well known results give

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

Scaling $\hat{\delta}$ by $T^{3/2}$, equation (1.2.3) is written as

$$[SSR^0_{TS} - SSR_{TS}(\lambda)] = [T^{3/2}\hat{\delta}]^2[T^{-1}\sum_{t=1}^{T} T^{-1}\tilde{DT}_t(\lambda)\tilde{DT}_t(\lambda)T^{-1}].$$

From the previous results, it follows that

$$\begin{aligned}
T^{3/2}\hat{\delta} &= [T^{-1}\sum_{t=1}^{T} T^{-1}\tilde{DT}_t(\lambda)\tilde{DT}_t(\lambda)T^{-1}]^{-1} \\
&\quad [T^{-1}\sum_{t=1}^{T} T^{-1}\tilde{DT}_t(\lambda)\tilde{DT}_t(\lambda^c)T^{-1}(T^{3/2}\delta)] + \\
&\quad [T^{-1}\sum_{t=1}^{T} T^{-1}\tilde{DT}_t(\lambda)\tilde{DT}_t(\lambda)T^{-1}]^{-1}[T^{-1/2}\sum_{t=1}^{T} T^{-1}\tilde{DT}_t(\lambda)u_t] \\
&\Rightarrow [\int_0^1 F(r,\lambda)^2 dr]^{-1}[\delta^* \int_0^1 F(r,\lambda)F(r,\lambda^c)dr] \\
&\quad +[\int_0^1 F(r,\lambda)^2 dr]^{-1}[d(1)\int_0^1 F(r,\lambda)dW(r)] \\
&= \frac{\delta^* \int_0^1 F(r,\lambda)F(r,\lambda^c)dr + d(1)\int_0^1 F(r,\lambda)dW(r)}{\int_0^1 F(r,\lambda)^2 dr},
\end{aligned}$$

and

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

which gives

$$[SSR^0_{TS} - SSR_{TS}(\lambda)] \Rightarrow \frac{[d(1)\int_0^1 F(r,\lambda)dW(r) + \delta^* \int_0^1 F(r,\lambda)F(r,\lambda^c)dr]^2}{\int_0^1 F(r,\lambda)^2 dr}.$$

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

$$\begin{aligned}
\hat{\lambda}_{TS} &= \arg\max_{\lambda\in\Lambda}\{SSR^0_{TS} - SSR_{TS}(\lambda)) \\
&\Rightarrow \arg\max_{\lambda\in\Lambda}\{\frac{[d(1)\int_0^1 F(r,\lambda)dW(r) + \delta^* \int_0^1 F(r,\lambda)F(r,\lambda^c)dr]^2}{\int_0^1 F(r,\lambda)^2 dr}\} \\
&= \arg\max_{\lambda\in\Lambda}\{\frac{[\int_0^1 F(r,\lambda)dW(r) + M\int_0^1 F(r,\lambda)F(r,\lambda^c)dr]^2}{\int_0^1 F(r,\lambda)^2 dr}\},
\end{aligned}$$

where $M \doteq \frac{\delta^*}{d(1)} \equiv \frac{\delta T^{3/2}}{d(1)}$.

A.5.2 Proof of part 2 in Theorem 1.6.2

Because SSR^0 does not depend on λ, we can equivalently define $\hat{\lambda}_{QS}$ as

$$\hat{\lambda}_{QS} = \arg\max_{\lambda \in \Lambda} \{SSR^0_{QS} - SSR_{QS}(\lambda)\}.$$

Using the Frisch and Waugh (1933) Theorem,

$$\hat{\delta} = [\sum_{t=1}^{T} \tilde{DQ}_t(\lambda)\tilde{DQ}_t(\lambda)]^{-1} \sum_{t=1}^{T} \tilde{DQ}_t(\lambda)\tilde{S}_t,$$

where $\{\tilde{DQ}_t(\lambda)\}$ and $\{\tilde{S}_t\}$ are the residuals from the OLS regressions of $\{DQ_t(\lambda)\}$ and $\{S_t\}$ on $[t \;\; t^2]'$. There is a standard result (See Sayginsoy and Vogelsang (2010)) that

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

Consider $T^{-2}\tilde{DT}_t(\lambda)$. Simple algebra gives $T^{-2}\tilde{DQ}_t(\lambda)$

