Estimating Break Points In Linear Models: A GMM Approach

A thesis submitted to the University of Manchester for the degree of PhD in the Faculty of Humanities

2015

Odaro Augustine-Ohwo

Economics, School of Social Sciences

Contents

List of Tables

List of Figures

Abstract

The University of Manchester

Odaro Augustine-Ohwo

Doctor of Philosophy in Economics

Break Point Estimation In Linear Models: A GMM Approach

2015

In estimating econometric time series models, it is assumed that the parameters remain constant over the period examined. This assumption may not always be valid when using data which span an extended period, as the underlying relationships between the variables in these models are exposed to various exogenous shifts. It is therefore imperative to examine the stability of models as failure to identify any changes could result in wrong predictions or inappropriate policy recommendations. This research proposes a method of estimating the location of break points in linear econometric models with endogenous regressors, estimated using Generalised Method of Moments (GMM).

The proposed estimation method is based on Wald, Lagrange Multiplier and Difference type test statistics of parameter variation. In this study, the equation which sets out the relationship between the endogenous regressor and the instruments is referred to as the Jacobian Equation (JE). The thesis is presented along two main categories: Stable JE and Unstable JE.

Under the Stable JE, models with a single and multiple breaks in the Structural Equation (SE) are examined. The break fraction estimators obtained are shown to be consistent for the true break fraction in the model. Additionally, using the fixed break approach, their T-convergence rates are established. Monte Carlo simulations which support the asymptotic properties are presented.

Two main types of Unstable JE models are considered: a model with a single break only in the JE and another with a break in both the JE and SE. The asymptotic properties of the estimators obtained from these models are intractable under the fixed break approach, hence the thesis provides essential steps towards establishing the properties using the shrinking breaks approach. Nonetheless, a series of Monte Carlo simulations conducted provide strong support for the consistency of the break fraction estimators under the Unstable JE.

A combined procedure for testing and estimating significant break points is detailed in the thesis. This method yields a consistent estimator of the true number of breaks in the model, as well as their locations. Lastly, an empirical application of the proposed methodology is presented using the New Keynesian Phillips Curve (NKPC) model for U.S. data. A previous study has found this NKPC model is unstable, having two endogenous regressors with Unstable JE. Using the combined testing and estimation approach, similar break points were estimated at 1975:2 and 1981:1. Therefore, using the GMM estimation approach proposed in this study, the presence of a Stable or Unstable JE does not affect estimations of breaks in the SE. A researcher can focus directly on estimating potential break points in the SE without having to pre-estimate the breaks in the JE, as is currently performed using Two Stage Least Squares.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the Copyright) and he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the Intellectual Property) and any reproductions of copyright works in the sis, for example graphs and tables (Reproductions), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy[1], in any relevant Thesis restriction declarations deposited in the University Library, The University Librarys regulations[2] and in The Universitys policy on presentation of Theses.

[1](see http://www.campus.manchester.ac.uk/medialibrary/policies/intellectualproperty.pdf)

[2](see http://www.manchester.ac.uk/library/aboutus/regulations)

Dedication

To my loving husband - Augustine,
my delightful children - Paul and Marho
and my precious parents - Professor and Dr. (Mrs) Otite.

Acknowledgments

Unconditional thanks to Almighty God "***The I am that I am!***"

I am extremely grateful to my principal supervisor, Professor Alastair Hall and my co-supervisor, Dr. Arthur Sinko. Alastair is unbelievably patient with astonishing wisdom. His time spent to painstakingly go through my work is remarkable. Needless to say, this research would not have been possible without him. Arthur's solution to anything and everything programming is extraordinary. I am grateful for his encouragement and 'motivational talks'. I could not have asked for a better supervisory team.

Thanks to the School of Social Sciences and Economics DA for sponsoring this research. I am grateful to Dr. Michael Croucher in EPS Applications Support for his assistance with programming in MATLAB and CONDOR usage. My thanks go to the ever-helpful Marie Waite at the SoSS postgraduate office. Huge thanks to Dr. Enegela and Dr. Yaduma for proofreading this thesis. Your time and useful comments are appreciated. Special thanks to Dr. Rabeya, James, Chuku and all the staff and colleagues of Economics DA.

To my family whose support and encouragement kept encouraging me, thanks a lot. Uncle Christopher, Fidelia, Kenneth, Edore, Afure and Tony, Jovi, Ovede, Edirin, Louisa, "My Toy" Beluchi and all my gorgeous nieces and nephews.

To all my friends who stood by me on this journey, Rev. Frs. Kenneth, Tega, James and Job, Philip, Uje, Natina, Jide, Bala, Landlord Demi, Ufoma, Pat, Rukky and Little Princess Gabrielle, I say a mighty thank you.

Special thanks to my parents, Prof. and Dr.(Mrs) Otite. You have always been there for me and have always encouraged me all through life. You know I can never be grateful enough. Best parents ever!

My lively and cheerful children, Paul and Marho... thank you for being considerate with mummy. I love you both so much.

As for my husband, Augustine...I have always been lost for words to describe your patience, love and support. You are truly exceptional!

Chapter 1

Introduction

A fundamental assumption when creating econometric models is that the parameters used in characterising the models remain stable or constant throughout the period examined. Such parameters are useful for economic forecasting, inference, budgeting, policy analyses and other economic decision-making. However, as these models usually span over long periods of time, the variables in these models are exposed to time-varying economic situations and various amendments in economic policies and decisions. Such amendments could be as a result of a change in government, key decision makers, technological innovations or even changes in consumers' preferences as they respond to modifications in interest rates, economic growth and other economic factors.

As noted in the famous policy evaluation by Lucas (1976): "...optimal decision rules vary systematically with changes in the structure of series relevant to the decision maker, it follows that any change in policy will systematically alter the structure of econometric models." Thus, the variables are subject to individual change or their relationships and interactions between them may change over time.

For example, Zhang et al. (2008) find changes in inflation dynamics in the United States over the period 1968 - 2005 which they attributed to the diverse monetary policies being followed by different chairmen of the Federal Reserve, specifically, the Volcker-Greenspan era. Additionally, within the same period, Bai (1997b) find changes in the relationship between the response of market interest rates to changes in the discount rates set by the Federal Reserve.

Structural changes are not limited only to economic variables, but also cut across a wide range of sectors. For example, in Medical Sciences, Erdman and Emerson (2008) and

Zhang and Siegmund (2007) detect changes in regions of DNA alteration which is important in cancer research. In the field of technology, Chernoff and Zacks (1964) examine the change point problem in means of independent and normally distributed random variables of a radar tracking system while Barry and Hartigan (1993) study the means of circular hollows cut by a milling machine. In Finance, Andreou and Ghysels (2002) detect multiple changes in the foreign exchange markets in Asia and Russia while Loschi et al. (2005) estimate a number of change points in the series of returns of stock indexes in Latin America.

The importance of detecting a change in economics cannot be over emphasised as failure to do so can be very costly, resulting in wrong predictions and inappropriate policy recommendations. Furthermore, if the timing of these parameter changes is identified, it can be taken into consideration when constructing models. This could also be useful in determining the causes for changes as well as analysing the effects of policy changes or other economic and exogenous shifts.

Thus, in developing econometric models, it is now commonplace to test models for stability, either of the model itself or of the parameters. In literature, this problem is commonly called the 'change-point' or 'structural break' problem and the location at which the change occurs is known as the 'change point' or 'break point'. The location of the break point may be known or unknown beforehand. It may also be considered to be within a specified time range in the sample. The problem is more specifically related to time series models whose observations are ordered by time. Models which have break points in them are referred to as unstable models, while those without are termed stable.

The main issues associated with a model that exhibits break points have been summarised in Bai (1997a). These are: (i) The determination of the number of breaks (ii) The estimation of the break points given this number and (iii) The statistical analysis of the resulting break point estimators. This research studies all three issues extensively, though more focus is placed on the latter two.

The determination of the number of breaks is conducted through hypothesis testing. This form of hypothesis test examines the null hypothesis of no break point or no structural change against an alternative of one or more break points. Various structural and parameter stability tests have been proposed in literature, notable amongst these are those for testing a single break in the model, for example Quandt (1958), Chow (1960), Andrews and Fair (1988) and Andrews (1993); and those developed for testing the presence of multiple breaks, for example, Bai and Perron (1998), Hall et al. (2012) and Hall et al. (2015). We discuss a variety of these tests in Chapter 6.

The hypothesis tests provide vital information about the stability of the model. Specifically, a model is termed unstable if the test statistic obtained is greater than the associated critical value. That is, a significant test statistic indicates the model changes at some point during the period under review but beyond that, it gives no other information. However, researchers interests may surpass merely ascertaining whether or not a model is unstable. They may also, perhaps more importantly, be interested in knowing the specific point in time the breaks occur. These could be vital inputs in economic policy analyses. This leads to the second issue associated with unstable models highlighted above.

The literature on estimating the locations of break points covers various methods of estimation. For example, from a Bayesian point of view, see Chernoff and Zacks (1964), Yao (1988), Barry and Hartigan (1993), Zhang and Siegmund (2007) and Erdman and Emerson (2008); for a maximum likelihood approach, see Hinkley and Hinkley (1970) and Bhattacharya (1987); for least squares, see Bai (1994a), Bai (1994b), Bai (1997a), Bai (1997b), Bai and Perron (1998), Chong (2001), Perron and Qu (2006) and Perron and Yamamoto (2015); for Two Stage Least Squares, see Hall et al. (2012) and Boldea et al. (2012); and for Generalised Method of Moments, see Li and Müller (2009) and Hall et al. (2012).

The instability in models can be caused by a single break point or by multiple break points. Bai (1994b) provides a method of estimating a single break point in the mean of a linear process within the context of Ordinary Least Squares (OLS). Also using least squares, Bai (1994a) proposes a method of estimating a break point in the regression parameters based on the Wald-type test statistic. The location of these break points need not be known beforehand. The asymptotic properties of the break point estimators were also established.

In estimating multiple break points on the other hand, two methods are generally used in literature - the Sequential and Simultaneous Estimation Methods. In the Sequential Estimation Method as proposed by Bai (1997a) and Chong (2001), the individual break points are estimated one at a time until all the break points are obtained, while in the Simultaneous Estimation Method, all break points are concurrently estimated alongside the regression parameters as laid out in Bai and Perron (1998).

These methods are effective in estimating the location of the break points in models; however, since they are all established within the OLS framework, they are based on the assumption that the independent variables in the regression models are exogenous. In other words, the regressors are assumed to be uncorrelated with the errors in the model. However, if these regressors are actually correlated with the errors, that is, if the regressors are

endogenous, then biased and inconsistent OLS parameter estimators would be obtained if OLS estimation is carried out on such a model. This problem of endogenous regressors, known as the 'endogeneity problem', is frequently encountered in research and the three usual causes[1] have been identified as: omitted variables, measurement errors and simultaneity.

One of the earliest and classical examples frequently used to illustrate the endogeneity problem is given in Wright (1928) who was interested in estimating the elasticities of demand and supply curves for the agricultural product, flaxseed. We present the system of equations as,

$$\begin{aligned} d_t &= \alpha p_t + u_t^d \\ s_t &= \beta_1 p_t + \beta_2 x_t + u_t^s, \end{aligned} \tag{1.1}$$

where d_t and s_t represent demand and supply quantities in year t respectively, p_t is the price of the commodity in year t, x_t is a vector containing factors relating to supply. The market is assumed to clear in year t and hence, $d_t = s_t$ and p_t is the equilibrium price. The difficulty lies in obtaining the elasticity of demand, that is, how to obtain an estimator of α in (1.1) when d_t and p_t are known. p_t is endogenous because producers change their price in response to demand and consumers change their demand in response to price. Hence, since the quantity demanded and price are simultaneously determined, then the OLS estimates are biased and Instrumental Variables (IV) are commonly used in literature to address this endogeneity problem.

These IV, which we denote as z_t have two main properties: (i) they are correlated with the endogenous variable, price p_t and (ii) they are uncorrelated with the error term u_t^d in (1.1). Thus, the endogenous variable p_t can be formed as,

$$p_t = \gamma_0 + \gamma z_t + u_t^p, \tag{1.2}$$

where $\gamma \neq 0$, $E[z_t u_t^d] = 0$, $E[u_t^d, u_t^p] = \Omega$, Ω has $Var[u_t^d]$ and $Var[u_t^p]$ as its main diagonal elements and $Cov[u_t^d, u_t^p]$ on its off-diagonal elements with $Cov[u_t^d, u_t^p] \neq 0$. As such, multiplying through by z_t, (1.1) now becomes,

$$E[z_t d_t] = \alpha E[z_t p_t] + E[z_t u_t^d],$$

where $E[z_t u_t^d] = 0$. Using yield per acre as an instrument, a consistent estimator of α,

[1] See Wooldridge (2009) and Verbeek (2008).

denoted $\hat{\alpha}$ is obtained as

$$\hat{\alpha} = \frac{\sum_{t=1}^{T} z_t d_t}{\sum_{t=1}^{T} z_t p_t}, \qquad since\ E[z_t u_t^d] = 0.$$

In econometrics, there are two main approaches to IV estimation, notably the Two Stage Least Squares (2SLS) and the Generalised Method of Moments (GMM). Using 2SLS, Hall et al. (2012) obtained consistent estimators of break points by minimising the residual sum of squares (RSS) in the second step of the 2SLS estimation process. Prior to this second step however, the reduced form for the endogenous regressor in the first step of the 2SLS procedure (as given in equation (1.2)) needs to be estimated and any breaks therein identified using the procedure in Bai and Perron (1998). The limiting distribution of these break fraction estimators are established in Boldea et al. (2012).

On the other hand, applying a similar minimisation procedure to the sums of partial GMM estimators over all partitions of the sample, Hall et al. (2012) obtained inconsistent break fraction estimators. They attribute this behaviour to the inherent structure of the GMM objective function[2]. Since the GMM minimand is the square of sums, then it is possible that the effect of any misspecification due to the estimation of a wrong break date is offset within these squares of sums. Consequently, this prevents the break points from being identified.

Li and Müller (2009) propose another GMM method which focuses on obtaining inference on only the stable subset of parameters in a partially stable GMM model with moderate magnitude of shift. Other GMM approaches focus on testing rather than estimating the break points. These are discussed later on in Chapter 6.

More recently, Perron and Yamamoto (2015) propose a method of estimation in the presence of endogenous variables still using OLS. Their approach involves reformulation of the model with the probability limits of the OLS parameter estimates in such a way as to make the regressors and errors contemporaneously uncorrelated. If this process is efficiently carried out, then the usual break point estimation method in Bai and Perron (1998) can be used to identify the breaks. They found this approach to yield consistent estimators of the break dates in the New Keynesian Hybrid Phillips Curve. To the best of our knowledge, no further research has been carried out within the GMM framework to estimate consistent break points. Although Zhang et al. (2008) provide a good reference for an

[2]The Monte Carlo simulations carried out also revealed these break point estimators are relatively dispersed over all the possible partitions in a manner similar to the estimators obtained from a model with no break point.

empirical application of our proposed methodology when they estimate the break point in the New Keynesian Philips Curve (NKPC), they however do not provide any theoretical justification for their results.

Estimation within the GMM framework, as introduced by Hansen (1982), is hinged on a population moment condition (pmc), $E[z_t u_t^d(\alpha_0)] = 0$, which is assumed to hold throughout the sample. The GMM objective function (or GMM minimand) is based on a quadratic form of the sample counterparts of this pmc. Based on the GMM minimand, the $J-test$, which is actually a test of the overidentifying restrictions[3], is frequently used as a diagnostic test for model specification within the GMM literature. However, Hall and Sen (1999) note that the pmc is actually made up of two orthogonal parts: the identifying restrictions which represent the part of the pmc that are used up during estimations of the parameters and the overidentifying restrictions which are essentially a measure of how far the restrictions are from zero and are thus, associated with the structural stability of the model. Therefore, the $J-test$ may not be effective in detecting a break in the parameters. This was the case when Ghysels and Hall (1990) find the asset pricing models of Hansen and Singleton (1982) and Dunn and Singleton (1986) which were structurally stable when the $J-test$ was used, actually exhibited parameter instability when subjected to other predictive and Likelihood Ratio tests. Thus, Ghysels and Hall (1990) conclude that the $J-test$ had no power against local alternatives that had a break in the parameters and consequently, parameter instability can still be undetected when using the $J-test$. This leaves a gap in the literature for an alternative GMM-based approach to estimating break points in linear models that exhibit endogeneity.

In this study, we propose a different statistic to that used in Hall et al. (2012). Our approach focuses on test statistics which have power to detect parameter variation. The statistics are based on sequences of the Wald, Difference and Lagrange Multiplier (LM) type tests. These tests are described in more detail in Chapter 6. Due to the complexity of the variance-covariance matrix used in the computation of the Wald and LM test statistics, the theoretical analysis and proofs of the asymptotic properties of the break fraction estimators cover only the Difference-type tests. However, these three tests statistics perform comparably in finite samples as seen in the Monte Carlo simulations in Chapter 5. Using our approach, the break fraction estimators are shown to be T-consistent which is coherent with existing OLS literature, as given in Bai (1994a), Bai (1997a), Bai (1997b), Bai and Perron (1998), Hall et al. (2012) and Boldea et al. (2012).

This research is broadly divided into two groups which are dependent on the endogenous

[3] See page 46 in Hall (2005).

regressors. First, within the GMM context as presented in this research, we refer to the reduced form equation in (1.2) as the Jacobian Equation (JE). This is motivated by the linear model used in this research. The Jacobian - which is the first derivative of the moment condition - is simply the relationship between the endogenous regressors and the instruments. We focus on the stability of this relationship between the endogenous regressor and the instrument. If the relationship is constant as in (1.2), we refer to it as the Stable Jacobian; otherwise, we refer to it as the Unstable Jacobian. We consider only the case of a single break Unstable Jacobian, that is,

$$\begin{aligned} p_t &= \gamma_0 + \gamma_1 z_t + u_t^p \quad before\ the\ break \\ &= \gamma_0 + \gamma_2 z_t + u_t^p \quad after\ the\ break. \end{aligned}$$

As stated earlier, when estimating break points within the 2SLS context, Hall et al. (2012) first of all estimate the break points in the reduced form given in (1.2) because ignoring any break points here would lead to inconsistent estimators of the break points in the structural equation (SE) in (1.1). Within GMM context using our proposed approach, we estimate break points in the SE directly as we prove in this research that a stable or unstable Jacobian has no effect on estimating a consistent break point in the SE. Thus, a break in the relationship between the endogenous regressor and instruments anywhere in the sample does not have any negative impact on the estimated break fractions in the SE. This finding is vital as it reduces the break point estimation process within the IV framework for models characterised by endogeneity.

The asymptotic properties of the break fraction estimators obtained from Stable Jacobian models are established based on the assumption that the parameter change is of fixed magnitude. For the Unstable Jacobian on the other hand, we consider the case where the magnitude of change of the JE parameters shrink with the sample size. Although the proofs for the Unstable Jacobian are not easily tractable, it is interesting to note that the results of the Monte Carlo simulations strongly indicate a break in the JE does not confound the estimations of a true break point in the SE.

We apply our proposed GMM method to the NKPC adopted in Hall et al. (2012). Using 2SLS, they establish two break points in the reduced form equation of the output gap and one break point in the reduced form of the expected inflation. The location of these break points in the reduced form was obtained first before estimating the current inflation in the main NKPC structural equation. Using the fixed break point tests, they find no evidence of an additional break point in the SE. Hence, they conclude the model has only two

break points. We show that applying our estimation method directly to the main SE of the NKPC yields the same two true break points in the model as identified in Hall et al. (2012). Thus, our approach potentially saves computational time and also allows researchers to focus on the main SE, as they no longer have to estimate the reduced form equation first. Furthermore, using 2SLS presupposes the reduced form equation is correctly specified, a requirement not necessary for GMM estimations and moreso, any instrument used in the 2SLS estimations is naturally a potential instrument for GMM estimations.

Besides these, the GMM estimation technique is appealing because it offers a way to handle a broader class of models, compared to OLS, which is limited to linear regression models only, or Nonlinear least squares[4] which is confined to nonlinear regression models. Thus, in the presence of an unstable relationship between the endogenous regressors and the instruments, the researcher does not have to check for stability in the two stages (reduced and structural form equations), but rather, estimates break points only in the main SE of interest. Consequently, instabilities in the JE can be ignored and consistent break point estimators of the SE can still be obtained.

Throughout this paper, $\Rightarrow$ denotes weak convergence, $\xrightarrow{p}$ denotes convergence in probability, $\xrightarrow{d}$ denotes convergence in distribution, $o_p(1)$ denotes a sequence of random variables converging to zero in probability and $O_p(1)$ denotes a sequence that is stochastically bounded. Also, for a sequence of matrices, V_T is $O_p(1)$ if each of its elements is $O_p(1)$, see White (2001).

The remaining part of this research is organised into seven chapters as follows. The first two chapters examine models where the JE is stable. Chapter 2 deals with a model which has a single break in the SE alone. This is a relatively simple model used to introduce our proposed estimation procedure. Detailed proofs of the asymptotic properties of the break fraction estimator obtained from the single break models are presented and shown to be consistent with those existing in the literature.

Chapter 3 considers a model with multiple breaks in the SE alone. The asymptotic properties of the break fraction estimators obtained are also established and shown to be consistent with those in Bai (1997a).

Chapter 4 investigates models with an unstable JE. Two main types of models with an unstable JE are studied. In the first, the model has one break point in the JE only, while in the second, the model has a break point in both the JE and the SE. The asymptotic behaviour of the break fraction estimators obtained from the unstable JE models are analysed using

[4] Boldea and Hall (2013) extend Bai and Perron's (1998) method for estimating unknown multiple break dates to nonlinear least squares estimation method.

the shrinking break approach.

Chapter 5 provides detailed results from the Monte Carlo simulations carried out to observe the finite sample behaviour of the break fraction estimators obtained from the models used in Chapters 2 to 4. These results support the theoretical analysis carried out in the previous chapters.

Chapter 6 discusses the determination of the true number of break points in a model. There is the possibility that the estimation procedure discussed in Chapters 2 and 3 can over-estimate the number of break points if there is no rule in place to terminate the estimation process. This chapter addresses that issue and proposes a combined method similar to Bai (1997a) for sequentially testing the significance of a break point before estimating its location.

Chapter 7 presents results for an empirical application. Our proposed estimation method is applied to the New Keynesian Phillips Curve model similar to that used in Zhang et al. (2008) and Hall et al. (2012). The break points in the reduced form have been established in Hall et al. (2012) and we only investigate to see if our proposed estimation method also identifies the same break points in the SE.

Chapter 8 summarises the findings of this research and sets out the conclusions. For easy reference, the relevant appendices are placed at the end of each chapter.

Chapter 2

Stable Jacobian - Single Break Model

This chapter considers issues relating to estimation of a single break point within the context of a linear model. The location of the break point to be estimated is unknown in the sample. The estimation process is based on the Wald, Lagrange Multiplier (LM) and Difference-type statistics which are generally used in the literature for testing parameter and structural change, see Andrews and Fair (1988) and Andrews (1993). However, for simplicity, only the Difference-type statistics are used for discussions and theoretical analyses.

The chapter is outlined as follows. Firstly, the model used alongside the assumptions are presented. Next, a description of the estimation process is given and the test statistic and break fraction estimator are discussed. This break fraction estimator indexes the break point which we want to identify. We then establish the asymptotic properties of the test statistic and break fraction estimator. Specifically, the break fraction estimators are shown to be T-consistent which is coherent with the existing literature; see Bai (1994a), Bai (1997a), Bai and Perron (1998), Hall et al. (2012), Boldea et al. (2012). Lastly, the chapter concludes with detailed proofs given in the appendix at the end of this chapter.

2.1 The Model and its Assumptions

The linear model with one break is the baseline model adopted in this research. It is a straightforward model, similar to that in Hall et al. (2012) and is presented thus,

$$\begin{aligned} y_t &= x_t^{'}\theta_0^{(1)} + u_t, \quad t = 1, 2, \ldots, k_0, \\ &= x_t^{'}\theta_0^{(2)} + u_t, \quad t = k_0 + 1, \ldots, T, \end{aligned} \tag{2.1}$$

where y_t is an observation on the dependent variable; x_t is the $p \times 1$ vector of endogenous regressors; u_t is the disturbance with zero mean; $\theta_0^{(i)}$, $i = 1, 2$ are the $p \times 1$ unknown parameters to be estimated; $k_0 = [T\lambda_0]$; $\lambda_0 \in \Lambda \subset (0, 1)$; $[\,\cdot\,]$ indicates the greatest integer function of the term within the square brackets and T is the total sample size. k_0 is defined as the true break point and λ_0 as the true break fraction. Hence, $\theta_0^{(1)}$ and $\theta_0^{(2)}$ symbolize the parameters before and after the true break point respectively. It is assumed that $\theta_0^{(1)} \neq \theta_0^{(2)}$ and by design, the break fraction, λ_0, is bounded away from the end points of the sample.

This situation is generally known in the literature as a pure structural change because all the p parameters in the model are subject to change.

As x_t is endogenous, it is correlated with the errors u_t and hence Instrumental Variables (IV) are required for the estimation. Recall these IV must be both correlated with x_t and uncorrelated with u_t. The endogenous variable is now constructed simply as

$$x_t^{'} = z_t^{'}\Delta_0 + v_t^{'}, \tag{2.2}$$

where z_t is the $q \times 1$ vector of instruments satisfying $E[z_t u_t(\theta_0)] = 0$, Δ_0 is the $q \times p$ coefficient matrix and $q > p$.

Throughout this research, we refer to the baseline model in (2.1) as the Structural Equation (SE) and the endogenous regressor equation in (2.2) as the Jacobian Equation (JE).

We make the following assumptions:

Assumption 1. $E[z_t x_t^{'}] = E[z_t z_t^{'}]\Delta_0 = Q_{zz}\Delta_0$, *where* $\|Q_{zz}\| < \infty$ *and positive definite. Also,* Δ_0 *has rank* p *and* $\|\Delta_0\| < \infty$.

Assumption 2. *The matrices* $\frac{1}{j}\sum_{t=k_0+1}^{k_0+j} z_t z_t^{'}$ *and* $\frac{1}{j}\sum_{t=k_0-j+1}^{k_0} z_t z_t^{'}$ *have minimum eigenvalues bounded away from zero in probability for all* $\lambda \in \Lambda$ *and* $j \geq q$.

Assumption 3. $sup_{\lambda\in\Lambda}\|T^{-1}\sum_{t=1}^{[T\lambda]} z_t x_t^{'} - \lambda Q_{zz}\Delta_0\| \xrightarrow{p} 0$.

Assumption 4. $sup_{\lambda\in\Lambda}\|Var\left[T^{-1/2}\sum_{t=1}^{[T\lambda]} z_t u_t\right] - \lambda\Omega\| \xrightarrow{p} 0$, *where* Ω *is a positive definite matrix of finite constants.*

Assumption 5. *The true number of breaks,* $m_0 = 1$.

Assumption 1 states a constant relationship exists between the endogenous regressor and its instruments. In Chapter 4 however, we examine a model where this stability assumption of the JE is relaxed. The standard rank condition necessary for identification[1] when

[1] See Identification Condition on page 35 in Hall (2005).

estimating using GMM is also given. Assumption 2 guarantees there are enough observations near the true break point, so it can be identified. Additionally, it assumes the matrices are invertible and stochastically bounded uniformly in j. Assumptions 3 and 4 present the uniform convergence across all break fractions of the partial sums of the Jacobian and variance, respectively. These are useful in establishing the asymptotic properties of the break fraction estimator. Lastly, we impose Assumption 5 in this chapter but consider cases when $m_0 \geq 2$ in Chapter 3 under the multiple break points scenario.

2.2 The Test Statistic

As mentioned earlier, for computational simplicity, the test statistic and other theoretical analysis presented here are centred on the Difference-type test of parameter variation. This test statistic is based on:

$$D_T(\lambda) = T\left\{Q_T(\bar{\theta}_T(\lambda), \lambda) - Q_T(\hat{\theta}_T(\lambda), \lambda)\right\}, \tag{2.3}$$

where $D_T(\lambda)$ is the Difference test statistic, $Q_T(\bar{\theta}_T(\lambda), \lambda)$ and $Q_T(\hat{\theta}_T(\lambda), \lambda)$ are the GMM minimands associated with the restricted and unrestricted estimations respectively. The vector of restricted parameters, $\bar{\theta}_T(\lambda) = [\bar{\theta}_{1,T}(\lambda)', \bar{\theta}_{2,T}(\lambda)']'$, where $\bar{\theta}_{1,T}(\lambda)$ is constrained to be equal to $\bar{\theta}_{2,T}(\lambda)$. On the other hand, the vector of the unrestricted parameters, $\hat{\theta}_T(\lambda) = [\hat{\theta}_{1,T}(\lambda)', \hat{\theta}_{2,T}(\lambda)']'$, where $\hat{\theta}_{1,T}(\lambda)$ and $\hat{\theta}_{2,T}(\lambda)$ are allowed to be different. Appendices A.1 and A.2 briefly demonstrate the minimisation process by which these parameters are obtained.

The minimands are constructed in the usual way consistent with GMM framework as,

$$Q_T(\theta_T(\lambda), \lambda) = g_T(\theta_T(\lambda), \lambda)' W_T(\lambda) g_T(\theta_T(\lambda), \lambda),$$

where

$$g_T(\theta_T(\lambda), \lambda) = \begin{bmatrix} T^{-1}\sum_{t=1}^{T} z_t u_t(\theta_1(\lambda))\mathcal{I}_{t,T}(\lambda) \\ T^{-1}\sum_{t=1}^{T} z_t u_t(\theta_2(\lambda))\{1 - \mathcal{I}_{t,T}(\lambda)\} \end{bmatrix},$$

$W_T(\lambda) = diag\{W_{1,T}(\lambda), W_{2,T}(\lambda)\}$, $W_{i,T}(\lambda)$ is a $q \times q$ deterministic positive semi-definite matrix and $\mathcal{I}_{t,T}(\lambda)$ is an indicator variable that takes the value one if $t \leq k$ and zero

otherwise. The partial sum GMM estimators are defined as,

$$\hat{\theta}_T(\lambda) = argmin_{\theta(\lambda)\in\Theta\times\Theta} Q_T(\theta_T(\lambda), \lambda),$$

where $\Theta \subset \Re^p$. Consistent with GMM estimation procedure, we make some additional assumptions about the data:

Assumption 6. $E[z_t u_t(\theta_0)] = 0$, *for* $\theta_0 = (\theta_0^{(1)'}, \theta_0^{(2)'})'$.

Assumption 7. $sup_{\lambda\in\Lambda}\|\hat{\theta}_{i,T}(\lambda) - \theta_*^{(i)}(\lambda)\| \xrightarrow{p} 0$, *where* $sup_{\lambda\in\Lambda}\|\theta_*^{(i)}(\lambda)\| < \infty$, *for* $i = 1, 2$.

Assumption 8. $sup_{\lambda\in\Lambda}\|W_{i,T}(\lambda) - W_i(\lambda)\| \xrightarrow{p} 0$ *for* $i = 1, 2$, *where* $W_1(\lambda) = \lambda^{-1}C$, $W_2(\lambda) = (1-\lambda)^{-1}C$ *and* C *is a nonsingular constant matrix.*

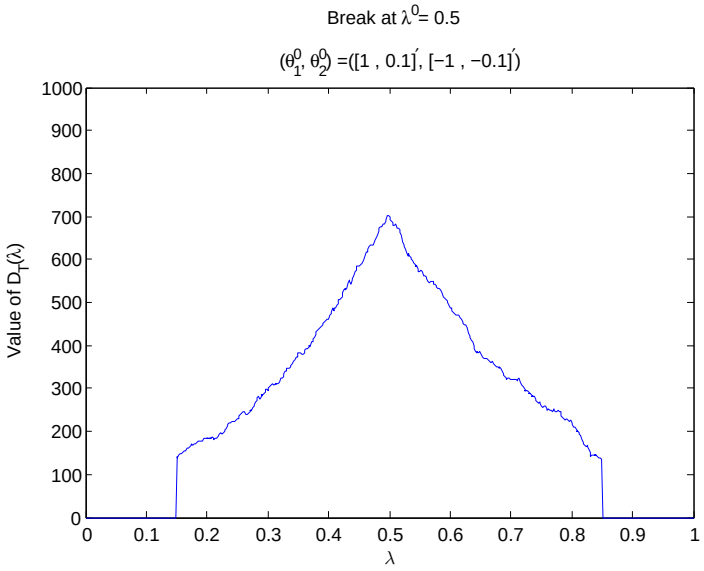
Assumption 6 states the identification condition which is the basis for GMM estimations. It infers that the population moment condition holds at the true break point and at the true parameter values. Assumption 7 imposes the uniform convergence of the estimators for all $\lambda \in \Lambda$. The individual structures of their limits, $\theta_*^{(i)}(\lambda)$, are given in the relevant sections. Assumption 8 displays the specific forms of the weighting matrices used. These have similar structures to the weights used in Hall et al. (2012).

Under the GMM framework as proposed by Hansen (1982), the convention is to estimate the parameters in a model using the two-step estimation procedure. In the first step, a preliminary weighting matrix (often the identity matrix) is used to obtain the first-step estimators. The inverses of the variances obtained from this first-step estimators are then used as weights in the second step estimations to obtain second-step GMM parameter estimators, which are known to be consistent and efficient. In our own estimation approach however, we consider only the first-step GMM estimators since the efficiency of the second-step estimators are based on the first-step estimators. Hence, any break in the model not identified by these first-step estimators would naturally undermine the estimations in the second-step[2]. This approach of using the first step estimators is reasonable as seen in the theoretical analysis and results of the Monte Carlo simulations.

The test statistic used for the break point estimation is constructed similar to Andrews (1993): a sequence made up of $D_T(\lambda)$ as defined in (2.3) is obtained for each candidate $\lambda \in \Lambda$. To illustrate its finite sample behaviour, two plots of $D_T(\lambda)$ across the full range

[2]Hall et al. (2012) attribute this to complications arising from centering of the estimated long run covariance matrix of the first-step estimators. See Hall (2005) page 125 for a detailed analysis of the impact of unstable models on the covariance matrix estimators.

Figure 2.1: $D_T(\lambda)$ for all $\lambda \in \Lambda$.
DGP has 1 break in SE (1 simulation)



of $\Lambda = [0.15, 0.85]$ for a sample size of $T = 600$ are presented in Figures 2.1 and 2.2. In these plots, a break is imposed halfway through the sample, that is $\lambda_0 = 0.5$. The parameters before and after the break are $\theta_0^{(1)} = (1, 0.1)'$ and $\theta_0^{(2)} = (-1, -0.1)'$ respectively[3]. As evidenced by the single simulation conducted in Figure 2.1, $D_T(\lambda_0)$ is clearly the global maximum and hence, $D_T(\lambda)$ peaks at this point. Repeating the estimation procedure a thousand times, the same shape is maintained as Figure 2.2 shows, highlighting the fact that $D_T(\lambda)$ monotonically decreases as the estimated break moves further away from the true break.

2.3 The Break Fraction Estimator

To obtain the break fraction estimator, the *supremum* of the sequence of $D_T(\lambda)$ is calculated and its position within the sample indicates the location of the true break point. The break fraction estimator, $\hat{\lambda}$, is thus defined as

$$\hat{\lambda} = argmax_{\lambda \in \Lambda_T} D_T(\lambda), \tag{2.4}$$

[3]More details of the simulation design can be found in Chapter 5 on Monte Carlo simulations.

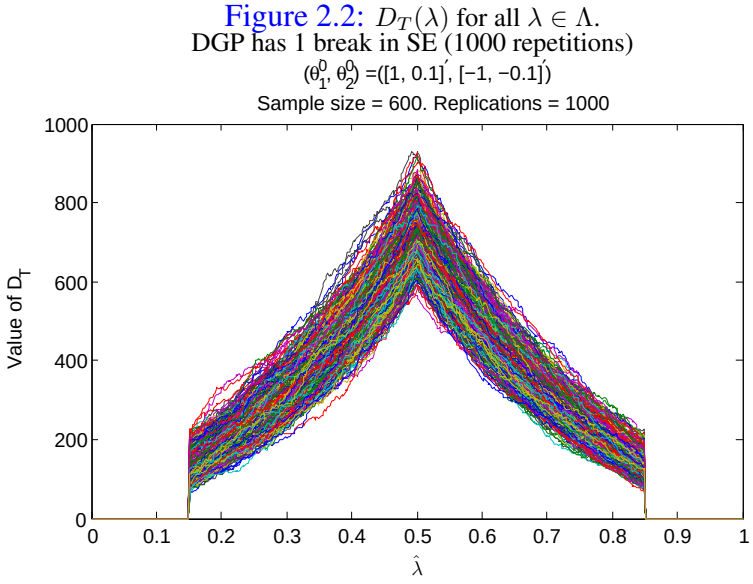


Figure 2.2: $D_T(\lambda)$ for all $\lambda \in \Lambda$.

where $\Lambda_T = \lambda$ such that $\lambda = k_i/T$ for $i = (\epsilon,\ 1-\epsilon)T$ and $\epsilon > 0$. Thus, $\hat{\lambda}$ estimates the location of the peak of $D_T(\lambda)$ as seen in Figures 2.1 and 2.2.

This method of estimating break points was suggested in Andrews (1993) and has previously been used in the literature. For example, Bai (1994a) propose a supWald statistics based on least squares to estimate the location of a break point. Conversely, minimising the residual sum of squares (RSS) across Λ has also been used in literature for OLS estimations; see Bai (1994b), Bai (1997a), Bai and Perron (1998) and Hall et al. (2012).

2.4 Consistency and Convergence Rate of $\hat{\lambda}$

This section presents some useful asymptotic properties of $\hat{\lambda}$ obtained from a single break model. Specifically, it establishes the consistency and rate of convergence of $\hat{\lambda}$. For the most part, our approach and propositions predominantly follow closely the arguments in Bai (1997a) and to a lesser degree, Bai (1994a) and Bai (1994b). To the best of our knowledge, these asymptotic properties of the break fraction estimator have not been established within the GMM framework. To help maintain focus on the salient issues, the detailed analysis of the proofs of lemmas used are relegated to the relevant appendices.

There are four vital things to note here before proceeding with the analysis. Firstly, the test statistic in (2.3) can be broken down analytically and re-expressed in a more tractable

form shown in Appendix A.3 as,

$$D_T(\lambda) = T(\hat{\theta}_{1,T}(\lambda) - \hat{\theta}_{2,T}(\lambda))' M_{*,T}(\lambda)^{-1} (\hat{\theta}_{1,T}(\lambda) - \hat{\theta}_{2,T}(\lambda)), \tag{2.5}$$

where $M_{*,T}(\cdot) = M_{1,T}(\cdot)^{-1} + M_{2,T}(\cdot)^{-1}$, $M_{i,T}(\cdot) = G_{i,T}(\cdot)' W_{i,T}(\cdot) G_{i,T}(\cdot)$ and $G_{i,T}(\cdot) = T^{-1} Z_{i,T}(\cdot)' X_{i,T}(\cdot)$, for $i = 1, 2$. In our study, we refer to $(\hat{\theta}_{1,T}(\lambda) - \hat{\theta}_{2,T}(\lambda))$ as the 'Parameter Difference' and $M_{*,T}(\lambda)^{-1}$ as the 'Centre Matrix'. In other words, the test statistic is composed of these two major parts and all our further analyses is based on this form of $D_T(\lambda)$ displayed in (2.5). Also, in this form, the test statistic can be thought of as a weighted average of the square of the Parameter Difference.

Secondly, there are only two possible regimes that can exist in a one-break model: the first regime covers the period before the break, $[1, k_0)$, while the second regime covers the period after the break, $[k_0 + 1, T]$. We occasionally refer to these as Subsample 1 and Subsample 2 respectively. Notice the SE is stable in each regime. Furthermore, observe that all break fractions in Subsample 1 are less than the true break, that is, $\lambda < \lambda_0$; the reverse is the case in the second subsample.

Thirdly, an additional assumption useful in establishing the convergence rate of the estimated break fraction needs to be made:

Assumption 9. *For some real number $r > 2$ and a constant $C > 0$,*

$$E\left| \sum_{t=i}^{j} z_t u_t \right|^r \leq C(j-i)^{r/2} \quad \textit{for all } 1 \leq i \leq j \leq T.$$

As noted in Bai (1994a), this assumption is a generalisation of the Hájek and Rényi (1955) inequality and is useful in the proof of limit theorems of independent summands seen in the appendices.

Lastly, we specify a scaled version of the test statistic as $D_T^*(\lambda)$, where $D_T^*(\lambda) = T^{-1} D_T(\lambda)$. This scaled version is also well defined for all Λ and useful for the asymptotic analysis of $\hat{\lambda}$.

2.4.1 Consistency of the Break Fraction Estimator

The proof strategy of the consistency of the break fraction estimator is as follows. Firstly, the nonstochastic limits of the test statistic are obtained for all $\lambda \in \Lambda$; secondly, we show the test statistic uniformly converges to these nonstochastic limits; and lastly, we show

these limits are uniquely maximised at the point where $\lambda = \lambda_0$. These steps are stated in the following three lemmas:

Lemma 1. *Under Assumptions 1 to 8, $D_T^*(\lambda)$ converges uniformly in probability to a nonstochastic function $D^*(\lambda)$ on $(0, 1)$.*

This nonstochastic function, $D^*(\lambda)$, exhibits three unique expressions across the range of Λ, one in each of the two regimes and one at the true break point, k_0. As detailed in Appendix B.1, these three expressions are given as,

$$D^*(\lambda_0) = \lambda_0(1-\lambda_0)\mathcal{P}, \qquad \textit{when } \lambda = \lambda_0 \tag{2.6}$$

$$D^*(\lambda) = \frac{\lambda}{1-\lambda}(1-\lambda_0)^2\mathcal{P}, \qquad \textit{for all } \lambda < \lambda_0 \tag{2.7}$$

$$D^*(\lambda) = \frac{1-\lambda}{\lambda}(\lambda_0)^2\mathcal{P}, \qquad \textit{for all } \lambda > \lambda_0, \tag{2.8}$$

where $\mathcal{P} = (\theta_0^{(1)} - \theta_0^{(2)})'\mathcal{B}(\theta_0^{(1)} - \theta_0^{(2)})$ and $\mathcal{B} = Q_{xz}CQ_{zx}$.

Lemma 2. *Under Assumptions 1 to 8,*

$$sup_{\lambda\in\Lambda}|D_T^*(\lambda) - D^*(\lambda)| = O_p(T^{-1/2}). \tag{2.9}$$

This lemma infers that for all λ, the test statistic $D_T^*(\lambda)$ is uniformly close to its expected nonstochastic function $D^*(\lambda)$, with high probability. The proof of this lemma is found in Appendix B.2 on page 44.

Lemma 3. *Under Assumptions 1 to 8, there exists an $\mathcal{E} > 0$ which depends on λ_0, $\theta_0^{(1)}$ and $\theta_0^{(2)}$ such that*

$$D^*(\lambda) - D^*(\lambda_0) \leq -\mathcal{E}|\lambda - \lambda_0| \quad \textit{for all } large\ T. \tag{2.10}$$

This lemma, proved in Appendix B.3 on page 48 indicates the test statistic is uniquely maximised at λ_0. In conjunction with Lemma 2, it follows that with high probability, the maximiser of $D_T^*(\lambda)$ will also be close to λ_0.

Based on these three lemmas, we state the consistency of the break fraction estimator in the following proposition followed by its proof,

Proposition 1. *Under Assumptions 1 to 8,*

$$\hat{\lambda} - \lambda_0 = O_p(T^{-1/2}). \tag{2.11}$$

The proposition implies the estimated break fraction, $\hat{\lambda}$, is asymptotically consistent for the true break fraction, λ_0.

Proof of Proposition 1

By adding and subtracting $D^*(\lambda)$ and $D^*(\lambda_0)$, we can write

$$D_T^*(\lambda) - D_T^*(\lambda_0) = \mathcal{D}_T^*\{\lambda, \lambda_0\} + \mathcal{D}^*\{\lambda, \lambda_0\}, \tag{2.12}$$

where

$$\mathcal{D}_T^*\{\lambda, \lambda_0\} = \{D_T^*(\lambda) - D^*(\lambda)\} - \{D_T^*(\lambda_0) - D^*(\lambda_0)\} \tag{2.13}$$

and

$$\mathcal{D}^*\{\lambda, \lambda_0\} = \{D^*(\lambda) - D^*(\lambda_0)\}. \tag{2.14}$$

Remark 1. *By definition, the test statistic, $D_T^*(\lambda)$ is a maximum so it must be that for $\mathcal{D}_T^*\{\lambda, \lambda_0\}$ in (2.13),*

$$\{D_T^*(\lambda) - D^*(\lambda)\} - \{D_T^*(\lambda_0) - D^*(\lambda_0)\} \leq 2\, sup_{\lambda\in\Lambda}|D_T^*(\lambda) - D^*(\lambda)|. \tag{2.15}$$

Remark 2. *Note also that $\mathcal{D}^*\{\lambda, \lambda_0\}$ given in (2.14) is defined for all $\lambda \in \Lambda$ and particularly for the estimated break fraction, $\hat{\lambda}$. Furthermore, the left hand side (LHS) of (2.12) is non-negative because by definition, $D_T^*(\hat{\lambda}) \geq D_T^*(\lambda_0)$.*

Combining $\mathcal{D}_T^*\{\lambda, \lambda_0\}$, $\mathcal{D}^*\{\lambda, \lambda_0\}$ and Remark 1 with Lemma 3, we obtain

$$\begin{aligned} D_T^*(\hat{\lambda}) - D_T^*(\lambda_0) &\leq 2\, sup_{\lambda\in\Lambda}|D_T^*(\lambda) - D^*(\lambda)| + \{D^*(\lambda) - D^*(\lambda_0)\} \\ &\leq 2\, sup_{\lambda\in\Lambda}|D_T^*(\lambda) - D^*(\lambda)| - \mathcal{E}|\hat{\lambda} - \lambda_0|. \end{aligned} \tag{2.16}$$

Hence, noting the latter part of Remark 2, it must be that

$$|\hat{\lambda} - \lambda_0| \leq \mathcal{E}^{-1} 2\, sup_{\lambda\in\Lambda}|D_T^*(\lambda) - D^*(\lambda)|. \tag{2.17}$$

This concludes the proof of Proposition 1 since the right hand side (RHS) of (2.17) is

$O_p(T^{-1/2})$ based on Lemmas 1 to 3.

2.4.2 Convergence Rate of the Break Fraction Estimator

The rate of convergence describes how fast the break fraction estimator converges to the true value. In this section, we show this rate to be the same as the sample size, T. In our analysis, we adopt the fixed break approach, that is, we assume the breaks are fixed as the sample size increases. To establish this rate for $\hat{\lambda}$, we make the following three vital definitions.

Firstly, define $V_1 = \{k : T\eta \leq k \leq T(1-\eta)\}$ for a small positive number η, such that $\lambda \in (\eta, 1-\eta)$. Thus, for all $k \in V_1$, this ensures that λ is both away from 0 and 1 for a positive number of the observations. From Proposition 1, since $\hat{\lambda}$ is consistent for λ_0, there is a high probability that $\hat{k}$ will fall into V_1. In other words, we write, $Pr(\hat{k} \notin V_1) < \epsilon$ for every $\epsilon > 0$ and all large T.

Secondly, define $V_2 = \{k : |k - k_0| \leq \mathcal{M}\}$, where $\mathcal{M} < \infty$ is constant. To prove the T-consistency of the break fraction, we show that $\hat{k}$ eventually falls into V_2 with large probability for large $\mathcal{M}$.

Lastly, define $V_3 = \{k : T\eta \leq k \leq T(1-\eta),\ |k - k_0| > \mathcal{M}\}$. That is, $V_3 = V_1 \cap V_2^c$, where V_2^c is the complement of V_2. We state in the following lemma, that the probability of the break point to lie within this set V_3 is very small,

Lemma 4. *Under Assumptions 1 to 9, there exists $\mathcal{M} < \infty$ such that for all large T and every $\epsilon > 0$,*

$$P\left(\sup_{k \in V_3} D_T(k) - D_T(k_0) > 0\right) < \epsilon. \tag{2.18}$$

With the proof of this lemma given in Appendix B.4, we now present our second proposition to state the rate T convergence of the break fraction estimator,

Proposition 2. *For every $\epsilon > 0$, there exists a finite $\mathcal{M}$ independent of T, such that for all large T,*

$$P(T|\hat{\lambda} - \lambda_0| > \mathcal{M}) < \epsilon. \tag{2.19}$$

Proof of Proposition 2

Using $Pr(A) \leq Pr(B^c) + Pr(B \cap A)$ and Lemma 4, we conclude that

$$\begin{aligned} P(T|\hat{\lambda} - \lambda_0| > \mathcal{M}) &= P(|\hat{k} - k_0| > \mathcal{M}) \\ &\leq Pr\big(\hat{k} \notin (T\eta, T(1-\eta)\big) + Pr\big(\hat{k} \in (T\eta, T(1-\eta)) \cap |\hat{k} - k_0| > \mathcal{M}\big) \\ &= Pr(\hat{k} \notin V_1) + Pr(\hat{k} \in V_3) \\ &\leq \epsilon + Pr\left(\sup_{k \in V_3} D_T(k) \geq D_T(k_0)\right) \\ &\leq 2\epsilon. \end{aligned}$$

This proof confirms our proposition that $\hat{\lambda}$ is T-consistent. Similar T-convergence rates are obtained in Bai (1994a), Bai (1997a), Bai and Perron (1998) and Hall et al. (2012) using Ordinary Least Squares (OLS).

2.5 Conclusion

In this chapter, issues relating to the estimation of a single unknown break point in a linear model are examined. We assume a pure structural change where all the parameters in the model are assumed to change at the same point in time. The estimation procedure is based on the *supremum* of a sequence of the Difference test statistic which is normally used to test parameter stability as given in Andrews (1993).

Additionally, the consistency and rate of convergence of the break fraction estimator were established and shown to be consistent with existing OLS literature provided in Bai (1994a), Bai (1997a), Bai and Perron (1998) and Hall et al. (2012). This implies our GMM approach of using parameter variation to estimate break points in linear models yields estimators with asymptotic properties consistent with existing literature. The detailed theoretical analysis carried out further adds to the body of literature within the GMM framework.

The finite sample behaviour of the break fraction estimator obtained from this one break model are presented through a series of Monte Carlo simulations in Chapter 5. In reality however, more than one break point may exist in a model. For example, Hall et al. (2012) find two breaks in the reduced form of the inflation forecast series between 1968 and 2001 while Hansen (2001) find three breaks in the U.S. labour productivity between 1947 to 2001. The next chapter considers models with multiple break points and the break fraction estimators obtained from such models are examined.

Appendix A

A.1 Unrestricted GMM estimator

This appendix reviews the process of obtaining GMM parameters from the unconstrained minimisation of the GMM minimand. Under this estimation technique[1], the unrestricted estimator solves the minimisation problem,

$$\hat{\theta}_T(\lambda) = argmin_{\theta(\lambda)\in\Theta\times\Theta} Q_T(\theta(\lambda);\lambda),$$

where $\Theta \subset \Re^p$; $\hat{\theta}_T(\lambda) = vec[\hat{\theta}_{1,T}(\lambda)', \hat{\theta}_{2,T}(\lambda)']'$; $Q_T(\theta(\lambda),\lambda)$ is the GMM minimand,

$$Q_T(\theta(\lambda),\lambda) = g_T(\theta_T(\lambda),\lambda)' W_T(\lambda) g_T(\theta_T(\lambda),\lambda);$$
$$g_T(\theta_T(\lambda),\lambda) = \begin{bmatrix} T^{-1}\sum_{t=1}^{T} z_t u_t(\theta_1(\lambda))\mathcal{I}_{t,T}(\lambda) \\ T^{-1}\sum_{t=1}^{T} z_t u_t(\theta_2(\lambda))\{1-\mathcal{I}_{t,T}(\lambda)\} \end{bmatrix};$$
$$W_T(\lambda) = \begin{bmatrix} W_{1,T}(\lambda) & 0 \\ 0 & W_{2,T}(\lambda) \end{bmatrix};$$

$W_{i,T}(\lambda)$ is a $q \times q$ deterministic positive semi-definite matrix; $\mathcal{I}_{t,T}(\lambda)$ is an indicator variable that takes the value one if $t \leq k$ and zero otherwise.

The first order conditions for this minimisation implies $\partial Q_T(\cdot)/\partial\theta(\cdot) = 0$, which results in the unrestricted GMM estimators:

$$\begin{aligned}\hat{\theta}_{i,T}(\lambda) &= (T^{-1}X_{i,T}'Z_{i,T}W_{i,T}T^{-1}Z_{i,T}'X_{i,T})^{-1}T^{-1}X_{i,T}'Z_{i,T}W_{i,T}T^{-1}Z_{i,T}'y_{i,T} \\ &= \theta_0^{(i)} + M_{i,T}^{-1}G_{i,T}'W_{i,T}T^{-1}Z_{i,T}'u_{i,T}, \end{aligned} \tag{A.1}$$

where $M_{i,T} = G_{i,T}'W_{i,T}G_{i,T}$, $G_{i,T} = T^{-1}Z_{i,T}'X_{i,T}$ for $i = 1, 2$ where 1 and 2 denote the

[1] See Hansen (1982), Andrews (1993) and Hall (2005).

subsample before and after the break point respectively. Thus, the parameters are allowed to be different across the two regimes.

A.2 Restricted GMM estimator

The restricted parameters on the other hand, are obtained from the constrained minimisation of the GMM minimand. That is,

$$\bar{\theta}_T(\lambda) = argmin_{\theta(\lambda)} Q_T(\theta(\lambda), \lambda), \qquad subject\ to\ \theta_1(\lambda) = \theta_2(\lambda),$$

where $\bar{\theta}_T(\lambda) = [\bar{\theta}_{1,T}(\lambda)', \bar{\theta}_{2,T}(\lambda)']'$. Forming the Lagrangean function, $\mathcal{L}(\cdot)$, the first order conditions imply $\partial\mathcal{L}(\cdot)/\partial\theta = 0$ and $\partial\mathcal{L}(\cdot)/\partial\kappa = 0$, where $\mathcal{L}(\cdot) = Q_T(\theta(\lambda), \lambda) + 2\kappa'(\theta_1(\lambda) - \theta_2(\lambda))$ and κ is the $p \times 1$ vector of Lagrange multipliers. Denote $\bar{\kappa}$ as the estimator of κ, these conditions result in the restricted parameter for Subsample 1 as,

$$\begin{aligned}\bar{\theta}_{1,T}(\lambda) &= (T^{-1}X_{1,T}'Z_{1,T}W_{1,T}T^{-1}Z_{1,T}'X_{1,T})^{-1}(T^{-1}X_{1,T}'Z_{1,T}W_{1,T}T^{-1}Z_{1,T}'y_{1,T} - \bar{\kappa}) \\ &= \hat{\theta}_{1,T} - M_{1,T}^{-1}\bar{\kappa}, \end{aligned} \tag{A.2}$$

and that of Subsample 2 as,

$$\bar{\theta}_{2,T}(\lambda) = \hat{\theta}_{2,T} + M_{2,T}^{-1}\bar{\kappa}. \tag{A.3}$$

Since $\bar{\theta}_{1,T}(\lambda) = \bar{\theta}_{2,T}(\lambda)$, equating the RHS of (A.2) and (A.3) results in the expression: $\bar{\kappa} = (M_{1,T}^{-1} + M_{2,T}^{-1})^{-1}(\hat{\theta}_{1,T} - \hat{\theta}_{2,T})$. Substituting $\bar{\kappa}$ back into the RHS of (A.2) and (A.3) gives the following expressions for the restricted estimators:

$$\bar{\theta}_{1,T}(\lambda) = \hat{\theta}_{1,T} - M_{1,T}^{-1}M_{*,T}^{-1}(\hat{\theta}_{1,T} - \hat{\theta}_{2,T}) \tag{A.4}$$

$$\bar{\theta}_{2,T}(\lambda) = \hat{\theta}_{2,T} + M_{2,T}^{-1}M_{*,T}^{-1}\big(\hat{\theta}_{1,T} - \hat{\theta}_{2,T}\big), \tag{A.5}$$

where $M_{*,T} = M_{1,T}^{-1} + M_{2,T}^{-1}$.

A.3 Analytical form of $D_T(\lambda)$

This appendix shows how the analytical form of the Difference test statistic is obtained from its main components - the restricted and unrestricted GMM minimands. For notational purposes, the dependence on T and λ are ignored.

To begin, we draw attention to two salient facts which are used in the analysis. Firstly, based on (A.4) and (A.5), the moment condition for the restricted parameters can be expressed in terms of the unrestricted parameters since,

$$\begin{aligned} T^{1/2}g_1(\bar{\theta}_1) &= T^{1/2}Z_1'u_1(\bar{\theta}_1) \\ &= T^{1/2}Z_1'y_1 - T^{1/2}Z_1'X_1\bar{\theta}_1 \\ &= T^{1/2}Z_1'y_1 - T^{1/2}Z_1'X_1\hat{\theta}_1 + T^{1/2}Z_1'X_1M_1^{-1}M_*^{-1}(\hat{\theta}_1 - \hat{\theta}_2) \\ &= T^{1/2}g_1(\hat{\theta}_1) + G_1M_1^{-1}M_*^{-1}T^{1/2}(\hat{\theta}_1 - \hat{\theta}_2). \end{aligned}$$

Similarly,

$$T^{1/2}g_2(\bar{\theta}_2) = T^{1/2}g_2(\hat{\theta}_2) - G_2M_2^{-1}M_*^{-1}T^{1/2}(\hat{\theta}_2 - \hat{\theta}_2).$$

Secondly, recall that the first order conditions for the unrestricted regressions imply $G_1'W_1T^{1/2}g_1(\hat{\theta}_1) = 0$ and $G_2'W_2T^{1/2}g_2(\hat{\theta}_2) = 0$.

With the GMM minimands associated with the restricted and unrestricted estimations defined as $Q_T(\bar{\theta}_1(\lambda), \bar{\theta}_2(\lambda); \lambda)$ and $Q_T(\hat{\theta}_1(\lambda), \hat{\theta}_2(\lambda); \lambda)$ respectively, the test statistic is broken down as follows,

$$\begin{aligned} D_T(\lambda) &= T\big\{Q_T(\bar{\theta}_1(\lambda), \bar{\theta}_2(\lambda); \lambda) - Q_T(\hat{\theta}_1(\lambda), \hat{\theta}_2(\lambda); \lambda)\big\} \\ &= T\big\{g_1(\bar{\theta}_1)'W_1g_1(\bar{\theta}_1) + g_2(\bar{\theta}_2)'W_2g_2(\bar{\theta}_2)\big\} \\ &\quad - T\big\{g_1(\hat{\theta}_1)'W_1g_1(\hat{\theta}_1) - g_2(\hat{\theta}_2)'W_2g_2(\hat{\theta}_2)\big\} \\ &= T\big\{g_1^*(\bar{\theta}_1)'W_1g_1^*(\bar{\theta}_1) + g_2^*(\bar{\theta}_2)'W_2g_2^*(\bar{\theta}_2)\big\} \\ &\quad - T\big\{g_1(\hat{\theta}_1)'W_1g_1(\hat{\theta}_1) - g_2(\hat{\theta}_2)'W_2g_2(\hat{\theta}_2)\big\} \\ &= T^{1/2}(\hat{\theta}_1 - \hat{\theta}_2)'M_*^{-1}T^{1/2}(\hat{\theta}_1 - \hat{\theta}_2), \end{aligned}$$

where $g_i^*(\bar{\theta}_i) = g_i(\hat{\theta}_i) + G_iM_i^{-1}M_*^{-1}(\hat{\theta}_1 - \hat{\theta}_2)$ for $i = 1, 2$.

Appendix B

B.1 Proof of Lemma 1

This appendix shows the uniform convergence of $D_T^*(\lambda)$ to its nonstochastic limit, $D^*(\lambda)$, for all $\lambda \in \Lambda$. First, notice that in the single break model, there are three different locations the candidate break fraction λ can be relative to the true break fraction, λ_0: (i) it can coincide with λ_0; (ii) it can lie before λ_0; (iii) it can lie after λ_0. Thus $D^*(\lambda)$ takes on three different expressions, depending on this location. In all cases, $D_T^*(\lambda)$ is split into its two main components: the Parameter Difference and the Centre Matrix. These are separately examined in detail (starting with the Parameter Difference) and their limits established before combining to get the form of $D^*(\lambda)$ for each of the cases.

Case 1: When $\lambda = \lambda_0$

This is a straightforward case since we get the structures of the individual parameters directly from (A.1) as

$$\hat{\theta}_1(\lambda_0) = \mathcal{H}_{1,T}(\lambda_0) T^{-1} \sum_{t=1}^{k_0} z_t y_t, \tag{B.1}$$

and

$$\hat{\theta}_2(\lambda_0) = \mathcal{H}_{2,T}(\lambda_0) T^{-1} \sum_{t=k_0+1}^{T} z_t y_t, \tag{B.2}$$

for Subsamples 1 and 2 respectively, where $\mathcal{H}_{i,T}(\lambda_0) = M_{i,T}(\lambda_0)^{-1} G_{i,T}(\lambda_0)' W_{i,T}(\lambda_0)$ for $i = 1, 2$; $y_t = x_t' \theta_0^{(1)} + u_t$ in (B.1) and $y_t = x_t' \theta_0^{(2)} + u_t$ in (B.2).

Since $\mathcal{H}_{1,T}(\lambda_0) T^{-1} \sum_{t=1}^{k_0} z_t u_t$ and $\mathcal{H}_{2,T}(\lambda_0) T^{-1} \sum_{t=k_0+1}^{T} z_t u_t$, are negligible based on

Assumptions 1, 3 and 4 to 8, this implies,

$$\hat{\theta}_i(\lambda_0) = \theta_0^{(i)} + o_p(1), \quad for i = 1, 2. \tag{B.3}$$

Therefore, to obtain the limit of the Parameter Difference when $\lambda = \lambda_0$, we write

$$\hat{\theta}_1(\lambda_0) - \hat{\theta}_2(\lambda_0) \xrightarrow{p} \theta_0^{(1)} - \theta_0^{(2)}. \tag{B.4}$$

For the Centre Matrix on the other hand, $M_{*,T}(\lambda_0) = M_{1,T}(\lambda_0)^{-1} + M_{2,T}(\lambda_0)^{-1}$ and from Assumptions 1, 3, 5 and 8,

$$\begin{aligned} M_{1,T}(\lambda_0) &= G_{1,T}(\lambda_0)' W_{1,T}(\lambda_0) G_{1,T}(\lambda_0) \\ &\xrightarrow{p} (\lambda_0 Q_{xz})(\lambda_0^{-1} C)(\lambda_0 Q_{zx}) \\ &= \lambda_0 (Q_{xz} C Q_{zx}). \end{aligned}$$

Similarly, $M_{2,T}(\lambda_0) = G_{2,T}(\lambda_0)' W_{2,T}(\lambda_0) G_{2,T}(\lambda_0) \xrightarrow{p} (1 - \lambda_0)(Q_{xz} C Q_{zx})$.Therefore,

$$M_{*,T}(\lambda_0)^{-1} \xrightarrow{p} \lambda_0(1 - \lambda_0)\mathcal{B}, \tag{B.5}$$

where $\mathcal{B} = Q_{xz} C Q_{zx}$. Combining the RHS of (B.4) and (B.5), we get that when $\lambda = \lambda_0$, the nonstochastic limit of the test statistic is given by the following expression,

$$D^*(\lambda_0) = \lambda_0(1 - \lambda_0)\mathcal{P}, \tag{B.6}$$

where $\mathcal{P} = (\theta_0^{(1)} - \theta_0^{(2)})' \mathcal{B} (\theta_0^{(1)} - \theta_0^{(2)})$.

Case 2: When $\lambda < \lambda_0$

The parameter obtained from Subsample 1 is from a stable model and hence it is still estimated at its true value,

$$\hat{\theta}_1(\lambda) = \mathcal{H}_{1,T}(\lambda) T^{-1} \sum_{t=1}^{k} z_t y_t \xrightarrow{p} \theta_0^{(1)}, \tag{B.7}$$

since $y_t = x_t'\theta_0^{(1)} + u_t$. Conversely, Subsample 2 is unstable and the estimated parameter is a weighted average of the two parameters[1]. This is given as,

$$\hat{\theta}_2(\lambda) = \mathcal{H}_{2,T}(\lambda)T^{-1}\sum_{t=k+1}^{T} z_t y_t, \tag{B.8}$$

where $T^{-1}\sum_{t=k+1}^{T} z_t y_t = T^{-1}\sum_{t=k+1}^{k_0} z_t y_t + T^{-1}\sum_{t=k_0+1}^{T} z_t y_t$.
Based on Assumptions 1, 3, 5 and 8,

$$\begin{aligned}
\mathcal{H}_{2,T}(\lambda) &= M_{2,T}(\lambda)^{-1}G_{2,T}(\lambda)'W_{2,T}(\lambda) \\
&= \big(G_{2,T}(\lambda)'W_{2,T}(\lambda)G_{2,T}(\lambda)\big)^{-1}G_{2,T}(\lambda)'W_{2,T}(\lambda) \\
&\xrightarrow{p} \big((1-\lambda)Q_{xz}(1-\lambda)^{-1}C(1-\lambda)Q_{zx}\big)^{-1}(1-\lambda)Q_{xz}(1-\lambda)^{-1}C \\
&= \{(1-\lambda)(Q_{xz}CQ_{zx})\}^{-1}Q_{xz}C.
\end{aligned}$$

Also, $T^{-1}\sum_{t=k+1}^{k_0} z_t y_t \xrightarrow{p} (\lambda_0 - \lambda)Q_{zx}\theta_0^{(1)}$ and $T^{-1}\sum_{t=k_0+1}^{T} z_t y_t \xrightarrow{p} (1-\lambda_0)Q_{zx}\theta_0^{(2)}$.
Thus we conclude that $\hat{\theta}_2(\lambda) \xrightarrow{p} \theta_*^{(2)}(\lambda)$, where

$$\theta_*^{(2)}(\lambda) = \mathcal{H}_2(\lambda)\Psi^{b1}, \tag{B.9}$$

$\mathcal{H}_2(\lambda) = \{(1-\lambda)(Q_{xz}CQ_{zx})\}^{-1}Q_{xz}C$ and $\Psi^{b1} = (\lambda_0 - \lambda)Q_{zx}\theta_0^{(1)} + (1-\lambda_0)Q_{zx}\theta_0^{(2)}$.
Notice (B.9) can be simplified further to,

$$\theta_*^{(2)}(\lambda) = \frac{1}{1-\lambda}\big\{(\lambda_0 - \lambda)\theta_0^{(1)} + (1-\lambda_0)\theta_0^{(2)}\big\}. \tag{B.10}$$

Combining $\theta_*^{(1)}(\lambda)$ and $\theta_*^{(2)}(\lambda)$ given in (B.7) and (B.10) respectively, we get the limit of the Parameter Difference,

$$\begin{aligned}
\theta_*^{(1)}(\lambda) - \theta_*^{(2)}(\lambda) &= \frac{1-\lambda}{1-\lambda}\theta_0^{(1)} - \frac{\lambda_0 - \lambda}{1-\lambda}\theta_0^{(1)} - \frac{1-\lambda_0}{1-\lambda}\theta_0^{(2)} \\
&= \frac{1-\lambda_0}{1-\lambda}\theta_0^{(1)} - \frac{1-\lambda_0}{1-\lambda}\theta_0^{(2)} \\
&= \frac{1-\lambda_0}{1-\lambda}(\theta_0^{(1)} - \theta_0^{(2)}).
\end{aligned} \tag{B.11}$$

On the other hand, the structure of the limit of the Centre Matrix does not vary since the Jacobian is stable. A change in the location of λ in the estimations only affects $M_{i,T}(\lambda)$

[1] Elliot and Müller (2014) note same structure while Hall et al. (2012) use similar expressions for the weighted combination of the parameters when they examine the performance of the GMM minimand.

by the number of terms being summed up in $G_{i,T}(\lambda)$ which all have identical limits based on Assumptions 1, 3 and 8. Hence, the Centre Matrix has a similar structure to (B.5) where we only need to replace λ_0 to reflect the location of the candidate break point, that is,

$$M_*(\lambda)^{-1} = \lambda(1-\lambda)\mathcal{B}. \tag{B.12}$$

Combining the limit of the Parameter Difference established on the RHS of (B.11) with that of the Centre Matrix in (B.12), gives the limit of the test statistic when $\lambda < \lambda_0$ as,

$$\begin{aligned} D^*(\lambda) &= \left\{\frac{1-\lambda_0}{1-\lambda}\right\}^2 \lambda(1-\lambda)(\theta_0^{(1)} - \theta_0^{(2)})'\mathcal{B}(\theta_0^{(1)} - \theta_0^{(2)}) \\ &= \frac{\lambda}{1-\lambda}(1-\lambda_0)^2\mathcal{P}. \end{aligned} \tag{B.13}$$

Case 3: When $\lambda > \lambda_0$

This scenario can be considered the reverse of the previous case when $\lambda < \lambda_0$; the second regime is stable, while the first regime now exhibits instability. Consequently, the estimated parameter in Subsample 2, $\hat{\theta}_2(\lambda) \xrightarrow{p} \theta_0^{(2)}$ while that in Subsample 1 is now the weighted average. Performing a similar analysis to Case 2, we write

$$\hat{\theta}_1(\lambda) = \mathcal{H}_{1,T}(\lambda)T^{-1}\sum_{t=1}^{k} z_t y_t, \tag{B.14}$$

where $T^{-1}\sum_{t=1}^{k} z_t y_t = T^{-1}\sum_{t=1}^{k_0} z_t y_t + T^{-1}\sum_{t=k_0+1}^{k} z_t y_t$.
By Assumptions 1, 3, 5 and 8 we obtain,

$$\begin{aligned} \mathcal{H}_{1,T}(\lambda) &= \left(G_{1,T}(\lambda)'W_{1,T}(\lambda)G_{1,T}(\lambda)\right)^{-1}G_{1,T}(\lambda)'W_{1,T}(\lambda) \\ &\xrightarrow{p} \left(\lambda Q_{xz}\lambda^{-1}C\lambda Q_{zx}\right)^{-1}\lambda Q_{xz}\lambda^{-1}C \\ &= \{\lambda(Q_{xz}CQ_{zx})\}^{-1}Q_{xz}C. \end{aligned}$$

Likewise, $T^{-1}\sum_{t=1}^{k_0} z_t y_t \xrightarrow{p} \lambda_0 Q_{xz}\theta_0^{(1)}$ and $T^{-1}\sum_{t=k_0+1}^{k} z_t y_t \xrightarrow{p} (\lambda-\lambda_0)Q_{xz}\theta_0^{(2)}$. Thus, the limit of the parameter before the break,

$$\theta_*^{(1)}(\lambda) = \mathcal{H}_1(\lambda)\Psi^{a1}, \tag{B.15}$$

where $\mathcal{H}_1(\lambda) = \{\lambda(Q_{xz}CQ_{zx})\}^{-1}Q_{xz}C$ and $\Psi^{a1} = \lambda_0 Q_{zx}\theta_0^{(1)} + (\lambda-\lambda_0)Q_{zx}\theta_0^{(2)}$.

The parameter limits on the RHS of (B.9) and (B.15) are used to deduce that the Parameter Difference in this case is,

$$\begin{aligned}\theta_*^{(1)}(\lambda) - \theta_*^{(2)}(\lambda) &= \frac{\lambda_0}{\lambda}\theta_0^{(1)} + \frac{\lambda - \lambda_0}{\lambda}\theta_0^{(2)} - \frac{\lambda}{\lambda}\theta_0^{(2)} \\ &= \frac{\lambda_0}{\lambda}(\theta_0^{(1)} - \theta_0^{(2)}). \end{aligned} \tag{B.16}$$

When combined with the Centre Matrix in (B.12) we obtain the following expression for the limit of the test statistic when $\lambda > \lambda_0$ as

$$\begin{aligned} D^*(\lambda) &= \left\{\frac{\lambda_0}{\lambda}\right\}^2 \lambda(1-\lambda)(\theta_0^{(1)} - \theta_0^{(2)})' \mathcal{B}(\theta_0^{(1)} - \theta_0^{(2)}) \\ &= \frac{1-\lambda}{\lambda}(\lambda_0)^2 \mathcal{P}. \end{aligned} \tag{B.17}$$

B.2 Proof of Lemma 2

This appendix shows that the test statistic, $D_T^*(\lambda)$, is uniformly close to its limit, $D^*(\lambda)$, in all the three cases presented in Appendix B.1. This is achieved by establishing the difference between them is $O_p(T^{-1/2})$ uniformly in λ.

To ease notation, let $\mu_{*,T}(\lambda) = \hat{\theta}_1(\lambda) - \hat{\theta}_2(\lambda)$ and $\mu_*(\lambda) = \theta_*^{(1)}(\lambda) - \theta_*^{(2)}(\lambda)$, such that $\hat{\theta}_i(\lambda) \xrightarrow{p} \theta_*^{(i)}(\lambda)$, then Lemma 2 can be re-expressed as

$$sup_{\lambda\in\Lambda}\left|\mu_{*,T}(\lambda)' M_{*,T}(\lambda)^{-1}\mu_{*,T}(\lambda) - \mu_*(\lambda)' M_*(\lambda)^{-1}\mu_*(\lambda)\right| = O_p(T^{-1/2}). \tag{B.18}$$

By adding and subtracting $\mu_{*,T}(\lambda)' M_{*,T}(\lambda)^{-1}\mu_*(\lambda)$ and $\mu_{*,T}(\lambda)' M_*(\lambda)^{-1}\mu_*(\lambda)$, we get

$$\mu_{*,T}(\lambda)' M_{*,T}(\lambda)^{-1}\mu_{*,T}(\lambda) - \mu_*(\lambda)' M_*(\lambda)^{-1}\mu_*(\lambda) = \breve{\mu}_*^T + \breve{M}_*^T,$$

where

$$\breve{\mu}_*^T = \mu_{*,T}(\lambda)' M_{*,T}(\lambda)^{-1}[\mu_{*,T}(\lambda) - \mu_*(\lambda)] + \mu_*(\lambda)' M_*(\lambda)^{-1}[\mu_{*,T}(\lambda) - \mu_*(\lambda)] \tag{B.19}$$

$$\breve{M}_*^T = \mu_{*,T}(\lambda)'[M_{*,T}(\lambda)^{-1} - M_*(\lambda)^{-1}]\mu_*(\lambda). \tag{B.20}$$

Notice the Parameter Differences, $\mu_{*,T}(\cdot)$ and $\mu_*(\cdot)$ are $O_p(1)$ and $O(1)$, respectively,

uniformly in λ because the parameters themselves are bounded. Likewise, the Centre Matrices, $M_{*,T}(\cdot)$ and $M_*(\cdot)$, since their individual components, $G_{i,T}(\cdot)$ and $W_{i,T}(\cdot)$ are bounded based on Assumptions 1, 3 and 8. We now need to show the terms in the square brackets on the RHS of B.19 and B.20 are $O_p(T^{-1/2})$ uniformly in λ. Starting with $\breve{M}_*^T$, we write[2]

$$M_{*,T}(\lambda)^{-1} - M_*(\lambda)^{-1} = M_*(\lambda)^{-1}[M_*(\lambda) - M_{*,T}(\lambda)]M_{*,T}(\lambda)^{-1}, \tag{B.21}$$

and expanding the terms in the square brackets we obtain,

$$\begin{aligned} M_*(\lambda) - M_{*,T}(\lambda) &= \{M_1(\lambda)^{-1} - M_{1,T}(\lambda)^{-1}\} + \{M_2(\lambda)^{-1} - M_{2,T}(\lambda)^{-1}\} \\ &= M_{1,T}(\lambda)^{-1}[M_{1,T}(\lambda) - M_1(\lambda)]M_1(\lambda)^{-1} \qquad \text{(B.22)} \\ &\quad + M_{2,T}(\lambda)^{-1}[M_{2,T}(\lambda) - M_2(\lambda)]M_2(\lambda)^{-1}. \qquad \text{(B.23)} \end{aligned}$$

As defined, the terms in the square brackets in (B.22),

$$M_{1,T}(\lambda) - M_1(\lambda) = G_{1,T}(\lambda)^{'}W_{1,T}(\lambda)G_{1,T}(\lambda) - G_1(\lambda)^{'}W_1(\lambda)G_1(\lambda),$$

where we have used $G_1(\lambda) = \lambda Q_{zx}$. Adding and subtracting $G_{1,T}(\lambda)^{'}W_{1,T}(\lambda)G_1(\lambda)$ and $G_{1,T}(\lambda)^{'}W_1(\lambda)G_1(\lambda)$, we get

$$\begin{aligned} M_{1,T}(\lambda) - M_1(\lambda) &= G_{1,T}(\lambda)^{'}W_{1,T}(\lambda)[G_{1,T}(\lambda) - G_1(\lambda)] \\ &\quad + G_{1,T}(\lambda)^{'}[W_{1,T}(\lambda) - W_1(\lambda)]G_1(\lambda) \\ &\quad + [G_{1,T}(\lambda) - G_1(\lambda)]^{'}W_1(\lambda)G_1(\lambda). \end{aligned}$$

Now consider the terms in the square brackets which apply to the Jacobian, G_1 and the weighting matrix, W_1. Without loss of generality, we state for the former,

$$\begin{aligned} G_{1,T}(\lambda) - G_1(\lambda) &= T^{-1}\sum_{t=1}^{[T\lambda]} z_t x_t^{'} - \lambda Q_{zx} \qquad \text{(B.24)} \\ &= T^{-1}\sum_{t=1}^{[T\lambda]} z_t x_t^{'} - T^{-1}\sum_{t=1}^{[T\lambda]} Q_{zx} \qquad \text{(B.25)} \\ &\quad + T^{-1}\sum_{t=1}^{[T\lambda]} Q_{zx} - \lambda Q_{zx}. \qquad \text{(B.26)} \end{aligned}$$

[2]Using matrix rule $A^{-1} - B^{-1} = B^{-1}[B - A]A^{-1}$.

Define $\mathcal{H}_{zx}(\lambda) = T^{-1/2}\sum_{t=1}^{[T\lambda]}(z_t x_t^{'} - Q_{zx})$, which is $O_p(1)$ by the Functional Central Limit Theorem[3] (FCLT) and Assumptions 1, 3 and 8, then it must be that the RHS of (B.25) is $O_p(T^{-1/2})$ uniformly in λ.

The RHS of (B.26) on the other hand is asymptotically negligible because $T^{-1}\sum_{t=1}^{[T\lambda]} Q_{zx} - \lambda Q_{zx} = \dfrac{[T\lambda]}{T}Q_{zx} - \lambda Q_{zx} \xrightarrow{p} 0$ as $T \to \infty$. We thus conclude $G_{1,T}(\lambda) - G_1(\lambda)$ is $O_p(T^{-1/2})$ uniformly in λ.

The difference in the weighting matrix is quite straight forward because from construction, $W_{1,T}(\lambda) = W_1(\lambda) = \dfrac{1}{\lambda}C$, where C is a constant matrix. Thus, $W_{1,T}(\lambda) - W_1(\lambda) = 0$.

A similar evaluation can be performed for the terms in the square brackets on the RHS of (B.23) to show it is also $O_p(T^{-1/2})$. Therefore we conclude $\breve{M}_*^T$ given in (B.20) is $O_p(T^{-1/2})$ uniformly in λ.

The proof for $\breve{\mu}_*^T$ in (B.19) is comparatively more involving as it involves the parameters which are dependent on the location of the candidate and true break fractions; hence the analysis covers the three different cases associated with a one break model introduced in Appendix B.1. We show that multiplying the Parameter Differences by $T^{1/2}$ in each of the three cases results in $O_p(1)$ terms and the results follows.

When $\lambda = \lambda_0$,

$$T^{1/2}\big(\mu_{*,T}(\lambda_0) - \mu_*(\lambda_0)\big) = T^{1/2}\{(\hat{\theta}_1(\lambda_0) - \theta_0^{(1)}) - (\hat{\theta}_2(\lambda_0) - \theta_0^{(2)})\}.$$

From $\hat{\theta}_1(\lambda_0)$ given in (B.1) and (B.3), we state

$$T^{1/2}(\hat{\theta}_1(\lambda_0) - \theta_0^{(1)}) = \mathcal{H}_{1,T}(\lambda_0)\mathcal{H}_{zu}(\lambda_0), \tag{B.27}$$

where $\mathcal{H}_{zu}(\lambda_0) = T^{-1/2}\sum_{t=1}^{k_0} z_t u_t \xrightarrow{d} \Omega^{1/2}B_q(\lambda_0)$ by the FCLT; $\Omega = Var[T^{-1/2}\sum_{t=1}^{k_0} z_t u_t]$ and B_q is a $q-$dimensional Brownian motion. Thus, the LHS of (B.27) is bounded in probability. An identical analysis yields a similar outcome for the second parameter $(\hat{\theta}_2(\lambda_0) - \theta_0^{(2)})$. We conclude that in this case when $\lambda = \lambda_0$, then $\breve{\mu}_*^T$ as given in (B.19) is $O_p(T^{-1/2})$.

When $\lambda < \lambda_0$,

From the limits of $\hat{\theta}_1(\lambda)$ established in (B.7), the Parameter Differences for the parameters before the break are similar to the RHS of (B.27), but holds uniformly for all $\lambda < \lambda_0$. That

[3] See page 176 White (2001).

is,

$$T^{1/2}(\hat{\theta}_1(\lambda) - \theta_0^{(1)}) = \mathcal{H}_{1,T}(\lambda)\mathcal{H}_{zu}(\lambda), \tag{B.28}$$

For the parameters after the break on the other hand, we use $\hat{\theta}_2(\lambda)$ in (B.8) and $\theta_*^{(2)}(\lambda)$ in (B.9) to deduce that for all $\lambda > \lambda_0$,

$$\begin{aligned}
&T^{1/2}(\hat{\theta}_2(\lambda) - \theta_*^{(2)}(\lambda)) \\
&\quad = \mathcal{H}_{2,T}(\lambda)\Big\{T^{-1/2}\sum_{t=k+1}^{T} z_t u_t + T^{-1/2}\sum_{t=k+1}^{k_0} z_t x_t^{'}\theta_0^{(1)} + T^{-1/2}\sum_{t=k_0+1}^{T} z_t x_t^{'}\theta_0^{(2)}\Big\} \\
&\qquad - T^{1/2}\theta_*^{(2)}(\lambda) \\
&\quad = \mathcal{H}_{2,T}(\lambda)\Big\{T^{-1/2}\sum_{t=k+1}^{T} z_t u_t \\
&\qquad + T^{-1/2}\sum_{t=k+1}^{k_0}(z_t x_t^{'} - Q_{zx})\theta_0^{(1)} + T^{-1/2}\sum_{t=k_0+1}^{T}(z_t x_t^{'} - Q_{zx})\theta_0^{(2)}\Big\} \\
&\qquad + \mathcal{H}_{2,T}(\lambda)\Big\{T^{-1/2}\sum_{t=k+1}^{k_0} Q_{zx}\theta_0^{(1)} + T^{-1/2}\sum_{t=k_0+1}^{T} Q_{zx}\theta_0^{(2)}\Big\} - T^{1/2}\mathcal{H}_2(\lambda)\Psi^{b1} \\
&\quad = \mathcal{H}_{2,T}(\lambda)[\mathcal{H}_{zu}(1) - \mathcal{H}_{zu}(\lambda)] && \text{(B.29)} \\
&\qquad + \mathcal{H}_{2,T}(\lambda)\big\{[\mathcal{H}_{zx}(\lambda_0) - \mathcal{H}_{zx}(\lambda)]\theta_0^{(1)} + [\mathcal{H}_{zx}(1) - \mathcal{H}_{zx}(\lambda_0)]\theta_0^{(2)}\big\} && \text{(B.30)} \\
&\qquad + T^{1/2}[\mathcal{H}_{2,T}(\lambda) - \mathcal{H}_2(\lambda)]\Psi^{b1} + T^{1/2}\zeta^{b1}, && \text{(B.31)}
\end{aligned}$$

where $\zeta^{b1} = \mathcal{H}_{2,T}(\lambda)\{T^{-1}\sum_{t=k+1}^{k_0} Q_{zx}\theta_0^{(1)} + T^{-1}\sum_{t=k_0+1}^{T} Q_{zx}\theta_0^{(2)} - \Psi^{b1}\}$,
$\mathcal{H}_{zx}(1) = T^{-1/2}\sum_{t=1}^{T}(z_t x_t^{'} - Q_{zx})$ and $\mathcal{H}_{zu}(1) = T^{-1/2}\sum_{t=1}^{T} z_t u_t$.
The terms on the RHS of (B.29) and (B.30) are $O_p(1)$ uniformly in λ. Similarly, (B.31) is also $O_p(1)$ because

$$\begin{aligned}
\mathcal{H}_{2,T}(\lambda) - \mathcal{H}_2(\lambda) &= M_{2,T}(\cdot)^{-1}G_{2,T}(\cdot)^{'}W_{2,T}(\cdot) - M_2(\cdot)^{-1}G_2(\cdot)^{'}W_2(\cdot) \\
&= M_{2,T}(\cdot)^{-1}[G_{2,T}(\cdot)^{'} - G_2(\cdot)^{'}]W_{2,T}(\cdot) \\
&\quad + [M_{2,T}(\cdot)^{-1} - M_2(\cdot)^{-1}]G_2(\cdot)^{'}W_{2,T}(\cdot) \\
&\quad + M_2(\cdot)^{-1}G_2(\cdot)^{'}[W_{2,T}(\cdot) - W_2(\cdot)],
\end{aligned}$$

which is $O_p(T^{-1/2})$ by similar arguments to (B.24). Lastly, ζ^{b1} is $o_p(1)$ because

$$\begin{aligned}\zeta^{b1} &= \mathcal{H}_{2,T}(\lambda)\Big\{\frac{T[\lambda_0-\lambda]}{T}Q_{zx}-(\lambda_0-\lambda)Q_{zx}\Big\}\theta_0^{(1)}\\ &\quad+\mathcal{H}_{2,T}(\lambda)\Big\{\frac{T[1-\lambda_0]}{T}Q_{zx}-(1-\lambda_0)Q_{zx}\Big\}\theta_0^{(2)}\\ &\xrightarrow{p} 0, \quad \text{uniformly in } \lambda \text{ as } T\to\infty.\end{aligned}$$

When $\lambda > \lambda_0$

In this case, since the parameters after the break point are stable, we focus only on the parameters from the first subsample. Using $\hat{\theta}_1(\lambda)$ in (B.14) and $\theta_*^{(1)}(\lambda)$ in (B.15), we perform an identical analysis to the preceeding case to obtain,

$$\begin{aligned}&T^{1/2}(\hat{\theta}_1(\lambda)-\theta_*^{(1)}(\lambda))\\ &=\mathcal{H}_{1,T}(\lambda)\Big\{T^{-1/2}\sum_1^k z_tu_t+T^{-1/2}\sum_1^{k_0}(z_tx_t'-Q_{zx})\theta_0^{(1)}+T^{-1/2}\sum_{k_0+1}^{k}(z_tx_t'-Q_{zx})\theta_0^{(2)}\Big\}\\ &\quad+\mathcal{H}_{1,T}(\lambda)\Big\{T^{-1/2}\sum_1^{k_0}Q_{zx}\theta_0^{(1)}+T^{-1/2}\sum_{k_0+1}^{k}Q_{zx}\theta_0^{(2)}\Big\}-T^{-1/2}\mathcal{H}_1(\lambda)\Psi^{a1}\\ &=\mathcal{H}_{1,T}(\lambda)\big\{\mathcal{H}_{zu}(\lambda)+\mathcal{H}_{zx}(\lambda_0)\theta_0^{(1)}+[\mathcal{H}_{zx}(\lambda)-\mathcal{H}_{zx}(\lambda_0)]\theta_0^{(2)}\big\}\\ &\quad+T^{1/2}[\mathcal{H}_{1,T}(\lambda)-\mathcal{H}_1(\lambda)]\Psi^{a1}+T^{1/2}\zeta^{a1},\end{aligned}\tag{B.32}$$

where $\zeta^{a1}=\mathcal{H}_{1,T}(\lambda)\{T^{-1}\sum_1^{k_0}Q_{zx}\theta_0^{(1)}+T^{-1}\sum_{k_0+1}^{k}Q_{zx}\theta_0^{(2)}-\Psi^{a1}\}$, which is $o_p(1)$. This concludes the proof.

B.3 Proof of Lemma 3

Using the limits of the test statistic established in (B.6), (B.13) and (B.17), an expression for $\mathcal{E}$ is obtained and shown to be positive for all $\lambda \neq \lambda_0$.

When $\lambda < \lambda_0$,

Use $D^*(\lambda)$ in (B.13) and multiply $D^*(\lambda_0)$ in (B.6) by $(1-\lambda_0)/(1-\lambda_0)$ to get,

$$\begin{aligned} D^*(\lambda) - D^*(\lambda_0) &= \Big(\lambda\frac{(1-\lambda_0)^2}{1-\lambda} - \lambda_0\frac{(1-\lambda_0)^2}{1-\lambda_0}\Big)\mathcal{P} \\ &\leq (\lambda-\lambda_0)\frac{(1-\lambda_0)^2}{(1-\lambda_0)}\mathcal{P} \qquad \text{(B.33)} \\ &= -|\lambda-\lambda_0|(1-\lambda_0)\mathcal{P} \\ &= -\mathcal{E}|\lambda-\lambda_0|, \qquad \text{(B.34)} \end{aligned}$$

where $\mathcal{E} = (1-\lambda_0)\mathcal{P}$, $\mathcal{P} = (\theta_0^{(1)} - \theta_0^{(2)})'\mathcal{B}(\theta_0^{(1)} - \theta_0^{(2)})$ and $\mathcal{B} = Q_{xz}CQ_{zx}$. Since $\mathcal{P}$ is positive and $\lambda_0 < 1$, then it must be that $\mathcal{E} > 0$.

When $\lambda > \lambda_0$,
Multiply $D^*(\lambda_0)$ in (B.6) by $(\lambda_0)^2/(\lambda_0)^2$ and $D^*(\lambda)$ in (B.17) by λ/λ to get,

$$\begin{aligned} D^*(\lambda) - D^*(\lambda_0) &= \Big(\lambda\frac{(1-\lambda)}{\lambda^2}(\lambda_0)^2 - \lambda_0\frac{(1-\lambda_0)}{(\lambda_0)^2}(\lambda_0)^2\Big)\mathcal{P} \\ &\leq (\lambda_0-\lambda)\frac{(1-\lambda_0)}{(\lambda_0)^2}(\lambda_0)^2\mathcal{P} \qquad \text{(B.35)} \\ &= -|\lambda_0-\lambda|(1-\lambda_0)\mathcal{P} \\ &= -\mathcal{E}|\lambda_0-\lambda|. \qquad \text{(B.36)} \end{aligned}$$

B.4 Proof of Lemma 4

The analysis in this appendix is carried out with reference to the break point, k, rather than the break fraction, λ. This does not change the fundamental structure of the previous analysis since $\lambda = k/T$. We express the test statistic as

$$D_T(k) = \big(\hat{\theta}_{1,T}(k) - \hat{\theta}_{2,T}(k)\big)'\tilde{M}_{*,T}(k)^{-1}\big(\hat{\theta}_{1,T}(k) - \hat{\theta}_{2,T}(k)\big),$$

where $\tilde{M}_{*,T}(k) = \tilde{M}_{1,T}(k)^{-1} + \tilde{M}_{2,T}(k)^{-1}$; $\tilde{M}_{1,T}(k) = \sum_{t=1}^{k} x_t z_t' \tilde{W}_{1,T}(k) \sum_{t=1}^{k} z_t x_t'$; $\tilde{M}_{2,T}(k) = \sum_{t=k+1}^{T} x_t z_t' \tilde{W}_{2,T}(k) \sum_{t=k+1}^{T} z_t x_t'$; $\tilde{W}_{1,T}(k) = \frac{1}{k}C$ and $\tilde{W}_{2,T}(k) = \frac{1}{T-k}C$. We insert '˜' above the terms to differentiate them from those used in previous appendices; the distinction being the scaling factor T^{-1}. Furthermore, without loss of generality, we set $C = I_q$ for this proof, where I_q is an identity matrix.

The proof is carried out in three steps. Firstly, new expressions for the test statistics,

$D_T(k)$ and $D_T(k_0)$ are obtained. Secondly, the difference between these two test statistics are separated into two distinct parts - a stochastic and a nonstochastic component. Thirdly, Lemma 4 is stated in terms of the stochastic component only in such a way that to complete the proof we only need to show that this stochastic component is asymptotically negligible.

The following proof covers only the case when $\lambda < \lambda_0$ but as highlighted in earlier appendices, the analysis for the case of $\lambda > \lambda_0$ follows a similar argument. Hence it is omitted.

To obtain the expressions for $D_T(\cdot)$ used in this proof, we start with the Parameter Difference. For the estimated parameter from the first subsample,

$$\hat{\theta}_1(k) = \theta_0^{(1)} + \tilde{\mathcal{H}}_{1,T}(k) \sum_{1}^{k} z_t u_t, \tag{B.37}$$

where $\tilde{\mathcal{H}}_{1,T}(k) = \tilde{M}_{1,T}(k)^{-1} \sum_{t=1}^{k} x_t z_t' \tilde{W}_{1,T}(k)$. With some adjustments, the parameter from the second subsample can be expressed as,

$$\begin{aligned}
\hat{\theta}_2(k) &= \tilde{\mathcal{H}}_{2,T}(k) \left\{ \sum_{t=k+1}^{k_0} z_t x_t' \theta_0^{(1)} + \sum_{t=k_0+1}^{T} z_t x_t' \theta_0^{(2)} + \sum_{t=k+1}^{T} z_t u_t \right\} \\
&= \tilde{\mathcal{H}}_{2,T}(k) \left\{ \sum_{t=k+1}^{k_0} z_t x_t' \theta_0^{(1)} + \sum_{t=k_0+1}^{T} z_t x_t' \theta_0^{(1)} \right. \\
&\qquad \left. - \sum_{t=k_0+1}^{T} z_t x_t' \theta_0^{(1)} + \sum_{t=k_0+1}^{T} z_t x_t' \theta_0^{(2)} + \sum_{t=k+1}^{T} z_t u_t \right\} \\
&= \theta_0^{(1)} - \tilde{\mathcal{H}}_{2,T}(k) \sum_{t=k_0+1}^{T} z_t x_t' \left(\theta_0^{(1)} - \theta_0^{(2)}\right) + \tilde{\mathcal{H}}_{2,T}(k) \sum_{t=k+1}^{T} z_t u_t,
\end{aligned} \tag{B.38}$$

where $\tilde{\mathcal{H}}_{2,T}(k) = \tilde{M}_{2,T}(k)^{-1} \sum_{t=k+1}^{T} x_t z_t' \tilde{W}_{2,T}(k)$. From (B.37) and (B.38), the Parameter Difference is,

$$\hat{\theta}_1(k) - \hat{\theta}_2(k) = A(k) + B(k),$$

where

$$A(k) = \tilde{\mathcal{H}}_{2,T}(k) \sum_{t=k_0+1}^{T} z_t x_t^{'}\big(\theta_0^{(1)} - \theta_0^{(2)}\big) \tag{B.39}$$

$$B(k) = B_1(k) - B_2(k) \tag{B.40}$$

$$B_1(k) = \tilde{\mathcal{H}}_{1,T}(k) \sum_{t=1}^{k} z_t u_t \tag{B.41}$$

$$B_2(k) = \tilde{\mathcal{H}}_{2,T}(k) \sum_{t=k+1}^{T} z_t u_t. \tag{B.42}$$

An analogous analysis for the parameter estimated at the true break point yields, $\hat{\theta}_1(k_0) - \hat{\theta}_2(k_0) = A(k_0) + B(k_0)$, where

$$A(k_0) = \big(\theta_0^{(1)} - \theta_0^{(2)}\big) \tag{B.43}$$

and $B(k_0)$ is identical to $B(k)$ in (B.40), albeit evaluated at k_0. Based on $A(k)$ and $B(k)$, the test statistic for all $\lambda < \lambda_0$ is expressed as,

$$\begin{aligned} D_T(k) &= \big(A(k) + B(k)\big)^{'} \tilde{M}_{*,T}(k)^{-1} \big(A(k) + B(k)\big) \\ &= A(k)^{'} \tilde{M}_{*,T}(k)^{-1} A(k) + 2A(k)^{'} \tilde{M}_{*,T}(k)^{-1} B(k) + B(k)^{'} \tilde{M}_{*,T}(k)^{-1} B(k). \end{aligned}$$

The difference between the two test statistics is given as,

$$D_T(k) - D_T(k_0) = \mathcal{A}^*(k, k_0) + h^*(k, k_0), \tag{B.44}$$

where

$$\mathcal{A}^*(k, k_0) = A(k)^{'} \tilde{M}_{*,T}(k)^{-1} A(k) - A(k_0)^{'} \tilde{M}_{*,T}(k_0)^{-1} A(k_0) \tag{B.45}$$

$$h^*(k, k_0) = \mathcal{B}^*(k, k_0) + 2\mathcal{C}^*(k, k_0) \tag{B.46}$$

$$\mathcal{B}^*(k, k_0) = B(k)^{'} \tilde{M}_{*,T}(k)^{-1} B(k) - B(k_0)^{'} \tilde{M}_{*,T}(k_0)^{-1} B(k_0) \tag{B.47}$$

$$\mathcal{C}^*(k, k_0) = A(k)^{'} \tilde{M}_{*,T}(k)^{-1} B(k) - A(k_0)^{'} \tilde{M}_{*,T}(k_0)^{-1} B(k_0). \tag{B.48}$$

We refer to $\mathcal{A}^*(k, k_0)$ and $h^*(k, k_0)$ as the nonstochastic and stochastic parts respectively. From Lemmas 2 and 3, $\mathcal{A}^*(k, k_0)$ is maximised near k_0.

Before we move on to the third step of the proof process, we mention two things. First,

we define

$$\gamma(k) = \frac{A(k_0)'\tilde{M}_{*,T}(k_0)^{-1}A(k_0) - A(k)'\tilde{M}_{*,T}(k)^{-1}A(k)}{|k_0 - k|} \tag{B.49}$$

and note that by way of its construction, $\gamma(k) > 0$ for all $\lambda < \lambda_0$.

Second, notice that $\gamma(k) = -\big(\mathcal{A}^*(k, k_0)/|k_0 - k|\big)$ and hence, the RHS of (B.44) can also be expressed as,

$$D_T(k) - D_T(k_0) = -|k_0 - k|\gamma(k) + h^*(k, k_0). \tag{B.50}$$

Noting Remark 2 on page 34, $D_T(\hat{k}) \geq D_T(k_0)$ by virtue of the test statistic being the maximum and hence the LHS of (B.50) must be non-negative. This implies that for the RHS of (B.50), it must be that $h^*(k, k_0)/|k_0 - k| \geq \gamma(k)$. We can use this fact to deduce that Lemma 4,

$$P\left(\underset{k\in V_3}{sup}\, D_T(k) \geq D_T(k_0)\right) \leq P\left(\underset{k\in V_3}{sup}\, \frac{h^*(k, k_0)}{|k_0 - k|} \geq \underset{k\in V_3}{inf}\, \gamma(k)\right). \tag{B.51}$$

Expressing Lemma 4 in this way enables us to limit our analysis only to the stochastic component, $h^*(k, k_0)$ which we now show to be asymptotically negligible when divided by $|k_0 - k|$ as long as $k \in V_3$.

Since $inf_{k\in V_3}\gamma(k) > 0$ and bounded from Asssumptions 1, 3, 7 and 8 and Lemmas 2 and 3, it suffices to show that for any fixed $\mathcal{F} > 0$,

$$P\left(\underset{k\in V_3}{sup}\, \frac{h^*(k, k_0)}{|k_0 - k|} \geq \mathcal{F}\right) < \epsilon. \tag{B.52}$$

We now analyse the two terms of $h^*(k, k_0)$ given in (B.46) and show they are very small in probability when divided by $|k_0 - k|$. To ease the analysis, we first obtain the order of magnitudes of the main terms.

Notice that from Assumptions 1 to 4, $\sum_{t=1}^{k} x_t z_t^{'}$ is $O_p(T)$, $\frac{1}{k}\sum_{t=1}^{k} z_t x_t^{'}$ is $O_p(1)$ and

$\sum_{t=1}^{k} z_t u_t$ is $O_p(T^{1/2})$ uniformly in k. Similarly,

$$\tilde{M}_{1,T}(k) = \frac{1}{k}\sum_{t=1}^{k} x_t z_t^{'} I \sum_{t=1}^{k} z_t x_t^{'} = O_p(T) \tag{B.53}$$

$$\tilde{M}_{2,T}(k) = \frac{1}{T-k}\sum_{t=k+1}^{T} x_t z_t^{'} I \sum_{t=k+1}^{T} z_t x_t^{'} = O_p(T) \tag{B.54}$$

$$\tilde{M}_{*,T}(k) = \tilde{M}_{1,T}(k)^{-1} + \tilde{M}_{2,T}(k)^{-1} = O_p(T^{-1}) \tag{B.55}$$

$$\tilde{\mathcal{H}}_{1,T}(k) = \tilde{M}_{1,T}(k)^{-1}\frac{1}{k}\sum_{t=1}^{k} x_t z_t^{'} I = O_p(T^{-1}) \tag{B.56}$$

$$\tilde{\mathcal{H}}_{2,T}(k) = \tilde{M}_{2,T}(k)^{-1}\frac{1}{T-k}\sum_{t=k+1}^{T} x_t z_t^{'} I = O_p(T^{-1}) \tag{B.57}$$

$$A(k) = \tilde{\mathcal{H}}_{2,T}(k)\sum_{t=k_0+1}^{T} z_t x_t^{'}(\theta_0^{(1)} - \theta_0^{(2)}) = O_p(1) \tag{B.58}$$

$$B(k) = \tilde{\mathcal{H}}_{1,T}(k)\sum_{t=1}^{k} z_t u_t - \tilde{\mathcal{H}}_{2,T}(k)\sum_{t=k+1}^{T} z_t u_t = O_p(T^{-1/2}). \tag{B.59}$$

The same order of magnitudes apply to the counterpart terms, when $k = k_0$. Starting with $\mathcal{B}^*(k, k_0)$, we use $\tilde{M}_{*,T}(k)$ in (B.55) and $B(k)$ in (B.59) to deduce

$$B(\cdot)^{'}\tilde{M}_{*,T}(\cdot)^{-1}B(\cdot) = O_p(T^{-1/2})O_p(T)O_p(T^{-1/2}) = O_p(1).$$

Therefore,

$$\frac{\mathcal{B}^*(k, k_0)}{|k_0 - k|} = \frac{O_p(1)}{|k_0 - k|}. \tag{B.60}$$

For the second stochastic component, $\mathcal{C}^*(k, k_0)$ given in (B.48), we rearrange and write,

$$\mathcal{C}^*(k, k_0) = \Upsilon_1(k, k_0) + \Upsilon_2(k, k_0) + \Upsilon_3(k, k_0),$$

where

$$\Upsilon_1(k, k_0) = A(k)^{'}\{\tilde{M}_{*,T}(k)^{-1} - \tilde{M}_{*,T}(k_0)^{-1}\}B(k_0) \tag{B.61a}$$

$$\Upsilon_2(k, k_0) = A(k)^{'}\tilde{M}_{*,T}(k)^{-1}\{B(k) - B(k_0)\} \tag{B.61b}$$

$$\Upsilon_3(k, k_0) = \{A(k) - A(k_0)\}^{'}\tilde{M}_{*,T}(k_0)^{-1}B(k_0). \tag{B.61c}$$

The order of magnitudes for the individual terms are given in (B.56), (B.58) and (B.59), hence we only examine the terms in the curly brackets. Also, note the following remark which we employ subsequently in the proofs.

Remark 3. *In this case when* $\lambda < \lambda_0$*, notice that* $\sum_{t=1}^{k_0}(\ldots) = \sum_{t=1}^{k}(\ldots) + \sum_{t=k+1}^{k_0}(\ldots)$ *and* $\sum_{t=k_0}^{T}(\ldots) = \sum_{t=k}^{T}(\ldots) - \sum_{t=k+1}^{k_0}(\ldots)$.

Starting with $\Upsilon_1(k, k_0)$, we break it down as follows,

$$\tilde{M}_{*,T}(k)^{-1} - \tilde{M}_{*,T}(k_0)^{-1} = \tilde{M}_{*,T}(k_0)^{-1}\{\tilde{M}_{*,T}(k_0) - \tilde{M}_{*,T}(k)\}\tilde{M}_{*,T}(k)^{-1}, \tag{B.62}$$

where the terms within the curly brackets,

$$\tilde{M}_{*,T}(k_0) - \tilde{M}_{*,T}(k) = \{\tilde{M}_{1,T}(k_0)^{-1} - \tilde{M}_{1,T}(k)^{-1}\} + \{\tilde{M}_{2,T}(k_0)^{-1} - \tilde{M}_{2,T}(k)^{-1}\}$$

and for $i = 1, 2$,

$$\tilde{M}_{i,T}(k_0)^{-1} - \tilde{M}_{i,T}(k)^{-1} = \tilde{M}_{i,T}(k_0)^{-1}\{\tilde{M}_{i,T}(k) - \tilde{M}_{i,T}(k_0)\}\tilde{M}_{i,T}(k)^{-1}. \tag{B.63}$$

With some algebra, the term within the curly brackets in (B.63),

$$\begin{aligned}\tilde{M}_{1,T}(k_0) - \tilde{M}_{1,T}(k) &= \sum_{t=1}^{k_0} x_t z_t^{'} \frac{1}{k_0} \sum_{t=1}^{k_0} z_t x_t^{'} - \sum_{t=1}^{k} x_t z_t^{'} \frac{1}{k} \sum_{t=1}^{k} z_t x_t^{'} \\ &= \mathcal{R}_1 + \mathcal{R}_2 + \mathcal{R}_3,\end{aligned}$$

where

$$\mathcal{R}_1 = \frac{1}{k_0} \sum_{t=1}^{k_0} x_t z_t^{'} \sum_{t=k+1}^{k_0} z_t x_t^{'} \tag{B.64a}$$

$$\mathcal{R}_2 = \sum_{t=k+1}^{k_0} x_t z_t^{'} \frac{1}{k} \sum_{t=1}^{k} z_t x_t^{'} \tag{B.64b}$$

$$\mathcal{R}_3 = \frac{1}{k_0} \sum_{t=1}^{k_0} x_t z_t^{'} \frac{1}{k} \sum_{t=1}^{k} z_t x_t^{'}(k - k_0). \tag{B.64c}$$

Similarly,

$$\begin{aligned}\tilde{M}_{2,T}(k_0) - \tilde{M}_{2,T}(k) &= \sum_{t=k_0+1}^{T} x_t z_t^{'} \frac{1}{T-k_0} \sum_{t=k_0+1}^{T} z_t x_t^{'} - \sum_{t=k+1}^{T} x_t z_t^{'} \frac{1}{T-k} \sum_{t=k+1}^{T} z_t x_t^{'} \\ &= \mathcal{S}_1 + \mathcal{S}_2 + \mathcal{S}_3,\end{aligned}$$

where

$$\mathcal{S}_1 = -\frac{1}{T-k_0} \sum_{t=k_0+1}^{T} x_t z_t^{'} \sum_{t=k+1}^{k_0} z_t x_t^{'} \tag{B.65a}$$

$$\mathcal{S}_2 = -\sum_{t=k+1}^{k_0} x_t z_t^{'} \frac{1}{T-k} \sum_{t=k+1}^{T} z_t x_t^{'} \tag{B.65b}$$

$$\mathcal{S}_3 = \frac{1}{T-k_0} \sum_{t=k_0+1}^{T} x_t z_t^{'} \frac{1}{T-k} \sum_{t=k+1}^{T} z_t x_t^{'}(k_0-k). \tag{B.65c}$$

Notice $\mathcal{R}_i$ and $\mathcal{S}_i$ are all $O_p(|k_0-k|)$ for $i=1,2,3$ from the order of magnitudes given in (B.53) and (B.54). It follows from (B.53), (B.54), (B.55), and (B.64a) to (B.65c) that $\tilde{M}_{*,T}(k)^{-1} - \tilde{M}_{*,T}(k_0)^{-1}$ in (B.62) is also $O_p(|k_0-k|)$. Consequently, $\Upsilon_1(k,k_0)$ is $O_p(T^{-1/2})O_p(|k_0-k|)$ and we conclude that,

$$\frac{\Upsilon_1(k,k_0)}{|k_0-k|} = O_p(T^{-1/2}). \tag{B.66}$$

For $\Upsilon_2(k,k_0)$ in (B.61b) on page 53, we state that $B(k) - B(k_0) = \{B_1(k) - B_1(k_0)\} +$

$\{B_2(k) - B_2(k_0)\}$ using (B.40). With some algebra,

$$
\begin{aligned}
B_1(k) - B_1(k_0) &= \{\tilde{M}_{1,T}(k)^{-1} - \tilde{M}_{1,T}(k_0)^{-1}\}\frac{1}{k}\sum_{t=1}^{k} x_t z_t^{'}\sum_{t=1}^{k} z_t u_t \\
&\quad + \tilde{M}_{1,T}(k_0)^{-1}\left\{\frac{1}{k}\sum_{t=1}^{k} x_t z_t^{'} - \frac{1}{k_0}\sum_{t=1}^{k_0} x_t z_t^{'}\right\}\sum_{t=1}^{k} z_t u_t \\
&\quad - \tilde{M}_{1,T}(k_0)^{-1}\frac{1}{k_0}\sum_{t=1}^{k_0} x_t z_t^{'}\sum_{t=k+1}^{k_0} z_t u_t \\
&= \tilde{M}_{1,T}(k_0)^{-1}\{\mathcal{R}_1 + \mathcal{R}_2 + \mathcal{R}_3\}\tilde{M}_{1,T}(k)^{-1}\frac{1}{k}\sum_{t=1}^{k} x_t z_t^{'}\sum_{t=1}^{k} z_t u_t \qquad \text{(B.67)} \\
&\quad + \tilde{M}_{1,T}(k_0)^{-1}\left\{\frac{1}{k}\sum_{t=1}^{k} x_t z_t^{'} - \frac{1}{k_0}\sum_{t=1}^{k_0} x_t z_t^{'}\right\}\sum_{t=1}^{k} z_t u_t \qquad \text{(B.68)} \\
&\quad - \tilde{\mathcal{H}}_{1,T}(k_0)\sum_{t=k+1}^{k_0} z_t u_t. \qquad \text{(B.69)}
\end{aligned}
$$

So also,

$$
\begin{aligned}
&B_2(k) - B_2(k_0) \\
&= \tilde{M}_{2,T}(k_0)^{-1}\{\mathcal{S}_1 + \mathcal{S}_2 + \mathcal{S}_3\}\tilde{M}_{2,T}(k)^{-1}\frac{1}{T-k}\sum_{t=k+1}^{T} x_t z_t^{'}\sum_{t=k+1}^{T} z_t u_t \qquad \text{(B.70)} \\
&\quad + \tilde{M}_{2,T}(k_0)^{-1}\left\{\frac{1}{T-k}\sum_{t=k+1}^{T} x_t z_t^{'} - \frac{1}{T-k_0}\sum_{t=k_0+1}^{T} x_t z_t^{'}\right\}\sum_{t=k+1}^{T} z_t u_t \qquad \text{(B.71)} \\
&\quad + \tilde{\mathcal{H}}_{2,T}(k_0)\sum_{t=k+1}^{k_0} z_t u_t. \qquad \text{(B.72)}
\end{aligned}
$$

From (B.53) and (B.64a)-(B.64c), we deduce the RHS of (B.67) is $O_p(T^{-1})O_p(|k - k_0|)O_p(T^{-1/2})$; and from (B.56), it follows that (B.69) is $O_p(T^{-1})\sum_{t=k+1}^{k_0} z_t u_t$. Similar orders are obtained for their counterparts in (B.70) and (B.72), respectively.