$$\begin{aligned}
&= T^{-2}DQ_t(\lambda) - \sum_{t=1}^{T} T^{-2}DQ_t(\lambda)[t \;\; t^2] \begin{bmatrix} T^{-1} & 0 \\ 0 & T^{-2} \end{bmatrix} \times \\
&\quad \left\{ \sum_{t=1}^{T} \begin{bmatrix} T^{-1} & 0 \\ 0 & T^{-2} \end{bmatrix} \begin{bmatrix} t \\ t^2 \end{bmatrix} \begin{bmatrix} t & t^2 \end{bmatrix} \begin{bmatrix} T^{-1} & 0 \\ 0 & T^{-2} \end{bmatrix} \right\}^{-1} \begin{bmatrix} T^{-1} & 0 \\ 0 & T^{-2} \end{bmatrix} \\
&\quad \begin{bmatrix} t \\ t^2 \end{bmatrix} \\
&\Rightarrow \frac{(r-\lambda)^2}{2}1(r > \lambda) - \int_0^1 \frac{(r-\lambda)^2}{2}1(r > \lambda)[r \;\; r^2]dr [\int_0^1 \begin{bmatrix} r \\ r^2 \end{bmatrix} \begin{bmatrix} r & r^2 \end{bmatrix} dr]^{-1} \\
&\quad \begin{bmatrix} r \\ r^2 \end{bmatrix} \\
&= \frac{(r-\lambda)^2}{2}1(r > \lambda) - [\int_0^1 \frac{(r-\lambda)^2}{2}1(r > \lambda)r dr \;\; \int_0^1 \frac{(r-\lambda)^2}{2}1(r > \lambda)r^2 dr] \\
&\quad \begin{bmatrix} 48 & -60 \\ -60 & 80 \end{bmatrix} \begin{bmatrix} r \\ r^2 \end{bmatrix}.
\end{aligned}$$

Because

$$\begin{aligned}\int_0^1 \frac{(r-\lambda)^2}{2}1(r>\lambda)rdr &= \frac{1}{8}-\frac{\lambda}{3}+\frac{\lambda^2}{4}-\frac{\lambda^4}{24},\\ \int_0^1 \frac{(r-\lambda)^2}{2}1(r>\lambda)r^2dr &= \frac{1}{10}-\frac{\lambda}{4}+\frac{\lambda^2}{6}-\frac{\lambda^5}{60},\end{aligned}$$

we have

$$\begin{aligned}&T^{-2}\tilde{DQ}_t(\lambda)\\ &\Rightarrow\\ &\frac{(r-\lambda)^2}{2}1(r>\lambda)-(-\lambda+2\lambda^2-2\lambda^4+\lambda^5)r-(\frac{1}{2}-\frac{5\lambda^2}{3}+\frac{5\lambda^4}{2}-\frac{4\lambda^5}{3})r^2.\end{aligned}$$

For simplicity, we define

$$\begin{aligned}&Q(r,\lambda)\\ \doteq\ &\frac{(r-\lambda)^2}{2}1(r>\lambda)-(-\lambda+2\lambda^2-2\lambda^4+\lambda^5)r-(\frac{1}{2}-\frac{5\lambda^2}{3}+\frac{5\lambda^4}{2}-\frac{4\lambda^5}{3})r^2.\end{aligned}$$

Because $\{u_t\}$ is I(1),

$$T^{-1/2}u_{[rT]} \Rightarrow d(1)W(r),$$

where $W(r)$ is the standard Wiener process. Well known results give

$$T^{-3/2}\sum_{t=1}^{T}T^{-1}\tilde{DQ}_t(\lambda)u_t \Rightarrow d(1)\int_0^1 \tilde{Q}(r,\lambda)W(r)dr.$$

Scaling (0.5.25) by T^{-2} gives

$$T^{-2}[SSR^0_{QS}-SSR_{QS}(\lambda)] = [T^{3/2}\hat{\delta}]^2[T^{-1}\sum_{t=1}^{T}T^{-2}\tilde{DQ}_t(\lambda)\tilde{DQ}_t(\lambda)T^{-2}].$$

The rest part of the proof is straight forward and follows the proof of Theorem 1.5.1 part 1.

B.1 Proofs and Additional Results of Chapter 2

Proof of Theorem 2.5.3. The result for the numerator of $LM(m,T_b)$ follows directly from Sayginsoy and Vogelsang (2010). All that is needed to complete the proof is the fixed-b

limit of $\tilde{\sigma}^2(m)$. Because the fixed-b algebra for $\tilde{\sigma}^2(m)$ is the same as the algebra used by Hashimzade and Vogelsang (2008), once we derive the limit of the partial sums of $\tilde{u}_t$ the fixed-b limits follow directly using arguments in Kiefer and Vogelsang (2005), Hashimzade and Vogelsang (2008) and Sayginsoy and Vogelsang (2010). Define

$$\tilde{S}_{[rT]} = \sum_{t=1}^{[rT]} \tilde{u}_t.$$

where,

$$\tilde{u}_t = y_t - \overline{y} = \delta(DU_t(T_b^0) - \overline{DU}(T_b^0)) + u_t - \overline{u},$$

giving

$$\tilde{S}_{[rT]} = \delta \sum_{t=1}^{[rT]} (DU_t(T_b^0) - \overline{DU}(T_b^0)) + \sum_{t=1}^{[rT]} (u_t - \overline{u}).$$