For the terms in the curly brackets in (B.68), we write

$$
\frac{1}{k}\sum_{t=1}^{k} x_t z_t^{'} - \frac{1}{k_0}\sum_{t=1}^{k_0} x_t z_t^{'} = \frac{1}{k}\sum_{t=1}^{k}(x_t z_t^{'} - Q_{xz}) - \frac{1}{k_0}\sum_{t=1}^{k_0}(x_t z_t^{'} - Q_{xz}) \qquad \text{(B.73)}
$$

and show that multiplying by $T^{1/2}$ results in terms that are bounded in probability.

$$T^{1/2}\frac{1}{k_0}\sum_{t=1}^{k_0}(x_t z_t^{'} - Q_{xz}) = \frac{1}{\lambda_0}\Big\{\frac{1}{T^{1/2}}\sum_{t=1}^{k_0}(x_t z_t^{'} - Q_{xz})\Big\} = {\lambda_0}^{-1}\mathcal{H}_{xz}(\lambda_0) \tag{B.74}$$

and

$$T^{1/2}\frac{1}{k}\sum_{t=1}^{k}(x_t z_t^{'} - Q_{xz}) = \lambda^{-1}\mathcal{H}_{xz}(\lambda) \tag{B.75}$$

From Appendix B.2, $\mathcal{H}_{xz}(\cdot)$ is $O_p(1)$ uniformly in λ and consequently we say the RHS of (B.74) and (B.75) are $O_p(T^{-1/2})$ uniformly in λ. Thus, the RHS of (B.68) - and likewise (B.71) - are both $O_p(T^{-1})$.

We now combine (B.67) - (B.72) to obtain

$$\begin{aligned} B(k) - B(k_0) = {} & O_p(T^{-1})O_p(|k_0 - k|)O_p(T^{-1/2}) + O_p(T^{-1}) \\ & + O_p(T^{-1})\sum_{t=k+1}^{k_0} z_t u_t. \end{aligned}$$

With thus conclude,

$$\Upsilon_2(k, k_0) = O_p(|k_0 - k|)O_p(T^{-1/2}) + O_p(1) + O_p(1)\sum_{t=k+1}^{k_0} z_t u_t,$$

where the second $O_p(1)$ term is

$$\Xi_{1,T} = A(k)^{'}\tilde{M}_{*,T}(k)^{-1}[\tilde{\mathcal{H}}_{2,T}(k_0) - \tilde{\mathcal{H}}_{1,T}(k_0)]. \tag{B.76}$$

We state $\Xi_{1,T} = \Xi_1 + O_p(T^{-1/2})$, where $\Xi_1 < \infty$, therefore,

$$\frac{\Upsilon_2(k, k_0)}{|k_0 - k|} = O_p(T^{-1/2}) + \frac{O_p(1)}{|k_0 - k|} + \frac{\Xi_1}{|k_0 - k|}\sum_{t=k+1}^{k_0} z_t u_t + \frac{O_p(T^{-1/2})}{|k_0 - k|}. \tag{B.77}$$

Lastly, for $\Upsilon_3(k,k_0)$ in (B.61c) on page 53, we write

$$A(k) = \tilde{\mathcal{H}}_{2,T}(k)\sum_{t=k_0+1}^{T} z_t x_t^{'}(\theta_0^{(1)} - \theta_0^{(2)})$$

$$= A(k_0) + \{\tilde{\mathcal{H}}_{2,T}(k) - \tilde{\mathcal{H}}_{2,T}(k_0)\}\sum_{t=k_0+1}^{T} z_t x_t^{'}(\theta_0^{(1)} - \theta_0^{(2)}),$$

where

$$\tilde{\mathcal{H}}_{2,T}(k) - \tilde{\mathcal{H}}_{2,T}(k_0) = \{\tilde{M}_{2,T}(k)^{-1} - \tilde{M}_{2,T}(k_0)^{-1}\}\frac{1}{T-k}\sum_{t=k+1}^{T} z_t x_t^{'}$$

$$+ \tilde{M}_{2,T}(k_0)^{-1}\left\{\frac{1}{T-k}\sum_{t=k+1}^{T} z_t x_t^{'} - \frac{1}{T-k_0}\sum_{t=k_0+1}^{T} z_t x_t^{'}\right\}$$

$$= \tilde{M}_{2,T}(k_0)^{-1}\{\mathcal{S}_1 + \mathcal{S}_2 + \mathcal{S}_3\}\tilde{M}_{2,T}(k)^{-1}\frac{1}{T-k}\sum_{t=k+1}^{T} z_t x_t^{'} \tag{B.78}$$

$$+ \tilde{M}_{2,T}(k_0)^{-1}\left\{\frac{1}{T-k}\sum_{t=k+1}^{T} z_t x_t^{'} - \frac{1}{T-k_0}\sum_{t=k_0+1}^{T} z_t x_t^{'}\right\} \tag{B.79}$$

and $\mathcal{S}_i$ are as defined in (B.65a) - (B.65c) for $i = 1,2,3$. In union with (B.54) and (B.65a) - (B.65c), it follows that (B.78) is $O_p(T^{-1})O_p(|k_0-k|)O_p(T^{-1})O_p(1)$. Similarly, combining (B.54) and (B.73) - (B.75), it follows that (B.79) is $O_p(T^{-3/2})$.

Thus, $A(k) - A(k_0)$ is $O_p(T^{-1})O_p(|k_0-k|) + O_p(T^{-1/2})$ and in combination with (B.40) and (B.55), it follows that

$$\frac{\Upsilon_3(k,k_0)}{|k_0-k|} = \frac{O_p(1)}{|k_0-k|} + O_p(T^{-1/2}). \tag{B.80}$$

Combining $\Upsilon_1(k,k_0), \Upsilon_2(k,k_0)$ and $\Upsilon_3(k,k_0)$ in (B.66), (B.77) and (B.80) respectively, provides the order of the second part of the stochastic component in (B.48), that is,

$$\frac{\mathcal{C}^*(k,k_0)}{|k_0-k|} = \frac{\Xi_1}{|k_0-k|}\sum_{t=k+1}^{k_0} z_t u_t + \frac{O_p(1)}{|k_0-k|}, \tag{B.81}$$

where we have left out the $O_p(T^{-1/2})$ terms as they are asymptotically negligible.

Finally, with the orders of $\mathcal{B}^*(k,k_0)$ and $\mathcal{C}^*(k,k_0)$ as given in (B.60) and (B.81) respectively, we conclude the proof for the stochastic component as follows[4],

$$
\begin{aligned}
&P\left(\sup_{k\in V_3}\frac{h^*(k,k_0)}{|k_0-k|}\geq\mathcal{F}\right)\\
&\qquad\leq P\left(\sup_{k\in V_3}\left\{\frac{|O_p(1)|}{|k_0-k|}+2\left|\frac{\Xi_1}{|k_0-k|}\sum_{t=k+1}^{k_0}z_tu_t\right|+2\frac{|O_p(1)|}{|k_0-k|}\right\}\geq\mathcal{F}\right)\\
&\qquad\leq P\left(\left\{\frac{|O_p(1)|}{\mathcal{M}}+2\|\Xi_1\|\sup_{k\in V_3}\left\|\frac{1}{|k_0-k|}\sum_{t=k+1}^{k_0}z_tu_t\right\|+2\frac{|O_p(1)|}{\mathcal{M}}\right\}\geq\mathcal{F}\right)\\
&\qquad\leq \Upsilon_4+\Upsilon_5,
\end{aligned}
$$

where

$$
\Upsilon_4=P\left(\frac{|O_p(1)|}{\mathcal{M}}\geq\frac{\mathcal{F}}{3}\right)+P\left(2\frac{|O_p(1)|}{\mathcal{M}}\geq\frac{\mathcal{F}}{3}\right), \tag{B.82}
$$

$$
\Upsilon_5=P\left(2\|\Xi_1\|\sup_{k\in V_3}\left\|\frac{1}{|k_0-k|}\sum_{t=k+1}^{k_0}z_tu_t\right\|\geq\frac{\mathcal{F}}{3}\right) \tag{B.83}
$$

and we have used $O_p(1)/|k_0-k|\leq O_p(1)/\mathcal{M}$ since $|k_0-k|\geq\mathcal{M}$, where $\mathcal{M}<\infty$ as given in Section 2.4.

Υ_4 is asymptotically negligible as we can get a large enough $\mathcal{M}$ such that the probabilities will be less than $\epsilon/3$. For Υ_5, Wang (2007) illustrates the Hájek and Rényi (1955) inequality is as a result of choosing the correct maximal inequality for a sequence of cumulative sums. We follow identical steps[5] given in Bai (1994a) and state that by Assumption 9, there exists an $L>0$, such that for any $\dfrac{\mathcal{F}}{\|\Xi_1\|}>0$,

$$
P\left(\sup_{k\in V_3}\left\|\frac{1}{|k_0-k|}\sum_{t=k+1}^{k_0}z_tu_t\right\|\geq\frac{\mathcal{F}}{\|\Xi_1\|}\right)<\frac{L}{\mathcal{F}^r\mathcal{M}^{r/2}} \tag{B.84}
$$

which is small for large $\mathcal{M}$.

This concludes the proof of Lemma 4.

[4] To obtain (B.82) and (B.83), we use the fact that for $\alpha>0$ and j positive random variables, ω_j, the $P(\sum_i^j\omega_i>\alpha)\leq\sum_i^j P\big(\omega_i>\alpha/j\big)$ for $i=1,\ldots,j$.

[5] See Lemma A.3 in Bai (1994a). We use identical generalisation of the Hájek and Rényi (1955) inequality in this research. More details can be found in Bai (1994a), Bai (1994b) and Wang (2007).

Chapter 3

Stable Jacobian - Multiple Break Model

In this chapter, linear models with two or more break points are examined. These types of multiple break models are common in economic and financial time series models. For example, to assess the constancy of the ex-ante real interest rate, Garcia and Perron (1996) use a Markov switching autoregressive model to estimate the time path of the ex-post[1] real interest rate over the period 1961 to 1986. Allowing for an arbitrary number of break points at unknown locations in the sample, they conclude that the ex-post real interest rate series is characterised by three states or two break points within the period. The first break in 1973 was attributed to the sudden jump in oil prices while the second break in 1981 was due to the rise in the federal budget deficit.

Similarly, Altissimo and Corradi (2003) propose a sequential testing procedure to obtain consistent estimation of the location and true number of break points in the mean of a single series. Applying their method to the spot Eurodollar rate, they detect five break points in the mean of this process between 1973 and 1995. These break points coincide with the changes in the US monetary policy.

In an attempt to determine the impact of the Fiscal Responsibility Act[2] on the fiscal performance of some countries, Caceres et al. (2010) regress the country's primary fiscal balance on its output gap and commodity price index. Two break points were discovered in the fiscal balance of Australia between the period 1974 to 2009.

Additionally, Banerjee et al. (2002) propose a method of estimating the location of multiple break points by first estimating the breaks in marginal models before imposing these

[1] Garcia and Perron (1996) note that under the assumption that available information is used efficiently, this is equivalent to analysing the ex-ante real interest rate.

[2] Amongst other targets, this act aims at ensuring long term economic stability, prudent management of the nations resources as well as securing accountability and transparency in fiscal operations.

breaks on the conditional model. Using this procedure, they detected five break points in the relationship between money demand, log of real income and real interest rates in the UK between 1969 and 1996.

However, for the purpose of our analysis in this chapter, we assume the true number of break points in the model is known beforehand but their locations are unknown; hence we focus on their estimation. However, Chapter 6 provides a method of determining the true number of breaks in a model.

The outline of the chapter follows a similar sequence as the previous chapter. The multiple break model is first introduced alongside the additional assumptions made. Secondly, the multiple breaks estimation procedure is presented. Thirdly, the asymptotic properties of the test statistic and break fraction estimators are established. Lastly, we conclude and the the proofs are presented in the appendix at the end of the chapter.

3.1 The Model and its Assumptions

The Structural Equation (SE) of the multiple break model used is presented thus,

$$\begin{aligned}
y_t &= x_t^{'}\theta_0^{(1)} + u_t, \quad t = 1, 2, \ldots, k_0^{(1)}, \\
y_t &= x_t^{'}\theta_0^{(2)} + u_t, \quad t = k_0^{(1)} + 1, \ldots, k_0^{(2)}, \\
&\vdots \qquad \vdots \qquad\qquad \vdots \\
y_t &= x_t^{'}\theta_0^{(m+1)} + u_t, \quad t = k_0^{(m)} + 1, \ldots, T,
\end{aligned} \tag{3.1}$$

where m is the total number of breaks in the model and $k_0^{(1)}, \ldots, k_0^{(m)}$ are the m true break points. Correspondingly, the true break fractions are $\lambda_0^{(1)}, \ldots, \lambda_0^{(m)}$. Notice there are $m + 1$ regimes in an m-break model.

The endogenous regressors, x_t are constructed in the same way as in the single break model given in (2.2) on page 27. We state some additional assumptions:

Assumption 10. $\theta_0^{(i)} \neq \theta_0^{(i+1)}$ *for* $i = 1, \ldots, m$.

Assumption 11. $E[z_t u_t(\theta_0)] = 0$, *for* $\theta_0 = (\theta_0^{(i)'}, \ldots, \theta_0^{(i+1)'})'$.

Assumption 12. $0 < \lambda_0^{(1)} < \ldots < \lambda_0^{(m)} < 1$.

Assumption 13. $inf(k_0^{(i+1)} - k_0^{(i)}) \geq max\{q, [T\epsilon]\}$ *for* $\epsilon > 0$ *and* $i = 1, \ldots, m-1$.

Assumption 14. *There are* $m \geq 2$ *number of breaks in the model.*

Assumption 12 allows each break point to be asymptotically distinct and bounded away from the boundaries of the sample while Assumption 13 ensures there are enough observations for estimations in each regime. Similar assumptions are also imposed in Bai (1997a), Bai and Perron (1998) and Hall et al. (2012).

3.2 The Multiple Break Points Estimation Procedure

In estimating multiple break points, it is common to use either the Sequential or the Simultaneous Estimation Method. When using the Simultaneous Estimation Method as introduced in Bai and Perron (1998), all the m break points are estimated at the same time across all the possible regimes in the model. Contrarily, in using the Sequential method as presented in Bai (1997a), the m break points are individually estimated one after the other.

In this section, we base our discussions and theoretical analysis on the Sequential Estimation Method only. However, some Monte Carlo simulation results for the finite sample behaviour of the Simultaneous Estimation Method are presented in Chapter 5. We have demonstrated that results from both methods are comparable.

Since the m breaks are estimated one at a time, the Sequential process discussed here is similar to that used in the single break scenario in Section 2.2; the main difference being the part of the sample used in the estimations. In other words, the multiple break points are obtained by simply estimating single break points repeatedly using different sections of the sample. Hence, the break fraction estimator remains same as given in (2.4) on page 30 albeit over a different range of Λ as discussed in the estimation procedure below.

The multiple break points estimation process is carried out in distinct steps. Each step involves a round of estimations which results in an estimated break point. In the initial step, the first of the m break points is obtained using all the observations in the sample, $[1, T]$, for the estimations. Denote this first estimated break point as $\hat{k}_1$, we show in the next section that it is consistent for one of the true break points.

In the second step, we condition on $\hat{k}_1$ and split the sample into two subsamples, $[1, \hat{k}_1]$ and $[\hat{k}_1 + 1, T]$. The estimation procedure is then repeated separately on these two subsamples and a second break point, $\hat{k}_2$, is estimated as the location of the higher of the *suprema* of $D_T(\cdot)$ obtained from both subsamples.

In the third step, the subsample which nests this second break point - say the second subsample from the first step above - is split into two further subsamples based on $\hat{k}_2$ and

the estimation procedure is again repeated on the three subsamples, $[1, \hat{k}_1]$, $[\hat{k}_1 + 1, \hat{k}_2]$ and $[\hat{k}_2 + 1, T]$, to determine the third break point, $\hat{k}_3$.

Restricted by the minimum number of observations allowed in a regime as given in Assumption 13, the estimation process is repeated until all the m break points in the model have been estimated (if this number is known) or until there are no more significant break points in the model (we discuss this procedure in Chapter 6).

3.3 Consistency and Convergence Rates

The asymptotic properties of the break fraction estimators obtained from a model with multiple breaks in its SE is examined in this subsection. The subsection is divided into two main parts. In the first part, the consistency and convergence rates of the break fraction estimators from a two break model are analysed. Presenting the main arguments and theoretical analysis in terms of a two break model only simplifies the analysis and there is no loss of generality as seen in the second part where we consider a model with more than two break points. The approach and propositions used follow Bai (1997a) closely; the proofs are placed in the relevant appendices.

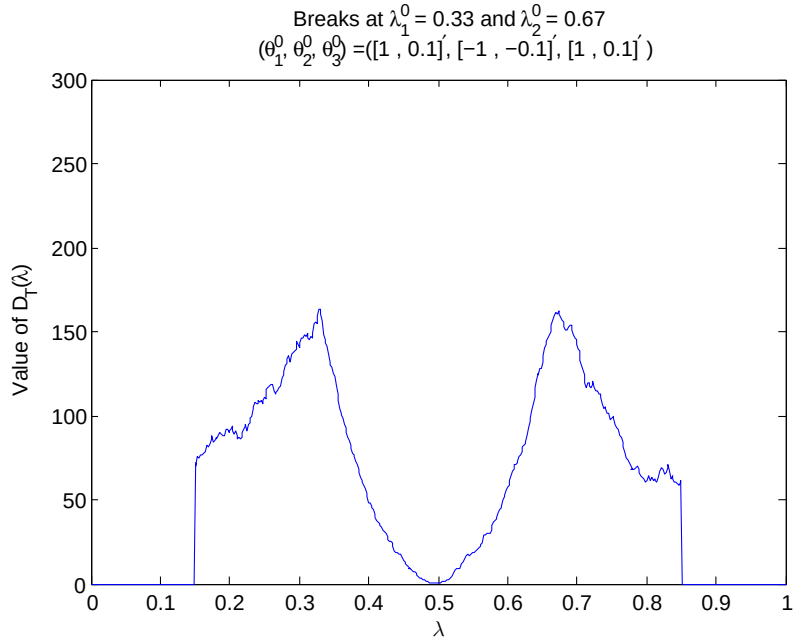
When the Model has Two Breaks; $m = 2$

First, let the two true breaks be fixed at $\lambda_0^{(1)}$ and $\lambda_0^{(2)}$, then the test statistic $D_T(\lambda)$ has two local maxima at these two points where the true breaks are located. To illustrate this in a finite sample, Figure 3.1 shows a single simulation of $D_T(\lambda)$ across all $\lambda \in \Lambda = [0.15, 0.85]$. Using a sample size of 600, the two true breaks are imposed at $\lambda_0^{(1)} = 0.33$ and $\lambda_0^{(2)} = 0.67$. The three true parameter values are $\theta_0^{(i)} = -1^{(i+1)}[1, 0.1]'$ for $i = 1, 2, 3$. The peaks of $D_T(\lambda)$ at these two points is clearly seen even when the simulation is repeated a thousand times as shown in Figure 3.2.

3.3.1 Consistency of the Break Fraction Estimator

To prove the consistency, we follow the three steps used in the one break model in Section 2.4 on page 31. First, the nonstochastic limits of the test statistic are obtained; secondly, the uniform convergence of the test statistic to its nonstochastic limit is established and lastly, these limits are shown to be uniquely maximised at the true break points. These are stated in the following three lemmas:

Figure 3.1: $D_T(\lambda)$ for all $\lambda \in \Lambda$.
DGP has 2 breaks in SE



Lemma 5. *Under Assumptions 1 to 4, 7 to 8 and 10 to 14, $D_T^*(\lambda)$ converges uniformly in probability to a nonstochastic function $D^*(\lambda)$ on $(0, 1)$.*

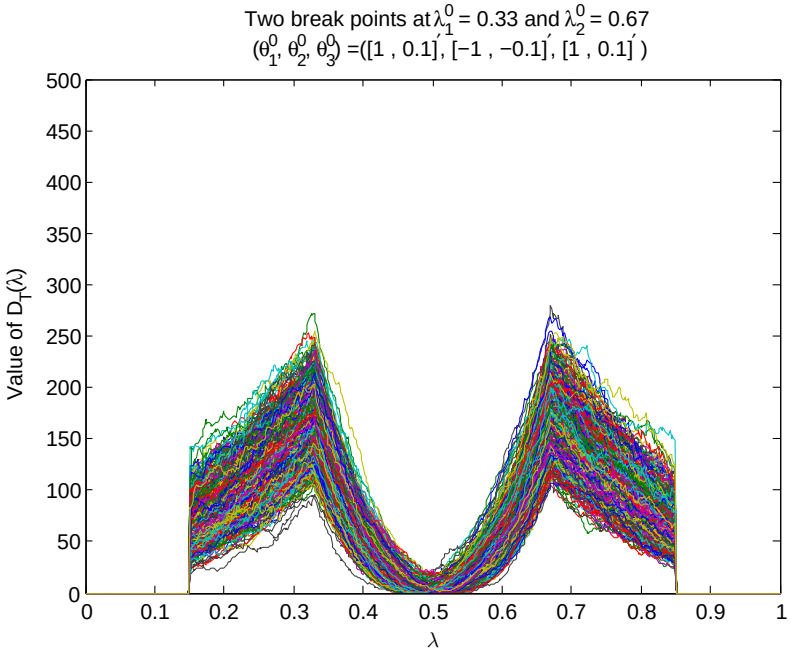
For a model with two breaks, the nonstochastic function $D^*(\lambda)$, exhibits five unique expressions across the range of Λ: one in each of the three regimes described above and one at each of the two true break points. These are examined in Appendix C.1 and given as,

$$D^*(\lambda) = \frac{\lambda}{1-\lambda}\{\tilde{w}_1 + \tilde{w}_2\}'\mathcal{B}\{\tilde{w}_1 + \tilde{w}_2\}, \quad for\ all\ \lambda < \lambda_0^{(1)} \tag{3.2}$$

$$D^*(\lambda_0^{(1)}) = \frac{\lambda_0^{(1)}}{1-\lambda_0^{(1)}}\{\tilde{w}_1 + \tilde{w}_2\}'\mathcal{B}\{\tilde{w}_1 + \tilde{w}_2\} \quad for\ \lambda = \lambda_0^{(1)} \tag{3.3}$$

$$D^*(\lambda) = \frac{1-\lambda}{\lambda}w_1^{*\prime}\mathcal{B}w_1^* + \frac{\lambda}{1-\lambda}\tilde{w}_2'\mathcal{B}\tilde{w}_2 + 2w_1^{*\prime}\mathcal{B}\tilde{w}_2 \ \ for\ all\ \lambda \in (\lambda_0^{(1)},\ \lambda_0^{(2)}) \tag{3.4}$$

Figure 3.2: $D_T(\lambda)$ for all $\lambda \in \Lambda$.
DGP has 2 breaks in SE



$$D^*(\lambda_0^{(2)}) = \frac{1-\lambda_0^{(2)}}{\lambda_0^{(2)}}\{w_1^* + w_2^*\}'\mathcal{B}\{w_1^* + w_2^*\} \quad for\ \lambda = \lambda_0^{(2)} \tag{3.5}$$

$$D^*(\lambda) = \frac{1-\lambda}{\lambda}\{w_1^* + w_2^*\}'\mathcal{B}\{w_1^* + w_2^*\} \quad for\ all\ \lambda > \lambda_0^{(2)}, \tag{3.6}$$

where $\tilde{w}_i = (1-\lambda_0^{(i)})w_i$, $w_i^* = \lambda_0^{(i)} w_i$, $w_i = \theta_0^{(i)} - \theta_0^{(i+1)}$ *for* $i = 1, 2$ *and* $\mathcal{B} = Q_{xz}CQ_{zx}$.

Lemma 6. *Under Assumptions 1 to 4, 7 to 8 and 10 to 14,*

$$sup_{\lambda\in\Lambda}|D_T^*(\lambda) - D^*(\lambda)| = O_p(T^{-1/2}). \tag{3.7}$$

This lemma, which is analogous to Lemma 2 in the one break model, implies that with high probability, $D_T^*(\lambda)$ is uniformly close to its nonstochastic functions given in (3.2) to (3.6). The proof of this lemma is found in Appendix C.2 on page 79.

We now make an additional assumption about the nonstochastic limit of $D_T^*(\lambda)$ at the two true break points. Notice both $D_T^*(\lambda_0^{(1)})$ and $D_T^*(\lambda_0^{(2)})$ are local maxima and have an equal chance of being estimated if the magnitude of the breaks are the same. For simplicity, we make the following assumption to guarantee the maximum value of $D^*(\lambda)$ is unique.

Assumption 15.

$$D^*(\lambda_0^{(1)}) > D^*(\lambda_0^{(2)}). \tag{3.8}$$

This assumption ensures $D^*(\lambda_0^{(1)})$ is the global maximum and hence, $\lambda_0^{(1)}$ dominates $\lambda_0^{(2)}$ in terms of the relative magnitudes of the shifts between the regimes. Thus, if Assumption 15 holds, then $\|\hat{\theta}_1(\lambda) - \hat{\theta}_2(\lambda)\| > \|\hat{\theta}_2(\lambda) - \hat{\theta}_3(\lambda)\|$. As such, the first break fraction estimator converges in probability to $\lambda_0^{(1)}$. At the end of this section, we briefly look at the two possible outcomes when this assumption is relaxed.

Lemma 7. *Under Assumptions 1 to 4, 7 to 8 and 10 to 15, there exists an $\mathcal{E} > 0$ depending only on $\lambda_0^{(i)}$ and $\theta_0^{(j)}$, for $i = 1, 2$ and $j = 1, 2, 3$, such that*

$$D^*(\lambda) - D^*(\lambda_0^{(1)}) \leq -|\lambda - \lambda_0^{(1)}|\mathcal{E} \quad for\ all\ large\ T. \tag{3.9}$$

This lemma which is proved in Appendix C.3 is comparable to Lemma 3 in the one break model. It implies $D^*(\lambda)$ is maximised at $\lambda_0^{(1)}$. Hence, with high probability, the maximiser of $D_T^*(\lambda)$ will also be close to $\lambda_0^{(1)}$, based on Lemma 6 and Assumption 15.

This consistency of the break fraction estimator $\hat{\lambda}$ for the true break fraction $\lambda_0^{(1)}$ is stated in the following proposition:

Proposition 3. *Under Assumptions 1 to 4, 7 to 8 and 10 to 15,*

$$\hat{\lambda} - \lambda_0^{(1)} = O_p(T^{-1/2}). \tag{3.10}$$

Proof of Proposition 3

The proof follows similar argument to that of the one break model in (2.13) to (2.17) on page 34, so we equally state

$$\begin{aligned} D_T^*(\lambda) - D_T^*(\lambda_0^{(1)}) &= \{D_T^*(\lambda) - D^*(\lambda)\} - \{D_T^*(\lambda_0^{(1)}) - D^*(\lambda_0^{(1)})\} \\ &\quad + \{D^*(\lambda) - D^*(\lambda_0^{(1)})\} \\ &\leq 2\ sup_{\lambda\in\Lambda}|D_T^*(\lambda) - D^*(\lambda)| + \{D^*(\lambda) - D^*(\lambda_0^{(1)})\} \\ &\leq 2\ sup_{\lambda\in\Lambda}|D_T^*(\lambda) - D^*(\lambda)| - |\lambda - \lambda_0^{(1)}|\mathcal{E}. \end{aligned} \tag{3.11}$$

By Remark 2 on page 34, $D_T^*(\hat{\lambda}) \geq D_T^*(\lambda_0^{(1)})$ so it must be that for the RHS of (3.11),

$$|\hat{\lambda} - \lambda_0^{(1)}| \leq \mathcal{E}^{-1} 2\ sup_{\lambda\in\Lambda}|D_T^*(\lambda) - D^*(\lambda)|. \tag{3.12}$$

By Lemma 5, $D_T^*(\lambda)$ converges to its nonstochastic limit $D^*(\lambda)$. By Lemma 6, we have that $sup_{\lambda\in\Lambda}|D_T^*(\lambda) - D^*(\lambda)|$ is $O_p(T^{-1/2})$ and by Lemma 7, $\mathcal{E}$ is positive. Hence, Proposition 3 is obtained by using Lemmas 5 to 7.

3.3.2 Convergence Rate of the Break Fraction Estimator

With the consistency of the break fraction estimator established, we now obtain the convergence rate by examining a restricted range of $D_T(\lambda)$ for the break point, k. The approach is again similar to that adopted in the one break model in Subsection 2.4.

We make the following three definitions:

i. $V_1^* = \{k : T\eta \leq k \leq T\lambda_0^{(2)}(1-\eta)\}$ for a small positive number η, such that $\lambda_0^{(1)} \in (\eta, \lambda_0^{(2)}(1-\eta))$. This guarantees that for all $k \in V_1^*$, k is bounded away from both 0 and $k_0^{(2)}$ for a positive fraction of the observations. Since we have shown that $\hat{\lambda}$ is consistent for $\lambda_0^{(1)}$ in Proposition 3, then there is a high probability that $\hat{k}$ will fall into V_1^*. Hence it follows that $Pr(\hat{k} \notin V_1^*) < \epsilon$ for every $\epsilon > 0$ and all large T.

ii. $V_2^* = \{k : |k - k_0^{(1)}| \leq \mathcal{M}\}$, where $\mathcal{M} < \infty$ is a constant. We show that $\hat{k}$ falls into V_2^* with high probability for large $\mathcal{M}$.

iii. $V_3^* = \{k : T\eta \leq k \leq T\lambda_0^{(2)}(1-\eta),\ |k-k_0^{(1)}| > \mathcal{M}\}$. In other words, $V_3^* = V_1^* \cap V_2^{*c}$, where V_2^{*c} is the complement of V_2^*. The probability of the estimator $\hat{k}$ to be in V_3^*, that is to be in V_1^* but not in V_2^*, is very small. We state this in the following lemma,

Lemma 8. *Under Assumptions 1 to 4, 7 to 8 and 10 to 15, there exists $\mathcal{M} < \infty$ such that for all large T and every $\epsilon > 0$,*

$$P\left(\sup_{k\in V_3^*} D_T(k) - D_T(k_0^{(1)}) \geq 0\right) < \epsilon. \tag{3.13}$$

The proof of this lemma is presented in Appendix C.4. As noted in Remark 2 on page 34, $D_T(\hat{k}) \geq D_T(k_0^{(1)})$ by way of definition. Hence, if $\hat{k}$ lies within V_3^*, then it must be that $sup_{k\in V_3^*} D_T(\hat{k}) \geq D_T(k_0^{(1)})$. By Lemma 8, the probability of this occurrence is arbitrarily small. The T-consistency of $\hat{\lambda}$ is now stated in the following proposition,

Proposition 4. *Under Assumptions 1 to 4, 7 to 8 and 10 to 15, for every $\epsilon > 0$, there exists a finite $\mathcal{M}$ independent of T, such that for all large T,*

$$P(T|\hat{\lambda} - \lambda_0^{(1)}| > \mathcal{M}) < \epsilon. \tag{3.14}$$

Proof of Proposition 4

The proof follows similar arguments used in the one break model in Proposition 2 presented on 35. We likewise conclude,

$$\begin{aligned} P(|\hat{k} - k_0^{(1)}| > \mathcal{M}) &\leq Pr(\hat{k} \notin V_1^*) + Pr(\hat{k} \in V_3^*) \\ &\leq \epsilon + Pr\left(\sup_{k \in V_3^*} D_T(k) \geq D_T(k_0^{(1)}) \right) \\ &\leq 2\epsilon. \end{aligned}$$

This concludes the proof and confirms the T-consistency of the first break fraction estimator.

The Second Break Fraction Estimator

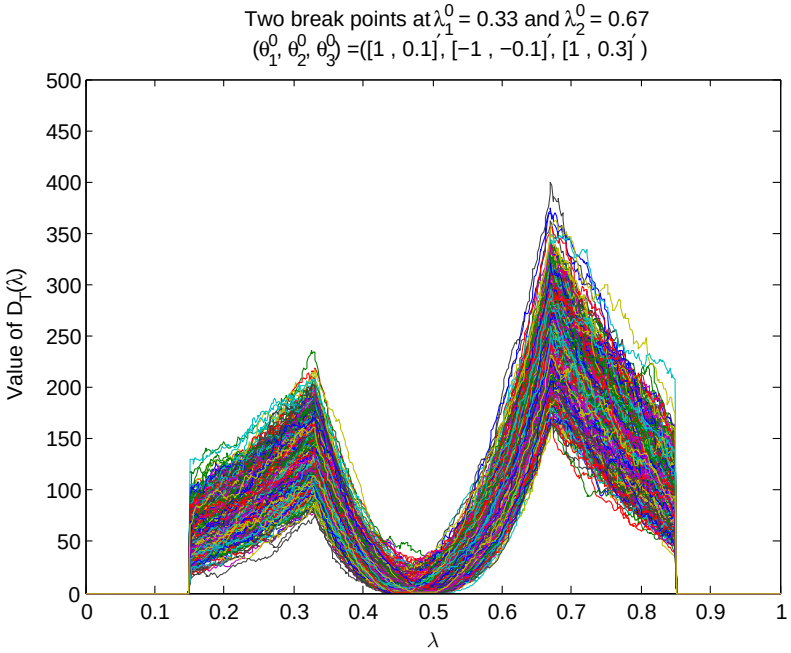
As described in the estimation procedure in Subsection 3.2, with $\hat{\lambda}$ obtained, the sample is split into two subsamples - Subsample 1, made up of observations in $[1, \hat{k}]$, and Subsample 2, made of observations in $[\hat{k} + 1, T]$. The same estimation technique is applied to both subsamples to obtain an estimator for the second break fraction. Since $\lambda_0^{(1)} < \lambda_0^{(2)}$, then the estimator would lie within Subsample 2. This estimator, denoted $\hat{k}_2$ is itself consistent for $\lambda_0^{(2)}$ because it is the most dominating break[3] in the model.

Relaxing Assumption 15

There are two possible outcomes if Assumption 15 on page 65 does not hold; either $D^*(\lambda_0^{(1)}) < D^*(\lambda_0^{(2)})$ or $D^*(\lambda_0^{(1)}) = D^*(\lambda_0^{(2)})$. The former is a bit more straightforward. Since the inequality is reversed, then $\hat{\lambda} \xrightarrow{p} \lambda_0^{(2)}$ because $D_T(\lambda_0^{(2)})$ will now be the global maximum. By symmetry, therefore, the same consistency arguments discussed above follow through. To support this, we present a thousand simulations of $D_T^*(\lambda)$ in Figure 3.3. In this model, the magnitude of the shift at the second break point is larger which makes $\lambda_0^{(2)}$ the dominant break point.

[3]The Monte Carlo simulations support this but we do not show theoretical proofs in this research. However for OLS case, Bai (1997a) showed that $\hat{k}_2$ has the same limiting distribution as $\hat{k}$, though asymptotically independent of it.

Figure 3.3: $D_T(\lambda)$ for all $\lambda \in \Lambda$.
DGP has 2 breaks in SE



The latter outcome if Assumption 15 is relaxed is a bit more complicated as $\hat{\lambda}$ converges to either $\lambda_0^{(1)}$ or $\lambda_0^{(2)}$. To prove the break fraction obtained here is still consistent and converges at the same rate, we follow a similar procedure to that used for $\hat{\lambda}$ above. First, notice that Lemmas 5 and 6 are independent of Assumption 15 and as such they apply in this case as well. We only need to provide analogous counterparts to Lemmas 7 and 8 to complete the proof.

Define k_0^* to be any number between $k_0^{(1)}$ and $k_0^{(2)}$ but bounded away from both of them for a positive fraction of observations, for example $0.5(k_0^{(1)} + k_0^{(2)})$. Let λ_0^* denote the corresponding break fraction. Then we state the following lemma which is comparable to Lemma 7,

Lemma 9. *Under Assumptions 1 to 4, 7 to 8, 10 to 14 and if* $D^*(\lambda_0^{(1)}) = D^*(\lambda_0^{(2)})$*, there exists an* $\mathcal{E} > 0$ *depending only on* $\lambda_0^{(i)}$ *and* $\theta_0^{(j)}$*, for* $i = 1, 2$ *and* $j = 1, 2, 3$*, such that for all large T,*

$$D^*(\lambda) - D^*(\lambda_0^{(1)}) \leq -|\lambda - \lambda_0^{(1)}|\mathcal{E} \quad uniformly\ in\ \lambda \leq \lambda_0^*, \tag{3.15}$$

$$D^*(\lambda) - D^*(\lambda_0^{(2)}) \leq -|\lambda - \lambda_0^{(2)}|\mathcal{E} \quad uniformly\ in\ \lambda \geq \lambda_0^*. \tag{3.16}$$

This Lemma proved in Appendix C.5 states that as long as $k \leq k_0^*$, then with high probability, the maximiser of $D_T^*(\lambda)$ will be close to $\lambda_0^{(1)}$ and analogously for $\lambda_0^{(2)}$ if $k \geq k_0^*$.

Now, let $T_1 = [1, k_0^*]$ and $T_2 = [k_0^* + 1, T]$ and define $\hat{k}_1^* = argmax_{k \in T_1} D_T^*(\lambda)$ and $\hat{k}_2^* = argmax_{k \in T_2} D_T^*(\lambda)$. Since $\hat{k}$ is the global maximiser of $D_T(\lambda)$, then it must be that

$$\hat{k} = \begin{cases} \hat{k}_1^* & if\ D_T^*(\hat{\lambda}_1^*) > D_T^*(\hat{\lambda}_2^*) \\ \hat{k}_2^* & if\ D_T^*(\hat{\lambda}_1^*) < D_T^*(\hat{\lambda}_2^*), \end{cases} \tag{3.17}$$

where $\hat{\lambda}_i^* = \hat{k}_i^*/T$ for $i = 1, 2$. The next proposition states the consistency of these break fraction estimators,

Proposition 5. *Under Assumptions 1 to 4, 7 to 8 and 10 to 14 and if* $D^*(\lambda_0^{(1)}) = D^*(\lambda_0^{(2)})$,

$$\hat{\lambda}_i^* - \lambda_0^{(i)} = O_p(T^{-1/2}), \quad for\ i = 1, 2. \tag{3.18}$$

Proof of Proposition 5

Lemmas 5, 6 and 9 are used to prove Proposition 5 following similar arguments to the proof of Proposition 3 carried out in (3.10) to (3.12) on pages 66 to 66. Thus, we conclude that when $D^*(\lambda_0^{(1)}) = D^*(\lambda_0^{(2)})$, the break fraction estimators, $\hat{\lambda}_1^*$ and $\hat{\lambda}_2^*$ are also consistent for the true break fractions, $\lambda_0^{(1)}$ and $\lambda_0^{(2)}$, respectively.

To prove the T-consistency of these break fraction estimators, we define the following:

1. $\ddot{V}_{1a} = \{k : T\eta \leq k \leq k_0^*\}$,
2. $\ddot{V}_{1b} = \{k : |k - k_0^{(1)}| \leq \mathcal{M}\}$, where $\mathcal{M} < \infty$ is a constant,
3. $\ddot{V}_{2a} = \{k : k_0^* + 1 \leq k \leq T(1 - \eta)\}$,
4. $\ddot{V}_{2b} = \{k : |k - k_0^{(2)}| \leq \mathcal{M}\}$,
5. $\ddot{V}_3^{(1)} = \ddot{V}_{1a} \cap \ddot{V}_{1b}^c$, where $\ddot{V}_{1b}^c$ is the complement of $\ddot{V}_{1b}$,
6. $\ddot{V}_3^{(2)} = \ddot{V}_{2a} \cap \ddot{V}_{2b}^c$

Therefore, analogous to Lemma 8, we state

Lemma 10. *Under Assumptions 1 to 4, 7 to 8 and 10 to 14 and if* $D^*(\lambda_0^{(1)}) = D^*(\lambda_0^{(2)})$, *there exists* $\mathcal{M} < \infty$ *such that for all large* T *and every* $\epsilon > 0$,

$$P\left(\sup_{k \in \ddot{V}_3^{(i)}} D_T(k) - D_T(k_0^{(i)}) \geq 0 \right) < \epsilon \quad for\ i = 1, 2, \tag{3.19}$$

With this lemma proved in Appendix C.6 on page 100, we now state our proposition that when $D^*(\lambda_0^{(1)}) = D^*(\lambda_0^{(2)})$, the break fraction estimator would converge to either $\lambda_0^{(1)}$ or $\lambda_0^{(2)}$ at the same rate of T like the other estimators earlier discussed.

Proposition 6. *Under Assumptions 1 to 4, 7 to 8 and 10 to 14 and if* $D^*(\lambda_0^{(1)}) = D^*(\lambda_0^{(2)})$, *for every* $\epsilon > 0$, *there exists a finite* $\mathcal{M}$ *independent of* T, *such that for all large* T,

$$P\Big(|T(\hat{\lambda} - \lambda_0^{(1)})| > \mathcal{M} \text{ and } |T(\hat{\lambda} - \lambda_0^{(2)})| > \mathcal{M}\Big) < \epsilon. \tag{3.20}$$

Proof of Proposition 6

Using Lemma 10 along with the consistency of $\hat{\lambda}_i^*$ given in (3.18), this proposition is proved in a similar way to Proposition 4 on page 67. Thus, in the presence of equal magnitude of breaks, the break fraction estimator is still T-consistent.

Lastly, we conjecture that under Assumptions 1 to 4, 7 to 8 and 10 to 14 and if $D^*(\lambda_0^{(1)}) = D^*(\lambda_0^{(2)})$, the break fraction estimator $\hat{\lambda}$ converges to a random variable with equal mass of 0.5 at $\lambda_0^{(1)}$ and $\lambda_0^{(2)}$. Although Bai (1997a) proves this for the OLS case, we do not provide any proofs in this study. However, the results of the Monte Carlo simulations present reasonable evidence to conjecture it holds. This is also seen in the similarity of the peaks of $D_T(\lambda)$ presented earlier in Figure 3.2 on page 65, where the magnitude of shifts in the two breaks are equivalent.

When the Model has more than Two Breaks; $m > 2$

The analyses and theoretical results presented above for a two break model also extend to models with more than two break points in their SE, albeit with more terms to manage in the analyses. For example, the different expressions of the nonstochastic function in Lemma 5 increases as the number of break points increase. Specifically, when looking for the first break point in an m-break model, the nonstochastic limit of $D_T^*(\lambda)$ exhibits $2m+1$ different expressions. In the next proposition, we present a formula to obtain these different expressions of $D^*(\lambda)$ for models with m breaks.

Proposition 7. *If Assumptions 1 to 4, 7 to 8 and 10 to 14 hold, then* $D_T^*(\lambda)$ *uniformly converges to the nonstochastic function* $D^*(\lambda)$ *on (0,1) given below*

$$D^*(\lambda) = \frac{1-\lambda}{\lambda}\Phi_A(j) + \frac{\lambda}{1-\lambda}\Phi_B(j) + \Phi_C(j), \tag{3.21}$$

where $\Phi_A(j) = \phi_A(\cdot)'\mathcal{B}\phi_A(\cdot)$; $\phi_A(j) = \sum_{i=1}^{j} w_i^*$; $\Phi_B(j) = \phi_B(\cdot)'\mathcal{B}\phi_B(\cdot)$; $\phi_B(j) = \sum_{i=j+1}^{m} \tilde{w}_i$; $\Phi_C(j) = \phi_A(\cdot)'\mathcal{B}\phi_B(\cdot)$; $w_i^* = \lambda_0^{(i)} w_i$; $\tilde{w}_i = (1 - \lambda_0^{(i)}) w_i$; $w_i = \theta_0^{(i)} - \theta_0^{(i+1)}$; i *is the break point being estimated;* j *is the number of true breaks before or at this candidate break point and* m *is the total number of true breaks in the model.*

This proposition which is proved in Appendix C.7 provides a generic formula by which the nonstochastic function can be obtained uniformly for all $\lambda \in \Lambda$.

To ensure the uniqueness of the maximum value of $D^*(\lambda)$, we state the following assumption:

Assumption 16.

$$D^*(\lambda_0^{(i)}) > D^*(\lambda_0^{(j)}), \quad for\ all\ j \neq i. \tag{3.22}$$

This assumption guarantees $k_0^{(i)}$ is the dominating break point.

Proposition 8. *If Assumptions 1 to 4, 7 to 8, 10 to 14 and 16 hold, the estimated break fraction,* $\hat{\lambda}$ *is* T*-consistent for* $\lambda_0^{(i)}$.

We conjecture Proposition 8 holds and do not present any theoretical proofs for this. However, the results of the Monte Carlo simulations in Chapter 5 provide strong evidence that this conjecture is consistent with what is found in finite samples.

Generally, if a break point is estimated within a subsample $[T_1,\ T_2]$, it must be bounded away from T_1 and T_2 for a positive fraction of observations. As long as Assumption 16 holds in this subsample, the break point estimator must be T-consistent for one of the m true breaks because it is the location at which $D_T^*(\lambda)$ is maximised over the range $[T_1,\ T_2]$. Thus, if all the m break points are estimated, then their corresponding break fractions are all T-consistent[4] as in the two break model.

3.4 Conclusion

This chapter considers issues associated with estimating the unknown location of multiple break points in linear models using GMM. The Sequential Estimation Method where the multiple break points are individually estimated one at a time as proposed in Bai (1997a) is adopted in this study. For each round of estimations, we used the Difference-type test of parameter variation based on GMM parameters.

[4]Similar to Bai and Perron (1998).

As in the single break model discussed in the previous chapter, the break fraction estimators obtained from these multiple break models are also shown to be consistent for the true break points in the model. Furthermore, the T convergence rate of the estimators are also established. Similar rates have been proven in the literature for the OLS case; see Bai (1997a), Bai and Perron (1998) and Hall et al. (2012). To the best of our knowledge, these asymptotic properties of the break fraction estimators have not been done within the GMM framework using the approach of parameter variation proposed in this study.

Series of Monte Carlo simulations carried out on multiple break models are presented in Chapter 5. The results support the theoretical proofs in this chapter. However, these theoretical properties are established based on the fact that the Jacobian Equation (JE) is stable; that is, the relationship between the endogenous regressors and instruments remain unchanged throughout the sample. In the next chapter, we consider models where this JE changes at an unknown location in the sample.

Appendix C

C.1 Proof of Lemma 5

This appendix shows the uniform convergence of $D_T^*(\lambda)$ to its nonstochastic limit $D^*(\lambda)$ for all the five expressions associated with a two break model. As in the single break case, the Parameter Difference and Centre Matrices are examined separately before combining them together to get $D^*(\lambda)$.

To simplify the analysis, we note the following three points. First, since the Jacobian is stable, then the structure of the Centre Matrix does not change throughout the analysis. It remains $M_*(\lambda)^{-1} = \lambda(1-\lambda)\mathcal{B}$ as given in (B.12) on page 43.

Secondly, we express the parameters before and after the break in all five cases as,

$$\hat{\theta}_i(\lambda) = \mathcal{H}_{i,T}(\lambda)\mathcal{H}_{zy}^{(i)}, \quad for\ i = 1, 2, \tag{C.1}$$

where as given in (B.1) and (B.2) on page 40,
$\mathcal{H}_{zy}^{(1)} = T^{-1}\sum_{t=1}^{k} z_t y_t$ and $\mathcal{H}_{zy}^{(2)} = T^{-1}\sum_{t=k+1}^{T} z_t y_t$.

Thirdly, based on Assumptions 1, 3 and 8, since the Jacobian is stable, then in all five cases[1], $\mathcal{H}_{1,T}(\lambda) \xrightarrow{p} \mathcal{H}_1(\lambda)$ and $\mathcal{H}_{2,T}(\lambda) \xrightarrow{p} \mathcal{H}_2(\lambda)$, where
$\mathcal{H}_1(\lambda) = \{\lambda(Q_{xz}CQ_{zx})\}^{-1}Q_{xz}C$ and $\mathcal{H}_2(\lambda) = \{(1-\lambda)(Q_{xz}CQ_{zx})\}^{-1}Q_{xz}C$.
Thus, only $\mathcal{H}_{zy}^{(i)}$ changes in all five cases and so we focus on it in the proofs below to obtain the limits of the parameter differences.

[1] See analysis done in the single break model for (B.8) and (B.14) on pages 42 and 43.

Case 1: When $\lambda < \lambda_0^{(1)}$

Since λ lies within the first regime, then the parameter before the candidate break point is estimated at the true value of the parameter in the first regime, similar to (B.7). Thus,

$$\hat{\theta}_1(\lambda) \xrightarrow{p} \theta_0^{(1)}, \quad uniformly\ in\ \lambda < \lambda_0^{(1)}. \tag{C.2}$$

On the other hand, the parameter after the candidate break point is a weighted average of the three parameters in the model, hence

$$\mathcal{H}_{zy}^{(2)} = f_{2,T} + f_{3,T} + f_{4,T}, \tag{C.3}$$

where $f_{2,T} = T^{-1}\sum_{t=k+1}^{k_0^{(1)}} z_t y_t^{'}$, $f_{3,T} = T^{-1}\sum_{t=k_0^{(1)}+1}^{k_0^{(2)}} z_t y_t$ *and* $f_{4,T} = T^{-1}\sum_{t=k_0^{(2)}+1}^{T} z_t y_t$. Based on Assumptions 1, 3 and 8,

$$f_{2,T} = T^{-1}\sum_{t=k+1}^{k_0^{(1)}} z_t x_t^{'}\theta_0^{(1)} + T^{-1}\sum_{t=k+1}^{k_0^{(1)}} z_t u_t \xrightarrow{p} f_2,$$

where $f_2 = (\lambda_0^{(1)} - \lambda)Q_{zx}\theta_0^{(1)}$. Similarly, $f_{3,T} \xrightarrow{p} f_3 = (\lambda_0^{(2)} - \lambda_0^{(1)})Q_{zx}\theta_0^{(2)}$ and $f_{4,T} \xrightarrow{p} f_4 = (1 - \lambda_0^{(2)})Q_{zx}\theta_0^{(3)}$.

Define $\Psi^{a2} = f_2 + f_3 + f_4$, then the limit of the second parameter,

$$\begin{aligned}\theta_*^{(2)}(\lambda) &= \mathcal{H}_2(\lambda)\Psi^{a2} \\ &= \frac{1}{1-\lambda}\big\{(\lambda_0^{(1)} - \lambda)\theta_0^{(1)} + (\lambda_0^{(2)} - \lambda_0^{(1)})\theta_0^{(2)} + (1 - \lambda_0^{(2)})\theta_0^{(3)}\big\},\end{aligned} \tag{C.4}$$

uniformly in λ. Using the RHS of (C.2) and (C.4), we get the Parameter Difference uniformly in $\lambda < \lambda_0^{(1)}$ as,

$$\begin{aligned}&\hat{\theta}_1(\lambda) - \hat{\theta}_2(\lambda) \\ &\xrightarrow{p} \frac{1}{1-\lambda}\big\{(1-\lambda)\theta_0^{(1)} - (\lambda_0^{(1)} - \lambda)\theta_0^{(1)} - (\lambda_0^{(2)} - \lambda_0^{(1)})\theta_0^{(2)} - (1 - \lambda_0^{(2)})\theta_0^{(3)}\big\} \\ &= \frac{1}{1-\lambda}\big\{(1 - \lambda_0^{(1)})(\theta_0^{(1)} - \theta_0^{(2)}) + (1 - \lambda_0^{(2)})(\theta_0^{(2)} - \theta_0^{(3)})\big\} \\ &= \frac{1}{1-\lambda}\big\{\tilde{w}_1 + \tilde{w}_2\big\},\end{aligned} \tag{C.5}$$

where $\tilde{w}_i = (1 - \lambda_0^{(i)})w_i$ and $w_i = \theta_0^{(i)} - \theta_0^{(i+1)}$ for $i = 1, 2$.