For $I(0)$ errors recall that under H_A we have $\delta = T^{-1/2}\delta_0$ and it follows that

$$\begin{aligned} T^{-1/2}\tilde{S}_{[rT]} &= \delta_0 T^{-1} \sum_{t=1}^{[rT]} (DU_t(T_b^0) - \overline{DU}(T_b^0)) + T^{-1/2} \sum_{t=1}^{[rT]} (u_t - \overline{u}) \\ &\Rightarrow \delta_0[(r-\lambda_0)\mathbf{1}(r>\lambda_0) - r(1-\lambda_0)] + \sigma[W(r) - rW(1)] \\ &= \sigma\left(\frac{\delta_0}{\sigma}[(r-\lambda_0)\mathbf{1}(r>\lambda_0) - r(1-\lambda_0)] + W(r) - rW(1)\right) \equiv \sigma Q_0(r). \end{aligned}$$

For $I(1)$ errors $\delta = T^{-1/2}\delta_0$ giving

$$\begin{aligned} &T^{-3/2}\tilde{S}_{[rT]} \\ &= \delta_0 T^{-1} \sum_{t=1}^{[rT]} (DU_t(T_b^0) - \overline{DU}(T_b^0)) + T^{-3/2} \sum_{t=1}^{[rT]} (u_t - \overline{u}) \\ &\Rightarrow \delta_0[(r-\lambda_0)\mathbf{1}(r>\lambda_0) - r(1-\lambda_0)] + d(1)[\int_0^r V_c(s)ds - r\int_0^1 V_c(s)ds] \\ &= d(1)\left(\frac{\delta_0}{d(1)}[(r-\lambda_0)\mathbf{1}(r>\lambda_0) - r(1-\lambda_0)] + \int_0^r V_c(s)ds - r\int_0^1 V_c(s)ds\right) \\ &\equiv d(1)Q_1(r). \end{aligned}$$

C.1 Proof of Theorem 3.4.4

1) Derive the asymptotic distribution with underspecified break number

We have

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

Under the assumptions of Model (3.2.1), the OLS estimate of δ 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

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

When the break number is under estimated, 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

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

Because

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

and

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

and

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

we obtain

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

where

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

From this result, it immediately follows that

$$\begin{aligned} & RSS_{MS}^0 - RSS_{MS}(\lambda) \\ = \; & [T^{-1}\sum_{t=1}^{T}\tilde{DU}_t(\lambda)^2][T^{1/2}\hat{\delta}_{MS}]^2 \\ \Rightarrow \; & \frac{1}{\sqrt{\lambda(1-\lambda)}}[d(1)(\lambda W(1)-W(\lambda))+\delta_1^*\Psi(\lambda,\lambda_1^c)+\delta_2^*\Psi(\lambda,\lambda_2^c)]^2. \end{aligned}$$

Applying the CMT theorem gives

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

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

Let's further take a look at the $M_1G2(\lambda,\lambda_1^c)+M_2G2(\lambda,\lambda_2^c)$. First take the first derivative of $G2_{MS}$ w.r.t. λ.

$$G2'_{MS}(\lambda,\lambda^c)=\frac{(1-\lambda^c)\lambda}{2(1-\lambda)\sqrt{\lambda(1-\lambda)}}, \text{ when } \lambda\le\lambda^c$$

and

$$G2'_{MS}(\lambda,\lambda^c)=\frac{(1-\lambda)\lambda^c}{2(1-\lambda)\sqrt{\lambda(1-\lambda)}}, \text{ when } \lambda\ge\lambda^c$$

Assume $\lambda_1^c<\lambda_2^c$,

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

$$(M_1 G2(\lambda, \lambda_1^c) + M_2 G2(\lambda, \lambda_2^c))' = M_1 \frac{(1-\lambda)\lambda_1^c}{2(1-\lambda)\sqrt{\lambda(1-\lambda)}} + M_2 \frac{(1-\lambda_2^c)\lambda}{2(1-\lambda)\sqrt{\lambda(1-\lambda)}},$$
$$\text{when } \lambda_1^c \leq \lambda \leq \lambda_2^c;$$

and

$$(M_1 G2(\lambda, \lambda_1^c) + M_2 G2(\lambda, \lambda_2^c))' = M_1 \frac{(1-\lambda)\lambda_1^c}{2(1-\lambda)\sqrt{\lambda(1-\lambda)}} + M_2 \frac{(1-\lambda)\lambda_2^c}{2(1-\lambda)\sqrt{\lambda(1-\lambda)}},$$
$$\text{when } \lambda \geq \lambda_2^c.$$

Through simple algebras, we can show that the peak values will be obtained at either λ_1^c or λ_2^c.

C.2 Proof of Theorem 3.4.5

1) asymptotic distribution of $\hat{\lambda^c}$

During underspecification of the break number, the only difference with the case of correct break number estimation in the form of the $RSS_0 - RSS_1(\lambda)$ is $\hat{\delta}_{TS}$. We can also get the standard result that

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

Consider $T^{-1}\tilde{DT}_t(\lambda)$, simple algebra as in Yang (2010) gives

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

Because $\{u_t\}$ is I(0),

$$T^{-1/2}\sum_{t=1}^{[rT]} u_t \Rightarrow d(1)W(r),$$

where $W(r)$ is the standard Wiener process. Well known results give

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

Now we estimate $\hat{\delta}_{TS}$. Scaling $\hat{\delta}_{TS}$ by $T^{3/2}$, equation (3.2.6) is written as

$$[SSR^0_{TS} - SSR_{TS}(\lambda)] = [T^{3/2}\hat{\delta}_{TS}]^2[T^{-1}\sum_{t=1}^{T} T^{-1}\tilde{DT}_t(\lambda)\tilde{DT}_t(\lambda)T^{-1}].$$

From the previous results, it follows that

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

Define the matrix X_0 and $DT(\lambda^c)$ be stacked $[1\ \ t]$ and $DT_t(\lambda^c)$ from $t = 1, \cdots, T$. If there are two breaks at λ^c_1 and λ^c_2 in the model (3.2.3)

$$\begin{aligned}
\tilde{Y} &= Y - X_0(X_0'X_0)^{-1}X_0'Y \\
&= (X_0[\alpha\ \ \beta]' + \delta_1 DT(\lambda^c_1) + \delta_2 DT(\lambda^c_2) + U) - \\
&\quad X_0(X_0'X_0)^{-1}X_0'(X_0[\alpha\ \ \beta]' + \delta_1 DT(\lambda^c_1) + \delta_2 DT(\lambda^c_2) + U) \\
&= \delta_1(I - X_0(X_0'X_0)^{-1}X_0')DT(\lambda^c_1) + \delta_2(I - X_0(X_0'X_0)^{-1}X_0')DT(\lambda^c_2) + U \\
&= \delta_1\tilde{DT}_t(\lambda^c_1) + \delta_2\tilde{DT}_t(\lambda^c_2) + U
\end{aligned}$$

Under I(0) errors I define the break magnitude within a $T^{-3/2}$ neighborhood of 0 as in Assumption (C2.b). Next I derive the asymptotic distribution of the $\hat{\lambda}_{TS}$ with underspecified break number.

From the previous results, it follows that

$$
\begin{aligned}
& T^{3/2}\hat{\delta}_{TS} \\
= \; & [T^{-1}\sum_{t=1}^{T} T^{-1}\tilde{DT}_t(\lambda)\tilde{DT}_t(\lambda)T^{-1}]^{-1}[T^{-1}\sum_{t=1}^{T} T^{-1}\tilde{DT}_t(\lambda)\tilde{DT}_t(\lambda_1^c)T^{-1}(T^{3/2}\delta_1) \\
& + T^{-1}\sum_{t=1}^{T} T^{-1}\tilde{DT}_t(\lambda)\tilde{DT}_t(\lambda_2^c)T^{-1}(T^{3/2}\delta_2)] + \\
& [T^{-1}\sum_{t=1}^{T} T^{-1}\tilde{DT}_t(\lambda)\tilde{DT}_t(\lambda)T^{-1}]^{-1}[T^{-1/2}\sum_{t=1}^{T} T^{-1}\tilde{DT}_t(\lambda)u_t] \\
\Rightarrow \; & [\int_0^1 F(r,\lambda)^2 dr]^{-1}[\delta_1^* \int_0^1 F(r,\lambda)F(r,\lambda_1^c)dr + \delta_2^* \int_0^1 F(r,\lambda)F(r,\lambda_2^c)dr] + \\
& [\int_0^1 F(r,\lambda)^2 dr]^{-1}[d(1)\int_0^1 F(r,\lambda)dW(r)] \\
= \; & \frac{[\delta_1^* \int_0^1 F(r,\lambda)F(r,\lambda_1^c)dr + \delta_2^* \int_0^1 F(r,\lambda)F(r,\lambda_2^c)dr] + d(1)\int_0^1 F(r,\lambda)dW(r)}{\int_0^1 F(r,\lambda)^2 dr},
\end{aligned}
$$

and

$$
T^{-1}\sum_{t=1}^{T} T^{-1}\tilde{DT}_t(\lambda)\tilde{DT}_t(\lambda)T^{-1} \;\Rightarrow\; \int_0^1 F(r,\lambda)^2 dr,
$$

which gives

$$
[SSR_{TS}^0 - SSR_{TS}(\lambda)] \Rightarrow
$$
$$
\frac{[d(1)\int_0^1 F(r,\lambda)dW(r) + \delta_1^* \int_0^1 F(r,\lambda)F(r,\lambda_1^c)dr + \delta_2^* \int_0^1 F(r,\lambda)F(r,\lambda_2^c)dr]^2}{\int_0^1 F(r,\lambda)^2 dr}.
$$