Combining this limit of the Parameter Difference on the RHS of (C.5) with the Centre Matrix, we obtain the nonstochastic limit of $D_T^*(\cdot)$ uniformly in λ for $\lambda < \lambda_0^{(1)}$ as

$$\begin{aligned} D^*(\lambda) &= \left(\frac{1}{1-\lambda}\right)^2 \lambda(1-\lambda)\{\tilde{w}_1 + \tilde{w}_2\}' \mathcal{B}\{\tilde{w}_1 + \tilde{w}_2\} \\ &= \frac{\lambda}{1-\lambda}\{\tilde{w}_1 + \tilde{w}_2\}' \mathcal{B}\{\tilde{w}_1 + \tilde{w}_2\}. \end{aligned} \tag{C.6}$$

Case 2: When $\lambda = \lambda_0^{(1)}$

The limit of the parameter before the break remains as in Case 1, while for the parameter after the break, $\mathcal{H}_{zy}^{(2)} = f_{3,T} + f_{4,T}$. Define $\Psi^{b2} = f_3 + f_4$, then the limit of the second parameter is given as,

$$\begin{aligned} \theta_*^{(2)}(\lambda) &= \mathcal{H}_2(\lambda_0^{(1)})\Psi^{b2} \\ &= \frac{1}{1-\lambda_0^{(1)}}\{(\lambda_0^{(2)} - \lambda_0^{(1)})\theta_0^{(2)} + (1-\lambda_0^{(2)})\theta_0^{(3)}\}. \end{aligned} \tag{C.7}$$

Using (C.2) and (C.7), the Parameter Difference is obtained as,

$$\begin{aligned} \hat{\theta}_1(\lambda) - \hat{\theta}_2(\lambda) &\xrightarrow{p} \frac{1}{1-\lambda_0^{(1)}}\{(1-\lambda_0^{(1)})\theta_0^{(1)} - (\lambda_0^{(2)} - \lambda_0^{(1)})\theta_0^{(2)} - (1-\lambda_0^{(2)})\theta_0^{(3)}\} \\ &= \frac{1}{1-\lambda_0^{(1)}}\{(1-\lambda_0^{(1)})(\theta_0^{(1)} - \theta_0^{(2)}) + (1-\lambda_0^{(2)})(\theta_0^{(2)} - \theta_0^{(3)})\} \\ &= \frac{1}{1-\lambda_0^{(1)}}\{\tilde{w}_1 + \tilde{w}_2\}' \mathcal{B}\{\tilde{w}_1 + \tilde{w}_2\}. \end{aligned} \tag{C.8}$$

Combining the RHS of (C.8) with the Centre Matrix, the nonstochastic limit of $D_T^*(\cdot)$ for this case when $\lambda = \lambda_0^{(1)}$ is

$$\begin{aligned} D^*(\lambda_0^{(1)}) &= \left(\frac{1}{1-\lambda_0^{(1)}}\right)^2 \lambda_0^{(1)}(1-\lambda_0^{(1)})\{\tilde{w}_1 + \tilde{w}_2\}' \mathcal{B}\{\tilde{w}_1 + \tilde{w}_2\} \\ &= \frac{\lambda_0^{(1)}}{1-\lambda_0^{(1)}}\{\tilde{w}_1 + \tilde{w}_2\}' \mathcal{B}\{\tilde{w}_1 + \tilde{w}_2\}. \end{aligned} \tag{C.9}$$

Case 3: When $\lambda_0^{(1)} < \lambda < \lambda_0^{(2)}$

In this scenario, neither of the parameters is estimated at its true value, rather they are both weighted averages. For the parameter before the estimated break point,

$\mathcal{H}_{zy}^{(1)} = f_{1,T} + f_{5,T}$, where $f_{1,T} = T^{-1}\sum_{t=1}^{k_0^{(1)}} z_t y_t$ and $f_{5,T} = T^{-1}\sum_{t=k_0^{(1)}+1}^{k} z_t y_t$.
For the parameter after the estimated break point,
$\mathcal{H}_{zy}^{(2)} = f_{6,T} + f_{4,T}$, where $f_{6,T} = T^{-1}\sum_{t=k+1}^{k_0^{(2)}} z_t y_t$.
From Assumptions 1, 3 and 8, $f_{1,T} \xrightarrow{p} f_1 = \lambda_0^{(1)} Q_{zx}\theta_0^{(1)}$, $f_{5,T} \xrightarrow{p} f_5 = (\lambda - \lambda_0^{(1)})Q_{zx}\theta_0^{(2)}$ and $f_{6,T} \xrightarrow{p} f_6 = (\lambda_0^{(2)} - \lambda)Q_{zx}\theta_0^{(2)}$. Thus, the limits of the parameters uniformly in λ are given as,

$$\theta_*^{(1)}(\lambda) = \mathcal{H}_1(\lambda)\Psi^{c2} \tag{C.10}$$

and

$$\theta_*^{(2)}(\lambda) = \mathcal{H}_2(\lambda)\Psi^{d2}, \tag{C.11}$$

where $\Psi^{c2} = f_1 + f_5$ and $\Psi^{d2} = f_6 + f_4$. The Parameter Difference is obtained by taking the difference between the RHS of (C.10) and (C.11) as,

$$\begin{aligned}\hat{\theta}_1(\lambda) - \hat{\theta}_2(\lambda) \xrightarrow{p} \frac{1}{\lambda(1-\lambda)}\big\{&(1-\lambda)\lambda_0^{(1)}\theta_0^{(1)} + (1-\lambda)(\lambda - \lambda_0^{(1)})\theta_0^{(2)} \\ &- (\lambda_0^{(2)} - \lambda)\lambda\theta_0^{(2)} - (1-\lambda_0^{(2)})\lambda\theta_0^{(3)}\big\} \\ = \frac{1}{\lambda(1-\lambda)}&\big\{(1-\lambda)\lambda_0^{(1)}w_1 + \lambda(1-\lambda_0^{(2)})w_2\big\} \\ = \frac{1}{\lambda(1-\lambda)}&\big\{(1-\lambda)w_1^* + \lambda\tilde{w}_2\big\}.\end{aligned} \tag{C.12}$$

where $w_1^* = \lambda_0^{(1)} w_1$. Now combine (C.12) with the Centre Matrix to get the nonstochastic limit of the test statistic uniformly in λ for $\lambda \in (\lambda_0^{(1)}, \lambda_0^{(2)})$ as,

$$\begin{aligned}D^*(\lambda) &= \left(\frac{1}{\lambda(1-\lambda)}\right)^2 \lambda(1-\lambda)\big\{(1-\lambda)w_1^* + \lambda\tilde{w}_2\big\}'\mathcal{B}\big\{(1-\lambda)w_1^* + \lambda\tilde{w}_2\big\} \\ &= \frac{1-\lambda}{\lambda}w_1^{*\prime}\mathcal{B}w_1^* + \frac{\lambda}{1-\lambda}\tilde{w}_2'\mathcal{B}\tilde{w}_2 + 2w_1^{*\prime}\mathcal{B}\tilde{w}_2.\end{aligned} \tag{C.13}$$

Case 4: When $\lambda = \lambda_0^{(2)}$

In this case, the parameter before the estimated break point is still a weighted average while the second parameter converges to the true parameter of the third regime. We have

that uniformly in λ, the second parameter,

$$\hat{\theta}_2(\lambda) \xrightarrow{p} \theta_0^{(3)}. \tag{C.14}$$

While for the parameter before the break we have $\mathcal{H}_{zy}^{(1)} = f_{1,T} + f_{3,T}$ with its limit uniformly in λ given as

$$\theta_*^{(1)}(\lambda) = \mathcal{H}_1(\lambda_0^{(2)})\Psi^{e2}, \tag{C.15}$$

where $\Psi^{e2} = f_1 + f_3$. The Parameter Difference using the RHS of (C.14) and (C.15) is obtained as,

$$\begin{aligned}\hat{\theta}_{1,T}(\lambda) - \hat{\theta}_{2,T}(\lambda) &\xrightarrow{p} \frac{1}{\lambda_0^{(2)}}\{\lambda_0^{(1)}\theta_0^{(1)} + (\lambda_0^{(2)} - \lambda_0^{(1)})\theta_0^{(2)} - \lambda_0^{(2)}\theta_0^{(3)}\} \\ &= \frac{1}{\lambda_0^{(2)}}\{\lambda_0^{(1)}(\theta_0^{(1)} - \theta_0^{(2)}) + \lambda_0^{(2)}(\theta_0^{(2)} - \theta_0^{(3)})\} \\ &= \frac{1}{\lambda_0^{(2)}}\{w_1^* + w_2^*\}. \end{aligned} \tag{C.16}$$

Therefore, combining the Parameter Difference in (C.16) with the Centre Matrix, the nonstochastic limit of $D_T^*(\lambda)$ when $\lambda = \lambda_0^{(2)}$ is

$$\begin{aligned} D^*(\lambda_0^{(2)}) &= \left(\frac{1}{\lambda_0^{(2)}}\right)^2 \lambda_0^{(2)}(1 - \lambda_0^{(2)})\{w_1^* + w_2^*\}'\mathcal{B}\{w_1^* + w_2^*\} \\ &= \frac{1 - \lambda_0^{(2)}}{\lambda_0^{(2)}}\{w_1^* + w_2^*\}'\mathcal{B}\{w_1^* + w_2^*\}. \end{aligned} \tag{C.17}$$

Case 5: When $\lambda > \lambda_0^{(2)}$

The candidate break point lies within the third regime, hence the parameter after the break remains as in (C.14) in Case 4 while for the parameter before the break, $\mathcal{H}_{zy}^{(1)} = f_{1,T} + f_{3,T} + f_{7,T}$, where $f_{7,T} = T^{-1}\sum_{t=k_0^{(2)}+1}^{k} z_t y_t$. From Assumptions 1, 3 and 8, it follows that $f_{7,T} \xrightarrow{p} f_7 = (\lambda - \lambda_0^{(2)})Q_{zx}\theta_0^{(3)}$. Hence, the limit of the parameter before the break,

$$\theta_*^{(1)}(\lambda) = \mathcal{H}_1(\lambda)\Psi^{f2}, \tag{C.18}$$

where $\Psi^{f2} = f_1 + f_3 + f_7$. Using the limits of the parameters as given on the RHS of (C.18) and (C.14), we obtain the Parameter Difference uniformly in λ as,

$$\begin{aligned}\hat{\theta}_{1,T}(\lambda) - \hat{\theta}_{2,T}(\lambda) \xrightarrow{p} & \frac{1}{\lambda}\{\lambda_0^{(1)}\theta_0^{(1)} + (\lambda_0^{(2)} - \lambda_0^{(1)})\theta_0^{(1)} + (\lambda - \lambda_0^{(2)})\theta_0^{(3)} - \lambda\theta_0^{(3)}\} \\ &= \frac{1}{\lambda}\{\lambda_0^{(1)}(\theta_0^{(1)} - \theta_0^{(2)}) + \lambda_0^{(2)}(\theta_0^{(2)} - \theta_0^{(3)})\} \\ &= \frac{1}{\lambda}\{w_1^* + w_2^*\}. \end{aligned} \tag{C.19}$$

Combining (C.19) with the Centre Matrix, the nonstochastic limit of $D_T^*(\lambda)$ uniformly in λ for $\lambda > \lambda_0^{(2)}$ is,

$$D^*(\lambda) = \frac{1-\lambda}{\lambda}\{w_1^* + w_2^*\}'\mathcal{B}\{w_1^* + w_2^*\}. \tag{C.20}$$

C.2 Proof of Lemma 6

In this appendix, the uniform convergence of the test statistic, $D_T^*(\lambda)$, to its nonstochastic function, $D^*(\lambda)$, is established by showing the *supremum* over λ of the difference between these two functions is $O_p(T^{-1/2})$. All the five expressions of $D_T^*(\lambda)$ associated with a two break model are examined.

Our approach follows similar steps to the one break model in Appendix B.2 where Lemma 6 is re-expressed as given in (B.18) to (B.20). In this form, the uniform convergence of the Parameter Differences and the Centre Matrices are examined separately.

Starting with the Centre Matrices in (B.20), since the Jacobian is stable, then the analysis carried out in the one break model extends to this two break model. Hence, we only state here that as established in (B.21) to (B.26),

$sup_{\lambda\in\Lambda}\|M_{*,T}(\lambda)^{-1} - M_*(\lambda)^{-1}\| = O_p(T^{-1/2})$.

In what follows, we establish the Parameter Differences, $\mu_{*,T}(\lambda) - \mu_*(\lambda)$, given in (B.19) on page 44 are also $O_p(T^{-1/2})$ uniformly over λ in each of the five cases associated with a two break model. To do this, we follow a similar approach used in the one break model in Appendix B.2 and equally show here that $T^{1/2}(\hat{\theta}_i(\lambda) - \theta_0^{(i)})$ is $O_p(1)$ uniformly over λ for $i = 1, 2$.

When $\lambda < \lambda_0^{(1)}$

From the parameter before the break given in (C.2) on page 75, we obtain

$$T^{1/2}(\hat{\theta}_1(\lambda) - \theta_*^{(1)}(\lambda)) = T^{1/2}(\hat{\theta}_1(\lambda) - \theta_0^{(1)}) = O_p(1). \tag{C.21}$$

For the parameter after the break, we rewrite (C.3) on page 75 and express it as

$$\hat{\theta}_2(\lambda) = \mathcal{H}_{2,T}(\lambda)\Big\{ f_{2,zx} + f_{3,zx} + f_{4,zx} + T^{-1}\sum_{t=k+1}^{T} z_t u_t \Big\}, \tag{C.22}$$

where $f_{2,zx} = T^{-1}\sum_{t=k+1}^{k_0^{(1)}} z_t x_t' \theta_0^{(1)}$, $f_{3,zx} = T^{-1}\sum_{t=k_0^{(1)}+1}^{k_0^{(2)}} z_t x_t' \theta_0^{(2)}$ and $f_{4,zx} = T^{-1}\sum_{t=k_0^{(2)}+1}^{T} z_t x_t' \theta_0^{(3)}$.

From the limit of $\hat{\theta}_2(\lambda)$ on the RHS of (C.4), we write

$$\begin{aligned} &T^{1/2}(\hat{\theta}_{2,T}(\lambda) - \theta_*^{(2)}(\lambda)) \\ &\quad = \mathcal{H}_{2,T}(\lambda)T^{-1/2}\Big\{ \sum_{t=k+1}^{T} z_t u_t + \sum_{t=k+1}^{k_0^{(1)}} z_t x_t' \theta_0^{(1)} + \sum_{t=k_0^{(1)}+1}^{k_0^{(2)}} z_t x_t' \theta_0^{(2)} + \sum_{t=k_0^{(2)}+1}^{T} z_t x_t' \theta_0^{(3)} \Big\} \\ &\qquad - T^{1/2}\mathcal{H}_2(\lambda)\Psi^{a2}. \end{aligned}$$

With some rearrangements, we can write

$$\begin{aligned} &T^{1/2}(\hat{\theta}_{2,T}(\lambda) - \theta_*^{(2)}(\lambda)) \\ &\quad = \mathcal{H}_{2,T}(\lambda)\Big\{ T^{-1/2}\sum_{t=k+1}^{T} z_t u_t + T^{-1/2}\sum_{t=k+1}^{k_0^{(1)}} (z_t x_t' - Q_{zx})\theta_0^{(1)} \\ &\qquad + T^{-1/2}\sum_{t=k_0^{(1)}+1}^{k_0^{(2)}} (z_t x_t' - Q_{zx})\theta_0^{(2)} + T^{-1/2}\sum_{t=k_0^{(2)}+1}^{T} (z_t x_t' - Q_{zx})\theta_0^{(3)} \Big\} \\ &\qquad + T^{1/2}[\mathcal{H}_{2,T}(\lambda) - \mathcal{H}_2(\lambda)]\Psi^{a2} + T^{1/2}\zeta_T^{a2}, \end{aligned}$$

where $\zeta_T^{a2} = \mathcal{H}_{2,T}(\lambda)\{f_{2,zx}^Q + f_{3,zx}^Q + f_{4,zx}^Q - \Psi^{a2}\}$ and Ψ^{a2} is as defined on page 75. Thus,

we conclude

$$
\begin{aligned}
&T^{1/2}(\hat{\theta}_{2,T}(\lambda) - \theta_*^{(2)}(\lambda)) \\
&\quad = \mathcal{H}_{2,T}(\lambda)\Big\{[\mathcal{H}_{zu}(1) - \mathcal{H}_{zu}(\lambda)] + [\mathcal{H}_{zx}(\lambda_0^{(1)}) - \mathcal{H}_{zx}(\lambda)]\theta_0^{(1)} \\
&\quad + [\mathcal{H}_{zx}(\lambda_0^{(2)}) - \mathcal{H}_{zx}(\lambda_0^{(1)})]\theta_0^{(2)} + [\mathcal{H}_{zx}(1) - \mathcal{H}_{zx}(\lambda_0^{(2)})]\theta_0^{(3)}\Big\} \\
&\quad + T^{1/2}[\mathcal{H}_{2,T}(\lambda) - \mathcal{H}_2(\lambda)]\Psi^{a2} + T^{1/2}\zeta_T^{a2},
\end{aligned}
\tag{C.23}
$$

where $\mathcal{H}_{zx}(\lambda) = T^{-1/2}\sum_{t=1}^{k}(z_t x_t^{'} - Q_{zx})$, $f_{2,zx}^{Q} = T^{-1}\sum_{t=k+1}^{k_0^{(1)}} Q_{zx}\theta_0^{(1)}$, $f_{3,zx}^{Q} = T^{-1}\sum_{t=k_0^{(1)}+1}^{k_0^{(2)}} Q_{zx}\theta_0^{(2)}$ and $f_{4,zx}^{Q} = T^{-1}\sum_{t=k_0^{(2)}+1}^{T} Q_{zx}\theta_0^{(3)}$.
As in the one break case, ζ_T^{a2} is $o_p(1)$ following similar arguments in (B.31) as detailed on page 47, $\mathcal{H}_{zx}(\cdot)$ is $O_p(1)$ by the FCLT while $\mathcal{H}_{2,T}(\cdot) - \mathcal{H}_2(\cdot)$ is $O_p(T^{-1/2})$ analogous to (B.31). Therefore, the terms on the RHS of (C.23) are all bounded in probability uniformly in $\lambda < \lambda_0^{(1)}$.

When $\lambda = \lambda_0^{(1)}$

The parameter before the break point is identical to (C.21) while for that after the break we have,

$$
\hat{\theta}_2(\lambda) = \mathcal{H}_{2,T}(\lambda_0^{(1)})\Big\{f_{3,zx} + f_{4,zx} + T^{-1}\sum_{t=k_0^{(1)}+1}^{T} z_t u_t\Big\}. \tag{C.24}
$$

Its limit is identical to that derived in (C.7) on page 76. Hence combining the RHS of (C.7) and (C.24), we have

$$
\begin{aligned}
&T^{1/2}(\hat{\theta}_{2,T}(\lambda) - \theta_*^{(2)}(\lambda)) \\
&= \mathcal{H}_{2,T}(\lambda_0^{(1)})T^{-1/2}\Big\{\sum_{t=k_0^{(1)}+1}^{T} z_t u_t + \sum_{t=k_0^{(1)}+1}^{k_0^{(2)}} (z_t x_t^{'} - Q_{zx})\theta_0^{(2)} + \sum_{t=k_0^{(2)}+1}^{T} (z_t x_t^{'} - Q_{zx})\theta_0^{(3)}\Big\} \\
&\quad + T^{1/2}[\mathcal{H}_{2,T}(\lambda) - \mathcal{H}_2(\lambda)]\Psi^{b2} + T^{1/2}\zeta_T^{b2} \\
&= \mathcal{H}_{2,T}(\lambda_0^{(1)})\Big\{[\mathcal{H}_{zu}(1) - \mathcal{H}_{zu}(\lambda_0^{(1)})] + [\mathcal{H}_{zx}(\lambda_0^{(2)}) - \mathcal{H}_{zx}(\lambda_0^{(1)})]\theta_0^{(2)} \\
&\quad + [\mathcal{H}_{zx}(1) - \mathcal{H}_{zx}(\lambda_0^{(2)})]\theta_0^{(3)}\Big\} \\
&\quad + T^{1/2}[\mathcal{H}_{2,T}(\lambda) - \mathcal{H}_2(\lambda)]\Psi^{b2} + T^{1/2}\zeta^{b2},
\end{aligned}
\tag{C.25}
$$

where $\zeta_T^{b2} = \mathcal{H}_{2,T}(\lambda_0^{(1)})\{f_{3,zx}^Q + f_{4,zx}^Q - \Psi^{b2}\}$ and Ψ^{b2} is as defined on page 76. The terms on the RHS of (C.25) are all bounded in probability and the result follows.

When $\lambda_0^{(1)} < \lambda < \lambda_0^{(2)}$

For the first parameter, we obtain

$$\hat{\theta}_1(\lambda) = \mathcal{H}_{1,T}(\lambda)\Big\{f_{1,zx} + f_{5,zx} + T^{-1}\sum_{t=1}^{k} z_t u_t\Big\}, \tag{C.26}$$

where $f_{1,zx} = T^{-1}\sum_{t=1}^{k_0^{(1)}} z_t x_t^{'}\theta_0^{(1)}$ and $f_{5,zx} = T^{-1}\sum_{t=k_0^{(1)}+1}^{k} z_t x_t^{'}\theta_0^{(2)}$.
From the limit of $\hat{\theta}_1(\lambda)$ given on the RHS of (C.10) on page 77, it follows that

$$\begin{aligned}
&T^{1/2}(\hat{\theta}_1(\lambda) - \theta_*^{(1)}(\lambda)) \\
&\quad= \mathcal{H}_{1,T}(\lambda)T^{-1/2}\Big\{\sum_{t=1}^{k} z_t u_t + \sum_{t=1}^{k_0^{(1)}}(z_t x_t^{'} - Q_{zx})\theta_0^{(1)} + \sum_{t=k_0^{(1)}+1}^{k}(z_t x_t^{'} - Q_{zx})\theta_0^{(2)}\Big\} \\
&\qquad + T^{1/2}[\mathcal{H}_{1,T}(\lambda) - \mathcal{H}_1(\lambda)]\Psi^{c2} + T^{1/2}\zeta^{c2} \\
&\quad= \mathcal{H}_{1,T}(\lambda)\big\{\mathcal{H}_{zu}(\lambda) + \mathcal{H}_{zx}(\lambda_0^{(1)})\theta_0^{(1)} + [\mathcal{H}_{zx}(\lambda) - \mathcal{H}_{zx}(\lambda_0^{(1)})]\theta_0^{(2)}\big\} \\
&\qquad + T^{1/2}[\mathcal{H}_{1,T}(\lambda) - \mathcal{H}_1(\lambda)]\Psi^{c2} + T^{1/2}\zeta_T^{c2},
\end{aligned} \tag{C.27}$$

where $\zeta_T^{c2} = \mathcal{H}_{1,T}(\lambda)\{f_{1,zx}^Q + f_{5,zx}^Q - \Psi^{c2}\}$, $f_{1,zx}^Q = T^{-1}\sum_{t=1}^{k_0^{(1)}} Q_{zx}\theta_0^{(1)}$,
$f_{5,zx}^Q = T^{-1}\sum_{t=k_0^{(1)}+1}^{k} Q_{zx}\theta_0^{(2)}$ and Ψ^{c2} is defined on page 77.
Similarly, for the estimated parameter after the break,

$$\hat{\theta}_2(\lambda) = \mathcal{H}_{2,T}(\lambda)\Big\{f_{6,zx} + f_{4,zx} + T^{-1}\sum_{t=k+1}^{T} z_t u_t\Big\}, \tag{C.28}$$

where $f_{6,zx} = T^{-1}\sum_{t=k+1}^{k_0^{(2)}} z_t x_t^{'}\theta_0^{(2)}$ and $f_{4,zx} = T^{-1}\sum_{t=k_0^{(2)}+1}^{T} z_t x_t^{'}\theta_0^{(3)}$.

Using the limit of $\hat{\theta}_2(\lambda)$ given on the RHS of (C.11) on page 77, we obtain

$$
\begin{aligned}
&T^{1/2}(\hat{\theta}_2(\lambda) - \theta_*^{(2)}(\lambda)) \\
&\quad = \mathcal{H}_{2,T}(\lambda)\big\{[\mathcal{H}_{zu}(1) - \mathcal{H}_{zu}(\lambda)] + [\mathcal{H}_{zx}(\lambda_0^{(2)}) - \mathcal{H}_{zx}(\lambda)]\theta_0^{(2)} \\
&\qquad + [\mathcal{H}_{zx}(1) - \mathcal{H}_{zx}(\lambda_0^{(2)})]\theta_0^{(3)}\big\} + T^{1/2}[\mathcal{H}_{2,T}(\lambda) - \mathcal{H}_2(\lambda)]\Psi^{d2} + T^{1/2}\zeta_T^{d2},
\end{aligned}
\tag{C.29}
$$

where $\zeta_T^{d2} = \mathcal{H}_{2,T}(\lambda)\{f_{6,zx}^Q + f_{4,zx}^Q - \Psi^{d2}\}$, $f_{6,zx}^Q = T^{-1}\sum_{t=k+1}^{k_0^{(2)}} Q_{zx}\theta_0^{(2)}$ and Ψ^{d2} is defined on page 77. The terms on the RHS of (C.27) and (C.29) are all bounded in probability which confirms Lemma 6 holds uniformly in $\lambda \in (\lambda_0^{(1)},\ \lambda_0^{(2)})$.

When $\lambda = \lambda_0^{(2)}$

For the first parameter, we have that

$$
\hat{\theta}_1(\lambda) = \mathcal{H}_{1,T}(\lambda_0^{(2)})\Big\{f_{1,zx} + f_{3,zx} + T^{-1}\sum_{t=1}^{k_0^{(2)}} z_t u_t\Big\}
$$

and with the limit given on the RHS of (C.15) on page 78 we obtain,

$$
\begin{aligned}
&T^{1/2}(\hat{\theta}_1(\lambda) - \theta_*^{(1)}(\lambda)) \\
&\quad = \mathcal{H}_{1,T}(\lambda_0^{(2)})T^{-1/2}\Big\{\sum_{t=1}^{k_0^{(2)}} z_t u_t + \sum_{t=1}^{k_0^{(1)}}\big(z_t x_t' - Q_{zx}\big)\theta_0^{(1)} + \sum_{t=k_0^{(1)}+1}^{k_0^{(2)}}\big(z_t x_t' - Q_{zx}\big)\theta_0^{(2)}\Big\} \\
&\qquad + T^{1/2}[\mathcal{H}_{1,T}(\lambda_0^{(2)}) - \mathcal{H}_1(\lambda_0^{(2)})]\Psi^{e2} + T^{1/2}\zeta^{e2} \\
&\quad = \mathcal{H}_{1,T}(\lambda_0^{(2)})\Big\{\mathcal{H}_{zu}(\lambda_0^{(2)}) + \mathcal{H}_{zx}(\lambda_0^{(1)})\theta_0^{(1)} + [\mathcal{H}_{zx}(\lambda_0^{(2)}) - \mathcal{H}_{zx}(\lambda_0^{(1)})]\theta_0^{(2)}\Big\} \\
&\qquad + T^{1/2}[\mathcal{H}_{1,T}(\lambda_0^{(2)}) - \mathcal{H}_1(\lambda_0^{(2)})]\Psi^{e2} + T^{1/2}\zeta_T^{e2},
\end{aligned}
$$

where $\zeta_T^{e2} = \mathcal{H}_{1,T}(\lambda_0^{(2)})\{f_{1,zx}^Q + f_{3,zx}^Q - \Psi^{e2}\}$ and $\Psi e2$ is given on page 78.
For the parameter after the break, $\hat{\theta}_2(\lambda) \xrightarrow{p} \theta_0^{(3)}$. Thus, in this case when $\lambda = \lambda_0^{(2)}$, we again see the Parameter Differences are $O_p(T^{-1/2})$.

When $\lambda > \lambda_0^{(2)}$

In this last scenario, the second parameter is similar to the preceding case above, while the first parameter is,

$$\hat{\theta}_1(\lambda) = \mathcal{H}_{1,T}(\lambda)\Big\{ f_{1,zx} + f_{3,zx} + f_{7,zx} + T^{-1}\sum_{t=1}^{k} z_t u_t \Big\}, \tag{C.30}$$

where $f_{7,zx} = T^{-1}\sum_{t=k_0^{(2)}}^{k} z_t x_t' \theta_0^{(3)}$. From the limit of $\hat{\theta}_1(\lambda)$ given on the RHS of (C.18) on page 78, we write

$$\begin{aligned} &T^{1/2}(\hat{\theta}_1(\lambda) - \theta_*^{(1)}(\lambda)) \\ &\quad = \mathcal{H}_{1,T}(\lambda)\Big\{ \mathcal{H}_{zu}(\lambda) + \mathcal{H}_{zx}(\lambda_0^{(1)})\theta_0^{(1)} + [\mathcal{H}_{zx}(\lambda_0^{(2)}) - \mathcal{H}_{zx}(\lambda_0^{(1)})]\theta_0^{(2)} \\ &\qquad + [\mathcal{H}_{zx}(\lambda) - \mathcal{H}_{zx}(\lambda_0^{(2)})]\theta_0^{(3)} \Big\} + T^{1/2}[\mathcal{H}_{1,T}(\lambda) - \mathcal{H}_1(\lambda)]\Psi^{f2} + T^{1/2}\zeta_T^{f2}, \end{aligned} \tag{C.31}$$

where $\zeta_T^{f2} = \mathcal{H}_{1,T}(\lambda)\{f_{1,zx}^Q + f_{3,zx}^Q + f_{7,zx}^Q - \Psi^{f2}\}$, $f_{7,zx}^Q = T^{-1}\sum_{k_0^{(2)}+1}^{k} Q_{zx}\theta_0^{(3)}$ and Ψ^{f2} is defined on page 78. The terms on the RHS of (C.31) are all bounded as in all the other four cases above.

This appendix shows the Parameter differences, $\mu_{*,T}(\lambda) - \mu_*(\lambda)$ given in (B.19) on page 44, are all $O_p(T^{-1/2})$ uniformly in λ for all $\lambda \in \Lambda$. In addition to the analysis of the Centre matrix in Appendix B.2, these proofs show the test statistic, $D_T^*(\lambda)$, uniformly converges to its nonstochastic limit, $D^*(\lambda)$ at the rate $O_p(T^{-1/2})$, thus implying Lemma 6.

C.3 Proof of Lemma 7

In this appendix, an expression for $\mathcal{E}$ is obtained and shown to be positive in all the five possibilities associated with a two break model. First, notice without loss of generality, the nonstochastic limits of the test statistics at the first and second break points given on the RHS of (3.3) and (3.5) on page 64 can also be expressed as,

$$D^*(\lambda_0^{(1)}) = \frac{1-\lambda_0^{(1)}}{\lambda_0^{(1)}}\Phi_A + \frac{\lambda_0^{(1)}}{1-\lambda_0^{(1)}}\Phi_B \tag{C.32}$$

and

$$D^*(\lambda_0^{(2)}) = \frac{1-\lambda_0^{(2)}}{\lambda_0^{(2)}}\Phi_A + \frac{\lambda_0^{(2)}}{1-\lambda_0^{(2)}}\Phi_B \tag{C.33}$$

respectively, where $\Phi_A = \phi_A^{'}\mathcal{B}\phi_A$, $\phi_A = w_1^* + w_2^*$, $\Phi_B = \phi_B^{'}\mathcal{B}\phi_B$, $\phi_B = \tilde{w}_1 + \tilde{w}_2$, $w_i^* = \lambda_0^{(i)}w_i$, $\tilde{w}_i = (1-\lambda_0^{(i)})w_i$ and $w_i = \theta_0^{(i)} - \theta_0^{(i+1)}$ for $i = 1, 2$.

From the RHS of the two expressions in (C.32) and (C.33), the implication of Assumption 15 on page 65 is that[2]

$$\frac{1-\lambda_0^{(1)}}{\lambda_0^{(1)}}\Phi_A + \frac{\lambda_0^{(1)}}{1-\lambda_0^{(1)}}\Phi_B > \frac{1-\lambda_0^{(2)}}{\lambda_0^{(2)}}\Phi_A + \frac{\lambda_0^{(2)}}{1-\lambda_0^{(2)}}\Phi_B.$$

Thus, it follows that

$$\left\{\frac{1-\lambda_0^{(1)}}{\lambda_0^{(1)}} - \frac{1-\lambda_0^{(2)}}{\lambda_0^{(2)}}\right\}\Phi_A > \left\{\frac{\lambda_0^{(2)}}{1-\lambda_0^{(2)}} - \frac{\lambda_0^{(1)}}{1-\lambda_0^{(1)}}\right\}\Phi_B$$

and with a little rearrangement, we obtain the inequality,

$$\frac{\Phi_A}{\lambda_0^{(1)}\lambda_0^{(2)}} > \frac{\Phi_B}{(1-\lambda_0^{(1)})(1-\lambda_0^{(2)})}. \tag{C.34}$$

When $\lambda < \lambda_0^{(1)}$

We use the RHS of (3.2) and (3.3) to deduce,

$$\begin{aligned} D^*(\lambda) - D^*(\lambda_0^{(1)}) &= \left\{\frac{\lambda}{1-\lambda} - \frac{\lambda_0^{(1)}}{1-\lambda_0^{(1)}}\right\}\Phi_B \\ &\leq \left\{\frac{\lambda}{1-\lambda_0^{(1)}} - \frac{\lambda_0^{(1)}}{1-\lambda_0^{(1)}}\right\}\Phi_B \\ &= (\lambda - \lambda_0^{(1)})\frac{\Phi_B}{1-\lambda_0^{(1)}} \\ &= -|\lambda - \lambda_0^{(1)}|\mathcal{E}_1, \end{aligned} \tag{C.35}$$

where $\mathcal{E}_1 = \Phi_B/(1-\lambda_0^{(1)}) > 0$ since $\Phi_B > 0$ and $\lambda_0^{(1)} < 1$.

[2]Assumption 15 imposes the first break dominates the second, that is, $D^*(\lambda_0^{(1)}) > D^*(\lambda_0^{(2)})$.

When $\lambda_0^{(1)} < \lambda \leq \lambda_0^{(2)}$

From the limits expressed on the RHS of (3.3) and (C.32) we write,

$$\begin{aligned} D^*(\lambda) - D^*(\lambda_0^{(1)}) &= \left\{ \frac{1-\lambda}{\lambda} - \frac{1-\lambda_0^{(1)}}{\lambda_0^{(1)}} \right\} \Phi_A + \left\{ \frac{\lambda}{1-\lambda} - \frac{\lambda_0^{(1)}}{1-\lambda_0^{(1)}} \right\} \Phi_B \\ &= \left\{ \frac{\lambda_0^{(1)} - \lambda}{\lambda \lambda_0^{(1)}} \right\} \Phi_A + \left\{ \frac{\lambda - \lambda_0^{(1)}}{(1-\lambda)(1-\lambda_0^{(1)})} \right\} \Phi_B. \end{aligned} \tag{C.36}$$

We analyse these two terms on the RHS of (C.36). For the first term, notice the numerator is negative because $\lambda_0^{(1)} < \lambda$. Also, since $\lambda \leq \lambda_0^{(2)}$, we say

$$\left\{ \frac{\lambda_0^{(1)} - \lambda}{\lambda \lambda_0^{(1)}} \right\} \Phi_A \leq \left\{ \frac{\lambda_0^{(1)} - \lambda}{\lambda_0^{(2)} \lambda_0^{(1)}} \right\} \Phi_A. \tag{C.37}$$

Likewise, for the second term,

$$\left\{ \frac{\lambda - \lambda_0^{(1)}}{(1-\lambda)(1-\lambda_0^{(1)})} \right\} \Phi_B \leq \left\{ \frac{\lambda - \lambda_0^{(1)}}{(1-\lambda_0^{(2)})(1-\lambda_0^{(1)})} \right\} \Phi_B. \tag{C.38}$$

From the analysis in (C.37) and (C.38), it follows that the LHS of (C.36),

$$\begin{aligned} D^*(\lambda) - D^*(\lambda_0^{(1)}) &\leq \left\{ \frac{\lambda_0^{(1)} - \lambda}{\lambda_0^{(2)} \lambda_0^{(1)}} \right\} \Phi_A + \left\{ \frac{\lambda - \lambda_0^{(1)}}{(1-\lambda_0^{(2)})(1-\lambda_0^{(1)})} \right\} \Phi_B \\ &= (\lambda_0^{(1)} - \lambda) \left\{ \frac{\Phi_A}{\lambda_0^{(1)} \lambda_0^{(2)}} - \frac{\Phi_B}{(1-\lambda_0^{(1)})(1-\lambda_0^{(2)})} \right\} \\ &= -|\lambda_0^{(1)} - \lambda| \mathcal{E}_2, \end{aligned} \tag{C.39}$$

where $\mathcal{E}_2 = \left\{ \frac{\Phi_A}{\lambda_0^{(1)} \lambda_0^{(2)}} - \frac{\Phi_B}{(1-\lambda_0^{(1)})(1-\lambda_0^{(2)})} \right\} > 0$, from (C.34).

When $\lambda > \lambda_0^{(2)}$

By adding and subtracting $D^*(\lambda_0^{(2)})$, we write

$$D^*(\lambda) - D^*(\lambda_0^{(1)}) = \{D^*(\lambda) - D^*(\lambda_0^{(2)})\} + \{D^*(\lambda_0^{(2)}) - D^*(\lambda_0^{(1)})\}. \tag{C.40}$$

Taking the terms in the first curly brackets on the RHS of (C.40), we deduce from the expressions on the RHS of (3.5) and (3.6) that

$$\begin{aligned} D^*(\lambda) - D^*(\lambda_0^{(2)}) &= \frac{1-\lambda}{\lambda}\Phi_A - \frac{1-\lambda_0^{(2)}}{\lambda_0^{(2)}}\Phi_A \\ &= \frac{\lambda_0^{(2)} - \lambda}{\lambda\lambda_0^{(2)}}\Phi_A \\ &\leq \frac{\lambda_0^{(2)} - \lambda}{(\lambda_0^{(2)})^2}\Phi_A \\ &= -|\lambda - \lambda_0^{(2)}|\mathcal{E}_3, \end{aligned} \tag{C.41}$$

where $\mathcal{E}_3 = \Phi_A/(\lambda_0^{(2)})^2 > 0$. For the terms in the second curly brackets, a similar argument to that used in (C.36) to (C.39) is followed to obtain

$$D^*(\lambda_0^{(2)}) - D^*(\lambda_0^{(1)}) = -|\lambda_0^{(2)} - \lambda_0^{(1)}|\mathcal{E}_2. \tag{C.42}$$

Therefore, when $\lambda > \lambda_0^{(2)}$, we use the RHS of (C.41) and (C.42) to conclude that

$$\begin{aligned} D^*(\lambda) - D^*(\lambda_0^{(1)}) &\leq -|\lambda - \lambda_0^{(2)}|\mathcal{E}_3 - |\lambda_0^{(2)} - \lambda_0^{(1)}|\mathcal{E}_2 \\ &\leq -|\lambda - \lambda_0^{(1)}|\mathcal{E}_4, \end{aligned}$$

where $\mathcal{E}_4 = min\{\mathcal{E}_2, \mathcal{E}_3\}$.

C.4 Proof of Lemma 8

The approach adopted in this proof follows closely that used in the one break model in Appendix B.4. Expressions for $D_T(k)$ and $D_T(k_0)$ are first obtained and the difference between them separated into a stochastic and a nonstochastic component. While the nonstochastic component is maximised near the true break, the stochastic component is shown to be asymptotically negligible as long as we are in the set V_3^*. Again, the proof covers only the case where $\lambda < \lambda_0^{(1)}$; the case of $\lambda > \lambda_0^{(1)}$ is similar and hence omitted. We use a '¨' above the terms to differentiate them from the one break model analysed in Appendix B.4.

Starting with the Parameter Difference, the parameter before the break is similar to that

in the one break case given in (B.37) on page 50, that is,

$$\hat{\theta}_1(k) = \theta_0^{(1)} + \ddot{\mathcal{H}}_{1,T}(k) \sum_{1}^{k} z_t u_t, \tag{C.43}$$

where $\ddot{\mathcal{H}}_{1,T}(k) = \ddot{M}_{1,T}(k)^{-1} \sum_{t=1}^{k} x_t z_t' \ddot{W}_{1,T}(k)$, $\ddot{M}_{1,T}(k) = \sum_{t=1}^{k} x_t z_t' \ddot{W}_{1,T}(k) \sum_{t=1}^{k} z_t x_t'$ and $\ddot{W}_{1,T}(k) = (1/k)C$.

In this two break model, the parameter after the break is,

$$\hat{\theta}_2(k) = \ddot{\mathcal{H}}_{2,T}(k) \left\{ \sum_{t=k+1}^{T} z_t u_t + \sum_{t=k+1}^{k_0^{(1)}} z_t x_t' \theta_0^{(1)} + \sum_{t=k_0^{(1)}+1}^{k_0^{(2)}} z_t x_t' \theta_0^{(2)} + \sum_{t=k_0^{(2)}+1}^{T} z_t x_t' \theta_0^{(3)} \right\}.$$

With some arrangements we write

$$\begin{aligned}
\hat{\theta}_2(k) = \ddot{\mathcal{H}}_{2,T}(k) \sum_{t=k+1}^{T} z_t u_t &+ \ddot{\mathcal{H}}_{2,T}(k) \left\{ \sum_{t=k+1}^{k_0^{(1)}} z_t x_t' \theta_0^{(1)} + \sum_{t=k_0^{(1)}+1}^{T} z_t x_t' \theta_0^{(1)} \right\} \\
&- \ddot{\mathcal{H}}_{2,T}(k) \left\{ \sum_{t=k_0^{(1)}+1}^{k_0^{(2)}} z_t x_t' \theta_0^{(1)} - \sum_{t=k_0^{(1)}+1}^{k_0^{(2)}} z_t x_t' \theta_0^{(2)} \right\} \\
&- \ddot{\mathcal{H}}_{2,T}(k) \left\{ \sum_{t=k_0^{(2)}+1}^{T} z_t x_t' \theta_0^{(1)} - \sum_{t=k_0^{(2)}+1}^{T} z_t x_t' \theta_0^{(3)} \right\} \\
= \ddot{\mathcal{H}}_{2,T}(k) \sum_{t=k+1}^{T} z_t u_t &+ \ddot{\mathcal{H}}_{2,T}(k) \sum_{t=k+1}^{T} z_t x_t' \theta_0^{(1)} \\
- \ddot{\mathcal{H}}_{2,T}(k) \sum_{t=k_0^{(1)}+1}^{k_0^{(2)}} z_t x_t' \big(\theta_0^{(1)} - \theta_0^{(2)}\big) &- \ddot{\mathcal{H}}_{2,T}(k) \sum_{t=k_0^{(2)}+1}^{T} z_t x_t' \big(\theta_0^{(1)} - \theta_0^{(3)}\big),
\end{aligned}$$

where $\ddot{\mathcal{H}}_{2,T}(k) = \ddot{M}_{2,T}(k)^{-1} \sum_{t=k+1}^{T} x_t z_t' \ddot{W}_{2,T}(k)$, $\ddot{W}_{2,T}(k) = (1/((T-k))C$ and

$\ddot{M}_{2,T}(k) = \sum_{t=k+1}^{T} x_t z_t' \ddot{W}_{2,T}(k) \sum_{t=k+1}^{T} z_t x_t'$. With a little rearrangement, this gives,

$$\hat{\theta}_2(k) = \ddot{\mathcal{H}}_{2,T}(k) \sum_{t=k+1}^{T} z_t u_t + \theta_0^{(1)} - \ddot{\mathcal{H}}_{2,T}(k) \sum_{t=k_0^{(1)}+1}^{k_0^{(2)}} z_t x_t' \big(\theta_0^{(1)} - \theta_0^{(2)}\big) - \ddot{\mathcal{H}}_{2,T}(k) \sum_{t=k_0^{(2)}+1}^{T} z_t x_t' \big(\theta_0^{(1)} - \theta_0^{(3)}\big). \tag{C.44}$$

From $\hat{\theta}_1(k)$ and $\hat{\theta}_2(k)$ given in (C.43) and (C.44) respectively, we get the Parameter Difference as,

$$\hat{\theta}_1(k) - \hat{\theta}_2(k) = \ddot{A}(k) + \ddot{B}(k),$$

where

$$\ddot{A}(k) = \ddot{\mathcal{H}}_{2,T}(k) \sum_{t=k_0^{(1)}+1}^{k_0^{(2)}} z_t x_t' \big(\theta_0^{(1)} - \theta_0^{(2)}\big) + \ddot{\mathcal{H}}_{2,T}(k) \sum_{t=k_0^{(2)}+1}^{T} z_t x_t' \big(\theta_0^{(1)} - \theta_0^{(3)}\big) \tag{C.45}$$

$$\ddot{B}(k) = \ddot{B}_1(k) - \ddot{B}_2(k) \tag{C.46}$$

$$\ddot{B}_1(k) = \ddot{\mathcal{H}}_{1,T}(k) \sum_{t=1}^{k} z_t u_t \tag{C.47}$$

$$\ddot{B}_2(k) = \ddot{\mathcal{H}}_{2,T}(k) \sum_{t=k+1}^{T} z_t u_t. \tag{C.48}$$

Thus, the test statistic is expressed uniformly in $\lambda < \lambda_0^{(1)}$ as

$$\begin{aligned} D_T(k) &= \big\{\ddot{A}(k) + \ddot{B}(k)\big\}' \ddot{M}_{*,T}(k)^{-1} \big\{\ddot{A}(k) + \ddot{B}(k)\big\} \\ &= \ddot{A}(k)' \ddot{M}_{*,T}(k)^{-1} \ddot{A}(k) + 2\ddot{A}(k)' \ddot{M}_{*,T}(k)^{-1} \ddot{B}(k) + \ddot{B}(k)' \ddot{M}_{*,T}(k)^{-1} \ddot{B}(k), \end{aligned} \tag{C.49}$$

where $\ddot{M}_{*,T}(k) = \ddot{M}_{1,T}(k)^{-1} + \ddot{M}_{2,T}(k)^{-1}$.

When $\lambda = \lambda_0^{(1)}$, the parameter before the break,

$$\hat{\theta}_1(k_0^{(1)}) = \theta_0^{(1)} + \ddot{\mathcal{H}}_{1,T}(k_0^{(1)}) \sum_{1}^{k_0^{(1)}} z_t u_t, \tag{C.50}$$

while that after the break,

$$\begin{aligned}
\hat{\theta}_2(k_0^{(1)}) &= \ddot{\mathcal{H}}_{2,T}(k_0^{(1)}) \Bigg\{ \sum_{t=k_0^{(1)}+1}^{T} z_t u_t + \sum_{t=k_0^{(1)}+1}^{k_0^{(2)}} z_t x_t' \theta_0^{(2)} + \sum_{t=k_0^{(2)}+1}^{T} z_t x_t' \theta_0^{(3)} \Bigg\} \\
&= \ddot{\mathcal{H}}_{2,T}(k_0^{(1)}) \Bigg\{ \sum_{t=k_0^{(1)}+1}^{T} z_t u_t + \sum_{t=k_0^{(1)}+1}^{k_0^{(2)}} z_t x_t' \theta_0^{(2)} + \sum_{t=k_0^{(2)}+1}^{T} z_t x_t' \theta_0^{(2)} \Bigg\} \\
&\quad - \ddot{\mathcal{H}}_{2,T}(k_0^{(1)}) \Bigg\{ \sum_{t=k_0^{(2)}+1}^{T} z_t x_t' \theta_0^{(2)} - \sum_{t=k_0^{(2)}+1}^{T} z_t x_t' \theta_0^{(3)} \Bigg\} \\
&= \ddot{\mathcal{H}}_{2,T}(k_0^{(1)}) \Bigg\{ \sum_{t=k_0^{(1)}+1}^{T} z_t u_t + \sum_{t=k_0^{(1)}+1}^{T} z_t x_t' \theta_0^{(2)} - \sum_{t=k_0^{(2)}+1}^{T} z_t x_t' \big(\theta_0^{(2)} - \theta_0^{(3)}\big) \Bigg\} \\
&= \ddot{\mathcal{H}}_{2,T}(k_0^{(1)}) \sum_{t=k_0^{(1)}+1}^{T} z_t u_t + \theta_0^{(2)} - \ddot{\mathcal{H}}_{2,T}(k_0^{(1)}) \sum_{t=k_0^{(2)}+1}^{T} z_t x_t' \big(\theta_0^{(2)} - \theta_0^{(3)}\big).
\end{aligned} \tag{C.51}$$

The Parameter Difference between $\hat{\theta}_1(k_0^{(1)})$ and $\hat{\theta}_2(k_0^{(1)})$ in (C.50) and (C.51) respectively is,

$$\hat{\theta}_1(k_0^{(1)}) - \hat{\theta}_2(k_0^{(1)}) = \ddot{A}(k_0^{(1)}) + \ddot{B}(k_0^{(1)}),$$

where

$$\ddot{A}(k_0^{(1)}) = \big(\theta_0^{(1)} - \theta_0^{(2)}\big) + \ddot{\mathcal{H}}_{2,T}(k_0^{(1)}) \sum_{t=k_0^{(2)}+1}^{T} z_t x_t' \big(\theta_0^{(2)} - \theta_0^{(3)}\big) \tag{C.52}$$

and $\ddot{B}(k_0^{(1)})$ is identical to (C.46) but evaluated at $k = k_0^{(1)}$.

We express the difference between the test statistic evaluated at k and $k_0^{(1)}$ as,

$$D_T(k) - D_T(k_0^{(1)}) = \ddot{\mathcal{A}}(k, k_0^{(1)}) + \ddot{h}(k, k_0^{(1)}), \tag{C.53}$$

where

$$\ddot{\mathcal{A}}(k,k_0^{(1)}) = \ddot{A}(k)'\ddot{M}_{*,T}(k)^{-1}\ddot{A}(k) - \ddot{A}(k_0^{(1)})'\ddot{M}_{*,T}(k_0^{(1)})^{-1}\ddot{A}(k_0^{(1)}) \tag{C.54}$$

$$\ddot{h}(k,k_0^{(1)}) = \ddot{\mathcal{B}}(k,k_0^{(1)}) + 2\ddot{\mathcal{C}}(k,k_0^{(1)}) \tag{C.55}$$

$$\ddot{\mathcal{B}}(k,k_0^{(1)}) = \ddot{B}(k)'\ddot{M}_{*,T}(k)^{-1}\ddot{B}(k) - \ddot{B}(k_0^{(1)})'\ddot{M}_{*,T}(k_0^{(1)})^{-1}\ddot{B}(k_0^{(1)}) \tag{C.56}$$

$$\ddot{\mathcal{C}}(k,k_0^{(1)}) = \ddot{A}(k)'\ddot{M}_{*,T}(k)^{-1}\ddot{B}(k) - \ddot{A}(k_0^{(1)})'\ddot{M}_{*,T}(k_0^{(1)})^{-1}\ddot{B}(k_0^{(1)}). \tag{C.57}$$

$\ddot{\mathcal{A}}(k,k_0^{(1)})$ and $\ddot{h}(k,k_0^{(1)})$ are referred to as the nonstochastic and stochastic parts respectively. Define

$$\ddot{\gamma}(k,k_0^{(1)}) = -\frac{\ddot{\mathcal{A}}(k,k_0^{(1)})}{|k_0^{(1)}-k|} \tag{C.58}$$

where $\ddot{\gamma}(k,k_0^{(1)}) > 0$ uniformly in $\lambda < \lambda_0^{(1)}$ as in the one break model on page 52. Then we have the following equality,

$$D_T(k) - D_T(k_0^{(1)}) = -|k_0^{(1)}-k|\ddot{\gamma}(k,k_0^{(1)}) + \ddot{h}(k,k_0^{(1)}). \tag{C.59}$$

By Remark 2 on page 34, the LHS of (C.59) must be non-negative, hence it must be that

$$\frac{\ddot{h}(k,k_0^{(1)})}{|k_0^{(1)}-k|} \geq \ddot{\gamma}(k). \tag{C.60}$$

Using (C.60), we state the following relationship with Lemma 4 holds,

$$P\left(\sup_{k\in V_3^*} D_T(k) \geq D_T(k_0^{(1)})\right) \leq P\left(\sup_{k\in V_3^*} \frac{\ddot{h}(k,k_0^{(1)})}{|k_0^{(1)}-k|} \geq \inf_{k\in V_3^*} \ddot{\gamma}(k)\right), \tag{C.61}$$

where $inf_{k\in V_3^*}\ddot{\gamma}(k)$ is bounded away from zero by Assumptions 1, 3 and 8.

Thus to prove Lemma 8, it suffices to show that for any fixed $\mathcal{F} > 0$,

$$P\left(\sup_{k\in V_3^*} \frac{\ddot{h}(k,k_0^{(1)})}{|k_0^{(1)}-k|} \geq \mathcal{F}\right) < \epsilon. \tag{C.62}$$

We now analyse the components of $\ddot{h}(k,k_0^{(1)})$ given in (C.55) and show that $\ddot{\mathcal{B}}(k,k_0^{(1)})$ and $\ddot{\mathcal{C}}(k,k_0^{(1)})$ are very small in probability when divided by $|k_0^{(1)}-k|$. First, recall the orders of magnitudes which are uniform in $\lambda \leq \lambda_0^{(1)}$ are similar to that of the one break model:

$\sum_{t=1}^{k} z_t x_t^{'} = O_p(T)$, $\frac{1}{k}\sum_{t=1}^{k} z_t x_t^{'} = O_p(1)$ and $\sum_{t=1}^{k} z_t u_t = O_p(T^{1/2})$.
Also, from Assumptions 1, 3, 4 and 8, it follows that similar to the case of the one break model given in (B.53) - (B.59), we have

$$\ddot{M}_{i,T}(\cdot) = O_p(T) \tag{C.63}$$
$$\ddot{M}_{*,T}(\cdot) = O_p(T^{-1}) \tag{C.64}$$
$$\ddot{\mathcal{H}}_{i,T}(\cdot) = O_p(T^{-1}) \tag{C.65}$$
$$\ddot{A}(\cdot) = O_p(1) \tag{C.66}$$
$$\ddot{B}(\cdot) = O_p(T^{-1/2}). \tag{C.67}$$

Secondly, as $\lambda < \lambda_0^{(1)}$, the following equalities are used in the analysis,
$\sum_{t=1}^{k_0^{(1)}}(\cdot) = \sum_{t=1}^{k}(\cdot) + \sum_{t=k+1}^{k_0^{(1)}}(\cdot)$ and $\sum_{t=k_0^{(1)}+1}^{T}(\cdot) = \sum_{t=k+1}^{T}(\cdot) - \sum_{t=k+1}^{k_0^{(1)}}(\cdot)$.
Starting with $\ddot{\mathcal{B}}(k, k_0^{(1)})$, use the order of magnitudes in (C.64) and (C.67) to obtain
$\ddot{B}(\cdot)^{'}\ddot{M}_{*,T}(\cdot)^{-1}\ddot{B}(\cdot) = O_p(T^{-1/2})O_p(T)O_p(T^{-1/2}) = O_p(1)$.
Therefore,

$$\frac{\ddot{\mathcal{B}}(k, k_0^{(1)})}{|k_0^{(1)} - k|} = \frac{O_p(1)}{|k_0^{(1)} - k|}. \tag{C.68}$$

For $\ddot{\mathcal{C}}(k, k_0^{(1)})$ given in (C.57), we rewrite it as

$$\ddot{\mathcal{C}}(k, k_0^{(1)}) = \ddot{\Upsilon}_1(k, k_0^{(1)}) + \ddot{\Upsilon}_2(k, k_0^{(1)}) + \ddot{\Upsilon}_3(k, k_0^{(1)}),$$

where

$$\ddot{\Upsilon}_1(k, k_0^{(1)}) = \ddot{A}(k)^{'}\{\ddot{M}_{*,T}(k)^{-1} - \ddot{M}_{*,T}(k_0^{(1)})^{-1}\}\ddot{B}(k_0^{(1)}) \tag{C.69a}$$
$$\ddot{\Upsilon}_2(k, k_0^{(1)}) = \ddot{A}(k)^{'}\ddot{M}_{*,T}(k)^{-1}\{\ddot{B}(k) - \ddot{B}(k_0^{(1)})\} \tag{C.69b}$$
$$\ddot{\Upsilon}_3(k, k_0^{(1)}) = \{\ddot{A}(k) - \ddot{A}(k_0^{(1)})\}^{'}\ddot{M}_{*,T}(k_0^{(1)})^{-1}\ddot{B}(k_0^{(1)}). \tag{C.69c}$$

For all $\ddot{\Upsilon}_i(k, k_0^{(1)}), i = 1, 2, 3$, we consider only the terms in the curly brackets as order of magnitudes for the individual terms are already given on the RHS of (C.63) to (C.67).
For $\ddot{\Upsilon}_1(k, k_0^{(1)})$,

$$\ddot{M}_{*,T}(k)^{-1} - \ddot{M}_{*,T}(k_0^{(1)})^{-1} = \ddot{M}_{*,T}(k_0^{(1)})^{-1}\{\ddot{M}_{*,T}(k_0^{(1)}) - \ddot{M}_{*,T}(k)\}\ddot{M}_{*,T}(k)^{-1}, \tag{C.70}$$

where the terms within the curly brackets,

$$\ddot{M}_{*,T}(k_0^{(1)}) - \ddot{M}_{*,T}(k) = \{\ddot{M}_{1,T}(k_0^{(1)})^{-1} - \ddot{M}_{1,T}(k)^{-1}\} + \{\ddot{M}_{2,T}(k_0^{(1)})^{-1} - \ddot{M}_{2,T}(k)^{-1}\}.$$

For $i = 1, 2$, we have

$$\ddot{M}_{i,T}(k_0^{(1)})^{-1} - \ddot{M}_{i,T}(k)^{-1} = \ddot{M}_{i,T}(k_0^{(1)})^{-1}\{\ddot{M}_{i,T}(k) - \ddot{M}_{i,T}(k_0^{(1)})\}\ddot{M}_{i,T}(k)^{-1}. \tag{C.71}$$

When $i = 1$, the terms within the curly brackets,

$$\begin{aligned}
\ddot{M}_{1,T}(k_0^{(1)}) - \ddot{M}_{1,T}(k) &= \frac{1}{k_0^{(1)}} \sum_{t=1}^{k_0^{(1)}} x_t z_t^{'} \sum_{t=1}^{k_0^{(1)}} z_t x_t^{'} - \frac{1}{k} \sum_{t=1}^{k} x_t z_t^{'} \sum_{t=1}^{k} z_t x_t^{'} \\
&= \frac{1}{k_0^{(1)}} \sum_{t=1}^{k_0^{(1)}} x_t z_t^{'} \sum_{t=1}^{k} z_t x_t^{'} + \frac{1}{k_0^{(1)}} \sum_{t=1}^{k_0^{(1)}} x_t z_t^{'} \sum_{t=k+1}^{k_0^{(1)}} z_t x_t^{'} \\
&\quad - \sum_{t=1}^{k_0^{(1)}} x_t z_t^{'} \frac{1}{k} \sum_{t=1}^{k} z_t x_t^{'} + \sum_{t=k+1}^{k_0^{(1)}} x_t z_t^{'} \frac{1}{k} \sum_{t=1}^{k} z_t x_t^{'} \\
&= \ddot{\mathcal{R}}_1 + \ddot{\mathcal{R}}_2 + \ddot{\mathcal{R}}_3,
\end{aligned}$$

where

$$\ddot{\mathcal{R}}_1 = \frac{1}{k_0^{(1)}} \sum_{t=1}^{k_0^{(1)}} x_t z_t^{'} \sum_{t=k+1}^{k_0^{(1)}} z_t x_t^{'} \tag{C.72a}$$

$$\ddot{\mathcal{R}}_2 = \sum_{t=k+1}^{k_0^{(1)}} x_t z_t^{'} \frac{1}{k} \sum_{t=1}^{k} z_t x_t^{'} \tag{C.72b}$$

$$\ddot{\mathcal{R}}_3 = \frac{1}{k_0^{(1)}} \sum_{t=1}^{k_0^{(1)}} x_t z_t^{'} \frac{1}{k} \sum_{t=1}^{k} z_t x_t^{'} (k - k_0^{(1)}). \tag{C.72c}$$

Likewise, when $i = 2$,

$$\begin{aligned}\ddot{M}_{2,T}(k_0^{(1)}) - \ddot{M}_{2,T}(k) &= \sum_{t=k_0^{(1)}+1}^{T} x_t z_t^{'} \frac{1}{T-k_0^{(1)}} \sum_{t=k_0^{(1)}+1}^{T} z_t x_t^{'} - \sum_{t=k+1}^{T} x_t z_t^{'} \frac{1}{T-k} \sum_{t=k+1}^{T} z_t x_t^{'} \\ &= \sum_{t=k_0^{(1)}+1}^{T} x_t z_t^{'} \frac{1}{T-k_0^{(1)}} \sum_{t=k+1}^{T} z_t x_t^{'} - \sum_{t=k_0^{(1)}+1}^{T} x_t z_t^{'} \frac{1}{T-k_0^{(1)}} \sum_{t=k+1}^{k_0^{(1)}} z_t x_t^{'} \\ &\quad - \sum_{t=k+1}^{k_0^{(1)}} x_t z_t^{'} \frac{1}{T-k} \sum_{t=k+1}^{T} z_t x_t^{'} - \sum_{t=k_0^{(1)}+1}^{T} x_t z_t^{'} \frac{1}{T-k} \sum_{t=k+1}^{T} z_t x_t^{'} \\ &= \ddot{\mathcal{S}}_1 + \ddot{\mathcal{S}}_2 + \ddot{\mathcal{S}}_3,\end{aligned}$$

where

$$\ddot{\mathcal{S}}_1 = -\frac{1}{T-k_0^{(1)}} \sum_{t=k_0^{(1)}+1}^{T} x_t z_t^{'} \sum_{t=k+1}^{k_0^{(1)}} z_t x_t^{'} \tag{C.73a}$$

$$\ddot{\mathcal{S}}_2 = -\sum_{t=k+1}^{k_0^{(1)}} x_t z_t^{'} \frac{1}{T-k} \sum_{t=k+1}^{T} z_t x_t^{'} \tag{C.73b}$$

$$\ddot{\mathcal{S}}_3 = \frac{1}{T-k_0^{(1)}} \sum_{t=k_0^{(1)}+1}^{T} x_t z_t^{'} \frac{1}{T-k} \sum_{t=k+1}^{T} z_t x_t^{'} (k_0^{(1)} - k). \tag{C.73c}$$

From Assumptions 1 and 3, $\ddot{\mathcal{R}}_i$ and $\ddot{\mathcal{S}}_i$ are $O_p(|k_0^{(1)} - k|)$ for $i = 1, 2, 3$. Consequently, the RHS of (C.71) is $O_p(T^{-1})O_p(|k_0^{(1)} - k|)O_p(T^{-1}) = O_p(|k_0^{(1)} - k|)O_p(T^{-2})$, while the RHS of (C.70) is $O_p(T)O_p(|k_0^{(1)} - k|)O_p(T^{-2})O_p(T) = O_p(|k_0^{(1)} - k|)$.