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

$$
\begin{aligned}
& \hat{\lambda}_{TS} \\
= \; & \arg\max_{\lambda\in\Lambda}\{SSR_{TS}^0 - SSR_{TS}(\lambda)\} \\
= \; & \arg\max_{\lambda\in\Lambda}\{[\frac{\int_0^1 F(r,\lambda)dW(r)}{\sqrt{\int_0^1 F(r,\lambda)^2 dr}} + \\
& \frac{M_1 \int_0^1 F(r,\lambda)F(r,\lambda_1^c)dr + M_2 \int_0^1 F(r,\lambda)F(r,\lambda_2^c)dr}{\sqrt{\int_0^1 F(r,\lambda)^2 dr}}]^2\},
\end{aligned}
$$

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

C.3 Analysis of $G2_{TS}$ under two breaks

In this section, we analyze the $G2_{TS}$ to show how the break point estimation would be in the presence of under-specification of the break number.

First $G2_{TS}(\lambda, \lambda^c)$ is a function of λ and λ^c. Then I prove that in the presence of two break points λ_1^c and λ_2^c it would not always achieve the global maximum at any true break point λ_1^c or λ_2^c in $[0, 1]$. This result illustrate the inconsistency problem of the break point estimator for the trend shift model.

When M_1 and M_2 are big, the break point estimator $\hat{\lambda}_{TS}$ is dominated by the properties of $M_1 G2(\lambda, \lambda_1^c) + M_2 G2(\lambda, \lambda_2^c)$. Yang (2010) shows that the maximal value of $G2(\lambda, \lambda^c)$ is achieved at $\lambda = \lambda^c$, hence when there is only 1 break, $\hat{\lambda}_{TS}$ is consistent to the true break point. However, when the true break points are two and the break number is estimated as one, $M_1 G2(\lambda, \lambda_1^c) + M_2 G2(\lambda, \lambda_2^c)$ achieves the maximal values not at λ_1^c or λ_2^c, which causes the inconsistent break point estimator.

$$\begin{aligned} G2_{TS}(\lambda, \lambda^c) &= \frac{(1-\lambda)^2(1-\lambda^c)^2(\lambda + \lambda^c + 2\lambda\lambda^c + 2)/6}{\sqrt{(1-\lambda^3)(1-\lambda^c)^3/3}} \\ &= \frac{(1-\lambda^c)^2(\lambda + \lambda^c + 2\lambda\lambda^c + 2)/6}{\sqrt{(\lambda^2+\lambda+1)/3}} \\ &= \frac{(1-\lambda^c)^2}{2\sqrt{3}} \frac{(\lambda + \lambda^c + 2\lambda\lambda^c + 2)}{\sqrt{\lambda^2+\lambda+1}}. \end{aligned}$$

Taking the derivative of $G2_{TS}$ with respect to λ gives

$$G2'_{TS}(\lambda) = \frac{\sqrt{3}(1-\lambda^c)^2}{4} \frac{(\lambda^c - \lambda)}{(\lambda^2+\lambda+1)^{3/2}}.$$

We can see that $G2'_{TS}(\lambda) \geq 0$ when $\lambda \leq \lambda^c$, which proves that the maximum value of $G2_{TS}$ is obtained at $\lambda = \lambda^c$ for $\lambda \leq \lambda^c$.

Now take $\lambda \geq \lambda^c$, giving

$$G2_{TS}(\lambda, \lambda^c) = \frac{\sqrt{3}(\lambda^c)^2}{4} \frac{[(6 - 3\lambda^c) + (2\lambda^c - 3)\lambda]}{\sqrt{\lambda^2 - 3\lambda + 3}}.$$

The derivative of $G2_{TS}(\lambda, \lambda^c)$ with respect to λ is

$$G2'_{TS}(\lambda) = \frac{\sqrt{3}(\lambda^c)^2}{4} \frac{(\lambda^c - \lambda)}{(\lambda^2 - 3\lambda + 3)^{3/2}}.$$

The fact that $G2'_{TS}(\lambda) \leq 0$ shows that the maximum value of $G2_{TS}$ is obtained at $\lambda = \lambda^c$ when $\lambda \geq \lambda^c$.