It follows from (C.66) and (C.67) that, $\ddot{\Upsilon}_1(k, k_0^{(1)}) = O_p(1)O_p(|k_0^{(1)} - k|)O_p(T^{-1/2})$.

Thus, dividing by $|k_0^{(1)} - k|$ we conclude,

$$\frac{\ddot{\Upsilon}_1(k, k_0^{(1)})}{|k_0^{(1)} - k|} = \frac{O_p(|k_0^{(1)} - k|)O_p(T^{-1/2})}{|k_0^{(1)} - k|} = O_p(T^{-1/2}). \tag{C.74}$$

For $\ddot{\Upsilon}_2(k, k_0^{(1)})$ given in (C.69b), we write as[3]

$$\ddot{B}(k) - \ddot{B}(k_0^{(1)}) = \{\ddot{B}_1(k) - \ddot{B}_1(k_0^{(1)})\} + \{\ddot{B}_2(k) - \ddot{B}_2(k_0^{(1)})\}. \tag{C.75}$$

[3] $\ddot{B}_1(\cdot)$ and $\ddot{B}_2(\cdot)$ are defined in (C.47) and (C.48) respectively.

For the terms in the first curly brackets on the RHS of (C.75), we write

$$\begin{aligned}\ddot{B}_1(k) - \ddot{B}_1(k_0^{(1)}) &= \ddot{M}_{1,T}(k)^{-1}\frac{1}{k}\sum_{t=1}^{k} x_t z_t^{'}\sum_{t=1}^{k} z_t u_t - \ddot{M}_{1,T}(k_0^{(1)})^{-1}\frac{1}{k_0^{(1)}}\sum_{t=1}^{k_0^{(1)}} x_t z_t^{'}\sum_{t=1}^{k_0^{(1)}} z_t u_t \\ &= \left\{\ddot{M}_{1,T}(k)^{-1} - \ddot{M}_{1,T}(k_0^{(1)})^{-1}\right\}\frac{1}{k}\sum_{t=1}^{k} x_t z_t^{'}\sum_{t=1}^{k} z_t u_t \\ &\quad + \ddot{M}_{1,T}(k_0^{(1)})^{-1}\left\{\frac{1}{k}\sum_{t=1}^{k} x_t z_t^{'} - \frac{1}{k_0^{(1)}}\sum_{t=1}^{k_0^{(1)}} x_t z_t^{'}\right\}\sum_{t=1}^{k} z_t u_t \\ &\quad - \ddot{M}_{1,T}(k_0^{(1)})^{-1}\frac{1}{k_0^{(1)}}\sum_{t=1}^{k_0^{(1)}} x_t z_t^{'}\sum_{t=k+1}^{k_0^{(1)}} z_t u_t.\end{aligned}$$

From the analysis carried out in (C.71) to (C.72c), we write

$$\begin{aligned}&\ddot{B}_1(k) - \ddot{B}_1(k_0^{(1)}) \\ &\quad = \ddot{M}_{1,T}(k_0^{(1)})^{-1}\left\{\ddot{\mathcal{R}}_1 + \ddot{\mathcal{R}}_2 + \ddot{\mathcal{R}}_3\right\}\ddot{M}_{1,T}(k)^{-1}\frac{1}{k}\sum_{t=1}^{k} x_t z_t^{'}\sum_{t=1}^{k} z_t u_t && \text{(C.76)} \\ &\quad + \ddot{M}_{1,T}(k_0^{(1)})^{-1}\left\{\frac{1}{k}\sum_{t=1}^{k} x_t z_t^{'} - \frac{1}{k_0^{(1)}}\sum_{t=1}^{k_0^{(1)}} x_t z_t^{'}\right\}\sum_{t=1}^{k} z_t u_t && \text{(C.77)} \\ &\quad - \ddot{H}_{1,T}(k_0^{(1)})\sum_{t=k+1}^{k_0^{(1)}} z_t u_t. && \text{(C.78)}\end{aligned}$$

Using a similar procedure for the second curly brackets in (C.75) yields,

$$\begin{aligned}&\ddot{B}_2(k) - \ddot{B}_2(k_0^{(1)}) \\ &\quad = \ddot{M}_{2,T}(k_0^{(1)})^{-1}\left\{\ddot{\mathcal{S}}_1 + \ddot{\mathcal{S}}_2 + \ddot{\mathcal{S}}_3\right\}\ddot{M}_{2,T}(k)^{-1}\frac{1}{T-k}\sum_{t=k+1}^{T} x_t z_t^{'}\sum_{t=k+1}^{T} z_t u_t && \text{(C.79)} \\ &\quad + \ddot{M}_{2,T}(k_0^{(1)})^{-1}\left\{\frac{1}{T-k}\sum_{t=k+1}^{T} x_t z_t^{'} - \frac{1}{T-k_0^{(1)}}\sum_{t=k_0^{(1)}+1}^{T} x_t z_t^{'}\right\}\sum_{t=k+1}^{T} z_t u_t && \text{(C.80)} \\ &\quad + \ddot{\mathcal{H}}_{2,T}(k_0^{(1)})\sum_{t=k+1}^{k_0^{(1)}} z_t u_t. && \text{(C.81)}\end{aligned}$$

Notice that the RHS of (C.76) to (C.81) are analogous to to their counterparts (B.67)

to (B.72) in the one break model on page 56. Hence, the same orders of magnitudes extend to them. The RHS of (C.76) and (C.79) are $O_p(T^{-1})O_p(|k-k_0|)O_p(T^{-1/2})$, the RHS of (C.77) and (C.80) are both $O_p(T^{-1})$ while the RHS of (C.78) and (C.81) are $O_p(T^{-1})\sum_{t=k+1}^{k_0} z_t u_t$.

From the order of magnitudes given in (C.66) and (C.64) we can deduce for $\ddot{\Upsilon}_2(k, k_0^{(1)})$ defined in (C.69b) and (C.75),

$$\begin{aligned}\ddot{\Upsilon}_2(k, k_0^{(1)}) &= O_p(1)O_p(T)\big\{O_p(|k_0^{(1)}-k|)O_p(T^{-3/2}) + O_p(T^{-1}) \\ &\qquad + O_p(T^{-1})\sum_{t=k+1}^{k_0^{(1)}} z_t u_t\big\}, \\ &= O_p(|k_0^{(1)}-k|)O_p(T^{-1/2}) + O_p(1) + \ddot{\Xi}_{1,T}\sum_{t=k+1}^{k_0^{(1)}} z_t u_t,\end{aligned}$$

where, similar to the one break model given on page 57, $\ddot{\Xi}_{1,T} = \ddot{\Xi}_1 + \ddot{\Xi}_k$, $\ddot{\Xi}_1 < \infty$ is a matrix of constants and $\ddot{\Xi}_k$ is $O_p(T^{-1/2})$. Dividing by $|k_0^{(1)}-k|$, we therefore conclude that,

$$\frac{\ddot{\Upsilon}_2(k, k_0^{(1)})}{|k_0^{(1)}-k|} = O_p(T^{-1/2}) + \frac{O_p(1)}{|k_0^{(1)}-k|} + \frac{\ddot{\Xi}_1}{|k_0^{(1)}-k|}\sum_{t=k+1}^{k_0^{(1)}} z_t u_t + \frac{O_p(T^{-1/2})}{|k_0^{(1)}-k|}\sum_{t=k+1}^{k_0^{(1)}} z_t u_t. \tag{C.82}$$

Lastly, for $\ddot{\Upsilon}_3(k, k_0^{(1)})$ given in (C.69c), we write[4]

$$\begin{aligned}\ddot{A}(k) - \ddot{A}(k_0^{(1)}) &= \ddot{\mathcal{H}}_{2,T}(k)\sum_{t=k_0^{(1)}+1}^{k_0^{(2)}} z_t x_t'\big(\theta_0^{(1)} - \theta_0^{(2)}\big) + \ddot{\mathcal{H}}_{2,T}(k)\sum_{t=k_0^{(2)}+1}^{T} z_t x_t'\big(\theta_0^{(1)} - \theta_0^{(3)}\big) \\ &\quad - \big(\theta_0^{(1)} - \theta_0^{(2)}\big) - \ddot{\mathcal{H}}_{2,T}(k_0^{(1)})\sum_{t=k_0^{(2)}+1}^{T} z_t x_t'\big(\theta_0^{(2)} - \theta_0^{(3)}\big).\end{aligned}$$

[4] $\ddot{A}(k)$ and $\ddot{A}(k_0^{(1)})$ are defined in (C.52) and (C.45) respectively.

We make some rearrangements as follows,

$$\begin{aligned}
\ddot{A}(k) - \ddot{A}(k_0^{(1)}) &= \ddot{\mathcal{H}}_{2,T}(k) \sum_{t=k_0^{(1)}+1}^{k_0^{(2)}} z_t x_t'\big(\theta_0^{(1)} - \theta_0^{(2)}\big) - \ddot{\mathcal{H}}_{2,T}(k) \sum_{t=k_0^{(1)}+1}^{k_0^{(2)}} z_t x_t'\big(\theta_0^{(1)} - \theta_0^{(2)}\big) \\
&\quad + \ddot{\mathcal{H}}_{2,T}(k) \sum_{t=k_0^{(2)}+1}^{T} z_t x_t'\big(\theta_0^{(1)} - \theta_0^{(3)}\big) - \ddot{\mathcal{H}}_{2,T}(k) \sum_{t=k_0^{(2)}+1}^{T} z_t x_t'\big(\theta_0^{(1)} - \theta_0^{(2)}\big) \\
&\quad - \ddot{\mathcal{H}}_{2,T}(k_0^{(1)}) \sum_{t=k_0^{(2)}+1}^{T} z_t x_t'\big(\theta_0^{(2)} - \theta_0^{(3)}\big) - \ddot{\mathcal{H}}_{2,T}(k) \sum_{t=k+1}^{k_0^{(1)}} z_t x_t'\big(\theta_0^{(1)} - \theta_0^{(2)}\big) \\
&= \ddot{\mathcal{H}}_{2,T}(k) \sum_{t=k_0^{(2)}+1}^{T} z_t x_t'\big(\theta_0^{(2)} - \theta_0^{(3)}\big) \\
&\quad - \ddot{\mathcal{H}}_{2,T}(k_0^{(1)}) \sum_{t=k_0^{(2)}+1}^{T} z_t x_t'\big(\theta_0^{(2)} - \theta_0^{(3)}\big) - \ddot{\mathcal{H}}_{2,T}(k) \sum_{t=k+1}^{k_0^{(1)}} z_t x_t'\big(\theta_0^{(1)} - \theta_0^{(2)}\big).
\end{aligned}$$

Therefore we can now write,

$$\ddot{A}(k) - \ddot{A}(k_0^{(1)}) = [\ddot{\mathcal{H}}_{2,T}(k) - \ddot{\mathcal{H}}_{2,T}(k_0^{(1)})] \sum_{t=k_0^{(2)}+1}^{T} z_t x_t'\big(\theta_0^{(2)} - \theta_0^{(3)}\big) \tag{C.83}$$

$$- \ddot{\mathcal{H}}_{2,T}(k) \sum_{t=k+1}^{k_0^{(1)}} z_t x_t'\big(\theta_0^{(1)} - \theta_0^{(2)}\big), \tag{C.84}$$

where $\ddot{\mathcal{H}}_{2,T}(k) - \ddot{\mathcal{H}}_{2,T}(k_0^{(1)})$ on the RHS of (C.83) has identical orders of magnitude to (B.78) and (B.79) in the one break model[5].

Thus, we conclude the RHS of (C.83) is $O_p(T^{-1})O_p(|k_0^{(1)} - k|) + O_p(T^{-1/2})$ while that of (C.84) is $O_p(T^{-1})O_p(|k_0^{(1)} - k|)$.

From the orders of magnitudes given in (C.64) and (C.67), it follows that $\ddot{\Upsilon}_3(k, k_0^{(1)})$ defined in (C.69c),

$$\begin{aligned}
\ddot{\Upsilon}_3(k, k_0^{(1)}) &= \big\{O_p(T^{-1})O_p(|k_0^{(1)} - k|) + O_p(T^{-1/2}) + O_p(T^{-1})O_p(|k_0^{(1)} - k|)\big\} \\
&\quad \times O_p(T)O_p(T^{-1/2}) \\
&= O_p(T^{-1/2})O_p(|k_0^{(1)} - k|) + O_p(1) + O_p(T^{-1/2})O_p(|k_0^{(1)} - k|).
\end{aligned}$$

[5]This order as given on page 58 is $O_p(T^{-2})O_p(|k_0^{(1)} - k|) + O_p(T^{-3/2})$.

Hence, we conclude

$$\frac{\ddot{\Upsilon}_3(k, k_0^{(1)})}{|k_0^{(1)} - k|} = \frac{O_p(1)}{|k_0^{(1)} - k|} + O_p(T^{-1/2}). \tag{C.85}$$

Combining $\ddot{\Upsilon}_1(k, k_0^{(1)})$, $\ddot{\Upsilon}_2(k, k_0^{(1)})$ and $\ddot{\Upsilon}_3(k, k_0^{(1)})$ as given in (C.74), (C.82) and (C.85) respectively, we now deduce the order of magnitude of $\ddot{\mathcal{C}}(k, k_0^{(1)})$ defined in (C.57) as

$$\frac{\ddot{\mathcal{C}}(k, k_0^{(1)})}{|k_0^{(1)} - k|} = \frac{\ddot{\Xi}_1}{|k_0^{(1)} - k|} \sum_{t=k+1}^{k_0^{(1)}} z_t u_t + \frac{O_p(1)}{|k_0^{(1)} - k|}, \tag{C.86}$$

where the $O_p(T^{-1/2})$ terms are again left out as they are asymptotically negligible.

With the orders of magnitudes of $\ddot{\mathcal{B}}(k, k_0^{(1)})$ and $\ddot{\mathcal{C}}(k, k_0^{(1)})$ given in (C.68) and (C.86) respectively, the conclusion of the proofs for the stochastic component, $\ddot{h}(k, k_0^{(1)})$, is identical to that of the one break model given in (B.82) and (B.83) in Appendix B.4. In particular, one only needs to substitute k for $k_0^{(1)}$ in Υ_5 defined in (B.83) and the result follows through, hence it is not repeated here.

C.5 Proof of Lemma 9

In this appendix, an $\mathcal{E} > 0$ is obtained when $D^*(\lambda_0^{(1)}) = D^*(\lambda_0^{(2)})$. This $\mathcal{E}$ is shown to be positive uniformly in λ for $\lambda \leq \lambda_0^*$, where λ_0^* is as defined on page 69.

When $k < k_0^{(1)}$

Since $\lambda_0^{(1)}$ is strictly less than $\lambda_0^{(2)}$, then the analysis carried out when $D^*(\lambda_0^{(1)}) > D^*(\lambda_0^{(2)})$ under Assumption 15 holds here as well when $D^*(\lambda_0^{(1)}) = D^*(\lambda_0^{(2)})$. Hence we only state the result which is identical to that given on the RHS of (C.35) on page 85, $D^*(\lambda) - D^*(\lambda_0^{(1)}) = -|\lambda - \lambda_0^{(1)}|\mathcal{E}_1$, where $\mathcal{E}_1 = \Phi_B/(1 - \lambda_0^{(1)}) > 0$.

When $k_0^{(1)} < k \leq k_0^*$

When $D^*(\lambda_0^{(1)}) = D^*(\lambda_0^{(2)})$, the inequality in (C.34) now becomes,

$$\frac{\Phi_A}{\lambda_0^{(1)}\lambda_0^{(2)}} = \frac{\Phi_B}{(1 - \lambda_0^{(1)})(1 - \lambda_0^{(2)})}. \tag{C.87}$$

From the RHS of (C.36), it follows that

$$\begin{aligned} D^*(\lambda) - D^*(\lambda_0^{(1)}) \\ &= \left\{ \frac{\lambda_0^{(1)} - \lambda}{\lambda \lambda_0^{(1)}} \right\} \Phi_A + \left\{ \frac{\lambda - \lambda_0^{(1)}}{(1-\lambda)(1-\lambda_0^{(1)})} \right\} \Phi_B \\ &= (\lambda_0^{(1)} - \lambda) \left\{ \frac{\Phi_A}{\lambda \lambda_0^{(1)}} - \frac{\Phi_B}{(1-\lambda)(1-\lambda_0^{(1)})} \right\} \\ &= (\lambda_0^{(1)} - \lambda) \left\{ \frac{\lambda_0^{(2)}}{\lambda} \left(\frac{\Phi_A}{\lambda_0^{(1)} \lambda_0^{(2)}} \right) - \frac{1-\lambda_0^{(2)}}{1-\lambda} \left(\frac{\Phi_B}{(1-\lambda_0^{(1)})(1-\lambda_0^{(2)})} \right) \right\}. \end{aligned} \tag{C.88}$$

Using the equality in (C.87), it follows that the terms in the large brackets on the RHS of (C.88) are equal. Hence we choose either of them and write,

$$\begin{aligned} D^*(\lambda) - D^*(\lambda_0^{(1)}) &= (\lambda_0^{(1)} - \lambda) \left\{ \frac{\lambda_0^{(2)}}{\lambda} - \frac{1-\lambda_0^{(2)}}{1-\lambda} \right\} \Phi_A^* \\ &= (\lambda_0^{(1)} - \lambda) \left\{ \frac{\lambda_0^{(2)} - \lambda}{\lambda(1-\lambda)} \right\} \Phi_A^*, \end{aligned} \tag{C.89}$$

where $\Phi_A^* = \Phi_A / (\lambda_0^{(1)} \lambda_0^{(2)})$. For all $k \leq k_0^* = 0.5(k_0^{(1)} + k_0^{(2)})$, it implies $2\lambda \leq \lambda_0^{(1)} + \lambda_0^{(2)}$. Thus for the terms within the curly brackets on the RHS of (C.89),

$$\begin{aligned} \frac{2\lambda_0^{(2)} - 2\lambda}{2\lambda(1-\lambda)} &\leq \frac{2\lambda_0^{(2)} - \lambda_0^{(1)} - \lambda_0^{(2)}}{2\lambda(1-\lambda)} \\ &\leq \frac{\lambda_0^{(2)} - \lambda_0^{(1)}}{2\lambda_0^{(1)}(1-\lambda_0^{(2)})} \end{aligned}$$

Therefore, it follows that

$$D^*(\lambda) - D^*(\lambda_0^{(1)}) \leq -|\lambda - \lambda_0^{(1)}| \mathcal{E}_5, \tag{C.90}$$

where $\mathcal{E}_5 = \left\{ (\lambda_0^{(2)} - \lambda_0^{(1)}) / (2\lambda_0^{(1)}(1-\lambda_0^{(2)}) \right\} \Phi_A^* > 0$.

The analysis in this appendix covers the first equation, (3.15) on page 69, in Lemma 9 where $\lambda \leq \lambda_0^*$. By symmetry, a similar evaluation can be carried out for the second equation (3.16) when $\lambda \geq \lambda_0^*$.

C.6 Proof of Lemma 10

The proof of this lemma is identical to that carried out for Lemma 8 in Appendix C.4. With the consistency of the break fraction estimators $\hat{\lambda}_i^*$ established for $\lambda_0^{(i)}$, $i = 1, 2$, then (3.15) and (3.16) on page 69 in Lemma 9 are used and the same results obtained in Appendix C.4 follow through. Hence, the proofs are not repeated here.

C.7 Proof of Proposition 7

This appendix presents the generic formula for the nonstochastic limit of $D_T^*(\lambda)$ in a multiple break model. Recall the generic formula as presented in (3.21) on page 71 is

$$D^*(\lambda) = \frac{1-\lambda}{\lambda}\Phi_A(j) + \frac{\lambda}{1-\lambda}\Phi_B(j) + \Phi_C(j),$$

where $\Phi_A(j) = \phi_A(j)'\mathcal{B}\phi_A(j)$; $\phi_A(j) = \sum_{i=1}^{j} \lambda_0^{(i)} w_i$; $\Phi_B(j) = \phi_B(j)'\mathcal{B}\phi_B(j)$; $\phi_B(j) = \sum_{i=j+1}^{m}(1-\lambda_0^{(i)})w_i$; $\Phi_C(j) = \phi_A(j)'\mathcal{B}\phi_B(j)$; $w_i = \theta_0^{(i)} - \theta_0^{(i+1)}$; i is the break point being estimated; j is the number of true breaks before or at this candidate break point and m is the total number of true breaks in the model. We prove the results below for the $m = 3$ case, however, the results go through for any finite m.

The break fractions and the model's parameters are represented as $\lambda_0^{(i)}$ and $\theta_0^{(j)}$ respectively, for $i = 1, 2, 3$ and $j = 1, \ldots, 4$. In this three break model, there are seven possible expressions for $D^*(\lambda)$ across the range of Λ and we show how each of these expressions crystalises into the generic formula displayed above. The approach used is similar to that used in the one break and two break models in Appendices B.1 and C.1 respectively.

First, recall the test statistic,

$$D_T(\lambda) = T(\hat{\theta}_{1,T}(\lambda) - \hat{\theta}_{2,T}(\lambda))' M_{*,T}(\lambda)^{-1} (\hat{\theta}_{1,T}(\lambda) - \hat{\theta}_{2,T}(\lambda)),$$

where $M_{*,T}(\cdot) = M_{1,T}(\cdot)^{-1} + M_{2,T}(\cdot)^{-1}$, $M_{i,T}(\cdot) = G_{i,T}(\cdot)'W_{i,T}(\cdot)G_{i,T}(\cdot)$ and $G_{i,T}(\cdot) = T^{-1}Z_{i,T}(\cdot)'X_{i,T}(\cdot)$, for $i = 1, 2$.

Secondly, recall $(\hat{\theta}_{1,T}(\lambda) - \hat{\theta}_{2,T}(\lambda))$ is the Parameter Difference and $M_{*,T}(\lambda)^{-1}$ is the Centre Matrix. Thirdly, recall that since the Jacobian is stable, then the Centre Matrix does not change in all seven cases. As given in (B.12) in Appendix B.1, $M_{*,T}(\cdot) =$

$M_{1,T}(\cdot)^{-1} + M_{2,T}(\cdot)^{-1}$, where

$$\begin{aligned} M_{1,T}(\lambda) &= G_{1,T}(\lambda)' W_{1,T}(\lambda) G_{1,T}(\lambda) \\ &\xrightarrow{p} (\lambda Q_{xz})(\lambda^{-1} C)(\lambda Q_{zx}) \\ &= \lambda (Q_{xz} C Q_{zx}). \end{aligned}$$

Similarly, $M_{2,T}(\lambda) = G_{2,T}(\lambda)' W_{2,T}(\lambda) G_{2,T}(\lambda) \xrightarrow{p} (1-\lambda)(Q_{xz} C Q_{zx})$.

Therefore, the Centre Matrix, $M_{*,T}(\lambda)^{-1} \xrightarrow{p} \lambda(1-\lambda)\mathcal{B}$, where $\mathcal{B} = Q_{xz} C Q_{zx}$.

Fourthly, the proofs concentrate on the asymptotic properties of the estimated parameters before and after the candidate break point. It is worth noting that in all the seven cases below, as given in Appendix C.1,

$\mathcal{H}_{1,T}(\lambda) = M_{1,T}(\cdot)^{-1} G_{1,T}(\cdot)' W_{1,T}(\cdot) \xrightarrow{p} \mathcal{H}_1(\lambda)$ and

$\mathcal{H}_{2,T}(\lambda) = M_{2,T}(\cdot)^{-1} G_{2,T}(\cdot)' W_{2,T}(\cdot) \xrightarrow{p} \mathcal{H}_2(\lambda)$, where

$\mathcal{H}_1(\lambda) = \{\lambda(Q_{xz} C Q_{zx})\}^{-1} Q_{xz} C$ and $\mathcal{H}_2(\lambda) = \{(1-\lambda)(Q_{xz} C Q_{zx})\}^{-1} Q_{xz} C$.

Lastly, recall that from Assumptions 1, 3 and 8, the Jacobian, $\sum_{t=1}^{k} z_t x_t' \xrightarrow{p} \lambda Q_{zx}$.

For simplicity, we present the proof following these four steps:

(i) The formula for the GMM parameter before the break is stated and its limiting form derived based on Assumptions 1, 3 and 8.

(ii) Step (i) above is performed for the parameter after the break.

(iii) The limits of the Parameter Difference are obtained by subtracting the limit of the second parameter obtained in (ii) from that of the first parameter obtained in (i).

(iv) The nonstochastic limit, $D^*(\lambda)$, is now derived by combining the limit of the Parameter Difference obtained in (iii) with the limit of the Centre Matrix.

Remark 4. *There is no true break point in the first regime when* $\lambda < \lambda_0^{(1)}$ *hence* $j = 0$ *and we say* $\phi_A(j) = 0$. *Similarly, since* $j = m$ *in the last regime when* $\lambda > \lambda_0^{(m)}$ *we say* $\phi_B(j) = 0$ *because no true break point exists there.*

Case 1: When $0 < \lambda < \lambda_0^{(1)} < \lambda_0^{(2)} < \lambda_0^{(3)} < 1$

The parameter before the break,

$$\begin{aligned} \hat{\theta}_1(\lambda) &= \mathcal{H}_{1,T}(\lambda) \left\{ \sum_{t=1}^{k} z_t u_t + \sum_{t=1}^{k} z_t x_t' \theta_0^{(1)} \right\} \\ &\xrightarrow{p} \theta_0^{(1)}. \end{aligned}$$

The parameter after the break,

$$\hat{\theta}_2(\lambda) = \mathcal{H}_{2,T}(\lambda)\Bigg\{ \sum_{t=k+1}^{k_0^{(1)}} z_t x_t^{'}\theta_0^{(1)} + \sum_{t=k_0^{(1)}+1}^{k_0^{(2)}} z_t x_t^{'}\theta_0^{(2)} + \sum_{t=k_0^{(2)}+1}^{k_0^{(3)}} z_t x_t^{'}\theta_0^{(3)} + \sum_{t=k_0^{(3)}+1}^{T} z_t x_t^{'}\theta_0^{(4)} + \sum_{t=k+1}^{T} z_t u_t \Bigg\}$$

$$\xrightarrow{p} \frac{1}{1-\lambda}\big\{(\lambda_0^{(1)}-\lambda)\theta_0^{(1)} + (\lambda_0^{(2)}-\lambda_0^{(1)})\theta_0^{(2)} + (\lambda_0^{(3)}-\lambda_0^{(2)})\theta_0^{(3)} + (1-\lambda_0^{(3)})\theta_0^{(4)}\big\}.$$

The Parameter Difference,

$$\begin{aligned}
&\hat{\theta}_{1,T}(\lambda) - \hat{\theta}_{2,T}(\lambda) \\
&\xrightarrow{p} \frac{1}{1-\lambda}\big\{(1-\lambda)\theta_0^{(1)} - (\lambda_0^{(1)}-\lambda)\theta_0^{(1)} - (\lambda_0^{(2)}-\lambda_0^{(1)})\theta_0^{(2)} \\
&\quad - (\lambda_0^{(3)}-\lambda_0^{(2)})\theta_0^{(3)} - (1-\lambda_0^{(3)})\theta_0^{(4)}\big\} \\
&= \frac{1}{1-\lambda}\big\{(1-\lambda_0^{(1)})(\theta_0^{(1)}-\theta_0^{(2)}) + (1-\lambda_0^{(2)})(\theta_0^{(2)}-\theta_0^{(3)}) + (1-\lambda_0^{(3)})(\theta_0^{(3)}-\theta_0^{(4)})\big\} \\
&= \frac{1}{1-\lambda}\sum_{i=j+1}^{m}(1-\lambda_0^{(i)})(\theta_0^{(i)}-\theta_0^{(i+1)}) \\
&= \frac{1}{1-\lambda}\sum_{i=j+1}^{m}\tilde{w}_i \\
&= \frac{1}{1-\lambda}\phi_B(j).
\end{aligned}$$

Thus, the nonstochastic limit uniformly for $\lambda < \lambda_0^{(1)}$,

$$\begin{aligned}
D^*(\lambda) &= \frac{\lambda}{1-\lambda}\phi_B(j)^{'}\mathcal{B}\,\phi_B(j) \\
&= \frac{\lambda}{1-\lambda}\Phi_B(j) \\
&= \frac{1-\lambda}{\lambda}\Phi_A(j) + \frac{\lambda}{1-\lambda}\Phi_B(j) + \Phi_C(j),
\end{aligned} \tag{C.91}$$

where $j = 0, \Phi_A(j)$ and $\Phi_C(j)$ are zero from Remark 4.

Case 2: When $0 < \{\lambda = \lambda_0^{(1)}\} < \lambda_0^{(2)} < \lambda_0^{(3)} < 1$

The parameter before the break,

$$\hat{\theta}_1(\lambda) = \mathcal{H}_{1,T}(\lambda_0^{(1)})\left\{ \sum_{t=1}^{k_0^{(1)}} z_t u_t + \sum_{t=1}^{k_0^{(1)}} z_t x_t^{'} \theta_0^{(1)} \right\}$$
$$\xrightarrow{p} \theta_0^{(1)}.$$

The parameter after the break,

$$\hat{\theta}_2(\lambda) = \mathcal{H}_{2,T}(\lambda_0^{(1)})\left\{ \sum_{t=k_0^{(1)}+1}^{k_0^{(2)}} z_t x_t^{'} \theta_0^{(2)} + \sum_{t=k_0^{(2)}+1}^{k_0^{(3)}} z_t x_t^{'} \theta_0^{(3)} + \sum_{t=k_0^{(3)}+1}^{T} z_t x_t^{'} \theta_0^{(4)} + \sum_{t=k+1}^{T} z_t u_t \right\}$$
$$\xrightarrow{p} \frac{1}{1-\lambda_0^{(1)}} \big\{ (\lambda_0^{(2)} - \lambda_0^{(1)})\theta_0^{(2)} + (\lambda_0^{(3)} - \lambda_0^{(2)})\theta_0^{(3)} + (1 - \lambda_0^{(3)})\theta_0^{(4)} \big\}.$$

The Parameter Difference,

$$\begin{aligned}
&\hat{\theta}_{1,T}(\lambda) - \hat{\theta}_{2,T}(\lambda) \\
&\xrightarrow{p} \frac{1}{1-\lambda_0^{(1)}} \big\{ (1 - \lambda_0^{(1)})\theta_0^{(1)} - (\lambda_0^{(2)} - \lambda_0^{(1)})\theta_0^{(2)} - (\lambda_0^{(3)} - \lambda_0^{(2)})\theta_0^{(3)} - (1 - \lambda_0^{(3)})\theta_0^{(4)} \big\} \\
&= \frac{1}{1-\lambda_0^{(1)}} \big\{ (1 - \lambda_0^{(1)})(\theta_0^{(1)} - \theta_0^{(2)}) + (1 - \lambda_0^{(2)})(\theta_0^{(2)} - \theta_0^{(3)}) \\
&\quad + (1 - \lambda_0^{(3)})(\theta_0^{(3)} - \theta_0^{(4)}) \big\} \\
&= \frac{1}{1-\lambda_0^{(1)}} \sum_{i=1}^{3} (1 - \lambda_0^{(i)}) w_i.
\end{aligned}$$

Thus, the nonstochastic limit for $\lambda = \lambda_0^{(1)}$,

$$\begin{aligned}
D^*(\lambda) &= \frac{\lambda_0^{(1)}}{1-\lambda_0^{(1)}}\{\sum_{i=1}^{3}(1-\lambda_0^{(i)})w_i\}'\mathcal{B}\,\{\sum_{i=1}^{3}(1-\lambda_0^{(i)})w_i\} \\
&= (1-\lambda_0^{(1)})\lambda_0^{(1)}w_1'\mathcal{B}w_1 + \frac{\lambda_0^{(1)}}{1-\lambda_0^{(1)}}\Big\{\sum_{i=2}^{3}(1-\lambda_0^{(i)})w_i\Big\}'\mathcal{B}\Big\{\sum_{i=2}^{3}(1-\lambda_0^{(i)})w_i\Big\} \\
&= \frac{1-\lambda_0^{(1)}}{\lambda_0^{(1)}}\{\lambda_0^{(1)}w_1\}'\mathcal{B}\{\lambda_0^{(1)}w_1\} + \frac{\lambda_0^{(1)}}{1-\lambda_0^{(1)}}\phi_B(j)'\mathcal{B}\phi_B(j) + \Phi_C(j) \\
&= \frac{1-\lambda_0^{(1)}}{\lambda_0^{(1)}}\phi_A(j)'\mathcal{B}\phi_A(j) + \frac{\lambda_0^{(1)}}{1-\lambda_0^{(1)}}\Phi_B + \Phi_C(j) \\
&= \frac{1-\lambda_0^{(1)}}{\lambda_0^{(1)}}\Phi_A(j) + \frac{\lambda_0^{(1)}}{1-\lambda_0^{(1)}}\Phi_B(j) + \Phi_C(j), \qquad \text{(C.92)}
\end{aligned}$$

where $j = 1$ and $\Phi_C(j) = w_1^{*\prime}\mathcal{B}\{\sum_{i=2}^{3}\tilde{w}_i\}$.

Case 3: When $0 < \lambda_0^{(1)} < \lambda < \lambda_0^{(2)} < \lambda_0^{(3)} < 1$

The parameter before the break,

$$\begin{aligned}
\hat{\theta}_1(\lambda) &= \mathcal{H}_{1,T}(\lambda)\left\{\sum_{t=1}^{k_0^{(1)}} z_t x_t'\theta_0^{(1)} + \sum_{t=k_0^{(1)}}^{k} z_t x_t'\theta_0^{(2)} + \sum_{t=1}^{k} z_t u_t\right\} \\
&\xrightarrow{p} \frac{1}{\lambda}\{\lambda_0^{(1)}\theta_0^{(1)} + (\lambda-\lambda_0^{(1)})\theta_0^{(2)}\}.
\end{aligned}$$

The parameter after the break,

$$\begin{aligned}
\hat{\theta}_2(\lambda) &= \mathcal{H}_{2,T}(\lambda)\left\{\sum_{t=k+1}^{k_0^{(2)}} z_t x_t'\theta_0^{(2)} + \sum_{t=k_0^{(2)}+1}^{k_0^{(3)}} z_t x_t'\theta_0^{(3)} + \sum_{t=k_0^{(3)}+1}^{T} z_t x_t'\theta_0^{(4)} + \sum_{t=k+1}^{T} z_t u_t\right\} \\
&\xrightarrow{p} \frac{1}{1-\lambda}\{(\lambda_0^{(2)}-\lambda)\theta_0^{(2)} + (\lambda_0^{(3)}-\lambda_0^{(2)})\theta_0^{(3)} + (1-\lambda_0^{(3)})\theta_0^{(4)}\}.
\end{aligned}$$

The Parameter Difference,

$$
\begin{aligned}
&\hat{\theta}_1(\lambda) - \hat{\theta}_2(\lambda) \\
&\xrightarrow{p} \frac{1}{\lambda(1-\lambda)}\big\{(1-\lambda)\lambda_0^{(1)}\theta_0^{(1)} + (1-\lambda)(\lambda-\lambda_0^{(1)})\theta_0^{(2)} - (\lambda_0^{(2)}-\lambda)\lambda\theta_0^{(2)} \\
&\qquad - (\lambda_0^{(3)}-\lambda_0^{(2)})\lambda\theta_0^{(3)} - (1-\lambda_0^{(3)})\lambda\theta_0^{(4)}\big\} \\
&= \frac{1}{\lambda(1-\lambda)}\big\{(1-\lambda)\lambda_0^{(1)}\theta_0^{(1)} - (1-\lambda)\lambda_0^{(1)}\theta_0^{(2)} + (1-\lambda_0^{(2)})\lambda\theta_0^{(2)} - (1-\lambda_0^{(2)})\lambda\theta_0^{(3)} \\
&\qquad + (1-\lambda_0^{(3)})\lambda\theta_0^{(3)} - (1-\lambda_0^{(3)})\lambda\theta_0^{(4)}\big\} \\
&= \frac{1}{\lambda(1-\lambda)}\big\{(1-\lambda)\lambda_0^{(1)}(\theta_0^{(1)}-\theta_0^{(2)}) + (1-\lambda_0^{(2)})\lambda(\theta_0^{(2)}-\theta_0^{(3)}) \\
&\qquad + (1-\lambda_0^{(3)})\lambda(\theta_0^{(3)}-\theta_0^{(4)})\big\} \\
&= \frac{1}{\lambda(1-\lambda)}\big\{(1-\lambda)\lambda_0^{(1)}w_1 + \lambda\sum_{i=2}^{3}(1-\lambda_0^{(i)})w_i\big\}.
\end{aligned}
$$

Thus, the nonstochastic limit uniformly in $\lambda \in (\lambda_0^{(1)}, \lambda_0^{(2)})$ is,

$$
\begin{aligned}
D^*(\lambda) &= \frac{1-\lambda}{\lambda}\{\lambda_0^{(1)}w_1\}'\mathcal{B}\{\lambda_0^{(1)}w_1\} \\
&\qquad + \frac{\lambda}{1-\lambda}\{\sum_{i=2}^{3}(1-\lambda_0^{(i)})w_i\}'\mathcal{B}\{\sum_{i=2}^{3}(1-\lambda_0^{(i)})w_i\} + \Phi_C(j) \\
&= \frac{1-\lambda}{\lambda}\phi_A(j)'\mathcal{B}\phi_A(j) + \frac{\lambda}{1-\lambda}\,\phi_B(j)'\mathcal{B}\phi_B(j) + \Phi_C(j) \\
&= \frac{1-\lambda}{\lambda}\Phi_A(j) + \frac{\lambda}{1-\lambda}\Phi_B(j) + \Phi_C(j), \qquad \text{(C.93)}
\end{aligned}
$$

where $j = 1$ and $\Phi_C(j)$ is similar to that of Case 2.

Case 4: When $0 < \lambda_0^{(1)} < \{\lambda = \lambda_0^{(2)}\} < \lambda_0^{(3)} < 1$

The parameter before the break,

$$
\begin{aligned}
\hat{\theta}_1(\lambda) &= \mathcal{H}_{1,T}(\lambda_0^{(2)})\left\{\sum_{t=1}^{k_0^{(1)}} z_t x_t'\theta_0^{(1)} + \sum_{t=k_0^{(1)}+1}^{k_0^{(2)}} z_t x_t'\theta_0^{(2)} + \sum_{t=1}^{k_0^{(2)}} z_t u_t\right\} \\
&\xrightarrow{p} \frac{1}{\lambda_0^{(2)}}\big\{\lambda_0^{(1)}\theta_0^{(1)} + (\lambda_0^{(2)}-\lambda_0^{(1)})\theta_0^{(2)}\big\}.
\end{aligned}
$$

The parameter after the break,

$$\hat{\theta}_2(\lambda) = \mathcal{H}_{2,T}(\lambda_0^{(2)})\Bigg\{ \sum_{t=k_0^{(2)}+1}^{k_0^{(3)}} z_t x_t^{'}\theta_0^{(3)} + \sum_{t=k_0^{(3)}+1}^{T} z_t x_t^{'}\theta_0^{(4)} + \sum_{t=k_0^{(2)}+1}^{T} z_t u_t \Bigg\}$$
$$\xrightarrow{p} \frac{1}{1-\lambda_0^{(2)}}\big\{(\lambda_0^{(3)} - \lambda_0^{(2)})\theta_0^{(3)} + (1-\lambda_0^{(3)})\theta_0^{(4)}\big\}.$$

The Parameter Difference,

$$\begin{aligned}
&\hat{\theta}_1(\lambda) - \hat{\theta}_2(\lambda) \\
&\xrightarrow{p} \frac{1}{\lambda_0^{(2)}(1-\lambda_0^{(2)})}\big\{(1-\lambda_0^{(2)})\lambda_0^{(1)}\theta_0^{(1)} + (1-\lambda_0^{(2)})(\lambda_0^{(2)} - \lambda_0^{(1)})\theta_0^{(2)} \\
&\qquad\qquad - (\lambda_0^{(3)} - \lambda_0^{(2)})\lambda_0^{(2)}\theta_0^{(3)} - (1-\lambda_0^{(3)})\lambda_0^{(2)}\theta_0^{(4)}\big\} \\
&= \frac{1}{\lambda_0^{(2)}(1-\lambda_0^{(2)})}\big\{(1-\lambda_0^{(2)})\lambda_0^{(1)}\theta_0^{(1)} - (1-\lambda_0^{(2)})\lambda_0^{(1)}\theta_0^{(2)} + (1-\lambda_0^{(2)})\lambda_0^{(2)}\theta_0^{(2)} \\
&\qquad\qquad + \lambda_0^{(2)}\lambda_0^{(2)}\theta_0^{(3)} - \lambda_0^{(3)}\lambda_0^{(2)}\theta_0^{(3)} - (1-\lambda_0^{(3)})\lambda_0^{(2)}\theta_0^{(4)}\big\} \\
&= \frac{1}{\lambda_0^{(2)}(1-\lambda_0^{(2)})}\big\{(1-\lambda_0^{(2)})\lambda_0^{(1)}w_1 + (1-\lambda_0^{(2)})\lambda_0^{(2)}\theta_0^{(2)} - (1-\lambda_0^{(2)})\lambda_0^{(2)}\theta_0^{(3)} \\
&\qquad\qquad + (1-\lambda_0^{(3)})\lambda_0^{(2)}\theta_0^{(3)} - (1-\lambda_0^{(3)})\lambda_0^{(2)}\theta_0^{(4)}\big\} \\
&= \frac{1}{\lambda_0^{(2)}(1-\lambda_0^{(2)})}\big\{(1-\lambda_0^{(2)})\lambda_0^{(1)}w_1 + (1-\lambda_0^{(2)})\lambda_0^{(2)}w_2 + (1-\lambda_0^{(3)})\lambda_0^{(2)}w_3\big\}.
\end{aligned}$$

Thus, the nonstochastic limit,

$$\begin{aligned}
D^*(\lambda) &= \frac{1-\lambda_0^{(2)}}{\lambda_0^{(2)}}\Big\{\{\lambda_0^{(1)}w_1\}^{'}\mathcal{B}\{\lambda_0^{(1)}w_1\} + \{\lambda_0^{(2)}w_2\}^{'}\mathcal{B}\{\lambda_0^{(2)}w_2\}\Big\} \\
&\quad + \frac{\lambda_0^{(2)}}{1-\lambda_0^{(2)}}\{(1-\lambda_0^{(3)})w_3\}^{'}\mathcal{B}\{(1-\lambda_0^{(3)})w_3\} + \Phi_C(j) \\
&= \frac{1-\lambda_0^{(2)}}{\lambda_0^{(2)}}\{\sum_{i=1}^{2}\lambda_0^{(i)}w_i\}^{'}\mathcal{B}\{\sum_{i=1}^{2}\lambda_0^{(i)}w_i\} + \frac{\lambda_0^{(2)}}{1-\lambda_0^{(2)}}\phi_B(j)^{'}\mathcal{B}\phi_B(j) + \Phi_C(j) \\
&= \frac{1-\lambda_0^{(2)}}{\lambda_0^{(2)}}\Phi_A(j) + \frac{\lambda_0^{(2)}}{1-\lambda_0^{(2)}}\Phi_B(j) + \Phi_C(j), \qquad \text{(C.94)}
\end{aligned}$$

where $j = 2$ and $\Phi_C(j) = \{\sum_{i=1}^{2} w_2^*\}^{'}\mathcal{B}\tilde{w}_3\}$.

Case 5: When $0 < \lambda_0^{(1)} < \lambda_0^{(2)} < \lambda < \lambda_0^{(3)} < 1$

The parameter before the break,

$$\hat{\theta}_1(\lambda) = \mathcal{H}_{1,T}(\lambda)\left\{ \sum_{t=1}^{k_0^{(1)}} z_t x_t' \theta_0^{(1)} + \sum_{t=k_0^{(1)}+1}^{k_0^{(2)}} z_t x_t' \theta_0^{(2)} + \sum_{t=k_0^{(2)}+1}^{k} z_t x_t' \theta_0^{(3)} + \sum_{t=1}^{k} z_t u_t \right\}$$
$$\xrightarrow{p} \frac{1}{\lambda}\{\lambda_0^{(1)}\theta_0^{(1)} + (\lambda_0^{(2)} - \lambda_0^{(1)})\theta_0^{(2)} + (\lambda - \lambda_0^{(2)})\theta_0^{(3)}\}$$

The parameter after the break,

$$\hat{\theta}_2(\lambda) = \mathcal{H}_{2,T}(\lambda)\left\{ \sum_{t=k+1}^{k_0^{(3)}} z_t x_t' \theta_0^{(3)} + \sum_{t=k_0^{(3)}+1}^{T} z_t x_t' \theta_0^{(4)} + \sum_{t=k+1}^{T} z_t u_t \right\}$$
$$\xrightarrow{p} \frac{1}{1-\lambda}\{(\lambda_0^{(3)} - \lambda)\theta_0^{(3)} + (1 - \lambda_0^{(3)})\theta_0^{(4)}\}.$$

The Parameter Difference,

$$\begin{aligned}
&\hat{\theta}_1(\lambda) - \hat{\theta}_2(\lambda) \\
&\xrightarrow{p} \frac{1}{\lambda(1-\lambda)}\{(1-\lambda)\lambda_0^{(1)}\theta_0^{(1)} + (1-\lambda)(\lambda_0^{(2)} - \lambda_0^{(1)})\theta_0^{(2)} \\
&\qquad + (1-\lambda)(\lambda - \lambda_0^{(2)})\theta_0^{(3)} - (\lambda_0^{(3)} - \lambda)\lambda\theta_0^{(3)} - (1 - \lambda_0^{(3)})\lambda\theta_0^{(4)}\} \\
&= \frac{1}{\lambda(1-\lambda)}\{(1-\lambda)\lambda_0^{(1)}w_1 + (1-\lambda)\lambda_0^{(2)}w_2 + (1 - \lambda_0^{(3)})\lambda w_3\}.
\end{aligned}$$

Thus, the nonstochastic limit uniformly in λ for $\lambda \in (\lambda_0^{(2)}, \lambda_0^{(3)})$,

$$\begin{aligned}
D^*(\lambda) &= \frac{1-\lambda}{\lambda}\left\{\{\lambda_0^{(1)}w_1\}'\mathcal{B}\{\lambda_0^{(1)}w_1\} + \{\lambda_0^{(2)}w_2\}'\mathcal{B}\{\lambda_0^{(2)}w_2\}\right\} \\
&\quad + \frac{\lambda}{1-\lambda}\{(1-\lambda_0^{(3)})w_3\}'\mathcal{B}\{(1-\lambda_0^{(3)})w_3\} + \Phi_C(j) \\
&= \frac{1-\lambda}{\lambda}\{\sum_{i=1}^{2}\lambda_0^{(i)}w_i\}'\mathcal{B}\{\sum_{i=1}^{2}\lambda_0^{(i)}w_i\} + \frac{\lambda}{1-\lambda}\phi_B(j)'\mathcal{B}\phi_B(j) + \Phi_C(j) \\
&= \frac{1-\lambda}{\lambda}\phi_A(j)'\mathcal{B}\phi_A(j) + \frac{\lambda}{1-\lambda}\Phi_B(j) + \Phi_C(j) \\
&= \frac{1-\lambda}{\lambda}\Phi_A(j) + \frac{\lambda}{1-\lambda}\Phi_B(j) + \Phi_C(j), \qquad \text{(C.95)}
\end{aligned}$$

where $j = 2$ and $\Phi_C(j)$ is similar to Case 4.

Case 6: When $0 < \lambda_0^{(1)} < \lambda_0^{(2)} < \{\lambda = \lambda_0^{(3)}\} < 1$

The parameter before the break,

$$\hat{\theta}_1(\lambda) = \mathcal{H}_{1,T}(\lambda_0^{(3)})\left\{ \sum_{t=1}^{k_0^{(1)}} z_t x_t' \theta_0^{(1)} + \sum_{t=k_0^{(1)}+1}^{k_0^{(2)}} z_t x_t' \theta_0^{(2)} + \sum_{t=k_0^{(2)}+1}^{k_0^{(3)}} z_t x_t' \theta_0^{(3)} + \sum_{t=1}^{k_0^{(3)}} z_t u_t \right\}$$
$$\xrightarrow{p} \frac{1}{\lambda_0^{(3)}}\{\lambda_0^{(1)}\theta_0^{(1)} + (\lambda_0^{(2)} - \lambda_0^{(1)})\theta_0^{(2)} + (\lambda_0^{(3)} - \lambda_0^{(2)})\theta_0^{(3)}\},$$

The parameter after the break,

$$\hat{\theta}_2(\lambda) = \mathcal{H}_{2,T}(\lambda_0^{(3)})\left\{ \sum_{t=k_0^{(3)}+1}^{T} z_t x_t' \theta_0^{(4)} + \sum_{t=k_0^{(3)}+1}^{T} z_t u_t \right\}$$
$$\xrightarrow{p} \theta_0^{(4)}.$$

The Parameter Difference,

$$\begin{aligned}
\hat{\theta}_1(\lambda) - \hat{\theta}_2(\lambda) &\xrightarrow{p} \frac{1}{\lambda_0^{(3)}}\{\lambda_0^{(1)}\theta_0^{(1)} + (\lambda_0^{(2)} - \lambda_0^{(1)})\theta_0^{(2)} + (\lambda_0^{(3)} - \lambda_0^{(2)})\theta_0^{(3)} - \lambda_0^{(3)}\theta_0^{(4)}\} \\
&= \frac{1}{\lambda_0^{(3)}}\{\lambda_0^{(1)}(\theta_0^{(1)} - \theta_0^{(2)}) + \lambda_0^{(2)}(\theta_0^{(2)} - \theta_0^{(3)}) + \lambda_0^{(3)}(\theta_0^{(3)} - \theta_0^{(4)})\} \\
&= \frac{1}{\lambda_0^{(3)}}\{\lambda_0^{(1)}w_1 + \lambda_0^{(2)}w_2 + \lambda_0^{(3)}w_3\} \\
&= \frac{1}{\lambda_0^{(3)}}\sum_{i=1}^{3} w_i^* \\
&= \frac{1}{\lambda_0^{(3)}}\phi_A(j).
\end{aligned}$$

Thus, the nonstochastic limit,

$$\begin{aligned} D^*(\lambda) &= \frac{1-\lambda_0^{(3)}}{\lambda_0^{(3)}}\phi_A(j)'\mathcal{B}\phi_A(j) \\ &= \frac{1-\lambda_0^{(3)}}{\lambda_0^{(3)}}\Phi_A(j) \\ &= \frac{1-\lambda_0^{(3)}}{\lambda_0^{(3)}}\Phi_A(j) + \frac{\lambda_0^{(3)}}{1-\lambda_0^{(3)}}\Phi_B(j) + \Phi_C(j), \end{aligned} \tag{C.96}$$

where $\Phi_B(j) = 0$ and $\Phi_C(j) = 0$ because there is no true break after the estimated break fraction, since $j = m = 3$ in this case.

Case 7: When $0 < \lambda_0^{(1)} < \lambda_0^{(2)} < \lambda_0^{(3)} < \lambda < 1$

The parameter before the break,

$$\begin{aligned} \hat{\theta}_1(\lambda) = \mathcal{H}_{1,T}(\lambda)\Bigg\{ &\sum_{t=1}^{k_0^{(1)}} z_t x_t' \theta_0^{(1)} + \sum_{t=k_0^{(1)}+1}^{k_0^{(2)}} z_t x_t' \theta_0^{(2)} + \sum_{t=k_0^{(2)}+1}^{k_0^{(3)}} z_t x_t' \theta_0^{(3)} + \sum_{t=k_0^{(3)}+1}^{k} z_t x_t' \theta_0^{(4)} \\ &+ \sum_{t=1}^{k} z_t u_t \Bigg\} \\ \xrightarrow{p} \frac{1}{\lambda_0^{(3)}}\{&\lambda_0^{(1)}\theta_0^{(1)} + (\lambda_0^{(2)} - \lambda_0^{(1)})\theta_0^{(2)} + (\lambda_0^{(3)} - \lambda_0^{(2)})\theta_0^{(3)} + (\lambda - \lambda_0^{(3)})\theta_0^{(4)}\}. \end{aligned}$$

The parameter after the break,

$$\begin{aligned} \hat{\theta}_2(\lambda) &= \mathcal{H}_{2,T}(\lambda)\Bigg\{ \sum_{t=k+1}^{T} z_t x_t' \theta_0^{(4)} + \sum_{t=k+1}^{T} z_t u_t \Bigg\} \\ &\xrightarrow{p} \theta_0^{(4)}. \end{aligned}$$

The parameter difference,

$$\begin{aligned} \hat{\theta}_{1,T}(\lambda) - \hat{\theta}_{2,T}(\lambda) &\xrightarrow{p} \frac{1}{\lambda}\{\lambda_0^{(1)}w_1 + \lambda_0^{(2)}w_2 + \lambda_0^{(3)}w_3\} \\ &= \frac{1}{\lambda}\phi_A(j). \end{aligned}$$

Thus, the nonstochastic limit uniformly in λ for $\lambda < \lambda_0^{(3)}$ is,

$$\begin{aligned} D^*(\lambda) &= \frac{1-\lambda}{\lambda}\phi_A(j)'\mathcal{B}\phi_A(j) \\ &= \frac{1-\lambda}{\lambda}\Phi_A(j) \\ &= \frac{1-\lambda}{\lambda}\Phi_A(j) + \frac{\lambda}{1-\lambda}\Phi_B(j) + \Phi_C(j), \end{aligned} \tag{C.97}$$

where again $\Phi_B(j)$ and $\Phi_C(j)$ are zero since $j = m = 3$.

Chapter 4

Unstable Jacobian Model

The discussions in the previous chapters were premised on the assumption that a stable relationship exists between the endogenous regressors and their instruments as presented in the Jacobian Equation (JE) in (2.2) on page 27. However, in reality, such stability may not always be the case as evidenced in the New Keynesian Phillips Curve model for US data used in Hall et al. (2012) and Boldea et al. (2012), where both endogenous regressors - expected inflation and output gap - were found to possess breaks in their reduced form equations. In this chapter, therefore, we examine the behaviour of the test statistic and the break fraction estimator obtained from models with an unstable JE.