However, if there are two breaks, λ_1^c and λ_2^c, we need to analyze what the break point estimation should be.

$$\begin{aligned} G2_{TS}(\lambda, \lambda_1^c) &= \frac{(1-\lambda)^2(1-\lambda_1^c)^2(\lambda + \lambda_1^c + 2\lambda\lambda_1^c + 2)/6}{\sqrt{(1-\lambda^3)(1-\lambda_1^c)^3/3}} \\ &= \frac{(1-\lambda_1^c)^2(\lambda + \lambda_1^c + 2\lambda\lambda_1^c + 2)/6}{\sqrt{(\lambda^2+\lambda+1)/3}} \\ &= \frac{(1-\lambda_1^c)^2}{2\sqrt{3}} \frac{(\lambda + \lambda_1^c + 2\lambda\lambda_1^c + 2)}{\sqrt{\lambda^2+\lambda+1}}. \end{aligned}$$

$$\begin{aligned} G2_{TS}(\lambda, \lambda_2^c) &= \frac{(1-\lambda)^2(1-\lambda_2^c)^2(\lambda + \lambda_2^c + 2\lambda\lambda_2^c + 2)/6}{\sqrt{(1-\lambda^3)(1-\lambda_2^c)^3/3}} \\ &= \frac{(1-\lambda_2^c)^2(\lambda + \lambda_2^c + 2\lambda\lambda_2^c + 2)/6}{\sqrt{(\lambda^2+\lambda+1)/3}} \\ &= \frac{(1-\lambda_2^c)^2}{2\sqrt{3}} \frac{(\lambda + \lambda_2^c + 2\lambda\lambda_2^c + 2)}{\sqrt{\lambda^2+\lambda+1}}. \end{aligned}$$

For each $G2_{TS}$, we take the derivative of $G2_{TS}$ with respect to λ and get

$$[G2_{TS}]'|_\lambda = \frac{\sqrt{3}(1-\lambda^c)^2}{4} \frac{(\lambda^c - \lambda)}{(\lambda^2+\lambda+1)^{3/2}}.$$

We can see that $[G2_{TS}]'|_\lambda \geq 0$ when $\lambda \leq \lambda^c$, which proves that the maximum value of $G2_{TS}$ is obtained at $\lambda = \lambda^c$.

When $\lambda \geq \lambda^c$, we can get

$$G2_{TS}(\lambda, \lambda^c) = \frac{\sqrt{3}(\lambda^c)^2}{4} \frac{[(6-3\lambda^c)+(2\lambda^c-3)\lambda]}{\sqrt{\lambda^2-3\lambda+3}}.$$

The derivative of $G2_{TS}(\lambda, \lambda_1^c)$ over λ is

$$[G2_{TS}]'|_\lambda = \frac{\sqrt{3}(\lambda^c)^2}{4} \frac{(\lambda^c-\lambda)}{(\lambda^2-3\lambda+3)^{3/2}}.$$

The fact that $[G2_{TS}^2]'|_\lambda \leq 0$ shows that the maximum value of $G2_{TS}$ is obtained at $\lambda = \lambda^c$ when $\lambda \geq \lambda^c$.

For two break at λ_1^c and λ_2^c, define $G2_{TS}^* = [M_1 * G2_{TS}(\lambda, \lambda_1^c) + M_2 * G2_{TS}(\lambda, \lambda_2^c)]^2$, we have the following proof to show that the maximum will not necessarily to be achieved at λ_1^c or λ_2^c.

a) when $\lambda < \lambda_1^c$, there is

$$\begin{aligned}
[G2_{TS}^*]'|_\lambda &= M_1[G2_{TS}]'(\lambda, \lambda_1^c)|_\lambda + M_2[G2_{TS}]'(\lambda, \lambda_2^c)|_\lambda \\
&= M_1 \frac{\sqrt{3}(1-\lambda_1^c)^2}{4} \frac{(\lambda_1^c-\lambda)}{(\lambda^2+\lambda+1)^{3/2}} + M_2 \frac{\sqrt{3}(1-\lambda_2^c)^2}{4} \frac{(\lambda_2^c-\lambda)}{(\lambda^2+\lambda+1)^{3/2}}.
\end{aligned}$$

To obtain the maximum value of $G2_{TS}^*$, we have to make $[G2_{TS}^*]'|_\lambda = 0$.

$$\begin{aligned}
[G2_{TS}^*]'|_\lambda &= M_1[G2_{TS}]'(\lambda, \lambda_1^c)|_\lambda + M_2[G2_{TS}]'(\lambda, \lambda_2^c)|_\lambda \\
&= M_1 \frac{\sqrt{3}(1-\lambda_1^c)^2}{4} \frac{(\lambda_1^c-\lambda)}{(\lambda^2+\lambda+1)^{3/2}} + M_2 \frac{\sqrt{3}(1-\lambda_2^c)^2}{4} \frac{(\lambda_2^c-\lambda)}{(\lambda^2+\lambda+1)^{3/2}}.
\end{aligned}$$

We need

$$M_1 \frac{\sqrt{3}(1-\lambda_1^c)^2}{4} \frac{(\lambda_1^c-\lambda)}{(\lambda^2+\lambda+1)^{3/2}} + M_2 \frac{\sqrt{3}(1-\lambda_2^c)^2}{4} \frac{(\lambda_2^c-\lambda)}{(\lambda^2+\lambda+1)^{3/2}} = 0.$$

a.1 both $M_1 > 0$ and $M_2 > 0$ or both $M_1 < 0$ and $M_2 < 0$.

For this case, we can see that, there is no λ to satisfy this condition. Hence there would be no maximum under this condition.

a.2 $M_1 > 0$ and $M_2 < 0$ or $M_1 < 0$ and $M_2 > 0$.

Under this condition, we can solve the λ that

$$\lambda = \frac{M_1(1-\lambda_1^c)^2\lambda_1^c + M_2(1-\lambda_2^c)^2\lambda_2^c}{M_1(1-\lambda_1^c)^2 + M_2(1-\lambda_2^c)^2}.$$

b) when $\lambda > \lambda_2^c$

Similar, we need to get

$$M_1 G2'_{TS}(\lambda, \lambda_1^c) + M_2 G2'_{TS}(\lambda, \lambda_2^c) = 0$$

hence

$$M_1 \frac{\sqrt{3}(\lambda_1^c)^2}{4} \frac{(\lambda_1^c - \lambda)}{(\lambda^2 - 3\lambda + 3)^{3/2}} + M_2 \frac{\sqrt{3}(\lambda_2^c)^2}{4} \frac{(\lambda_2^c - \lambda)}{(\lambda^2 - 3\lambda + 3)^{3/2}} = 0.$$

b.1 both $M_1 > 0$ and $M_2 > 0$ or both $M_1 < 0$ and $M_2 < 0$.

Similar reason, there will be result for this condition.

a.2 $M_1 > 0$ and $M_2 < 0$ or $M_1 < 0$ and $M_2 > 0$.

$$\lambda = \frac{M_1(\lambda_1^c)^2\lambda_1^c + M_2(\lambda_2^c)^2\lambda_2^c}{M_1(\lambda_1^c)^2 + M_2(\lambda_2^c)^2}.$$

c) when $\lambda_1^c < \lambda < \lambda_2^c$

Under this condition, we will need to make

$$M_1 \frac{\sqrt{3}(\lambda_1^c)^2}{4} \frac{(\lambda_1^c - \lambda)}{(\lambda^2 - 3\lambda + 3)^{3/2}} + M_2 \frac{\sqrt{3}(1-\lambda_2^c)^2}{4} \frac{(\lambda_2^c - \lambda)}{(\lambda^2 + \lambda + 1)^{3/2}} = 0.$$

We can see that the solution depends M_1 and M_2.

Therefore our analysis shows that the break point estimator performs very different for different M_1 and M_2. whether M_1 and M_2 are positive or negative would be important. It shows the existence of the inconsistency problem for multiple trend shift estimation problem.

C.4 Proof of Theorem 3.5.6

C.4.1 Proof of part 1: $\hat{\lambda}_{TS}$

Under assumption (C1.b): $u_t = \rho u_{t-1} + \varepsilon_t$, where $\rho \doteq 1 - \frac{c}{T}$,

$$T^{-3/2} \sum_{t=1}^{T} T^{-1} \tilde{DT}_t(\lambda) u_t \Rightarrow d(1) \int_0^1 F(r, \lambda) V_c(r) dr,$$

where $V_c(r) = \int_0^r \exp(-c(r-s)) dW(s)$.

C.4.2 Proof of part 2: $\hat{\lambda}_{MS}$

Because $u_t = (1 - \frac{c}{T}) u_{t-1} + \varepsilon_t$, it follows that $\Delta u_t = -\frac{c}{T} u_{t-1} + \varepsilon_t$. This gives

$$\begin{aligned}(T-1)^{-1/2} \sum_{t=2}^{T} \tilde{DU}_t(\lambda) \Delta u_t &= (T-1)^{-1/2} \sum_{t=2}^{T} \tilde{DU}_t(\lambda) \varepsilon_t - \\ &\quad (T-1)^{-1/2} (T)^{-1} c \sum_{t=2}^{T} \tilde{DU}_t(\lambda) u_{t-1},\end{aligned}$$

where

$$(T-1)^{-1/2} \sum_{t=2}^{T} \tilde{DU}_t(\lambda) \varepsilon_t \Rightarrow d(1)[\lambda W(1) - W(\lambda)],$$

and

$$(T-1)^{-1/2} (T)^{-1} c \sum_{t=2}^{T} \tilde{DU}_t(\lambda) u_{t-1} \Rightarrow d(1) c \int_0^1 (1(r > \lambda) - (1-\lambda)) V_c(r) dr.$$

The rest of the proof is straightforward and follows the proof of Theorem 3.4.5 and Theorem 3.4.4.

BIBLIOGRAPHY

BIBLIOGRAPHY

Andrews, D. W. K.: (1991), Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* **59**, 817–854.

Andrews, D. W. K.: (1993), Tests for parameter instability and structural change with unknown change point, *Econometrica* **61**, 821–856.

Andrews, D. W. K. and Ploberger, W.: (1994), Optimal tests when a nuisance parameter is present only under the alternative, *Econometrica* **62**, 1383–1414.