Within the Two Stage Least Squares (2SLS) framework, Hall et al. (2012) suggest a three-step methodology for estimating the break points in models with an unstable JE. In their first step, the JE is estimated along with any break point in it using the sup$F$ type test proposed in Bai and Perron (1998). The Structural Equation (SE) is then split based on the break points identified in the first step and additional breaks are estimated in these subsamples. Lastly, a fixed break point test is used to check if the break points identified in the JE in the first step are also present in the SE. This procedure yields consistent break fraction estimators as established in Hall et al. (2012) and Boldea et al. (2012).

In using 2SLS however, it is crucial that the break points in the JE are consistently estimated in the first step. Otherwise, the resulting estimators would be inconsistent. More recently, Perron and Yamamoto (2015) propose a method to obtain consistent break fraction estimators in linear models with an unstable JE still using Ordinary Least Squares (OLS). However, their approach needs to be applied with caution as there is the possibility that spurious break points in the model would be detected due to the correlation between the endogenous regressors and the errors.

We show in this chapter that the GMM procedure proposed in this study still yields consistent break fraction estimators in the SE of models with an unstable JE. The estimation process follows the same Sequential Method using the Difference test statistic as carried out in the previous two chapters. The breaks in the SE are estimated directly unlike the 2SLS; hence there is no requirement to pre-estimate the break in the JE beforehand.

First, recall the test statistic as given in (2.5) on page 32,

$$D_T(\lambda) = T(\hat{\theta}_{1,T}(\lambda) - \hat{\theta}_{2,T}(\lambda))' M_{*,T}(\lambda)^{-1} (\hat{\theta}_{1,T}(\lambda) - \hat{\theta}_{2,T}(\lambda)),$$

where $\hat{\theta}_{1,T}(\lambda) - \hat{\theta}_{2,T}(\lambda)$ is the Parameter Difference, $M_{*,T}(\lambda)^{-1}$ is the Centre Matrix, $M_{*,T}(\cdot) = M_{1,T}(\cdot)^{-1} + M_{2,T}(\cdot)^{-1}$, $M_{i,T}(\cdot) = G_{i,T}(\cdot)' W_{i,T}(\cdot) G_{i,T}(\cdot)$ and $G_{i,T}(\cdot) = T^{-1} Z_{i,T}(\cdot)' X_{i,T}(\cdot)$, for $i = 1, 2$.

In the previous two chapters under the Stable Jacobian, the Centre Matrices were relatively simple to construct and for all $\lambda \in \Lambda$, converged in probability to $\lambda(1-\lambda)\mathcal{B}$, where $\mathcal{B} = Q_{xz} C Q_{zx}$, as shown in (B.12) in Appendix B.1. When the JE is unstable, however, the analysis becomes more complicated and the limiting properties of the Centre Matrices - and consequently, the test statistic and the break fraction estimator - are not as straightforward to obtain. In the remaining part of this chapter, we examine the limiting properties of the estimators obtained from two main types of models which exhibit unstable JEs.

In Section 4.1, we consider a model that has a break point only in the JE and none in the SE. This model provides a simple way to introduce the Unstable JE concept since the estimators have the same limit for all $\lambda \in \Lambda$. In Section 4.2, we extend the analysis to a model with two break points - one break point in the JE and the other in the SE. Throughout the chapter, the location of the break point in the JE is assumed to be unknown. The significance of this research and suggestions on extending this study are discussed in the conclusion in Section 4.3. As in the previous chapter, the main derivations are placed in the relevant appendices.

4.1 One Break in the JE but Stable SE

The presence of the break point in the JE makes it necessary to modify the model used in Chapter 2. Hence, this section starts with a description of the Unstable JE model alongside its assumptions. These assumptions surround the presentation of the Jacobian when a break exists in the JE. Following this, the test statistic and the break fraction

estimator are examined and their limiting properties established. The analysis carried out in this section assumes the magnitude of change between the parameters in the JE are fixed.

4.1.1 The Model and its Assumptions

Since this model has no break in the SE, we represent the SE thus,

$$y_t = x_t^{'}\theta_0 + u_t, \quad t = 1, 2, \ldots, T, \tag{4.1}$$

where θ_0 is the $p \times 1$ vector of coefficients and y_t and u_t are the dependent variable and error term, as defined earlier in (2.1) on page 26. On the other hand, to reflect the break point in the Jacobian, the JE in (2.2) is rewritten as:

$$\begin{aligned} x_t^{'} &= z_t^{'}\Delta_0^{(1)} + v_t^{'} \quad for \;\; t = 1, 2, \ldots, h_0 && (4.2)\\ &= z_t^{'}\Delta_0^{(2)} + v_t^{'} \quad for \;\; t = h_0 + 1, \ldots, T, && (4.3) \end{aligned}$$

where z_t is the $q \times 1$ vector of instruments as defined in (2.2), $\Delta_0^{(1)} \neq \Delta_0^{(2)}$, $h_0 = [T\pi_0]$, $\pi_0 \in \Pi \subset (0, 1)$, π_0 and h_0 are respectively defined as the true break fraction and true break point in the JE. Due to this break in the JE, it is crucial to change Assumptions 1 to 3 on page 27 which were used for the Stable Jacobian models,

Assumption 17.

$$\begin{aligned} E[z_t x_t^{'}] &= E[z_t z_t^{'}]\Delta_0^{(1)} = Q_{zz}\Delta_0^{(1)} \quad for \;\; t \leq h_0 && (4.4)\\ &= E[z_t z_t^{'}]\Delta_0^{(2)} = Q_{zz}\Delta_0^{(2)} \quad for \;\; t > h_0, && (4.5) \end{aligned}$$

where $\|Q_{zz}\| < \infty$ *and positive definite.*

Assumption 18. $rank\{\Delta_0^{(i)}\} = p$ *and* $\|\Delta_0^{(i)}\| < \infty$ *for* $i = 1, 2$.

Assumption 19. *The matrices* $\frac{1}{j}\sum_{t=h_0+1}^{h_0+j} z_t z_t^{'}$ *and* $\frac{1}{j}\sum_{t=h_0-j+1}^{h_0} z_t z_t^{'}$ *have minimum eigenvalues bounded away from zero in probability for all* $\pi \in \Pi$ *and* $j \geq q$.

Assumption 20. $sup_{\pi\in\Pi}\|T^{-1}\sum_{t=1}^{h} z_t x_t^{'} - \pi Q_{zz}\Delta_0^{(1)}\| \xrightarrow{p} 0$.

Assumption 17 indicates the expectation of the Jacobian before and after the break point. Assumption 18 is the standard rank condition for identification[1] when estimating using

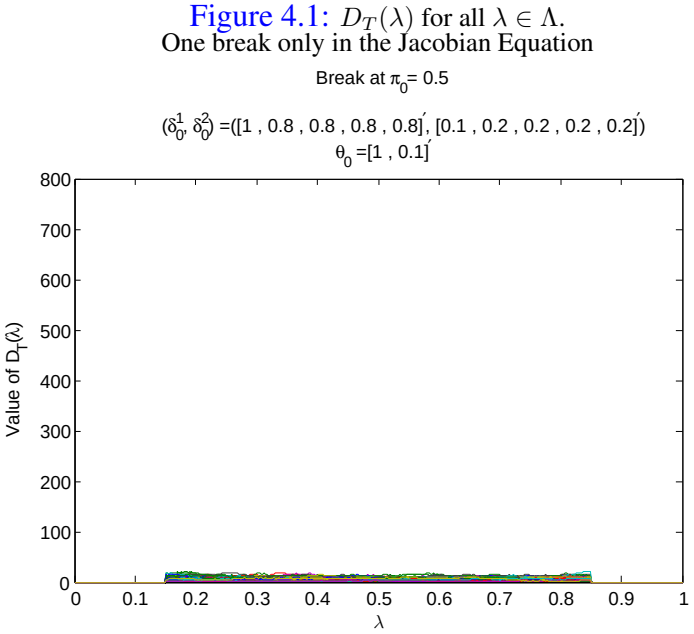
[1] See Identification Condition on page 35 in Hall (2005).

GMM. Assumption 19 requires that there be enough observations near the true break point in the Jacobian. Assumption 20 states the uniform convergence of the Jacobian and in union with Assumption 18, it implies $Q_{zz}\Delta_0^{(1)}$ in (4.4) and $Q_{zz}\Delta_0^{(2)}$ in (4.5) have rank equal to p for $\pi \in \Pi$.

4.1.2 The Test Statistic

The test statistic $D_T(\lambda)$ used here is identical to that used in the Stable Jacobian in Chapters 2 and 3. It is constructed as a sequence comprising of values of $D_T(\lambda)$ for all $\lambda \in \Lambda$. To illustrate the finite sample behaviour of $D_T(\lambda)$ from a model with one break point in the JE, we present a plot of $D_T(\lambda)$ across the full range of $\Lambda = [0.15, 0.85]$ from a sample of size $T = 600$. The break in the JE is imposed at $\pi_0 = 0.5$ with specific parameter values $\Delta_0^{(1)} = (1, 0.8, 0.8, 0.8, 0.8)'$ and $\Delta_0^{(2)} = (0.1, 0.2, 0.2, 0.2, 0.2)'$. As seen in the one thousand simulations displayed in Figure 4.1, $D_T(\lambda)$ has values very close to zero across all $\lambda \in \Lambda$. Comparing this behaviour to the plot of $D_T(\lambda)$ obtained from a model with a break in the SE displayed in Figure 2.2 on page 31, this indicates that $D_T(\lambda)$ is not influenced by the instability in the JE. Thus, our proposed estimation method is a promising option to estimate break points in the SE in the presence of breaks in the Jacobian.

Figure 4.1: $D_T(\lambda)$ for all $\lambda \in \Lambda$.



4.1.3 The Break Fraction Estimator

The break fraction estimator, $\hat{\lambda}$ and its estimation process work identically to that discussed under the Stable Jacobian in Section 2.2; after the sequence of $D_T(\lambda)$ is obtained, $\hat{\lambda}$ is estimated as the location of the *supremum* of this sequence.

Asymptotic Properties of the Break Fraction Estimator

In order to establish the asymptotic properties of the break fraction estimator, we first present the nonstochastic limit of the test statistic. Unlike the Stable Jacobian case given in Lemma 1, in this case when there is a break only in the JE, $D_T^*(\lambda)$ has identical nonstochastic limits uniformly in λ for $\lambda \in \Lambda$. This is stated in the following lemma:

Lemma 11. *If $\theta_0^{(1)} = \theta_0^{(2)}$ and Equations (4.1) to (4.3) and Assumptions 4, 6 to 8 and 17 to 20 hold, then $D_T^*(\lambda)$ converges to a nonstochastic function $D^*(\lambda)$ on (0,1), given as*

$$D^*(\lambda) = (\theta_0^{(1)} - \theta_0^{(2)})' M_*(\lambda)^{-1}(\theta_0^{(1)} - \theta_0^{(2)}) = 0, \quad uniformly\ in\ \lambda\ for\ \lambda \in \Lambda \tag{4.6}$$

where[2]

$$M_*(\lambda) = M_1(\lambda)^{-1} + M_2(\lambda)^{-1} \tag{4.7}$$
$$M_i(\lambda) = G_i^h(\lambda)' W_i(\lambda) G_i^h(\lambda), \quad for\ i = 1, 2, \tag{4.8}$$
$$G_1^h(\lambda) = min(\pi_0, \lambda) Q_{zz} \Delta_0^{(1)} + max(0, \lambda - \pi_0) Q_{zz} \Delta_0^{(2)} \tag{4.9}$$
$$G_2^h(\lambda) = max(\pi_0 - \lambda, 0) Q_{zz} \Delta_0^{(1)} + min(1 - \pi_0, 1 - \lambda) Q_{zz} \Delta_0^{(2)} \tag{4.10}$$
$$W_1(\lambda) = \lambda^{-1} C,\ W_2(\lambda) = (1 - \lambda)^{-1} C \text{ and } C \text{ is a nonsingular constant matrix.}$$

This lemma, which is proved in Appendix D.1, implies $D_T(\lambda)$ converges to zero uniformly in λ for $\lambda \in \Lambda$. Based on this asymptotic property of $D_T(\lambda)$, the break fraction estimator converges to a random variable, $\lambda^\dagger$. That is,

$$\hat{\lambda} \xrightarrow{p} \lambda^\dagger, \quad where\ \lambda^\dagger \in \Lambda. \tag{4.11}$$

Though we do not formally prove this asymptotic behaviour of $\hat{\lambda}$, the results of the Monte Carlo simulations presented in Chapter 5 provide strong evidence of this random behaviour.

[2]Note that $G_i^h(\lambda)$ is analogous to G_i in the Stable Jacobian model, for $i = 1, 2$. We use it here simply to differentiate the Stable Jacobian from the Unstable Jacobian.

4.2 One Break each in the JE and SE

In this section, we extend the analysis done above for the single break in the JE to a model that includes an additional break in the SE. Having a break point in both the JE and SE results in two possibilities: the break points may occur either at the same location or they may occur at different locations in the JE and SE. We refer to the former as the Coincidental Break scenario and the latter as the Separate Break scenario. For the remaining part of this chapter, the behaviour of the test statistic and break fraction estimator are examined under these two scenarios. Additionally, it is assumed that the magnitude of the parameter change in the JE shrinks with the sample size. Although we conjecture consistency also holds with fixed breaks in the JE, the proofs were intractable because of the complexity of the Centre Matrix as given in (4.7) to (4.10) on page 115. However, the Monte Carlo simulation results presented in Chapter 5 corroborate this conjecture. Consequently, the shrinking breaks framework adopted in this section is only required for the proofs of the limiting properties of the break fraction estimators.

4.2.1 The Model and its Assumptions

To reflect the break in the SE, we use the model with a single break in the SE given in (2.1) on page 26, while we use the model in (4.2) and (4.3) to indicate the break in the JE. That is,

$$\begin{aligned} y_t &= x_t'\theta_0^{(1)} + u_t, \quad t = 1, 2, \ldots, [T\lambda_0] \\ &= x_t'\theta_0^{(2)} + u_t, \quad t = [T\lambda_0] + 1, \ldots, T \end{aligned}$$

and (4.12)

$$\begin{aligned} x_t' &= z_t'\Delta_0^{(1)} + v_t' \quad t = 1, 2, \ldots, [T\pi_0] \\ &= z_t'\Delta_0^{(2)} + v_t' \quad t = [T\pi_0] + 1, \ldots, T, \end{aligned}$$

where the terms are as defined previously in (2.1), (4.2) and (4.3).

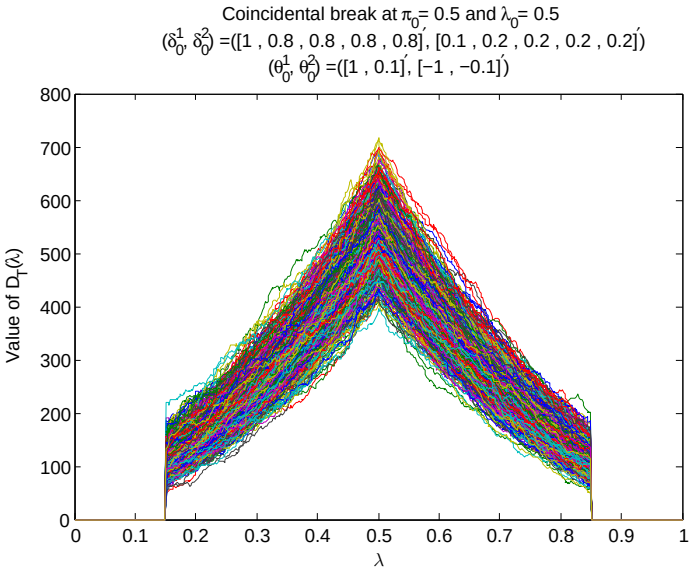
4.2.2 Coincidental Break Scenario

When the break points in the SE and JE are coincidental, then λ_0 and π_0 in (4.12) occur at the same point in the sample, that is, $\lambda_0 = \pi_0$. To discuss the characteristics of the break fraction estimator in this setting, we first present the finite sample behaviour of the test

statistic $D_T(\lambda)$ over the range of $\Lambda = [0.15, 0.85]$ in Figure 4.2.

In this model, we impose $\lambda_0 = \pi_0 = 0.5$ and for $t \leq \lambda_0$, the parameters used are $\theta_0^{(1)} = (1, 0.1)'$ and $\Delta_0^{(1)} = (1, 0.8, 0.8, 0.8, 0.8)'$; while for $t > \lambda_0$, $\theta_0^{(2)} = (-1, -0.1)'$ and $\Delta_0^{(2)} = (0.1, 0.2, 0.2, 0.2, 0.2)'$. $D_T(\lambda)$ exhibits similar behaviour across the range, for each of the one thousand simulations performed, attaining its maximum at the location of the true break point, when $\lambda_0 = 0.5$. When compared to $D_T(\lambda)$ obtained from a model with a single break in the JE displayed in Figure 4.1, it indicates the peak of $D_T(\lambda)$ in Figure 4.2 is due to the break in the SE and not the break in the JE. This is further confirmed in the theoretical analysis below and the Monte Carlo simulation results presented in Chapter 5.

Figure 4.2: $D_T(\lambda)$ for all $\lambda \in \Lambda$.
One break each in the Jacobian and Structural Equations



Some assumptions vital to establishing the limiting properties of the break fraction estimator are now stated. As highlighted earlier, we adopt the shrinking breaks approach[3] when dealing with the break in the JE. Under this approach, the parameter change in the JE shrinks and converges to zero as the sample size increases. This is stated in the following assumption,

Assumption 21.

$$\Delta_T^{(i)} = \Delta_0 + d_T^{(i)}, \quad for\ i = 1, 2, \tag{4.13}$$

[3]Using the fixed break approach as done in the Stable Jacobian proved intractable due to the structure of the Centre Matrices.

where $d_T^{(i)} = d^{(i)}\nu_T$, $d^{(i)}$ *is a constant term independent of* T, $\nu_T = T^{-\alpha}$ *and* $\alpha \in (0, 0.5)$.

Assumption 22. $rank\{\Delta_T^{(i)}\} = p$ *and* $\|\Delta_T^{(i)}\| < \infty$ *for* $i = 1, 2$.

The rate used for the shrinking breaks in Assumption 21 is simply a mathematical approach adopted in the literature[4] to obtain the asymptotic properties of estimators. These properties are otherwise difficult to obtain or they are obtained in complex forms using the fixed break approach.

To ease notation, let $G_{1,T}^h = E[z_t x_t^{'}]$ when $h \leq h_0$ and $G_{2,T}^h = E[z_t x_t^{'}]$ when $h > h_0$. Then based on Assumption 21, to reflect the shrinking breaks in the JE, we write

Assumption 23.

$$G_{1,T}^h = E[z_t z_t^{'}]\Delta_T^{(1)} = G_0 + Q_{zz} d_T^{(1)} \quad for \;\; t \leq h_0 \tag{4.14}$$

$$G_{2,T}^h = E[z_t z_t^{'}]\Delta_T^{(2)} = G_0 + Q_{zz} d_T^{(2)} \quad for \;\; t > h_0, \tag{4.15}$$

where $G_0 = Q_{zz}\Delta_0$.

Assumption 23 is analogous to Assumption 17 on page 113 used for the one break in the JE under the fixed break approach. It implies that under this shrinking breaks approach, the expectation of the Jacobian in the Unstable case is identical to that of its counterpart in the Stable case plus a term which is $O_p(\nu_T)$ that shrinks as the sample size increases.

Assumption 24.

$$sup_{\pi \in \Pi} \| T^{-1} \sum_{t=1}^{[T\pi]} \left(z_t x_t^{'} - G_{1,T}^h \right) \| = o_p(1), \quad when \; t \leq h_0 \tag{4.16}$$

$$sup_{\pi \in \Pi} \| T^{-1} \sum_{[T\pi]+1}^{T} \left(z_t x_t^{'} - G_{2,T}^h \right) \| = o_p(1), \quad when \; t > h_0. \tag{4.17}$$

Assumption 24 states the difference between the Jacobian and $G_{i,T}^h$ is very small in probability for $i = 1, 2$.

Consistency and Convergence Rate

This section considers issues relating to the limiting properties of the break fraction estimator obtained from a model with a coincidental break in the SE and JE. First, recall

[4] See Bai (1997a), Boldea et al. (2012) and Perron and Yamamoto (2015).

from Assumption 1 on page 27 that when the Jacobian is stable, $E[z_t x_t'] = G_0$. In the following lemma, we state that the difference between G_0 and the Jacobian obtained from a model with an unstable JE is asymptotically negligible.

Lemma 12. *If the model is generated by (4.12) and Assumptions 21 to 24 hold,*

$$sup_{\pi\in\Pi}\|T^{-1}\sum_{t=1}^{[T\pi]}\left(z_t x_t^{'} - G_0\right)\| = o_p(1). \tag{4.18}$$

Remark 5. *As seen in the proof of Lemma 12 presented in Appendix D.2 on page 126, the $o_p(1)$ term in Lemma 12 is $O(\nu_T)$ and hence, the difference between the Centre Matrix[5] in a Stable and Unstable JE is at most $O(\nu_T^2)$, which is asymptotically negligible.*

Based on Lemma 12 and Remark 5, the consistency of $\hat{\lambda}$ can be proved following a similar three-step procedure used in the Stable Jacobian case detailed in Subsection 2.4.1 on page 32. Although we do not present the theoretical proofs in this research, we conjecture in the following remark that (based on preliminary analysis) using this shrinking breaks approach.

Remark 6. *if y_t and x_t are generated as given in (4.12) and Assumptions 21 to 24 hold, then for every $\epsilon > 0$, there exists a finite $\mathcal{M}$ independent of T, such that for all large T,*

$$P(T|\hat{\lambda} - \lambda_0| > \mathcal{M}\nu_T^{-2}) < \epsilon. \tag{4.19}$$

This assumption infers the consistency of the break fraction estimator obtained from a model with a break in both the SE and JE and also provides its rate of convergence to the true break fraction in the SE. Furthermore, the results from the Monte Carlo simulations carried out in Chapter 5 support this consistency assumption made in Remark 6.

4.2.3 Separate Break Scenario

When the break in the SE does not occur at the same point as the break in the JE, then for the model given in (4.12), we have that $\lambda_0 \neq \pi_0$. Thus there are two possible events that can happen which we term Event 1 and Event 2. In Event 1, the break in the SE lies after the break in the JE, that is $\pi_0 < \lambda_0$ while in Event 2, the break in the SE lies before the break in the JE, that is $\lambda_0 < \pi_0$.

[5]Recall the Centre Matrix is essentially a quadratic of the Jacobian, $G_i^h(\cdot), i = 1, 2$.

In both events, there are five possible cases that can occur, dependent on the location of the candidate break fraction. Starting with Event 1, we have

Case 1: When $\lambda < \pi_0 < \lambda_0$

Case 2: When $\lambda = \pi_0 < \lambda_0$

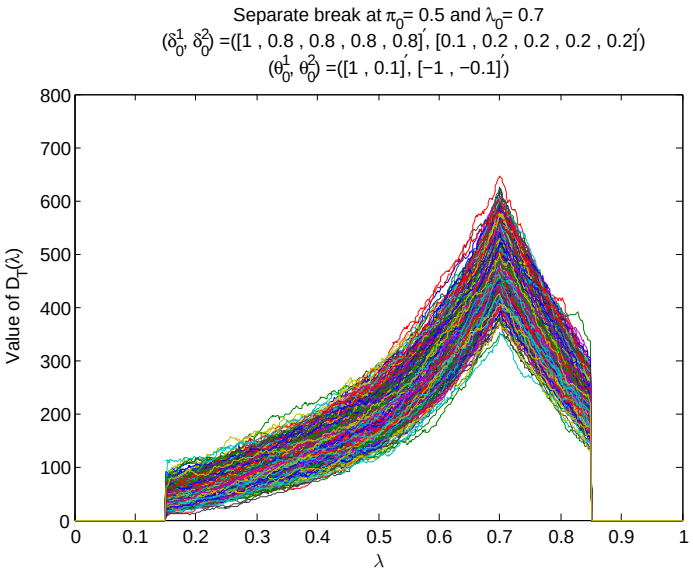
Case 3: When $\pi_0 < \lambda < \lambda_0$

Case 4: When $\pi_0 < \lambda = \lambda_0$

Case 5: When $\pi_0 < \lambda_0 < \lambda$

We show a plot of $D_T(\lambda)$ for Event 1 in Figure 4.3. The DGP is identical to that used in the Coincidental Break Scenario on page 117 except that in this case, the break in the SE, $\lambda_0 = 0.7$ while the break in the JE is left at $\pi_0 = 0.5$. The plot clearly shows the *supremum* of $D_T(\lambda)$ occurs when $\lambda = 0.7$.

Figure 4.3: $D_T(\lambda)$ for all $\lambda \in \Lambda$.
One break each in both the JE and SE (Event 1)



Likewise, for Event 2, the five possible cases are

Case 6: When $\lambda < \lambda_0 < \pi_0$

Case 7: When $\lambda = \lambda_0 < \pi_0$

Case 8: When $\lambda_0 < \lambda < \pi_0$

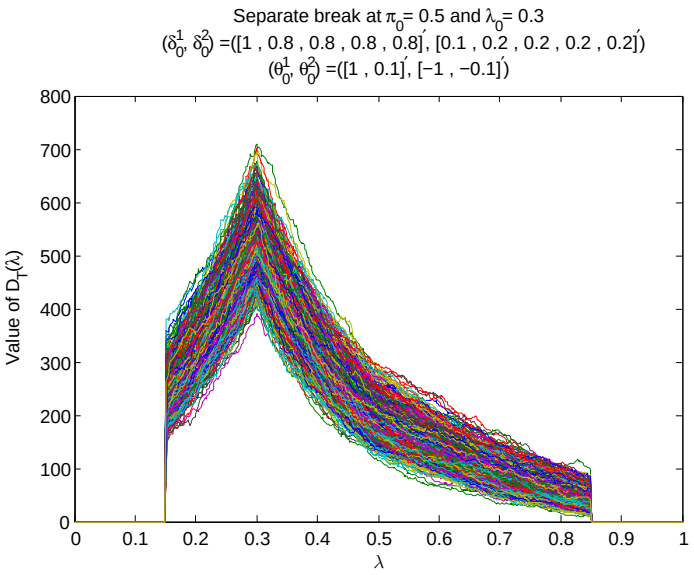
Case 9: When $\lambda_0 < \lambda = \pi_0$

Case 10: When $\lambda_0 < \pi_0 < \lambda$

Figure 4.4 shows $D_T(\lambda)$ across the range of Λ when Event 2 occurs. Apart from imposing the break in the SE at $\lambda_0 = 0.3$, the data generating process is identical to that in Event 1.

Yet again, the peak of $D_T(\lambda)$ when $\lambda = 0.3$ is clearly evident.

Figure 4.4: $D_T(\lambda)$ for all $\lambda \in \Lambda$.
One break each in both the JE and SE (Event 2)



Consistency and Convergence Rate

Based on Assumptions 21 to 24, Lemma 12 and Remark 5, the asymptotic behaviour of the test statistic and break fraction estimator obtained from a Separate Break model is identical to those in the Coincidental Break model. Consequently, the assumption of the consistency and convergence rate given in Remark 6 extend to both Events 1 and 2 of the Separate Break scenario as well.

4.3 Conclusion

This chapter considered break point estimation in models with an unstable Jacobian Equation, that is, models in which the relationship between the endogenous regressor and its instruments are not constant. Hall et al. (2012) provide a good example of this type of problem that exists in reality. In their study of the NKPC using US data, they identified two break points in the JE of the endogenous regressor, output gap. To determine the total number of break points in the model, they first estimated the breaks in the JE before those in the SE based on 2SLS. In our own estimation approach using GMM however, there is no need to pre-estimate the break points in the JE first. Our strategy which is based on parameter variation is focused on the main SE of interest.

Two main types of models with an unstable JE were considered. Firstly, a model with a single break in the JE only was examined. Using the fixed break approach, the break fraction estimator was shown to converge to a random variable uniformly in λ for $\lambda \in \Lambda$. Secondly, a model with a break in both the SE and JE were studied. These models were classified as either Coincidental or Separate break models, based on the location of the two break points in the sample. Using the shrinking breaks approach, we showed initial steps in the proofs of consistency and convergence rate of the break fraction estimator obtained from these models. Furthermore, the Monte Carlo simulations in Chapter 5 support the conjecture made that the break fraction obtained from the models with either a Coincidental or Separate break in the JE and SE is consistent for the true break fraction in the SE. This implies that an unstable relationship between the endogenous regressor and the instruments does not confound the estimations of a true break point in the SE, if one exists.

Thus, the estimation of a break point in the SE when the JE is unstable follows a similar process as the Stable Jacobian outlined in Chapters 2 and 3. This proposed methodology using GMM therefore provides fewer steps than the 2SLS approach in Hall et al. (2012). In this way, our method enables the researcher to focus entirely on any significant break point in the main Structural Equation of interest rather than on the Jacobian Equation.

Appendix D

D.1 Proof of Lemma 11

With the single break only in the Jacobian Equation (JE), there are three different expressions $D^*(\lambda)$ can take, dependent on the location of the true break fraction in the JE, π_0 and the break fraction being considered, λ. The first case is when they both occur at the same location, that is when $\lambda = \pi_0$; the second case is when $\lambda < \pi_0$ and the third case is when $\lambda > \pi_0$.

As in the Stable Jacobian proofs, $D_T(\lambda)$ is also analysed here based on its two main components - the Parameter Difference and the Centre Matrix. For the Parameter Difference, since there is no break in the Structural Equation[1], then it is straightforward to see that for all $\lambda \in \Lambda$,

$$\hat{\theta}_1(\lambda) - \hat{\theta}_2(\lambda) \xrightarrow{p} \theta_0^{(1)} - \theta_0^{(2)} = 0, \tag{D.1}$$

since $\theta_0^{(1)} = \theta_0^{(2)}$. This implies that irrespective of the location of π_0, the Parameter Difference is always zero as it is dependent only on a break point in the SE.

On the other hand, examining the Centre Matrix as given in (4.7) to (4.10), it is evidently dependent on π_0 and consequently, it is not invariant to the location of the break fraction

[1] Also by construction the estimated parameters are unaffected by only a break in the Jacobian (see GMM estimation in Appendix A.1).

being considered. Recall the Centre Matrix in this case is denoted as $M_*(\lambda)^{-1}$, where

$M_*(\lambda) = M_1(\lambda)^{-1} + M_2(\lambda)^{-1}$
$M_i(\lambda) = G_i^h(\lambda)^{'} W_i(\lambda) G_i^h(\lambda), \quad for\ i = 1, 2,$
$G_1^h(\lambda) = min(\pi_0, \lambda) Q_{zz} \Delta_0^{(1)} + max(0, \lambda - \pi_0) Q_{zz} \Delta_0^{(2)}$
$G_2^h(\lambda) = max(\pi_0 - \lambda, 0) Q_{zz} \Delta_0^{(1)} + min(1 - \pi_0, 1 - \lambda) Q_{zz} \Delta_0^{(2)}$
$W_1(\lambda) = \lambda^{-1} C$, $W_2(\lambda) = (1 - \lambda)^{-1} C$ and C is a nonsingular constant matrix.

This is unlike the Centre Matrix analysed earlier in models with a Stable Jacobian given in (B.12) in Appendix B.1. Based on the limit of the Parameter Difference given in (D.1), it is sufficient to show that the limit of the Centre Matrix is finite in order to determine the limiting behaviour of $D_T^*(\lambda)$. Assumptions 17 and 18[2] imply that both $M_1(\cdot)$ and $M_2(\cdot)$ are finite positive definite matrices with full rank. For the record, the specific forms of $G_1^h(\lambda)$ and $G_2^h(\lambda)$ for all $\lambda \in \Lambda$, covering all the three cases highlighted above are provided below.

Case 1: When $\lambda = \pi_0$

For the Jacobian before the break, we have

$$T^{-1} \sum_{t=1}^{k} z_t x_t^{'} = T^{-1} \sum_{t=1}^{k} z_t z_t^{'} \Delta_0^{(1)} \xrightarrow{p} G_1^h(\lambda),$$

where

$$G_1^h(\lambda) = \lambda Q_{zz} \Delta_0^{(1)}, \tag{D.2}$$

while for that after the break, we have

$$T^{-1} \sum_{t=k+1}^{T} z_t x_t^{'} = T^{-1} \sum_{t=k+1}^{T} z_t z_t^{'} \Delta_0^{(2)} \xrightarrow{p} G_2^h(\lambda),$$

where

$$G_2^h(\lambda) = (1 - \lambda) Q_{zz} \Delta_0^{(2)}. \tag{D.3}$$

[2]These assumptions state that $E[z_t x_t^{'}] = E[z_t z_t^{'}] \Delta_0^{(i)} = Q_{zz} \Delta_0^{(i)}$, $rank\{\Delta_0^{(i)}\} = p$, $\|Q_{zz}\| < \infty$ and $\|\Delta_0^{(i)}\| < \infty$ for $i = 1, 2$.

Case 2: When $\lambda < \pi_0$

The Jacobian before the break is identical to (D.2), while that after the break,

$$T^{-1}\sum_{t=k+1}^{T} z_t x_t^{'} = T^{-1}\sum_{t=k+1}^{h_0} z_t z_t^{'}\Delta_0^{(1)} + T^{-1}\sum_{t=h_0+1}^{T} z_t z_t^{'}\Delta_0^{(2)} \xrightarrow{p} G_2^h(\lambda),$$

where

$$G_2^h(\lambda) = (\pi_0 - \lambda)Q_{zz}\Delta_0^{(1)} + (1-\pi_0)Q_{zz}\Delta_0^{(2)}. \tag{D.4}$$

Case 3: When $\lambda > \pi_0$

In this case, the Jacobian after the break is identical to (D.3) while that before the break,

$$T^{-1}\sum_{t=1}^{k} z_t x_t^{'} = T^{-1}\sum_{t=1}^{h_0} z_t z_t^{'}\Delta_0^{(1)} + T^{-1}\sum_{t=h_0+1}^{k} z_t z_t^{'}\Delta_0^{(2)} \xrightarrow{p} G_1^h(\lambda),$$

where

$$G_1^h(\lambda) = \pi_0 Q_{zz}\Delta_0^{(1)} + (\lambda - \pi_0)Q_{zz}\Delta_0^{(2)}. \tag{D.5}$$

In all three cases, $G_1^h(\lambda)$ for $i = 1, 2$ given in (D.2) to (D.5) are seen to be finite. This confirms the Centre Matrix is finite and in combination with the Parameter Difference given in (D.1), proves the nonstochastic limit of $D_T^*(\lambda)$ as presented in Lemma 11.

D.2 Proof of Lemma 12

We consider the case where $\pi \leq h_0$. The analysis for $\pi > h_0$ is similar and hence is omitted. First, by adding and subtracting $G^h_{1,T}$, we write[3],

$$\begin{aligned}\|T^{-1}\sum_{t=1}^{[T\pi]}\big(z_t x_t^{'} - G_0\big)\| &= \|T^{-1}\sum_{t=1}^{[T\pi]}\big(z_t x_t^{'} - G^h_{1,T} + G^h_{1,T} - G_0\big)\| \\ &= \|T^{-1}\sum_{t=1}^{[T\pi]}\big(z_t x_t^{'} - G^h_{1,T}\big) + T^{-1}\sum_{t=1}^{[T\pi]}\big(G^h_{1,T} - G_0\big)\| \\ &\leq \|T^{-1}\sum_{t=1}^{[T\pi]}\big(z_t x_t^{'} - G^h_{1,T}\big)\| + \|T^{-1}\sum_{t=1}^{[T\pi]}\big(G^h_{1,T} - G_0\big)\|.\end{aligned}$$

Taking the *supremum* implies,

$$\begin{aligned}&sup_{\pi\in\Pi}\|T^{-1}\sum_{t=1}^{[T\pi]}\big(z_t x_t^{'} - G_0\big)\| \\ &\qquad\leq sup_{\pi\in\Pi}\Big\{\|T^{-1}\sum_{t=1}^{[T\pi]}\big(z_t x_t^{'} - G^h_{1,T}\big)\| + \|T^{-1}\sum_{t=1}^{[T\pi]}\big(G^h_{1,T} - G_0\big)\|\Big\} \\ &\qquad\leq \Gamma_{1,T} + \Gamma_{0,T},\end{aligned}$$

where

$$\Gamma_{1,T} = sup_{\pi\in\Pi}\|T^{-1}\sum_{t=1}^{[T\pi]}\big(z_t x_t^{'} - G^h_{1,T}\big)\| \tag{D.6}$$

and

$$\Gamma_{0,T} = sup_{\pi\in\Pi}\|T^{-1}\sum_{t=1}^{[T\pi]}\big(G^h_{1,T} - G_0\big)\|. \tag{D.7}$$

Recall that Assumption 24 on page 118 states,

$$sup_{\pi\in\Pi}\|T^{-1}\sum_{t=1}^{[T\pi]}\big(z_t x_t^{'} - G^h_{1,T}\big)\| = o_p(1), \;\; when\; t \leq h_0.$$

[3]Using the Triangle inequality.

This implies $\Gamma_{1,T}$ is asymptotically negligible. Also, from Assumption 23 on page 118,

$$G_{1,T}^{h} = E[z_t z_t^{'}]\Delta_T^{(1)} = G_0 + Q_{zz} d_T^{(1)} \quad for \;\; t \leq h_0,$$

and so we write

$$\Gamma_{0,T} = sup_{\pi \in \Pi} \| T^{-1} \sum_{t=1}^{[T\pi]} Q_{zz} d_T^{(1)} \| \;\; = \;\; sup_{\pi \in \Pi} \| Q_{zz} d^{(1)} T^{-1} \sum_{t=1}^{[T\pi]} \nu_T \|. \tag{D.8}$$

Since Q_{zz} and $d^{(1)}$ are both $O(1)$ and ν_T is $O(T^{-\alpha})$, then

$$\Gamma_{0,T} = sup_{\pi \in \Pi} \| O(1) O(1) O(T^{-\alpha}) \| = sup_{\pi \in \Pi} \| o(1) \| = o(1).$$

Combining $\Gamma_{1,T}$ and $\Gamma_{0,T}$ we therefore conclude,

$$sup_{\pi \in \Pi} \| T^{-1} \sum_{t=1}^{[T\pi]} \left(z_t x_t^{'} - G_0 \right) \| \leq o_p(1) + o(1) = o_p(1).$$

Chapter 5

Monte Carlo Simulation Results

This chapter presents detailed results of the Monte Carlo simulation experiments carried out to examine the performance of the break fraction estimators in finite samples. These simulation results support the underlying theory discussed in Chapters 2 to 4. Though most of the results presented are based on the Difference test statistic, $D_T(\lambda)$ on which the theoretical analysis is established, however, a few results from the simulations involving the Wald Test, $Wald_T(\lambda)$ and the Lagrange Multiplier Test, $LM_T(\lambda)$ are also presented. The similarities between all three tests are clearly evident, alluding to the conjecture that similar theoretical results derived for the $D_T(\lambda)$ in this research may also extend to the other two tests.

In addition, although most of the estimations from the multiple break models are carried out using the Sequential Estimation Method on which the theory is based, for comparisons, we also show some results obtained from the Simultaneous Estimation Method. As seen, both methods yield similar outcomes.

The chapter is divided into two broad sections. The first section presents results from the Stable Jacobian models while the second presents results from the Unstable Jacobian models. Throughout the simulations in this chapter, the true number of break points are assumed to be known and only their locations need to be estimated. However, in the next chapter, we investigate the break fraction estimators obtained from models where the true number of break points is not known *a priori*. The fundamental Data Generating Process (DGP) used is presented in the Stable Jacobian section and any relevant changes are made to it in the Unstable Jacobian section.

5.1 Simulation Results from Stable Jacobian models

In this section, the break fraction estimators obtained from four different models with a Stable Jacobian are examined. We start with the simplest model which is one with no breaks in its Structural Equation (SE), that is, a model where $m_0 = 0$. This model forms a good foundation in which we progress into the remaining models with one, two and three break points in the SE. In the first three types of models considered, that is, when $m = 0, 1, 2$, we sequentially estimate two break points irrespective of the true number of breaks. This provides an expanded view of the behaviour of the break fraction estimator.

5.1.1 The DGP for Stable Jacobian models

The DGP for all the models[1] is made up of two parts - the SE and the Jacobian Equation (JE). For models with a stable JE, the SE is unique to the number of break points in the models, hence, we present them alongside their respective models below.

Conversely, the JE is generic to all models with a Stable JE and is constructed as,

$$x_t = [1, z_t^{'}]\delta + v_t, \quad t = 1, 2, \ldots, T, \tag{5.1}$$

where (i) x_t is the $p \times 1$ endogenous regressor; (ii) z_t is the $q \times 1$ instrument vector constructed as $z_t \sim i.i.d\ N(0_{q\times 1}, I_q)$, $q = 4$; (iii) δ is a $(q+1) \times 1$ vector of the JE coefficients with specific values $(1, 0.5, 0.5, 0.5, 0.5)^{'}$, generated to yield a population R^2 of 0.5; (iv) the errors are generated thus: $(u_t, v_t)^{'} \sim IN(0_{2\times 1}, \Omega)$ where u_t is the error in the SE and $\Omega = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$; (v) a trimming of $\epsilon = 0.15$ is applied in all simulations, that is, the range of the break fraction estimator, $\Lambda = [0.15, 0.85]$ and consequently, all estimations are performed using the central 70% of the sample; (vi) the results cover five sample sizes of 140, 260, 300, 420 and 600; (vii) the results presented are based on 1000 repetitions; (viii) all simulations are programmed in MATLAB.

5.1.2 Simulation Results for a Model with No Break in the SE

In the next chapter, we discuss the behaviour of the test statistic $D_T(\lambda)$ in a model with no breaks in either its JE or SE. However, the simulations presented in this section investigate

[1] The DGP adopted in this study is similar to that used in Hall et al. (2012).

the behaviour of the break fraction estimator $\hat{\lambda}$ obtained from such models with no breaks. The results reported cover three sample sizes of 140, 300 and 600.

The DGP of the SE in a Model with No Breaks

The SE is generated thus,

$$y_t = [1, x_t]'\theta_0 + u_t, \quad for \;\; t = 1, 2, \ldots, T, \tag{5.2}$$

where y_t is the dependent variable; x_t and u_t are as given in (5.1); θ_0 is the $p \times 1$ vector of coefficients with specific values $\theta_0 = [1, 0.1]'$.

The Simulation Results

We present eight simulation results under two categories. The first category comprises four simulation results obtained from estimating one break point only while the second category covers results obtained from estimating two break points.

Figures 5.1 and 5.2 show the histograms of the estimated break fractions $\hat{\lambda}$ across Λ in models with sample sizes of 140 and 300 respectively. For comparisons, Figures 5.3 and 5.4 show the results obtained using $D_T(\lambda)$ and $LM_T(\lambda)$ respectively, both from a sample size of 600. The similarity of $\hat{\lambda}$ in all four plots is clearly evident. Since $m_0 = 0$, any break fraction within the range has an equal chance of being estimated, hence, $\hat{\lambda}$ is relatively diffuse over Λ.

When two breaks are estimated, the randomness of the break fraction estimators becomes more apparent. Figures 5.5 and 5.6 show the two estimated break fractions plotted against each other. Note that for the Sequential Estimation Method, we define $\hat{k}_1$ and $\hat{k}_2$ as the first and second sequentially estimated break fractions[2] and based on Assumption 12 on page 61, the first estimated break fraction, $\hat{\lambda}_1 = min\{\hat{k}_1, \hat{k}_2\}$. From the results displayed, it is clear that similar to the first estimated break fraction, the second is equally dispersed across Λ.

For comparison, Figures 5.7 and 5.8 show plots of the two estimated break fractions obtained through Sequential and Simultaneous Estimation Methods respectively, using the

[2]Note that $\hat{k}_1$ and $\hat{k}_2$ are used in these Sequential estimation simulations to define the first and second sequentially estimated break points only and they should not be substituted with those used to denote break points in Chapter 3.

$Wald_T(\lambda)$. The computation for the Simultaneous Estimation Method is somewhat different. We follow a similar algorithm used in Bai and Perron (2003) and allow identification of the two break points as global maximisers of the test statistic, $D_T(\lambda)$. The optimisation is taken over all partitions such that the observations in each partition, $inf(k_0^{(i+1)} - k_0^{(i)}) \geq max\{q, [T\epsilon]\}$, as set out in Assumption 13. As seen, both break fractions are likewise randomly estimated.

Figure 5.1: Histogram of $\hat{\lambda}$ using $D_T(\lambda)$
Estimating 1 break; DGP has no break in SE or JE (T=140)

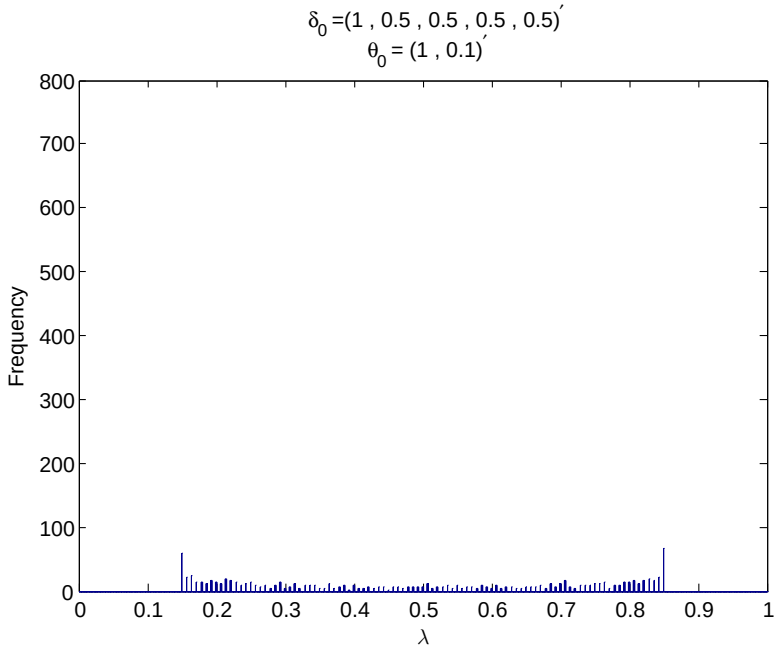


Figure 5.2: Histogram of $\hat{\lambda}$ using $D_T(\lambda)$
Estimating 1 break; DGP has no break in SE or JE (T=300)

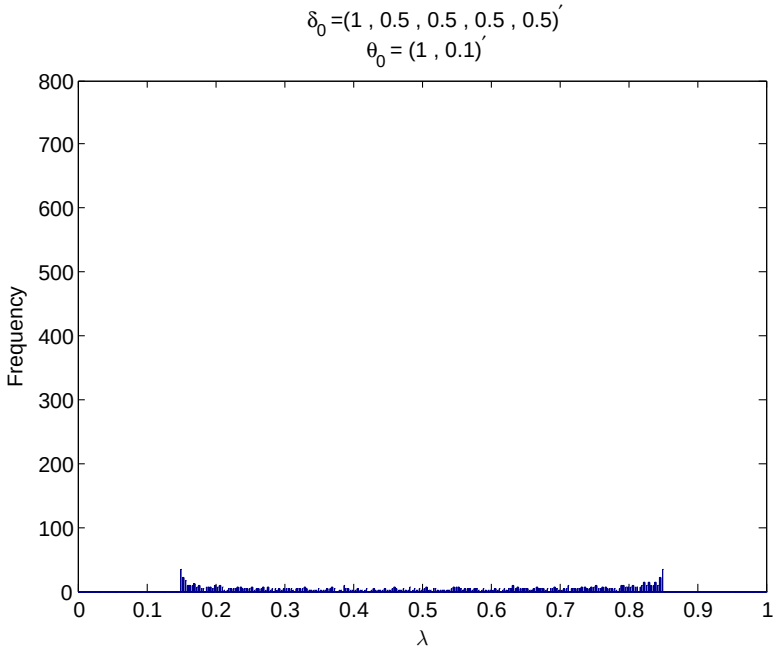


Figure 5.3: Histogram of $\hat{\lambda}$ using $D_T(\lambda)$
Estimating 1 break; DGP has no break in SE or JE (T=600)

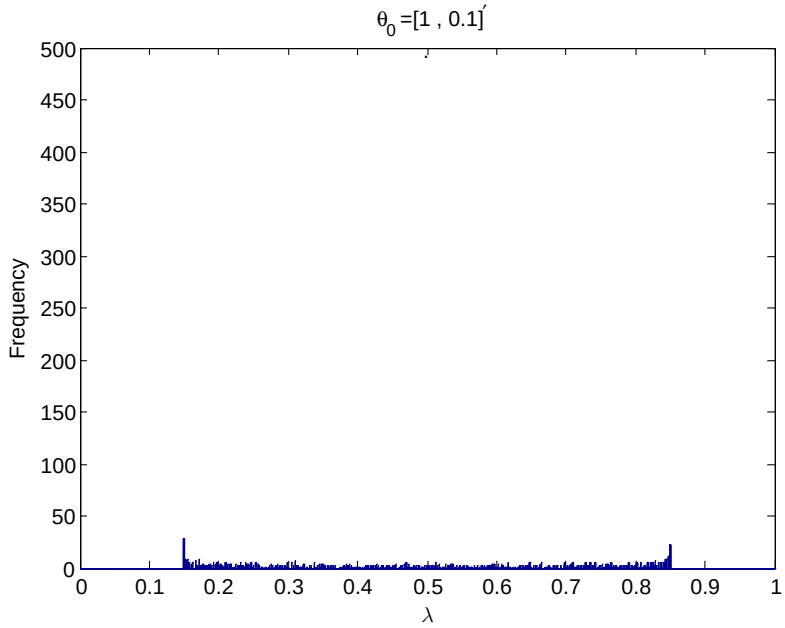


Figure 5.4: Histogram of $\hat{\lambda}$ using $LM_T(\lambda)$
Estimating 1 break; DGP has no break in SE or JE (T=600)

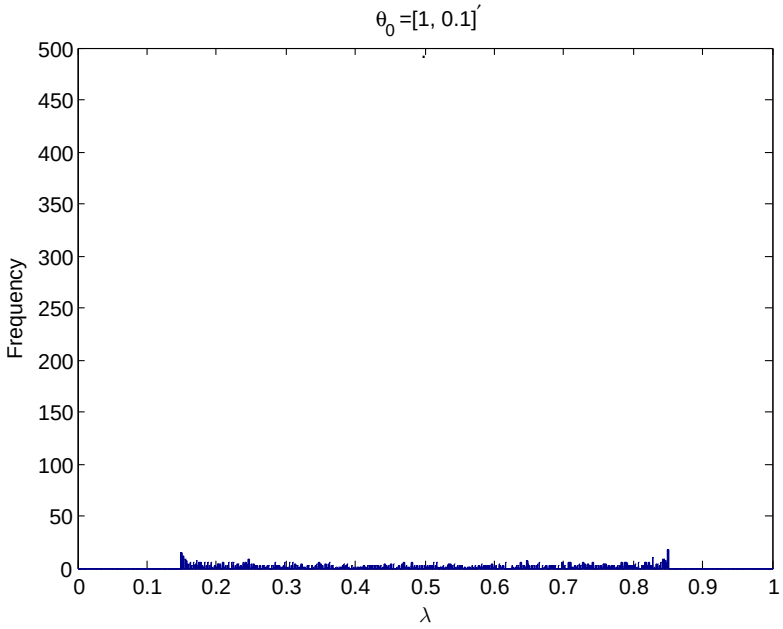


Figure 5.5: First against second estimated break fraction using $D_T(\lambda)$
Estimating 2 breaks; DGP has no break in SE or JE (T=140)

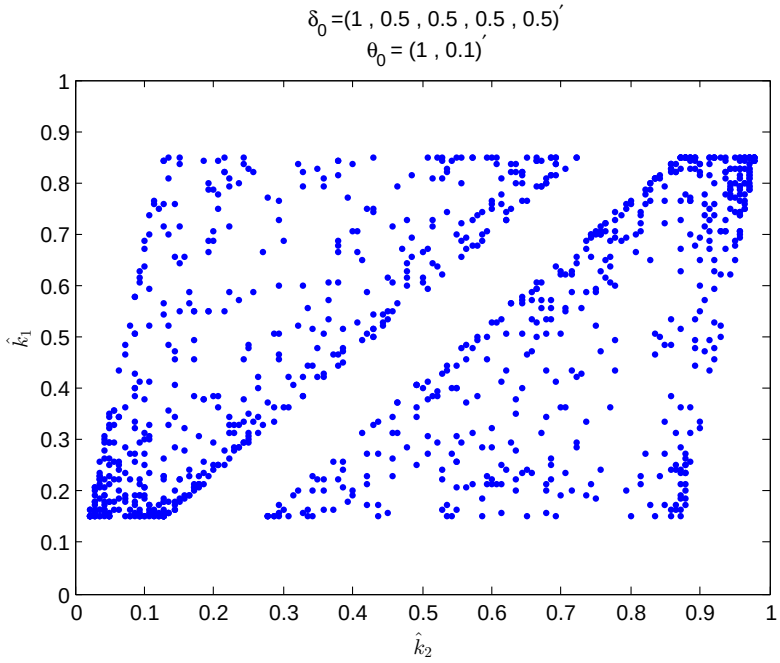


Figure 5.6: First against second estimated break fraction using $D_T(\lambda)$
Estimating 2 breaks; DGP has no break in SE or JE (T=300)

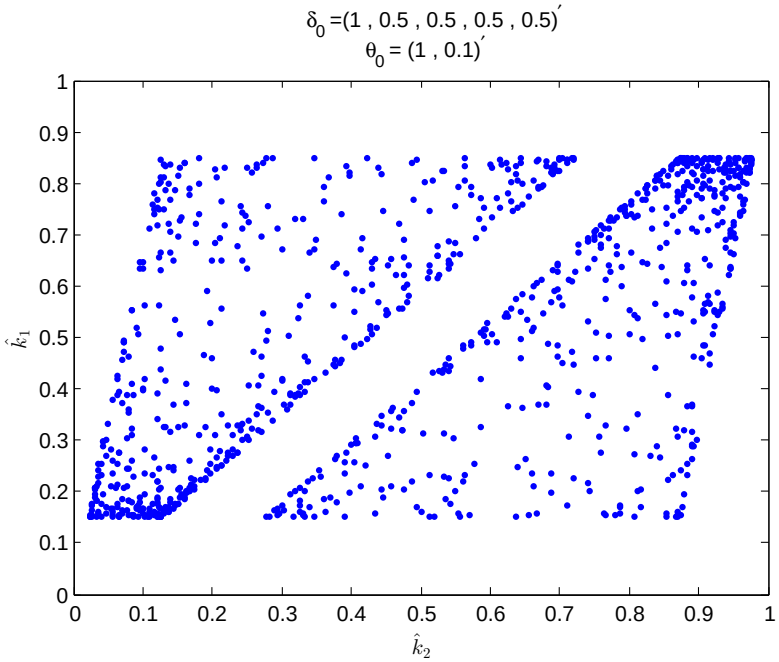


Figure 5.7: First against second estimated break fraction using $Wald_T(\lambda)$
(Sequential Estimation) Estimating 2 breaks; DGP has no break in SE or JE (T=600)

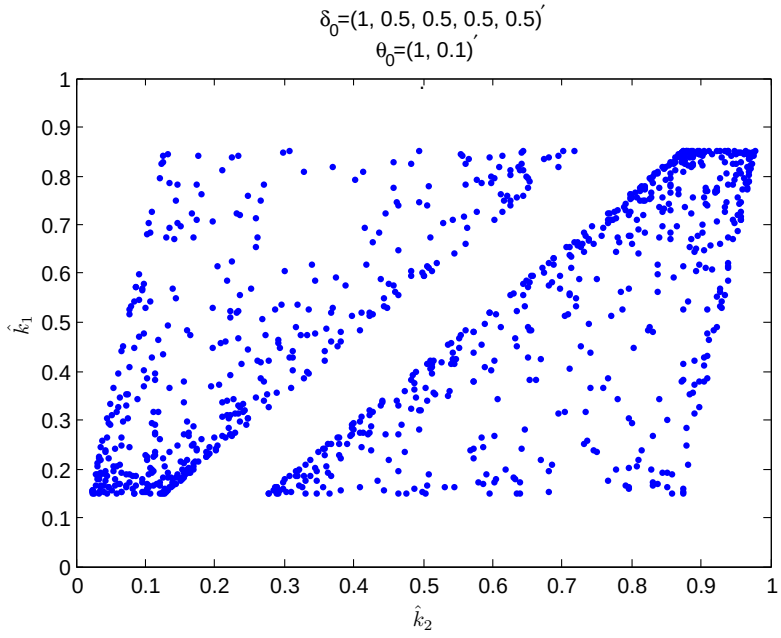
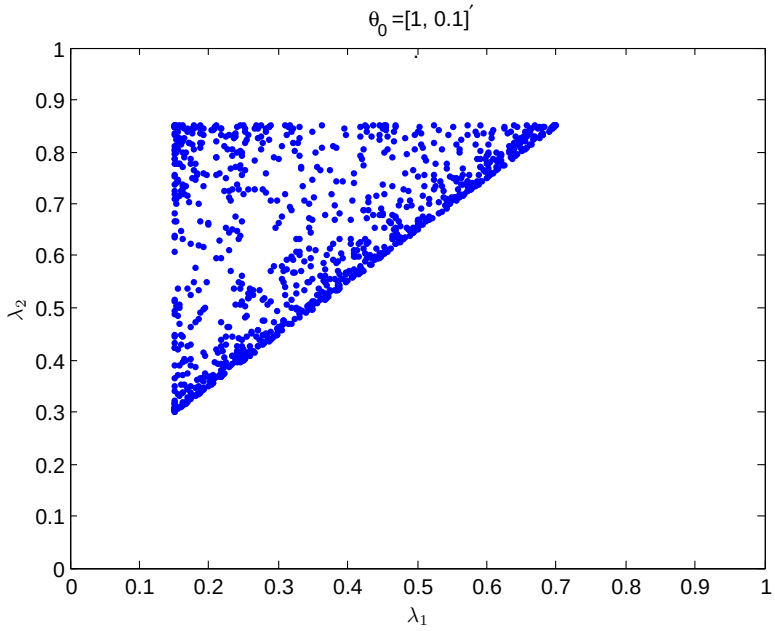


Figure 5.8: First against second estimated break fraction using $Wald_T(\lambda)$ (Simultaneous Est) Estimating 2 breaks; DGP has no break in SE or JE (T=600)



5.1.3 Simulation Results for a Model with One Break in the SE

This section presents results from models where the true break, $m_0 = 1$. The results shown are based on three sample sizes of 140, 300 and 600.