Bai, J.: (1998), A note on spirious breaks, *Econometric Theory* **14**, 663–669.

Bai, J. S.: (1994), Least squares estimation of a shift in linear process, *Journal of Time Series Analysis* **15**, 453–472.

Bai, J. S.: (1995), Least absolute deviation estimation of a shift, *Econometric Theory* **11**, 403–436.

Bai, J. S. and Perron, P.: (1998), Estimating and testing linear models with multiple structural breaks, *Econometrica* **66**, 47–78.

Bai, J. S. and Perron, P.: (2003), Computaion and analysis of multiple structural change models, *Journal of Applied Econometics* **18**, 1–22.

Bowman, A. W. and Azzalini, A.: (1997), *Applied smoothing techniques for data analysis: the kernel approach with S-plus illustrations*, Oxford University Press, United Kingdom. Various Issues.

Bunzel, H. and Vogelsang, T. J.: (2005), Powerful trend function tests that are robust to strong serial correlation with an application to the prebisch-singer hypothesis, *Journal of Business and Economic Statistics* **23**, 381–394.

Canjels, E. and Watson, M. W.: (1997), Estimating deterministic trends in the presence of serially correlated errors, *Review of Economics and Statistics* **May**, 184–200.

Chong, T.: (1994), Consistency of change-point estimators when the mumber of change points in structural change models is underspecified. working paper, Department of Economics, University of Rochester.

Chong, T.: (1995), Partial parameter consistency in a misspecified structural change model, *Economics Letters* **49**, 351–357.

Crainiceanu, C. and Vogelsang, T. J.: (2007), Non-monotonic power for tests of mean shift in a time series, *Journal of Statistical Computation and Simulation* **77**, 457–476.

Deng, A. and Perron, P.: (2006), A comparison of alternative asymptotic frameworks to analyze a structural change in a linear time trend, *Econometrics Journal* **9**, 423–447.

Deng, A. and Perron, P.: (2008), A non-local perspective on the power properties of the cusum and cusum of squares tests for structural change, *Journal of Econometrics* **141**, 212–240.

Frisch, R. and Waugh, F.: (1933), Partial time regressions as compared with individual trends, *Econometrica* **45**, 939–953.

Hashimzade, N. and Vogelsang, T. J.: (2008), Fixed-b asymptotic approximation of the sampling behavior of nonparametric spectral density estimators, *Journal of Time Series Analysis* **29**, 142–162.

Juhl, T. and Xiao, Z.: (2009), Tests for changing mean with monotonic power, *Journal of Econometrics* **148**, 14–24.

Kejriwal, M.: (2009), Tests for a mean shift with good size and monotonic power, *Economics Letters* **102**, 78–82.

Kejriwal, M. and Perron, P.: (2010), A sequential procedure to determine the number of breaks in trend with an integrated or stationary noise component. working paper, Department of Economics, Purdue University.

Kiefer, N. M. and Vogelsang, T. J.: (2005), A new asymptotic theory for heteroskedasticity-autocorrelation robust tests, *Econometric Theory* **21**, 1130–1164.

Ng, S. and Vogelsang, T. J.: (2002), Analysis of vector autoregressions in the presence of shifts in mean, *Econometric Reviews* **21**, 353–381.

Nunes, L. C., Kuan, C. M. and Newbold, P.: (1995), Spurious break, *Econometric Theory* **11**, 736–749.

Perron, P.: (1990), Testing for a unit root in a time series with a changing mean, *Journal of Business and Economic Statistics* **8**, 153–162.

Perron, P.: (1991), A test for changes in a polynomial trend function for a dynamic time series. manuscript, Princeton University.

Perron, P. and Zhu, X.: (2005), Structural breaks with stochastic and deterministic trends, *Journal of Econometrics* **129**, 65–119.

Phillips, P. C. B.: (1987), Time series regression with unit roots, *Econometrica* **55**, 277–302.

Sayginsoy, O. and Vogelsang, T. J.: (2010), Testing for a shift in trend at an unknown date: A fixed-b analysis of heteroskedasticity autocorrelation robust ols based tests, *Econometric Theory* . forthcoming.

Vogelsang, T. J.: (1997), Wald-type tests for detecting shifts in the trend function of a dynamic time series, *Econometric Theory* **13**, 818–849.

Vogelsang, T. J.: (1998), Testing for a shift in mean without having to estimate serial correlation parameters, *Journal of Business and Economic Statistics* **16**, 73–80.

Vogelsang, T. J.: (1999), Sources of nonmonotonic power when testing for a shift in mean of a dynamic time series, *Journal of Econometrics* **88**, 283–300.

Yang, J.: (2010), Break point estimates for a shift in trend: Levels versus first differences. Working paper, Department of Economics, Michigan State University.

Yao, Y.-C.: (1987), Approximating the distribution of the ml estimate of the change point in a sequence of independent r.v.'s, *Annals of Statistics* **4**, 1321–1328.