The DGP of the SE in a Model with One Break

The SE is generated thus,

$$y_t = [1, x_t]'\theta_0^{(1)} + u_t, \quad for \;\; t = 1, 2, \ldots, [\lambda_0 T] \tag{5.3}$$

$$= [1, x_t]'\theta_0^{(2)} + u_t, \quad for \;\; t = [\lambda_0 T] + 1, \ldots, T, \tag{5.4}$$

where $\theta_0^{(1)}$ and $\theta_0^{(2)}$ are the parameters before and after the true break point with specific values, $(\theta_0^{(1)}, \theta_0^{(2)}) = \left([1, 0.1]', [-1, -0.1]'\right)$. The results shown are based on two different locations of the break fraction in the model, when $\lambda_0 = 0.5$ and 0.7.

The Simulation Results

We present eight simulation results organised into four categories. The first two categories present the results obtained from estimating one break point only while the last

two categories reports results from estimating two break points.

In the first category, the histograms of $\hat{\lambda}$ obtained from models with sample sizes 140 and 600 are displayed in Figures 5.9 and 5.10 respectively. The break fraction estimators are clearly concentrated around the true break of $\lambda_0 = 0.5$, with the precision increasing as the sample size gets larger. The two histograms in the second category shown in Figures 5.11 and 5.12 also display $\hat{\lambda}$, albeit the break in the SE is moved to $\lambda_0 = 0.7$. Yet again, the histogram is clearly centred at this true break point.

In the third category, the histograms of the two estimated break fractions are displayed in Figures 5.13 and 5.14 for sample sizes of 140 and 300 respectively. The first estimated break fraction converges at the true break where $\lambda = 0.5$, while the second break fraction is randomly estimated across the range of Λ.

The last category compares the two estimated break fractions obtained from the Sequential Estimation Method shown in Figure 5.15 with those obtained from the Simultaneous Estimation Method shown in Figure 5.16. For the Sequential estimation method, the first break fraction is always the true break fraction, shown by $\hat{k}_1 = 0.5$, while the second break fraction is randomly selected, as seen by the spread of $\hat{k}_2$. This supports the theory as laid out in Chapter 3 that the dominant break fraction (which is the only true break fraction in this case) will always be estimated first.

In the Simultaneous estimations results, either of the two estimated break fractions always identifies the true break at $\lambda_0 = 0.5$. However, neither of them is consistent for it in particular. So when one break fraction estimates the true break, the other is arbitrarily estimated from any of the candidate break fractions within Λ.

The focus here however, should be on the fact that the true break point is always estimated as indeed if there was a true second break point in the model, it would be consistently estimated as the simulation results from the multiple break models show. Therefore, both the Sequential and Simultaneous Estimation Methods yield similar outcomes as they both convincingly estimated the location of the true break point in the model.

Figure 5.9: Histogram of $\hat{\lambda}$ using $D_T(\lambda)$
Estimating 1 break; DGP has 1 break in SE (T=140)

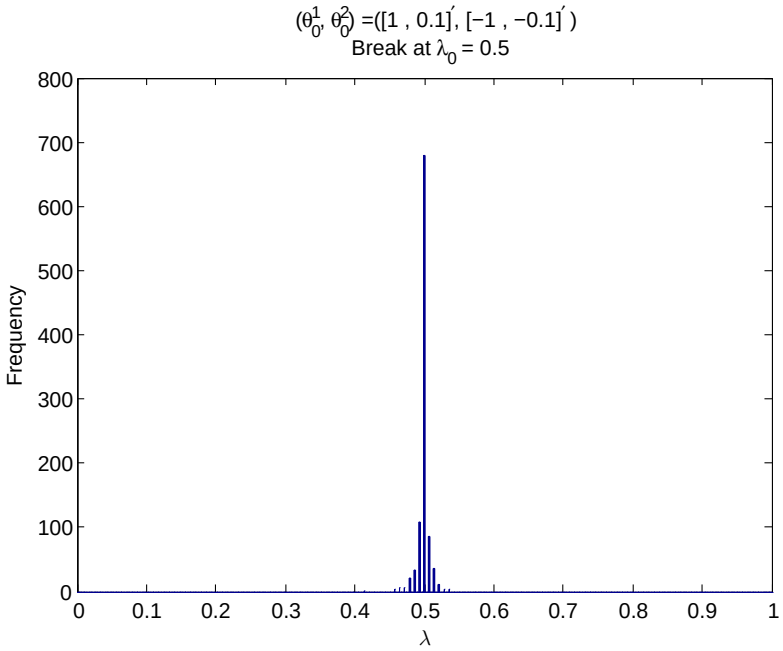


Figure 5.10: Histogram of $\hat{\lambda}$ using $D_T(\lambda)$
Estimating 1 break; DGP has 1 break in SE (T=600)

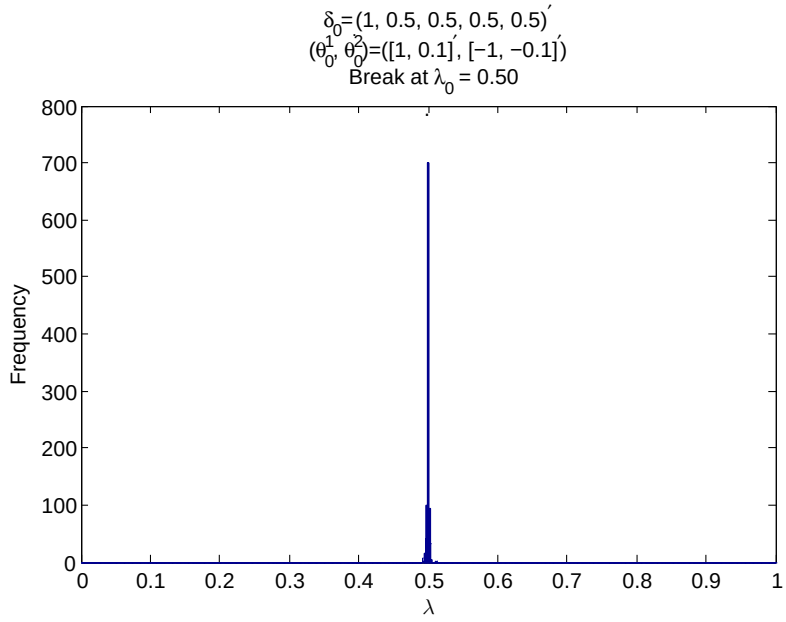


Figure 5.11: Histogram of $\hat{\lambda}$ using $D_T(\lambda)$
Estimating 1 break; DGP has 1 break in SE (T=140)

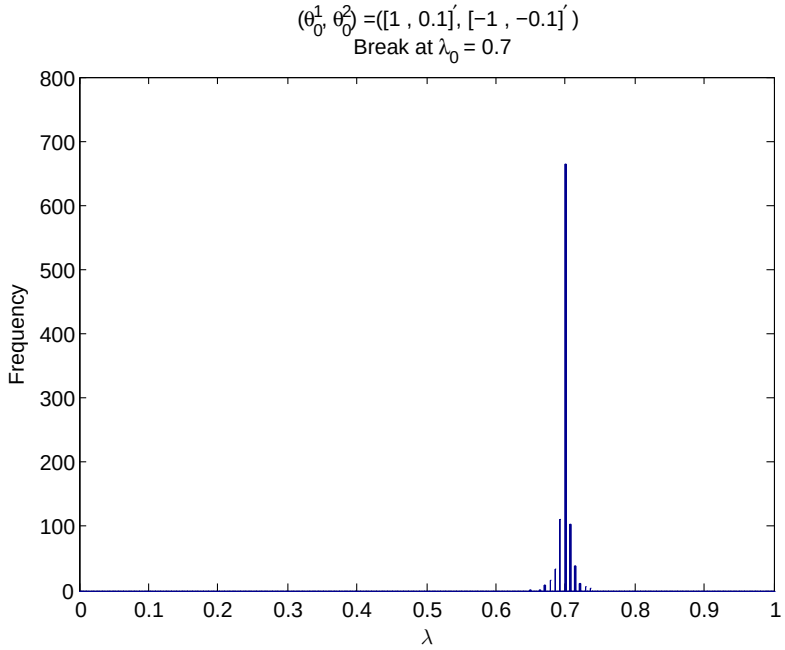


Figure 5.12: Histogram of $\hat{\lambda}$ using $D_T(\lambda)$
Estimating 1 break; DGP has 1 break in SE (T=600)

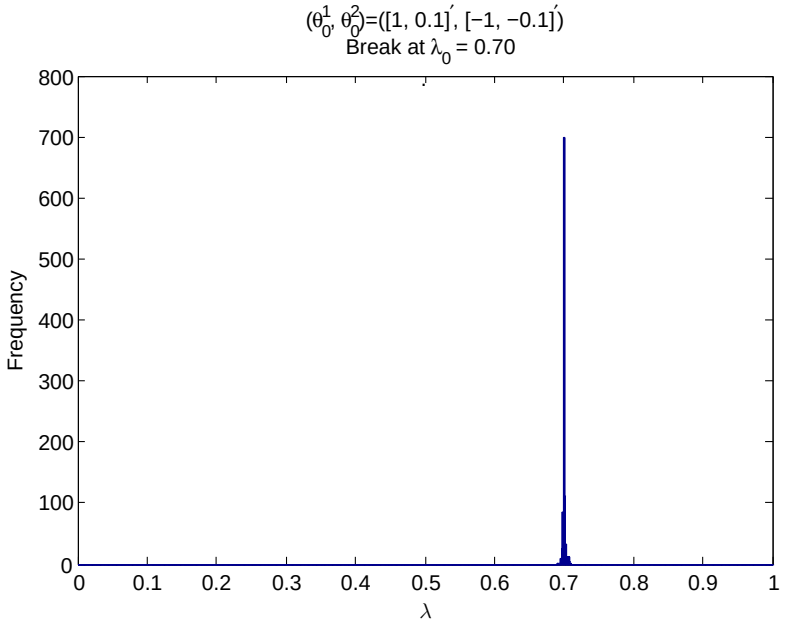


Figure 5.13: Histogram of $\hat{\lambda}$ using $D_T(\lambda)$
Estimating 2 breaks; DGP has 1 break in SE (T=140)

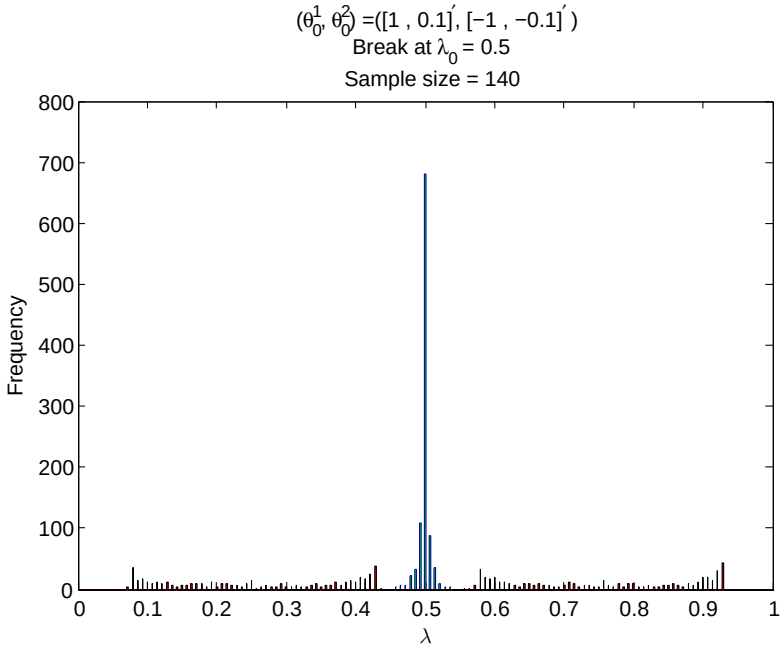


Figure 5.14: Histogram of $\hat{\lambda}$ using $D_T(\lambda)$
Estimating 2 breaks; DGP has 1 break in SE (T=300)

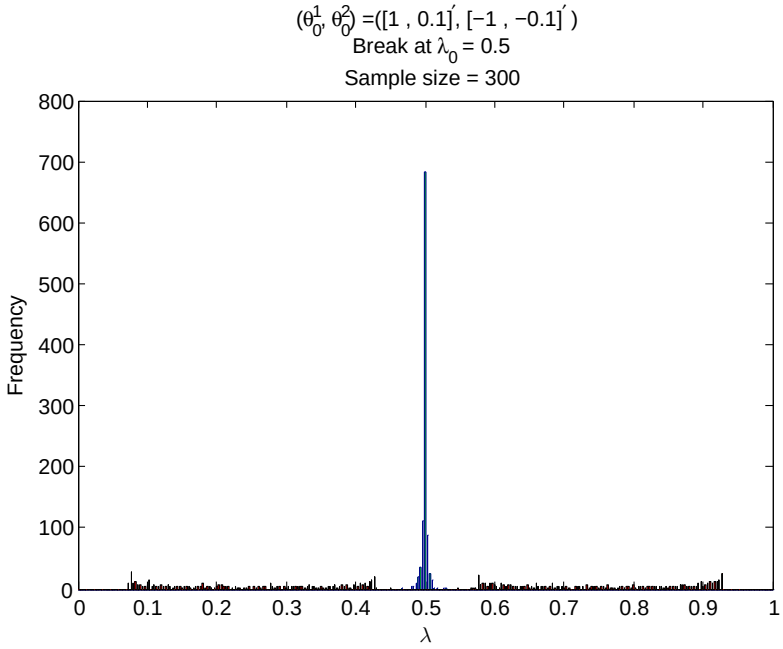


Figure 5.15: First against second estimated break fraction using $D_T(\lambda)$
(Sequential Estimation Method) Estimating 2 breaks; DGP has 1 break in SE (T=600)

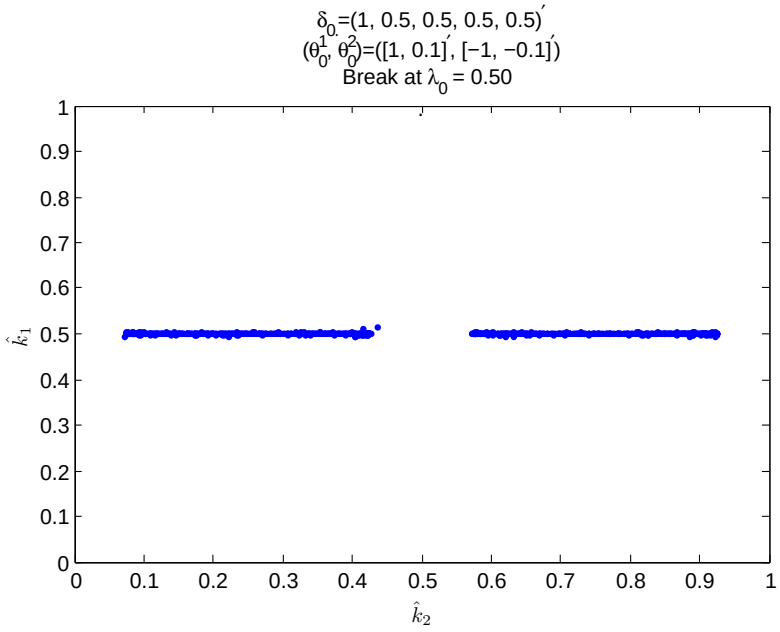
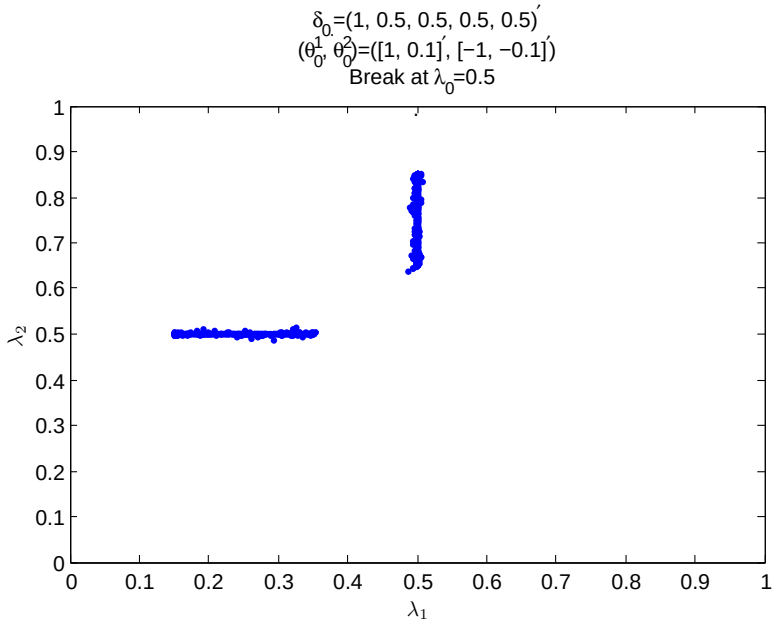


Figure 5.16: First against second estimated break fraction using $D_T(\lambda)$
(Simultaneous Est) Estimating 2 breaks; DGP has 1 break in SE (T=600)



5.1.4 Simulation Results for a Model with Two Breaks in the SE

The results reported in this section are for models where $m_0 = 2$. The results are split into two broad categories. The first considers models which have equal magnitude of shifts as displayed in DGP 1 below while the second looks at models with unequal shifts as given in DGP's 2 and 3. The simulation results in the first category are based on three sample sizes of 140, 260 and 420 while those in the second category use a sample size of 300 and 600.

As highlighted earlier in the results for the one break model in Subsection 5.1.3, by definition and based on Assumption 12 on page 61, we have that $\lambda_0^{(1)} < \lambda_0^{(2)}$. In line with the sequential procedure outlined in Section 3.2, we define $\hat{k}_1$ as the first estimated break fraction obtained when using the whole sample for estimations since this first break fraction will not always necessarily be at $\lambda_0^{(1)}$ or $\lambda_0^{(2)}$. Likewise, $\hat{k}_2$ is defined as the second estimated break fraction obtained from the second round of estimations using the subsamples. Finally, we define the first and second break fraction estimators simply as $\hat{\lambda}_1 = min\{\hat{k}_1, \hat{k}_2\}$ and $\hat{\lambda}_2 = max\{\hat{k}_1, \hat{k}_2\}$ respectively.

The DGP of the SE in a Model with Two Breaks

In both categories, the true break fractions are placed at $\lambda_0^{(1)} = 0.33$ and $\lambda_0^{(2)} = 0.67$. Since the magnitude of shifts differs in both categories, we present the DGP for a generic SE as

$$y_t = [1, x_t]'\theta_0^{(i)} + u_t, \quad for \;\; t = 1, 2, \ldots, T \;\; and \;\; i = 1, 2, 3. \tag{5.5}$$

Specifically,

For DGP 1: $(\theta_0^{(1)}, \theta_0^{(2)}, \theta_0^{(3)}) = \big([1, 0.1]', [-1, -0.1]', [1, 0.1]'\big)$.

For DGP 2: $(\theta_0^{(1)}, \theta_0^{(2)}, \theta_0^{(3)}) = \big([1, 0.1]', [-1, -0.1]', [1, 4.0]'\big)$.

For DGP 3: $(\theta_0^{(1)}, \theta_0^{(2)}, \theta_0^{(3)}) = \big([1, 2.0]', [-1, -0.1]', [1, 0.1]'\big)$.

The Simulation Results

Six simulation results are presented in the first category using DGP 1. In Figures 5.17 and 5.18, the histograms of the first and second estimated break fractions obtained from a model with sample size 420 are respectively displayed. As conjectured on page 71, since the breaks in the model are of equal magnitude, then no particular break fraction

dominates the other and hence, the break fraction estimator $\hat{\lambda}$ converges to a random variable with equal mass at $\lambda_0^{(1)}$ and $\lambda_0^{(2)}$.

Figures 5.19 and 5.20 show the histograms of the two estimated break fractions together in a model with sample size 140 and 260 respectively. The distribution is similar in both models with clustering around the two true break fractions.

In Figures 5.21 and 5.22, we present the bivariate histograms of the two estimated break fractions in models with sample sizes 260 and 420 respectively. As seen, the break fraction estimator is not consistent for a specific break but more importantly, they identify either of the two true break fractions. The similarities in the heights of the histograms in the two models indicate both break fractions are equally estimated.

For the second category, six simulations are presented to compare the break fraction estimators obtained from a model using DGP 1 with those obtained from DGPs 2 and 3. First, notice in DGP 2, the second break point, $\lambda_0^{(2)}$ is the dominant break because $\|\theta_0^{(2)} - \theta_0^{(3)}\| > \|\theta_0^{(1)} - \theta_0^{(2)}\|$; while in DGP 3, $\lambda_0^{(1)}$ is the dominant break fraction because $\|\theta_0^{(1)} - \theta_0^{(2)}\| > \|\theta_0^{(2)} - \theta_0^{(3)}\|$.

In Figures 5.23 and 5.24, we compare the histograms of the first estimated break point using DGP 1 and DGP 2 respectively, both with a sample size of 300. For the equal shifts in Figure 5.23, either of the two break fractions is estimated first, similar to category one above. This is in contrast to that in Figure 5.24 where the dominant break fraction, $\lambda_0^{(2)} = 0.67$, is estimated first. Similarly, for the second estimated break fraction displayed in Figures 5.25 and 5.26, the difference is again clearly seen. While either of the two true breaks is identified using the equal magnitude of shifts in DGP 1, on the other hand, $\lambda_0^{(1)} = 0.33$ is always estimated when using DGP 2. These results are consistent with Proposition 3 on page 66 because the first estimated break fraction always converges to the dominant break fraction, which is $\lambda_0^{(2)}$ in this case.

The last two results presented in this section compare the bivariate histograms of the two estimators obtained from DGPs 1 and 3, both using sample sizes of 600. These histograms immediately highlight the relationship between the first and second estimated break fractions. From the results of the model with equal breaks displayed in Figure 5.27, the two true break points are again equally identified by both break fraction estimators. Conversely, using DGP 3 with unequal shifts, Figure 5.28 clearly indicates that the more dominant break fraction, $\lambda_0^{(1)} = 0.33$ is always estimated first while $\lambda_0^{(2)} = 0.67$ is always estimated second.

The estimation results in the second category highlights one of the advantages of the Sequential Estimation Method; it affords the researcher the possibility of recognising

which of the breaks in a model is more dominant if this is a feature desired.

Figure 5.17: Histogram of the first estimated break fraction using $D_T(\lambda)$
First estimated break; DGP has 2 breaks in SE (T=420)

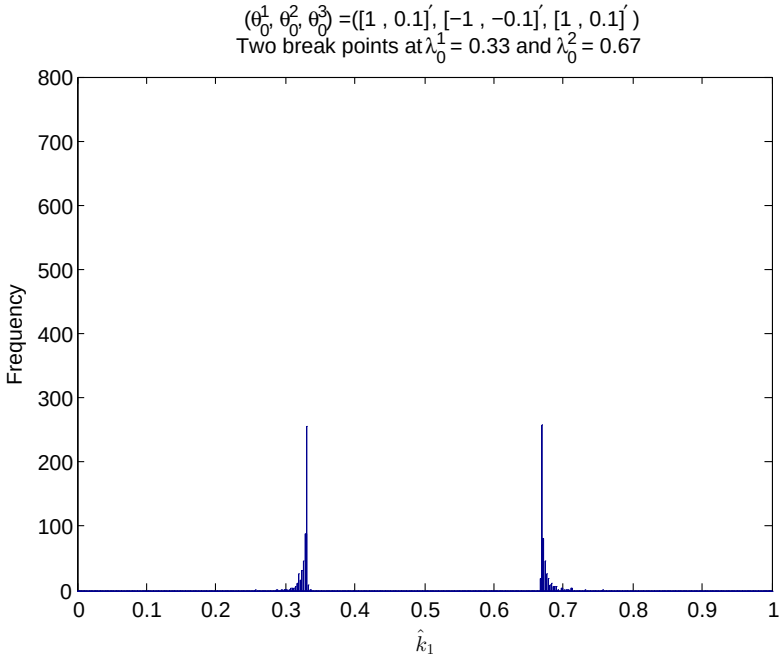


Figure 5.18: Histogram of the second estimated break fraction using $D_T(\lambda)$
Second estimated break; DGP has 2 breaks in SE (T=420)

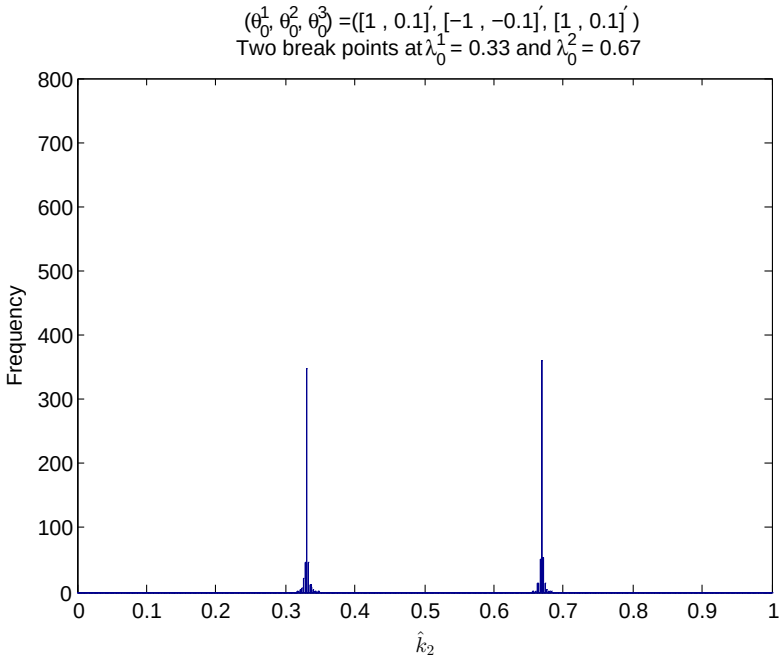


Figure 5.19: Histogram of the two estimated break fractions using $D_T(\lambda)$
Estimating 2 breaks; DGP has 2 breaks in SE (T=140)

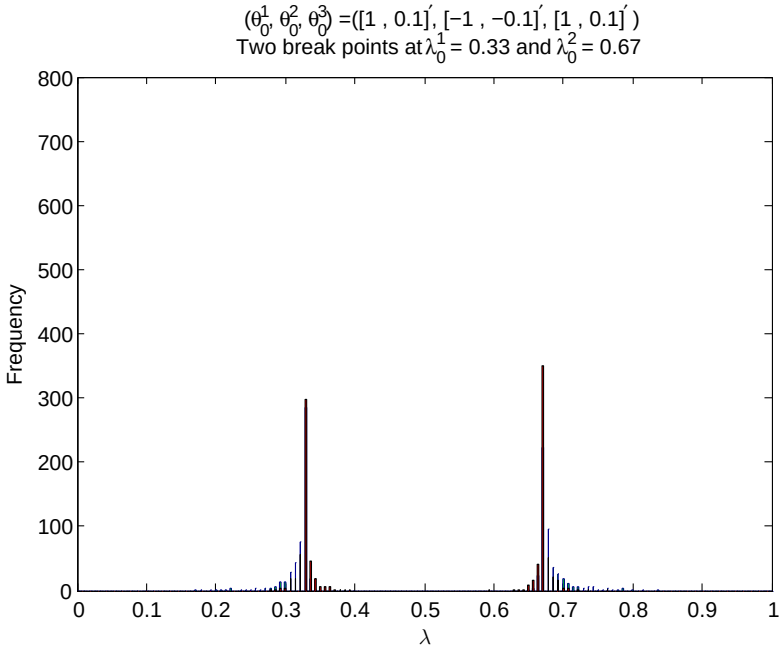


Figure 5.20: Histogram of the two estimated break fractions using $D_T(\lambda)$
Estimating 2 breaks; DGP has 2 breaks in SE (T=260)

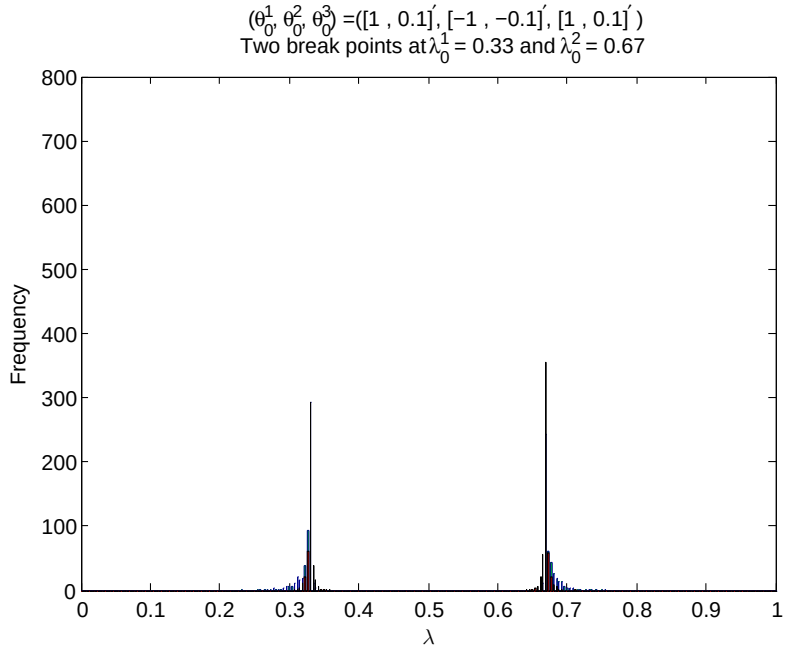


Figure 5.21: Bivariate Histogram of both break fractions using $D_T(\lambda)$
Estimating 2 breaks; DGP has 2 breaks in SE (T=260)

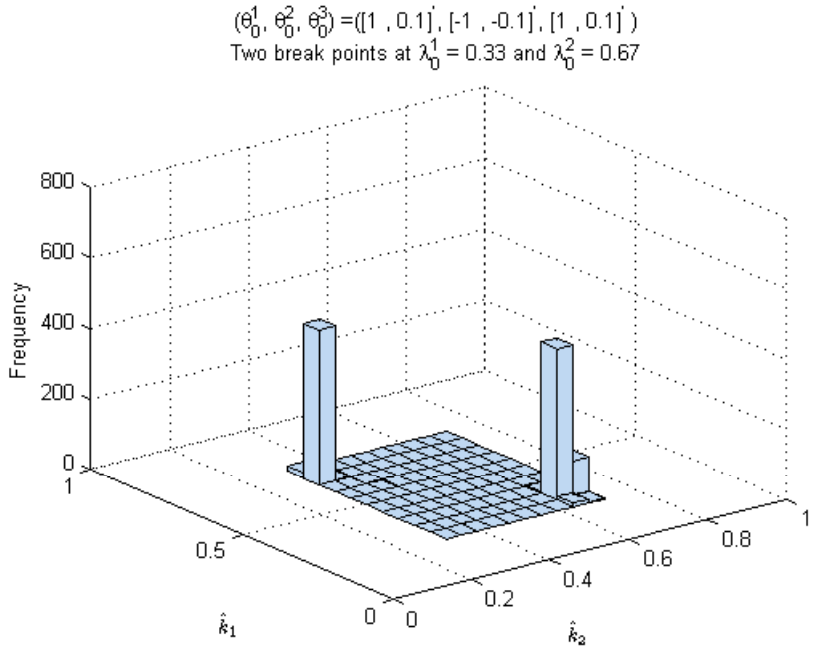


Figure 5.22: Bivariate Histogram of both break fractions using $D_T(\lambda)$
Estimating 2 breaks; DGP has 2 breaks in SE (T=420)

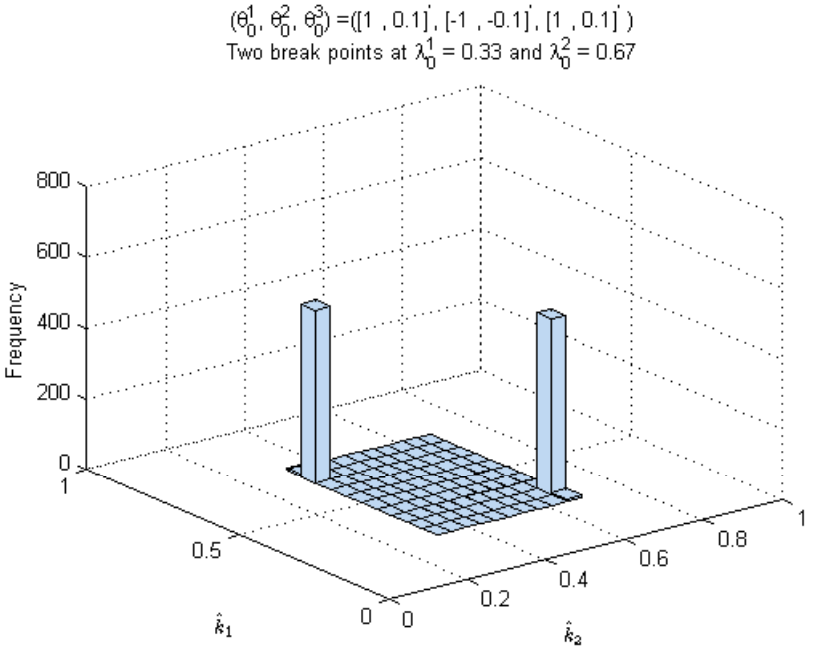


Figure 5.23: Histogram of the first estimated break fraction using $D_T(\lambda)$
DGP has 2 breaks of equal magnitude in SE (T=300)

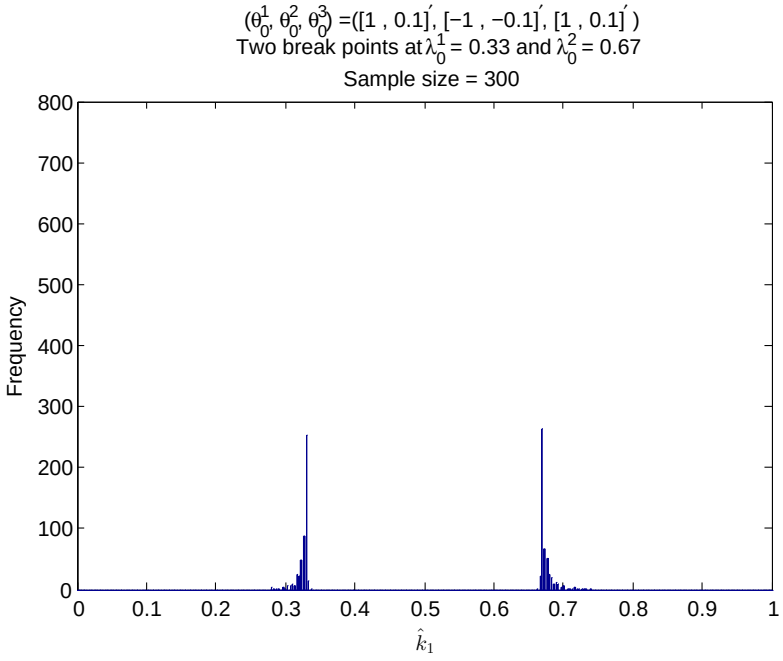


Figure 5.24: Histogram of the first estimated break fraction using $D_T(\lambda)$
DGP has 2 breaks of unequal magnitude in SE (T=300)

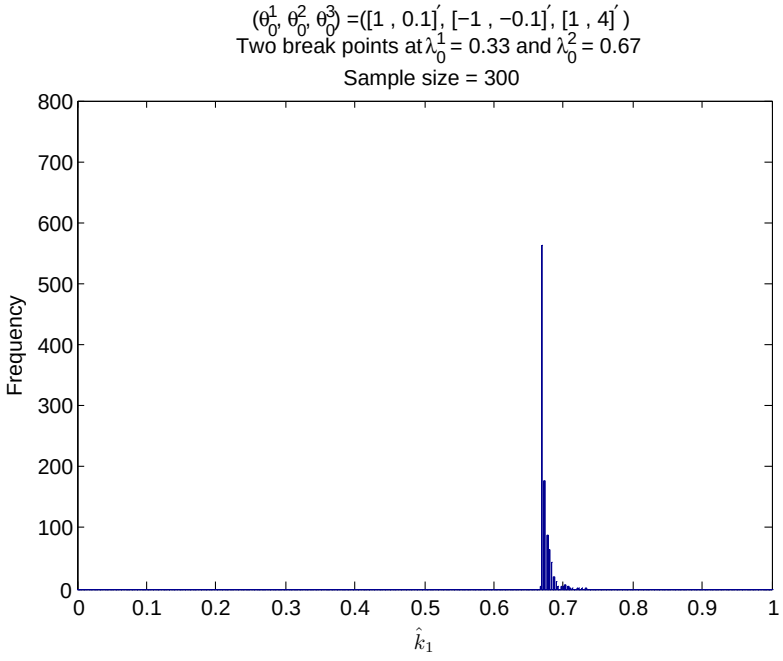


Figure 5.25: Histogram of the second estimated break fraction using $D_T(\lambda)$
DGP has 2 breaks of equal magnitude in SE (T=300)

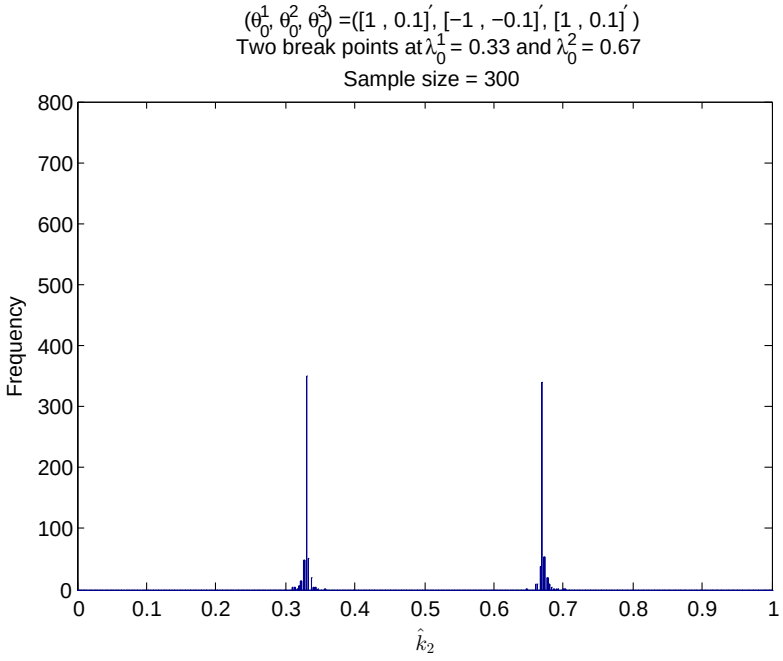


Figure 5.26: Histogram of the second estimated break fraction using $D_T(\lambda)$
DGP has 2 breaks of unequal magnitude in SE (T=300)

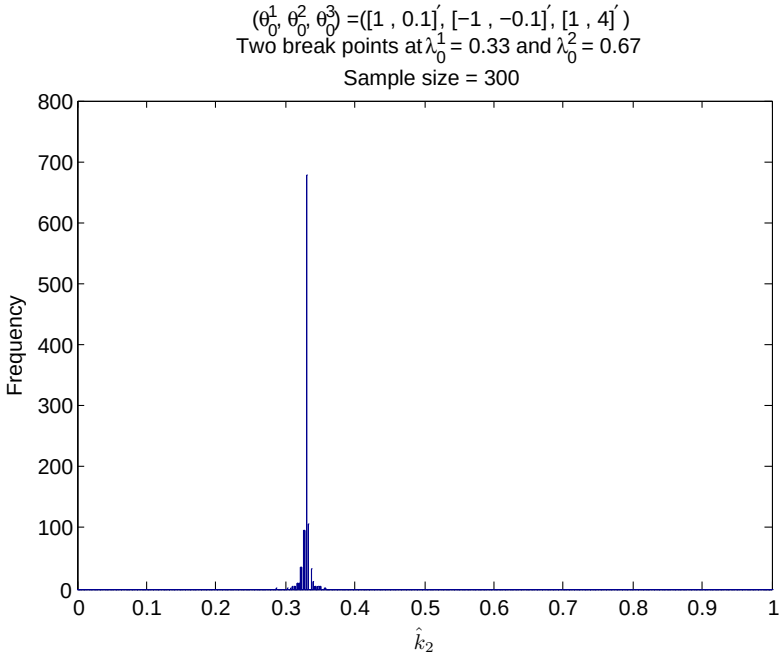


Figure 5.27: Bivariate Histogram of both break fractions using $D_T(\lambda)$
DGP has 2 breaks of equal magnitude in SE (T=600)

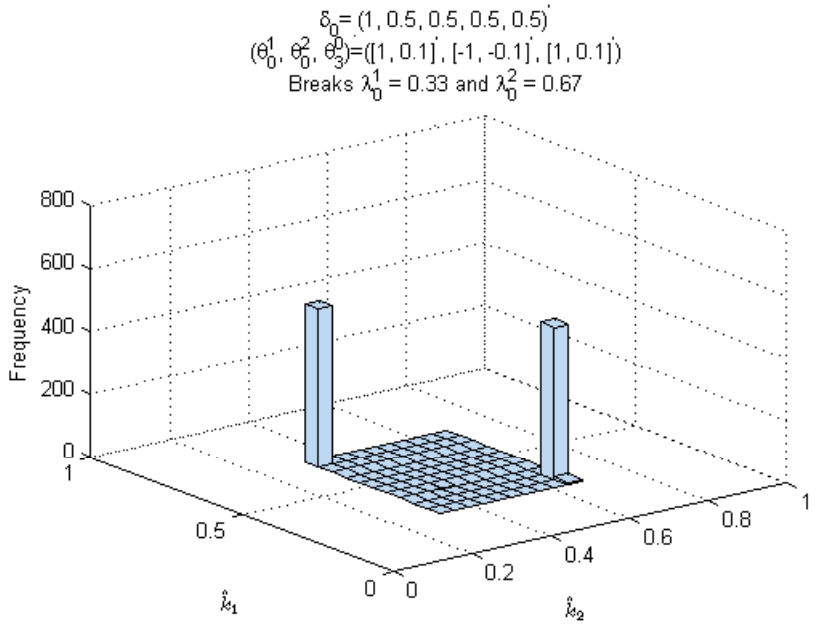
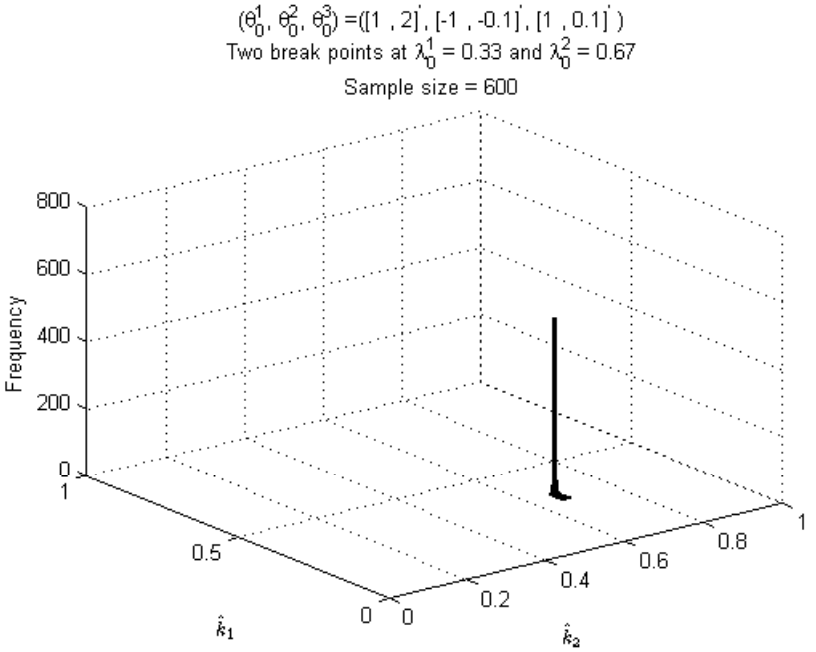


Figure 5.28: Bivariate Histogram of both break fractions using $D_T(\lambda)$
DGP has 2 breaks of unequal magnitude in SE (T=600)



5.1.5 Simulation Results for a Model with Three Breaks in the SE

The simulation results presented in this subsection where $m_0 = 3$ are organised in two categories. The first category presents results for models with equal magnitudes of shifts while the second presents models with unequal magnitudes of shifts. The results are based on three sample sizes of 140, 300 and 420.

The DGP of the SE in a Model with Three Breaks

The three true break fractions are placed at $\lambda_0^{(1)} = 0.25, \lambda_0^{(2)} = 0.50$ and $\lambda_0^{(3)} = 0.75$.

To cover the two categories, we again present two DGPs with the generic SE given as

$$y_t = [1, x_t]'\theta_0^{(i)} + u_t, \quad for \;\; t = 1, 2, \ldots, T, \;\; and \; i = 1, 2, 3, 4. \tag{5.6}$$

Specifically,

For DGP 1: $(\theta_0^{(1)}, \theta_0^{(2)}, \theta_0^{(3)}, \theta_0^{(4)}) = \big([1, 0.1]', [-1, -0.1]', [1, 0.1]', [-1, -0.1]'\big)$.

For DGP 2: $(\theta_0^{(1)}, \theta_0^{(2)}, \theta_0^{(3)}, \theta_0^{(4)}) = \big([1, 0.1]', [-1, -0.1]', [1, 2.0]', [-1, -0.5]'\big)$.

The Simulation Results

We present six simulation results across both categories. The first two display results for the first category using DGP 1 which has equal sizes of shifts. These are seen in Figures 5.29 and 5.30 where the histograms of all the three estimated break fractions obtained from models with sample sizes of 140 and 420 are respectively displayed. Consistent with Proposition 3 on page 66, the break fraction estimators are clearly seen to converge towards the true break fractions (0.25, 0.5 or 0.75) as the sample size T increases.

The second category which uses DGP 2 presents four results all from a model with a sample size of 300. First, notice the magnitudes of the breaks in DGP 2 have been ordered so $\|\theta_0^{(3)} - \theta_0^{(4)}\| > \|\theta_0^{(2)} - \theta_0^{(3)}\| > \|\theta_0^{(1)} - \theta_0^{(2)}\|$. Consequently, $\lambda_0^{(3)}$ is the most dominant, followed by $\lambda_0^{(2)}$. Hence in line with the theory, we expect $\lambda_0^{(3)}$ to be estimated first before $\lambda_0^{(2)}$ and then lastly, $\lambda_0^{(1)}$. Each of the three estimated break fractions are displayed in turn.

In Figures 5.31 to 5.33, we display the histograms of the first, second and third estimated break fractions respectively. The results are as expected because $\hat{k}_1$ distinctly converges to $\lambda_0^{(3)}$, the most dominant break as seen in Figure 5.31; $\hat{k}_2$ converges to $\lambda_0^{(2)}$ as seen in Figure 5.32 and $\hat{k}_3$ converges to $\lambda_0^{(1)}$ as seen in Figure 5.33.

Lastly, in Figure 5.34 we present a subplot which shows the second and third estiamted break fraction plotted against the first estimated break fraction. The first estimated break fraction is placed on the x axis, while the second and third estimated break fractions are placed on the top and bottom half of the y axis respectively. This subplot provides a more convenient way to view the three estimated breaks concurrently. One can readily see that the first break fraction is consistently around $\lambda = 0.75$, while the second and third break fractions are around $\lambda = 0.5$ and 0.25 respectively.

These simulation results displayed under the second category further verify Proposition 3 on page 66 because the estimated break fraction always converges to the dominant break fraction.

Figure 5.29: Histogram of the three estimated break fractions using $D_T(\lambda)$
Estimating 3 breaks; DGP has 3 breaks in SE (T=140)

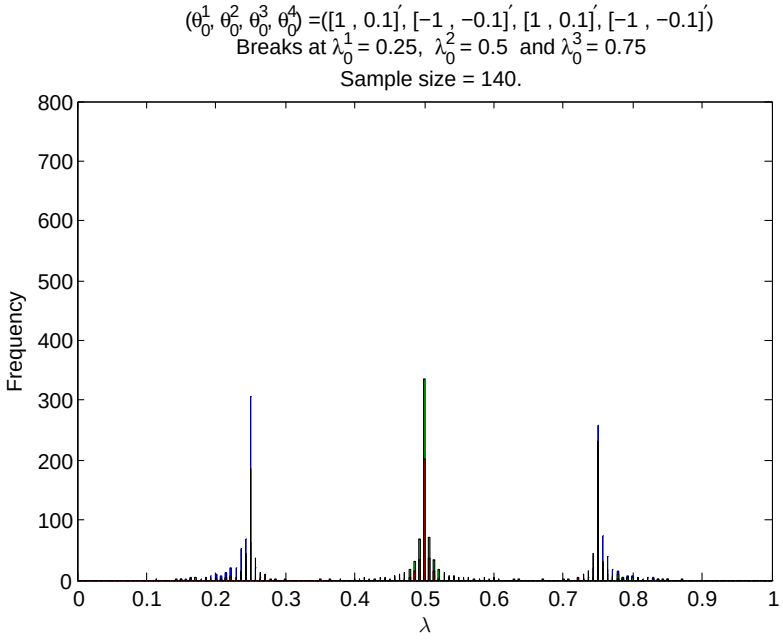


Figure 5.30: Histogram of the three estimated break fractions using $D_T(\lambda)$
Estimating 3 breaks; DGP has 3 breaks in SE (T=420)

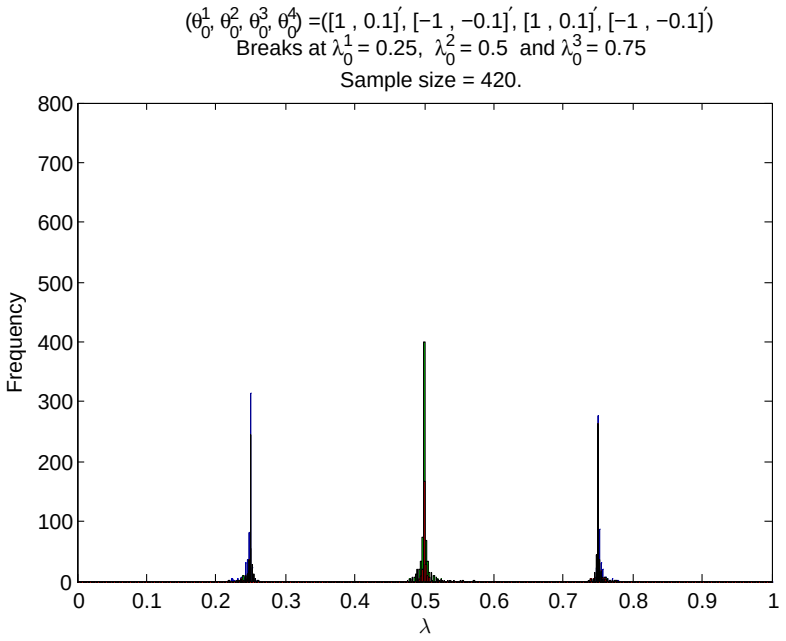


Figure 5.31: Histogram of the first estimated break fraction using $D_T(\lambda)$
DGP has 3 breaks of unequal magnitude in SE (T=300)

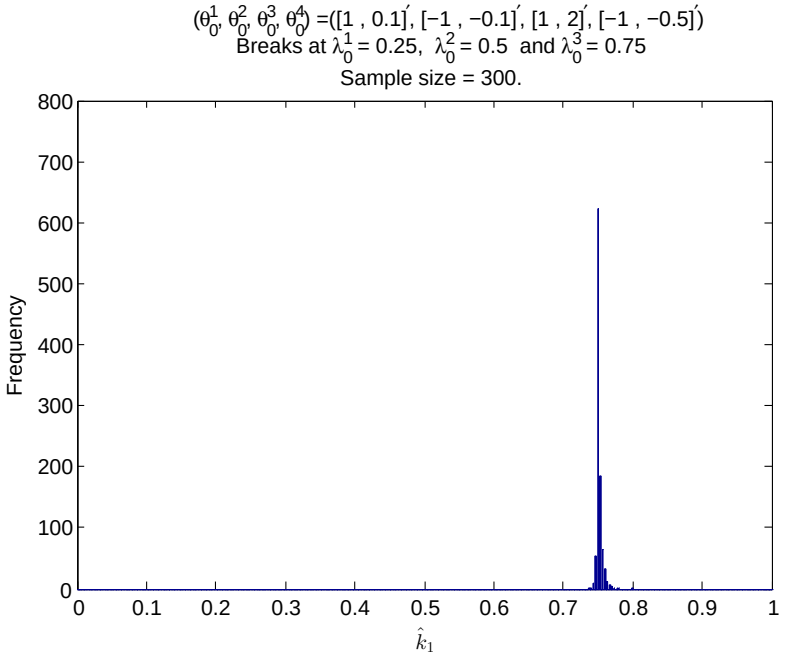


Figure 5.32: Histogram of the second estimated break fraction using $D_T(\lambda)$
DGP has 3 breaks of unequal magnitude in SE (T=300)

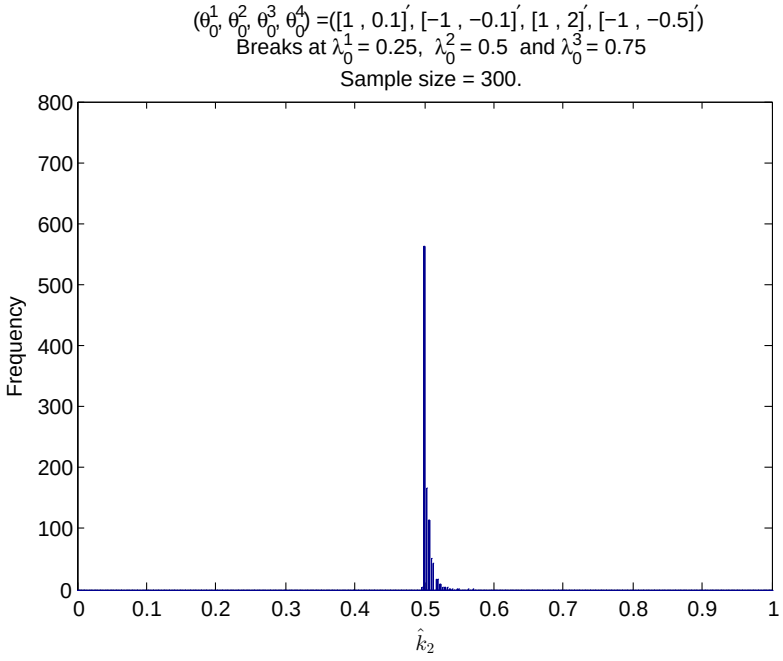


Figure 5.33: Histogram of the third estimated break fraction using $D_T(\lambda)$
DGP has 3 breaks of unequal magnitude in SE (T=300)

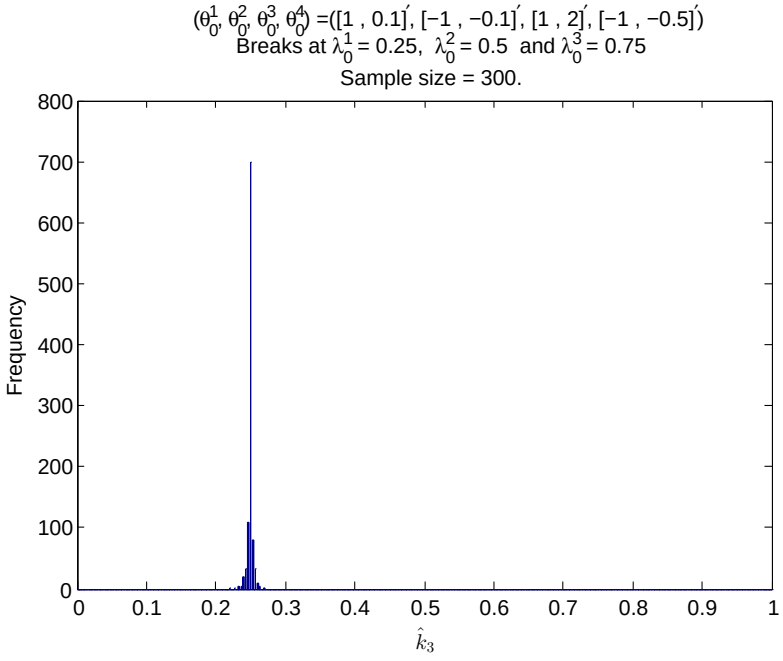
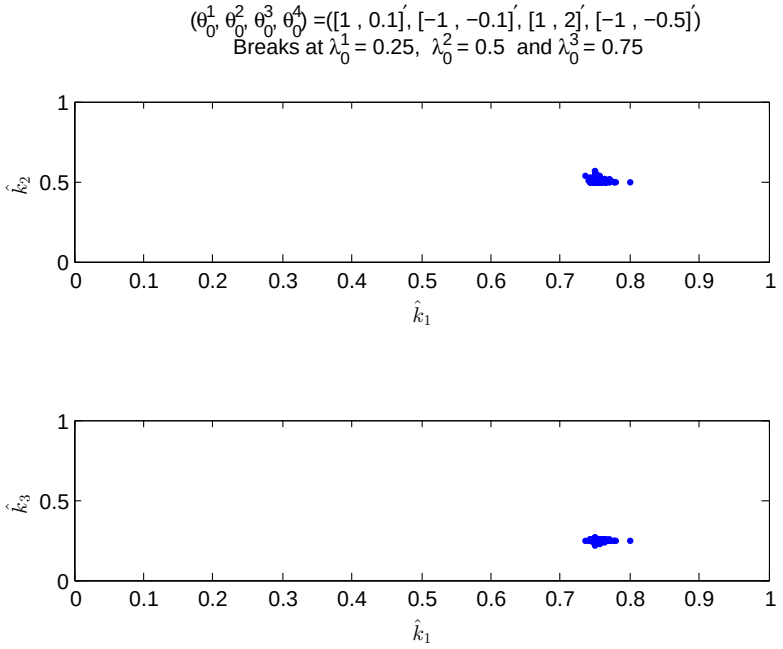


Figure 5.34: Subplot of the three estimated break fractions using $D_T(\lambda)$
DGP has 3 breaks of unequal magnitude in SE (T=300)



5.2 Simulation Results from Unstable Jacobian models

This section presents the Monte Carlo simulation results from three classes of models which have an unstable Jacobian, that is, models which have a break in the relationship between the endogenous regressors, x_t and the instruments, z_t. The first class of models has only one break in the JE and none in the SE while the second and third classes of models have one break each in both the JE as well as the SE. The breaks in the second class occur at the same location in the sample while those in the third class occur at different locations in the sample. As in Section 4.2, these are respectively referred to as Coincidental and Separate Breaks. The results cover models of different sample sizes. As in the Stable Jacobian simulations, we also report results from estimating two break points, irrespective of the true number of break points in the model.

5.2.1 The DGP for Unstable Jacobian Models

Due to the break in the Jacobian, it is compulsory the JE given in (5.1) on page 129 has to be modified. Throughout this section, we construct the JE with a single break as,

$$x_t = [1, z_t]' \delta_0^{(1)} + v_t, \quad for\ t = 1, 2, \ldots, [T\pi_0] \tag{5.7}$$
$$= [1, z_t]' \delta_0^{(2)} + v_t, \quad for\ t = [T\pi_0] + 1, \ldots, T, \tag{5.8}$$

where $(\delta_0^{(1)}, \delta_0^{(2)}) = ([1, 0.8, 0.8, 0.8, 0.8]', [0.1, 0.2, 0.2, 0.2, 0.2]')$ and the true break fraction in the JE, $\pi_0 = 0.5$. All other variables are constructed as outlined in the base model given in (5.1) on page 129.

The SE on the other hand, is presented alongside the classes of models being considered below.

5.2.2 Simulation Results for a Model with One Break in the JE only

The models in this subsection have breaks only in the JE, as such, the SE is stable and constructed as in the model with no breaks used in Subsection 5.1.2. We rewrite it here for easy reference,

$$y_t = [1, x_t]' \theta_0 + u_t, \quad for\ \ t = 1, 2, \ldots, T, \quad where\ \ \theta_0 = [1, 0.1]'.$$

The Simulation Results

We present four results from the Monte Carlo simulation experiment carried out on models with a break existing only in the JE. Figures 5.35 and 5.36 display the histograms of the first estimated break fraction in models with sample sizes of 140 and 600 respectively. Clearly, these break fraction estimators are dispersed throughout the range of Λ indicating it does not identify any particular break in the model. In Figures 5.37 and 5.38, the two estimated break fractions are plotted against each other and as expected, both break fraction estimators randomly identify any of the candidate break fractions hence, they are both spread across Λ.

The results in this section are similar to those obtained in the case of a model with no breaks in the SE presented earlier in Subsection 5.1.2. Thus, the simulation results presented here support the theoretical analysis carried out in Chapter 4. That is, the behaviour of the break fraction estimator when the model has no break in the SE is comparable to

its behaviour when there is a break in the JE alone.

Figure 5.35: Histogram of the estimated break fraction using $D_T(\lambda)$
Estimating 1 break; DGP has 1 break in JE only (T=140)

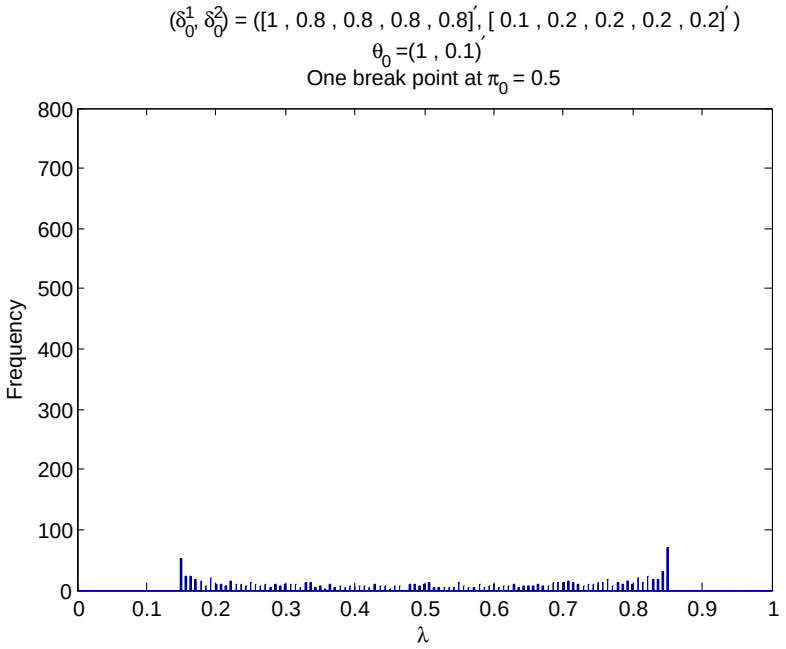


Figure 5.36: Histogram of the estimated break fraction using $D_T(\lambda)$
Estimating 1 break; DGP has 1 break in JE only (T=600)

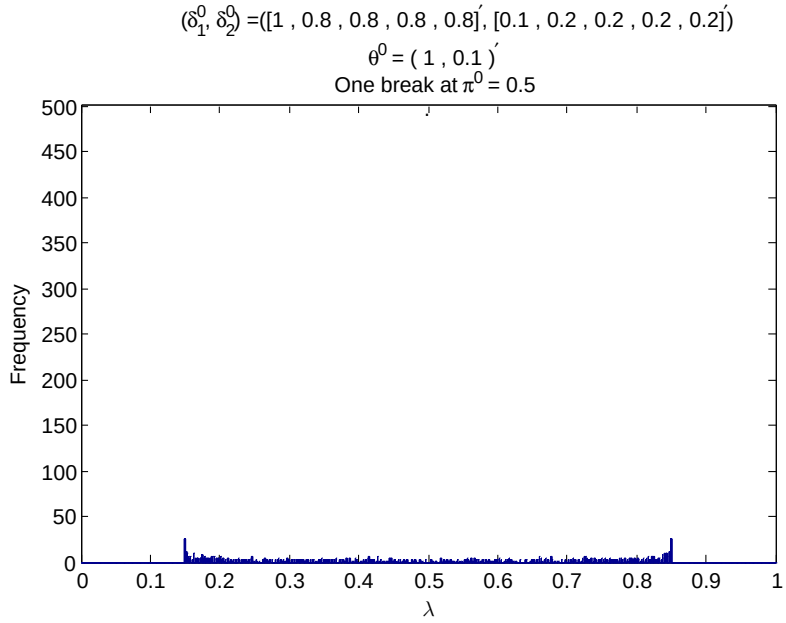


Figure 5.37: Plot of first and second estimated break fractions using $D_T(\lambda)$
Estimating 2 breaks; DGP has 1 break in JE only (T=140)

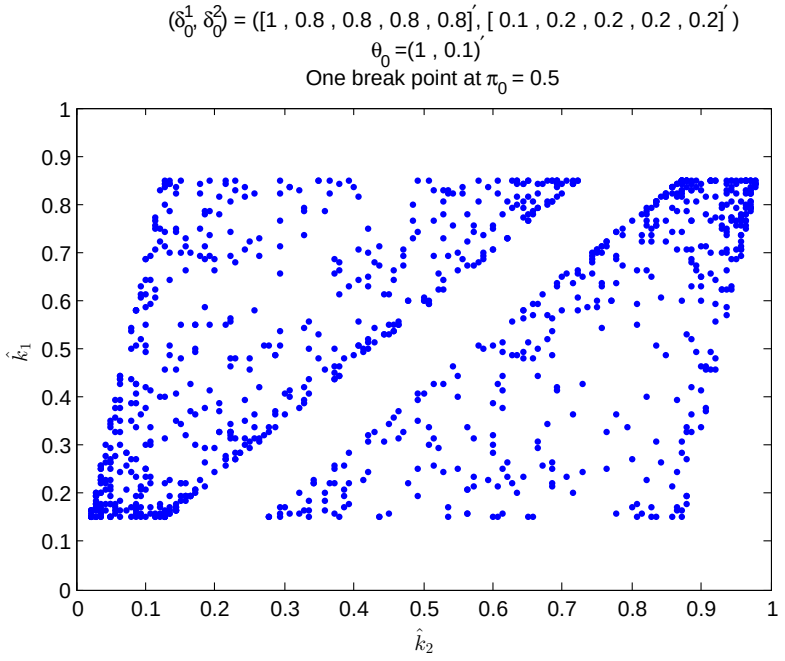
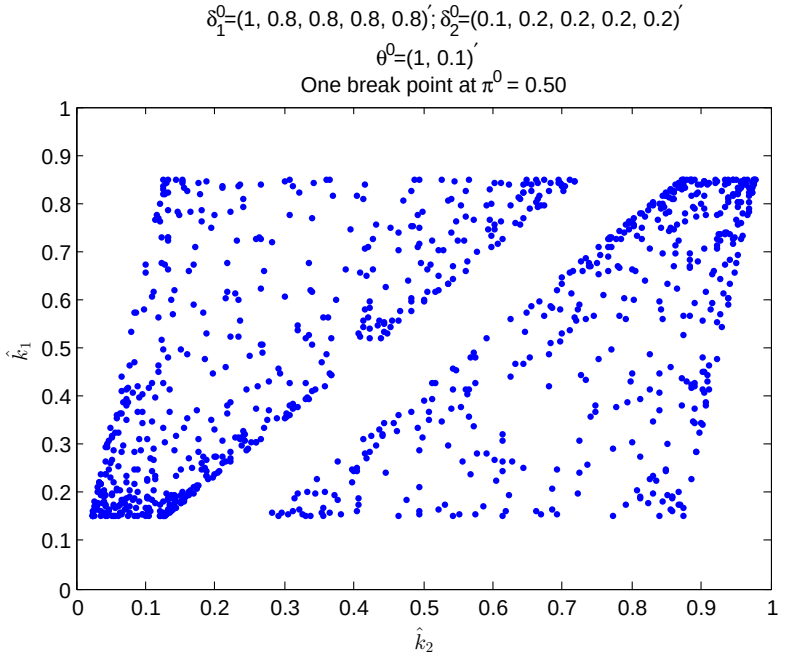


Figure 5.38: Plot of first and second estimated break fractions using $D_T(\lambda)$
Estimating 2 breaks; DGP has 1 break in JE only (T=600)



5.2.3 Simulation Results for a Model with a Coincidental Break

For the break in the SE, we use an identical model to that in (5.3) on page 135. That is,

$$y_t = [1, x_t]'\theta_0^{(i)} + u_t, \quad for \ \ t = 1, \ldots, T \ \ and \ \ i = 1, 2,$$

where $\theta_0^{(1)} = [1, 0.1]'$, $\theta_0^{(2)} = [-1, -0.1]'$ and the breaks, $\lambda_0 = \pi_0 = 0.5$.

The Simulation Results

Eight results are presented in three groups. The first group made up of Figures 5.39 to 5.42 show histograms of the first estimated break fractions obtained from models with sample sizes 140, 260, 300 and 600 respectively. In all samples, the break fraction estimator clearly converges to the true break, $\lambda_0 = 0.5$.

The second group comprising Figures 5.43 and 5.44 display histograms of the two estimated break fractions from models with sample sizes 300 and 420 respectively. While the first estimated break fraction distinctly converges to the break in the SE in both models, the second break fraction estimator on the other hand, is random and estimates any λ in the acceptable range.

Similar behaviour is seen in the last group consisting of Figures 5.45 and 5.46 where the first and second break fraction estimators are plotted against each other. It is again obvious that the first break fraction estimator clusters around the true break $\lambda_0 = 0.5$ while the second estimated break fraction is random.

These simulation results verify the diuscussions set out in Section 4.2.2 on page 116 and hence the break fraction estimator from a model with a coincidental break in the JE and SE will converge to the true break fraction in the SE.

Figure 5.39: Histogram of the first estimated break fraction using $D_T(\lambda)$
Estimating 1 break; DGP has Coincidental break in SE and JE (T=140)

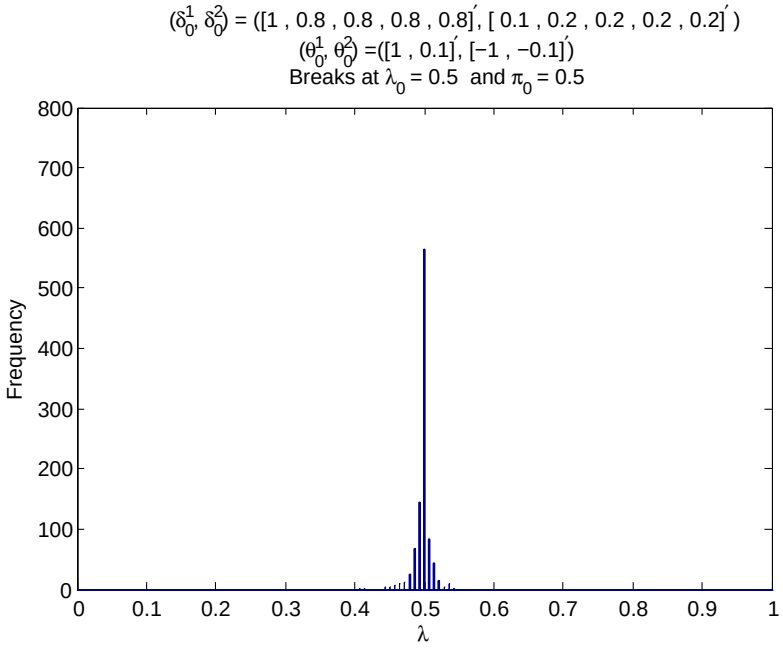


Figure 5.40: Histogram of the first estimated break fraction using $D_T(\lambda)$
Estimating 1 break; DGP has Coincidental break in SE and JE (T=260)

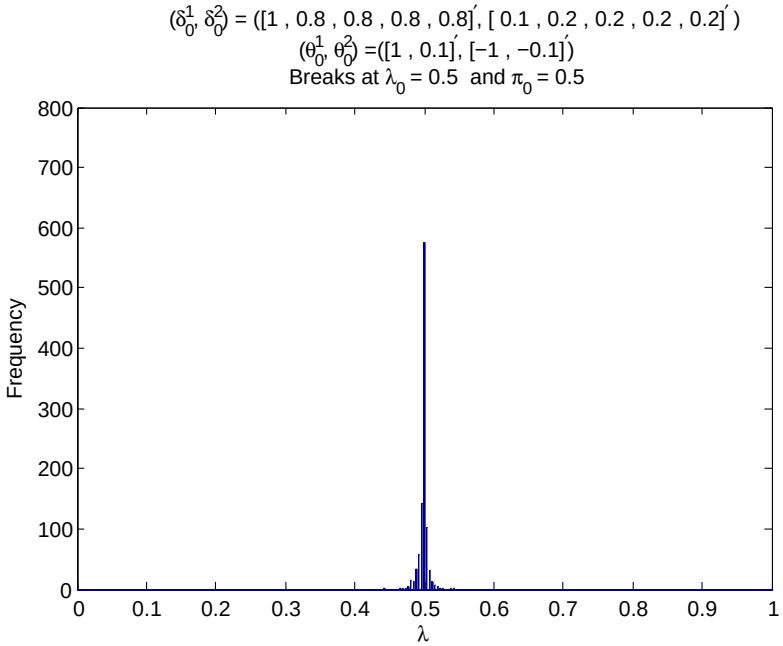


Figure 5.41: Histogram of the first estimated break fraction using $D_T(\lambda)$
Estimating 1 break; DGP has Coincidental break in SE and JE (T=300)

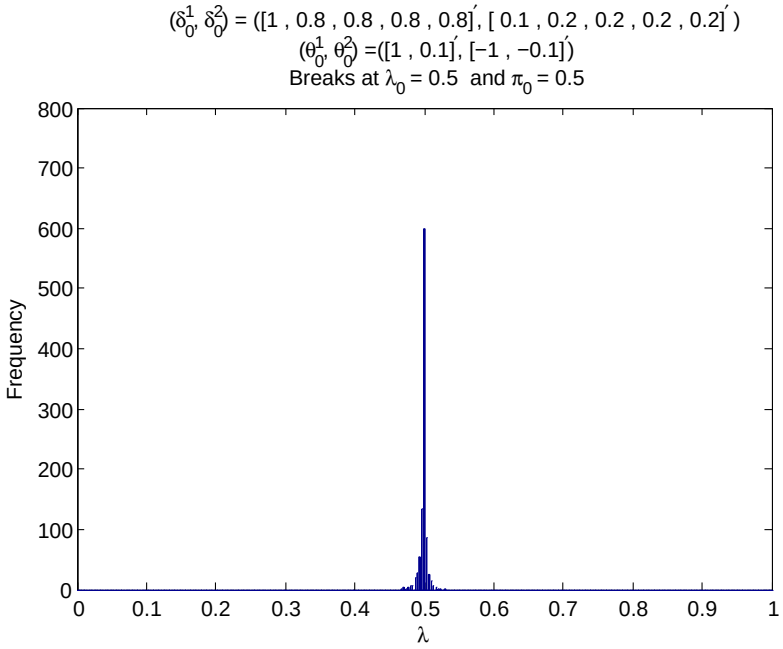


Figure 5.42: Histogram of the first estimated break fraction using $D_T(\lambda)$
Estimating 1 break; DGP has Coincidental break in SE and JE (T=600)

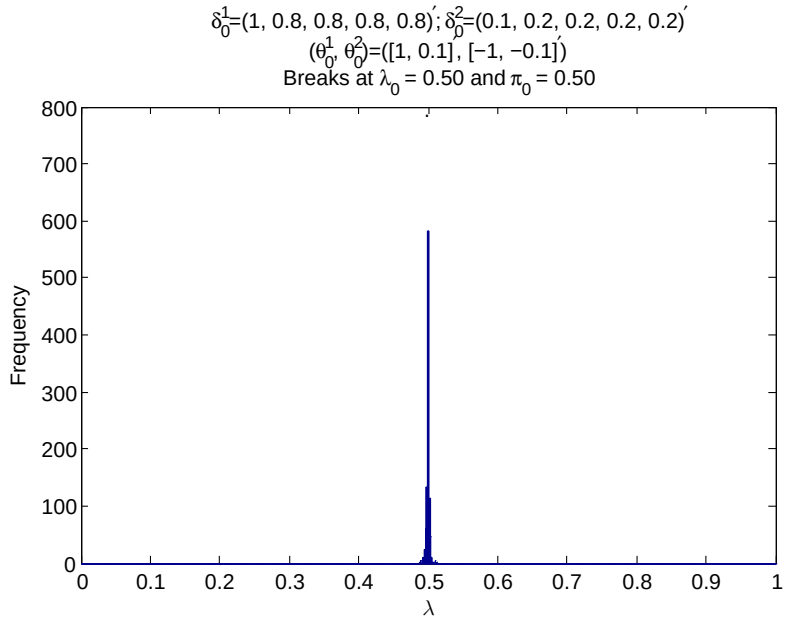


Figure 5.43: Histogram of the two estimated break fractions using $D_T(\lambda)$ Estimating 2 breaks; DGP has Coincidental break in SE and JE (T=300)

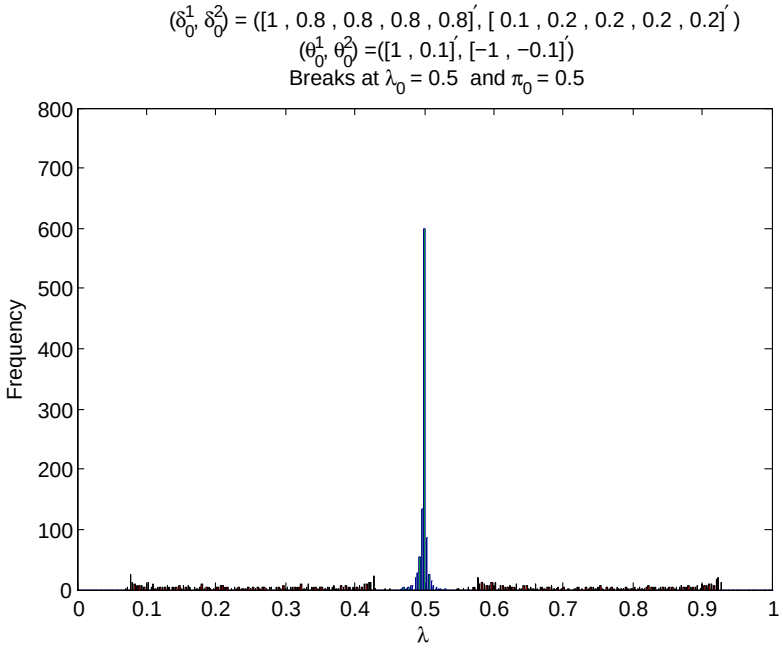


Figure 5.44: Histogram of the two estimated break fractions using $D_T(\lambda)$ Estimating 2 break; DGP has Coincidental break in SE and JE (T=420)

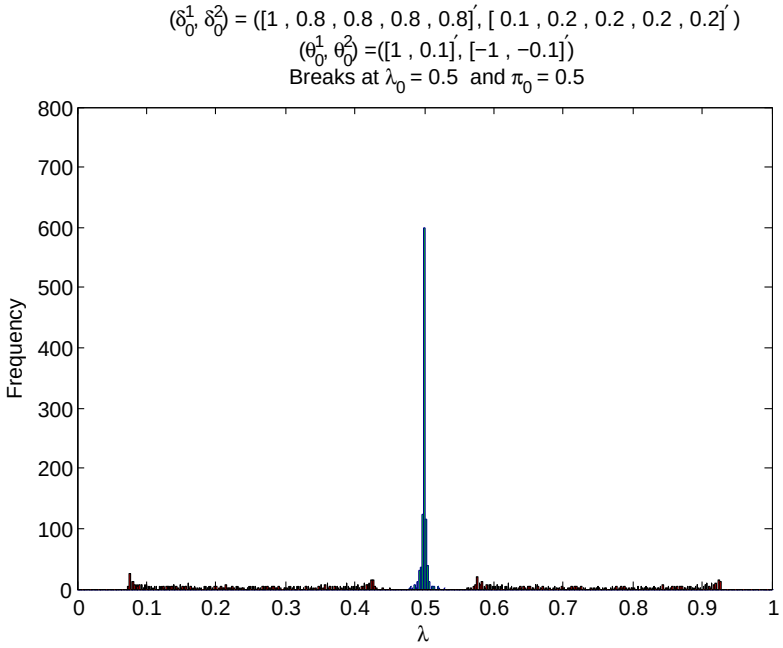


Figure 5.45: Plot of the two estimated break fractions using $D_T(\lambda)$
Estimating 2 breaks; DGP has Coincidental break in SE and JE (T=420)

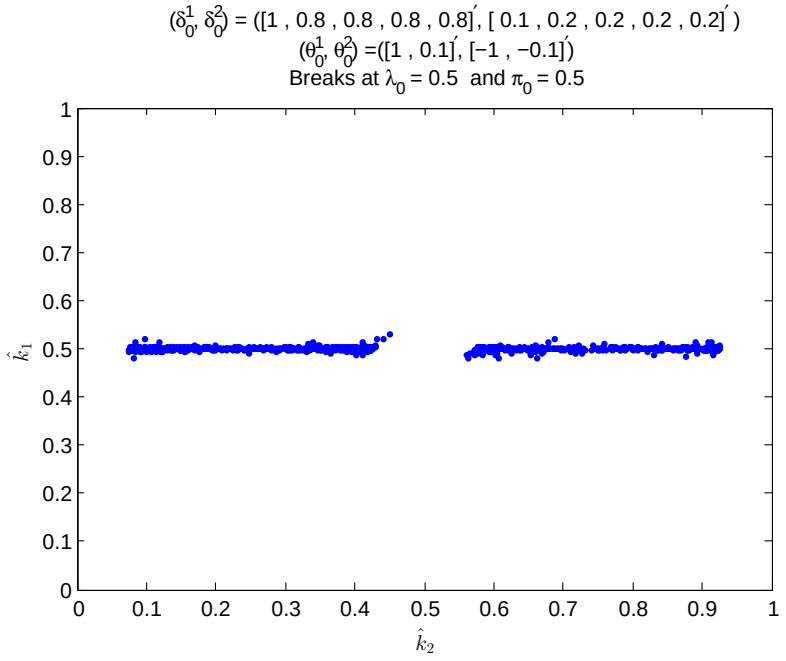
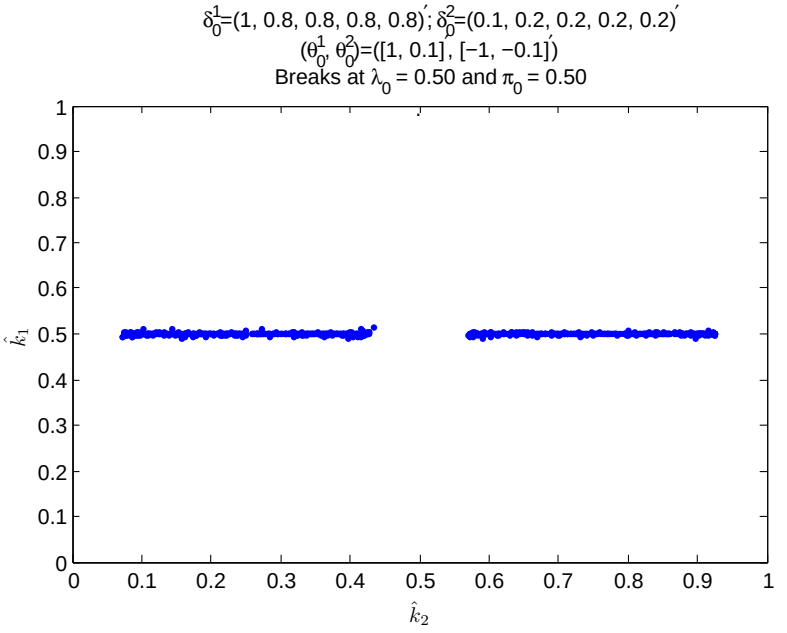


Figure 5.46: Plot of the two estimated break fractions using $D_T(\lambda)$
Estimating 2 break; DGP has Coincidental break in SE and JE (T=600)



5.2.4 Simulation Results for a Model with a Separate Break

We present the results in this section in two groups. The first displays results from estimating one break only, while the second covers results from estimating two breaks. Additionally, we consider two different break locations in the SE: when $\lambda_0 = 0.3$ and 0.7. All other parts of the DGP are left identical to the Coincidental break model discussed previously in Subsection 5.2.3, including the break in the JE at $\pi_0 = 0.5$.

The Simulation Results

We present four results in the first group and two in the second. In Figures 5.47 and 5.48, the histograms of the first estimated break fractions are displayed for models with sample sizes 260 and 420 respectively. These histograms converge to the true break fraction at $\lambda_0 = 0.3$. Similarly, when $\lambda_0 = 0.7$, Figures 5.49 and 5.50 show the break fraction estimators clearly gather again at this true break in the SE.

The last two figures show the simulation results when two break fractions are estimated, using a sample size of 600. In Figure 5.51 where both break fractions are plotted against each other, the first estimated break fraction noticeably converges to $\lambda_0 = 0.7$ for all one thousand repetitions. In Figure 5.52 which displays the histograms of the two break fraction estimators, the first estimated break fraction is distinct while the second estimated break fraction randomly estimates any candidate λ.

These simulation results presented for the Separate Breaks again supports the discussions in Section 4.2, that when there is a break in both the SE and the JE, the estimated break fraction would still be consistent for the true break in the SE. Therefore, estimating the location of a break point in a model is unaffected by an unstable relationship between the endogenous regressor and its instruments.

Figure 5.47: Histogram of the first estimated break fraction using $D_T(\lambda)$
Estimating 1 break; DGP has Separate breaks in SE and JE (T=260)

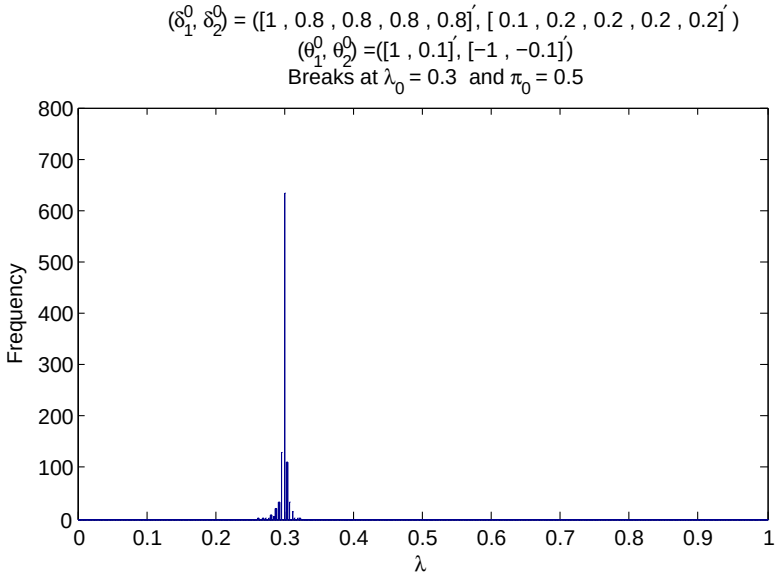


Figure 5.48: Histogram of the first estimated break fraction using $D_T(\lambda)$
Estimating 1 break; DGP has Separate breaks in SE and JE (T=420)

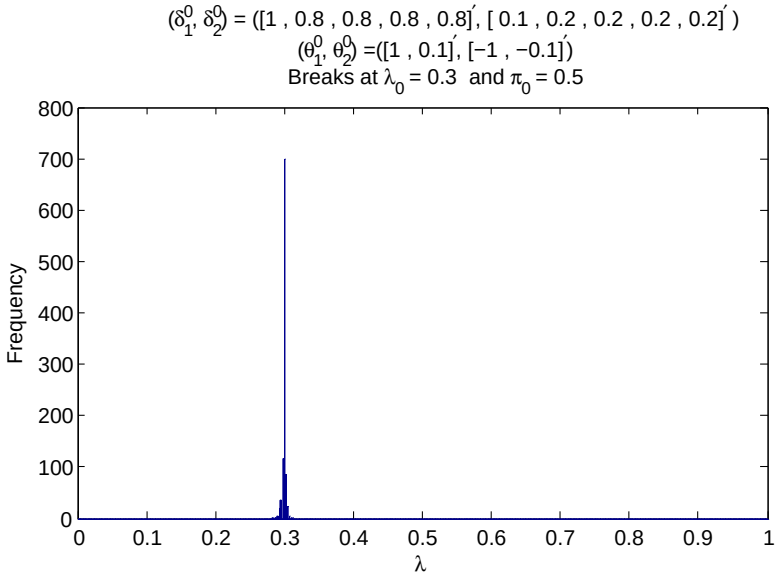


Figure 5.49: Histogram of the first estimated break fraction using $D_T(\lambda)$
Estimating 1 break; DGP has Separate breaks in SE and JE (T=260)

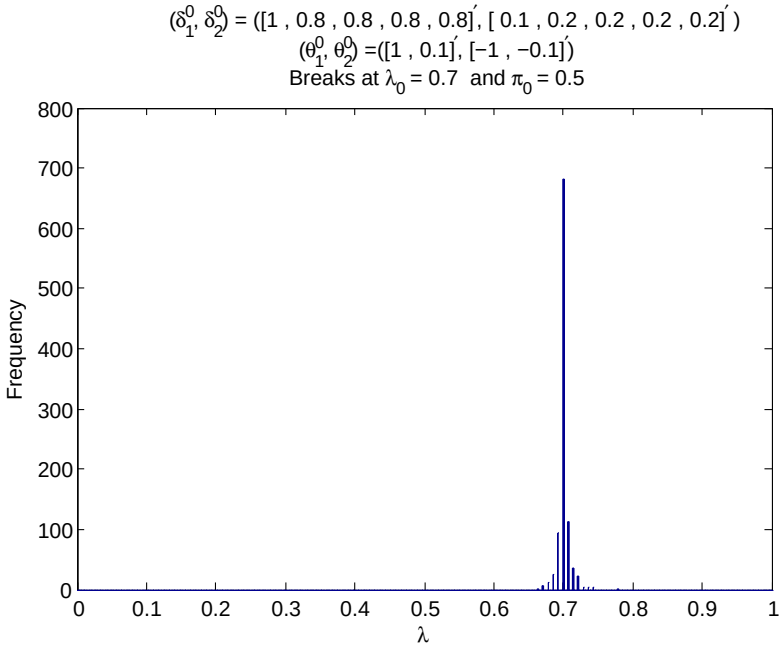


Figure 5.50: Histogram of the first estimated break fraction using $D_T(\lambda)$
Estimating 1 break; DGP has Separate breaks in SE and JE (T=420)

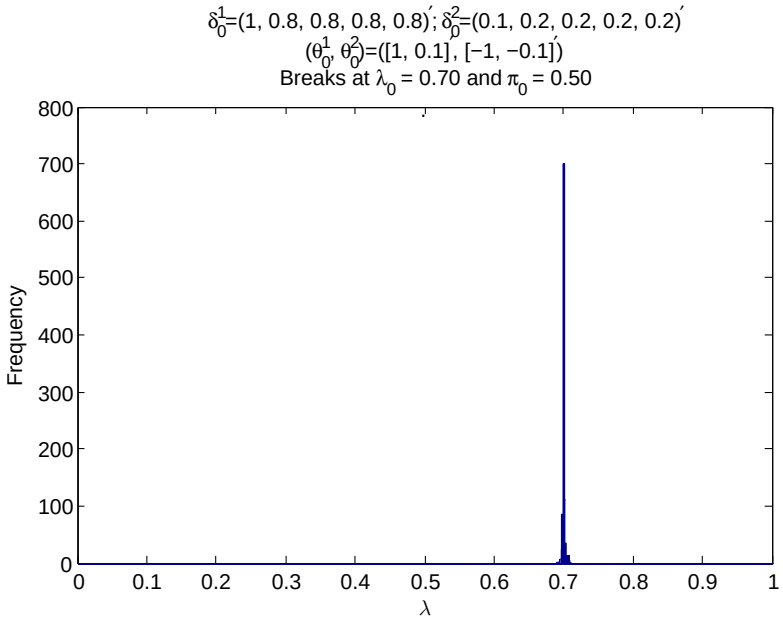


Figure 5.51: Plot of first against second estimated break fractions using $D_T(\lambda)$
Estimating 2 breaks; DGP has Separate breaks in SE and JE (T=600)

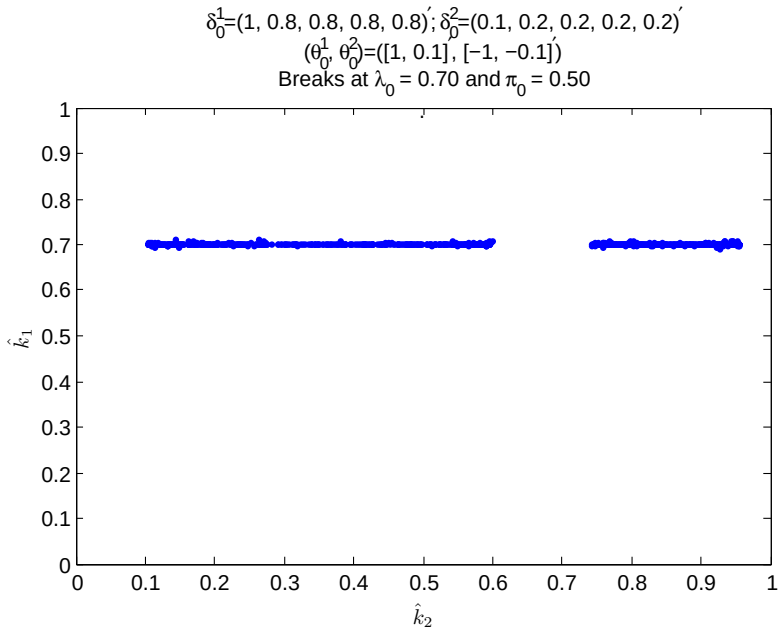
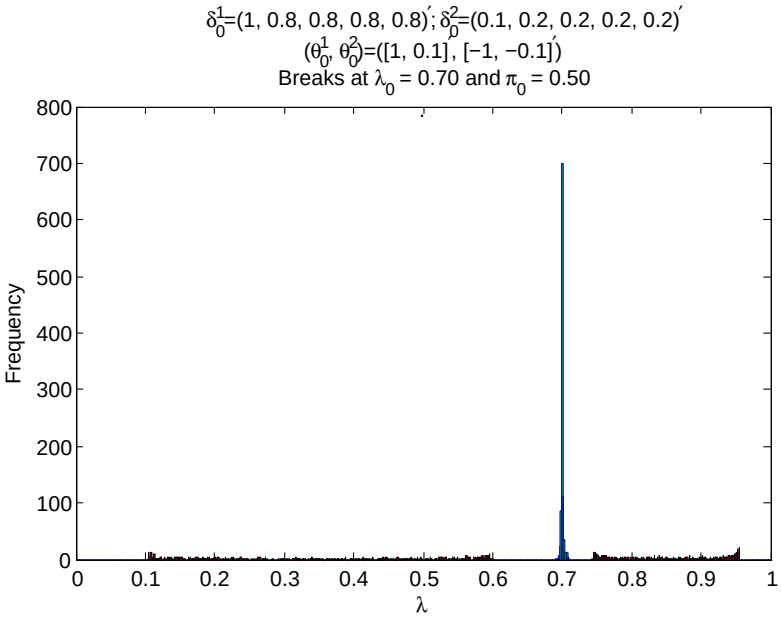


Figure 5.52: Histogram of the two estimated break fractions using $D_T(\lambda)$
Estimating 2 breaks; DGP has Separate breaks in SE and JE (T=600)



5.3 Conclusion

The finite sample properties of the break fraction estimators were examined in this chapter through the use of Monte Carlo simulations. The results were organised into two sections - Stable Jacobian and Unstable Jacobian. The results are based on 1000 repetitions, covering models with five different sample sizes.

Under the Stable Jacobian, models with zero, one, two and three break points in their SE were considered. In the models with no breaks, the break fraction estimator identified any candidate break fraction within the range and was thus randomly distributed across the whole range of Λ. Conversely, in the other three models, the break fraction estimator was consistent for the true breaks in the SE, as evidenced by the convergence to the true break locations as the sample size increased. This supports the consistency theory given in Propositions 1 and 3 on pages 33 and 66 for the break fraction estimators obtained from models with single and multiple break points respectively.

Furthermore, when different magnitudes of shifts were imposed in the multiple break models, the results show the first estimated break fraction is consistent for the dominant break in the model. Thus confirming Proposition 3 on page 66.

The simulation results from the Unstable Jacobian models likewise support the theoretical findings as set out in Chapter 4. Although there was a break in the JE, the break fraction estimator consistently identified the break points in the SE only. In this way, the estimation of a break point in the SE follows a similar process as the Stable Jacobian outlined in Chapters 2 and 3, irrespective of the stability of the JE.

This is particularly appealing as it indicates an Unstable Jacobian does not confound estimations of a true break in the SE. Thus, when a model has an endogenous regressor, the reduced form does not need to be estimated first to determine its stability, neither does the SE have to be partitioned into subsamples based on the stable reduced forms as proposed in Hall et al. (2012) using Two Stage Least Squares. This method using GMM thus permits the researcher to focus on and account for break points in the main SE of interest, rather than on the JE.

Chapter 6

Determining the True Number of Break Points

In the previous three chapters, the discussions and analyses were based on the premise that the true number of breaks, m_0, were known and only their locations needed to be estimated. These break points were estimated one after the other until they were all obtained. In reality however, knowledge of the true number of break points existing in a model may not be known beforehand. Furthermore, by the design of the Sequential Estimation procedure, there will always be a value of $supD_T(\lambda)$ estimated. This may not always indicate that it is a true break point. There needs to be a procedure to assess if the break at the *supremum* is big enough to indicate a true break occurs at that point. We highlight this using the case of a model with no break point in either the Structural Equation (SE) or the Jacobian Equation (JE).

The SE and JE of such a model would have the forms,

$$\begin{aligned} y_t &= [1, x_t]^{'}\theta_0 + u_t, \quad t = 1, 2, \ldots, T \quad and \\ x_t &= [1, z_t^{'}]\delta_0 + v_t, \quad t = 1, 2, \ldots, T \end{aligned} \tag{6.1}$$

respectively, where all the terms are as previously defined in (2.2) and (4.1). We examine the asymptotic behaviour of the test statistic in the usual way by decomposing it into its two main parts - the Parameter Difference and the Centre Matrix.

First, recall the test statistic as given in (2.5) on page 32 in the Stable Jacobian chapter,

$$D_T(\lambda) = T(\hat{\theta}_{1,T}(\lambda) - \hat{\theta}_{2,T}(\lambda))^{'} M_{*,T}(\lambda)^{-1} (\hat{\theta}_{1,T}(\lambda) - \hat{\theta}_{2,T}(\lambda)),$$

where $\hat{\theta}_{1,T}(\lambda) - \hat{\theta}_{2,T}(\lambda)$ is referred to as the Parameter Difference and $M_{*,T}(\lambda)^{-1}$ as the Centre Matrix. Starting with the Parameter Difference, notice that as there is no break in the SE, then $\theta_0^{(1)} = \theta_0^{(2)}$. Hence, similar to (D.1) on page 123, it must be that for all $\lambda \in \Lambda$,

$$\hat{\theta}_1(\lambda) - \hat{\theta}_2(\lambda) \xrightarrow{p} \theta_0^{(1)} - \theta_0^{(2)} = 0.$$

On the other hand, the limit of the Centre Matrix for all $\lambda \in \Lambda$ remains as $\lambda(1-\lambda)\mathcal{B}$, where $\mathcal{B} = Q_{xz}CQ_{zx}$, as shown in (B.12) in Appendix B.1. Therefore, we can conclude that when there is no break in the model, $D_T(\lambda)$ will exhibit a similar behaviour asymptotically for all $\lambda \in \Lambda$, that is,

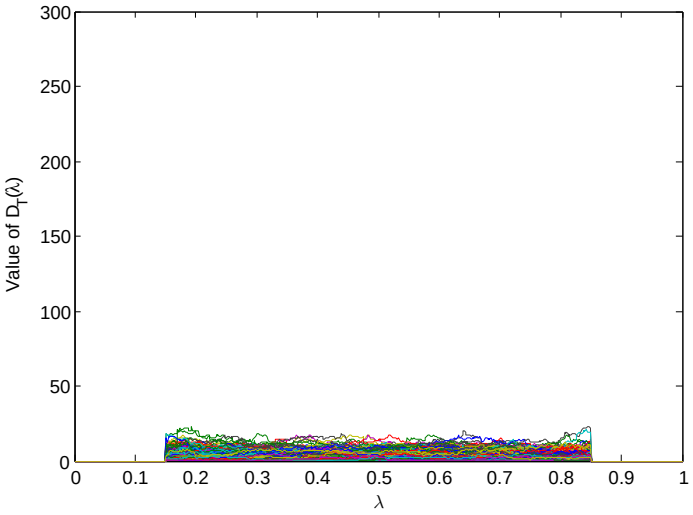
$$\begin{aligned} D_T^*(\lambda) &\xrightarrow{p} \lambda(1-\lambda)(\theta_0^{(1)} - \theta_0^{(2)})'\mathcal{B}(\theta_0^{(1)} - \theta_0^{(2)}) \\ &= 0. \end{aligned} \tag{6.2}$$

Figure 6.1 shows a thousand simulations of $D_T(\lambda)$ to illustrate its finite sample behaviour in these sort of models that have no true break point. We use a sample size of 600 with parameter values $\theta_0 = (1, 0.1)'$ and $\delta_0 = (1, 0.5, 0.5, 0.5, 0.5)'$. As in other simulations, the range of $\Lambda = [0.15, 0.85]$. The simulations clearly show there is no major difference in the values of $D_T(\lambda)$ across the whole range of Λ, as they are all approximately zero, consistent with the theory presented above. Nonetheless, the break fraction estimator, $\hat{\lambda}$, randomly selects any of these values as the location of the break point, as seen in the Monte Carlo simulation experiments reported in Chapter 5.

Thus, when the number of true break points is unknown, it becomes imperative to test for the significance of each of the estimated break points, as well as develop a rule to determine when to terminate the sequential estimations. This is useful as it ascertains whether the estimated break points are trivial and should be ignored or nontrivial and should be incorporated into the model.

The chapter is organised as follows. Section 6.1 presents a brief literature review of various conventional tests for structural and parameter stability. Section 6.2 details the method of testing for the true number of breaks adopted in this study. The finite sample performance of the test used is reported via a series of Monte Carlo simulations in Section 6.3. Conclusions are presented in Section 6.4 while relevant details of the Data Generating Process can be found in the appendix.

Figure 6.1: $D_T(\lambda)$ for all $\lambda \in \Lambda$.
DGP has no break in Structural or Jacobian Equations
$\theta_0 = [1 \ , \ 0.1]'$



6.1 Tests for Structural and Parameter Stability

This section reviews some frequently used tests for structural and parameter stability existing in the literature. It is divided into four main parts. The first and second parts review tests for single and multiple break points, respectively, while the third looks at tests for continuously varying parameters. Finally, the last part considers the Information Criteria approach to testing.

6.1.1 Tests for a Single Break Point

A commonly used test for a single break point is the sample-split test proposed by Chow (1960) where the sample is divided into two at a suspected break point determined by the researcher. The timing of this break point may be selected based on some events known beforehand such as the oil shock of 1973 or indeed any other exogenous event of interest capable of causing a structural change. The test is essentially an F test which compares the sum of squared residuals (SSR) of the whole sample with the combined individual SSR's obtained from the two subsamples. Though this test uses Ordinary Least Squares (OLS) estimation, Andrews and Fair (1988) extend it to other estimation methods using the Wald, Lagrange Multiplier (LM) and Likelihood Ratio (LR)-type tests. Due to the arbitrary break point, it is possible that different researchers end up with different conclusions from the same sample period.

In reality, however, the location of the break point may not be known beforehand and hence a test statistic which does not rely on a predetermined break point may be preferred. For example, Quandt (1960) LR test statistic which takes the maximal Chow (1960) statistic calculated over each candidate break point or the optimal test for structural change proposed by Andrews and Ploberger (1994) which uses sequences of the exponential averages of the Wald, LM and LR-like tests. Andrews (1993) focuses on instability in the parameters using the maximum of the Wald, LM and LR-like test. We discuss this test in detail in the next section. Comparing analytic and bootstrap procedures, Diebold and Chen (1996) extend Andrews (1993) test to a dynamic model, providing finite sample evaluation of the size of the test. More recently, Elliot and Müller (2014) suggest an alternative test for breaks in the means, variances and coefficients in linear and nonlinear parametric models. These tests developed for a break at an unknown location in the sample are more complicated and have a non-standard limit distribution because the break date is present only under the alternative hypothesis.

6.1.2 Tests for Multiple Break Points

Various methods have been proposed in the literature for testing the presence of multiple breaks, for example, see the LR tests proposed in Bai (1999) and the tests for a shift in mean proposed in Altissimo and Corradi (2003). Jouini and Boutahar (2005) and Antoshin et al. (2008) provide good reviews of various hypothesis tests. We discuss the three F-type tests proposed by Bai and Perron (1998) which are frequently used in practice. These tests do not require the location of the break points in the sample to be known.

The first test is the sup F-type test which is designed to test the null of parameter constancy against an alternative of a fixed number of break points, say k. This is suitable if a particular choice of k break points is of interest to the researcher. However, if k is unknown or unspecified beforehand, the second test known as the Double Maximum F-test, $DmaxF_T$, may be appropriate as it tests for different values of k, up to some maximum. Bai and Perron (1998) suggests an upper limit of 5 breaks is sufficient. The $DmaxF_T$ has some fixed weights attached to the breaks which may be a function of the asymptotic critical values or a reflection of prior knowledge of the number of breaks. If these weights are all set to one, then the test is known as the Unweighted $DmaxF_T$. Otherwise, it is referred to as the Weighted $DmaxF_T$.

The last test, denoted as $F_T(l+1|l)$, offers a sequential method of testing the stability of the model by comparing the SSR of the null of l breaks with the SSR of the alternative

hypothesis of $l+1$ breaks. If the SSR obtained from the $(l+1)$th break is significantly lower than that obtained from the SSR of the lth breaks, the null of stability is rejected and an additional break point is chosen in the subsample that produces the largest reduction in the SSR of the whole sample. The process terminates when the stability test fails to reject the null hypothesis. The critical values for these tests are given in Tables I and II in Bai and Perron (1998).

These three tests, though effective, are suitable only when using OLS, where the regressors are assumed to be uncorrelated with the errors. In situations where the regressors are endogenous, OLS yields inconsistent estimators. Hall et al. (2012) extend these three tests to models with endogenous regressors using Two Stage Least Squares (2SLS). Their method requires splitting the structural equation into $h+1$ subsamples based on the h break points in the reduced form; thus, each of the structural equation subsamples have stable reduced forms. The limiting properties of these tests are established and shown to be analogous to those in Bai and Perron (1998) and therefore the same critical values are used. Additionally, Hall et al. (2012) propose Wald-type test versions of the three F-tests which are more robust and allow for serial correlation in the errors.

6.1.3 Tests for Continuously Varying Parameters

An interesting approach to determining instability in models using graphs was introduced by Brown et al. (1975). The approach uses the cumulative sum (CUSUM) and CUSUM of squares (CUSUMSQ) and plots the recursive least squares estimates of the model's parameters which provides a means of identifying continuously changing parameters. The recursive estimation basically involves continuously estimating regressions with a subsample of the data, then sequentially adding one observation at a time and re-estimating the regression until the end of the sample is reached. Stability of the parameters is assumed if the plot shows they converge towards a common value. Nyblom (1989) extends the CUSUM to a Lagrange Multiplier (LM) test for joint stability of time-varying parameters while Hansen (1992) proposes a transformed test that examines both joint and individual stability of the parameters using an average of the squared CUSUM.

6.1.4 Information Criteria approach

More recently, some researchers have adapted the Information Criteria (IC), used in econometrics for model selection, to test for model stability (see Yao (1988), Jouini and

Boutahar (2005), Hall et al. (2013) and Hall et al. (2015)). Commonly used criteria are Akaike (AIC), Schwarz Bayesian (BIC) and Hannan-Quinn (HQIC) IC. These IC are made up of two terms - the model's log likelihood[1] which provides a natural assessment of the quality of the fit of the model and a penalty term which is a function of the number of parameters in the model. The main function of the penalty term is to avoid under and over-fitting since the RSS must decrease with an additional break point. This makes the IC approach a viable method to estimating the significant number of break points in a model.

Yao (1988) considers a version of the BIC which he establishes is consistent for the estimation of the number of break points in the mean of an independent normal sequence. However, Bai and Perron (2006) note Yao (1988) test performs badly under some specifications, for example, when serial correlation is present in the errors. Furthermore, Yao (1988) uses a penalty term which takes each break date estimation as an estimation of a coefficient of the model but Hall et al. (2013) observe this penalty term may not be sufficient to capture the impact of the break. Hence, they impose a more severe penalty function equivalent to the estimation of $3m$ coefficients rather than m employed by Yao (1988). They estimate the models using OLS and the results of their Monte Carlo simulations show their modified BIC performs well, irrespective of serial correlation in the errors. More recently, Hall et al. (2015) note that a similar improved performance of the modified BIC is maintained when it is extended to models with endogenous regressors estimated using 2SLS.

The multiple break tests discussed above require the break points to be simultaneously estimated. Thus, they are not appropriate for our sequential estimation procedure. In estimating break points sequentially, Bai (1997a) suggests estimation should be based on a dual process of hypothesis testing of the constancy of the model's parameters in each regime, alongside the sequential estimation of the break point. We adopt the combined technique of Bai (1997a) in this research and summarise the process below.

6.2 Combined Procedure of Hypothesis Testing and Break Point Estimation

The hypothesis test used in Bai (1997a) is the sup F-test which is based on the difference in the restricted and unrestricted SSR. The combined procedure is outlined in the

[1] Under normality, the log likelihood function reduces to the ln[RSS].

following five steps below:

Step 1: Test the entire sample $[1, T]$ for parameter constancy. If the test statistic is greater than the critical value[2], then the null hypothesis of parameter constancy is rejected. This implies there is at least one break point in the model which has to be estimated. Otherwise, the model is stable and has no break points.

Step 2: Estimate one break point using all observations $[1, T]$ in the sample. Since this is an estimation for a shift in the mean of a linear process, the break point estimator, $\hat{k}_1$, is obtained by minimising the RSS among all candidate break points across the sample.

Step 3: Split the sample into two subsamples, $[1, \hat{k}_1]$ and $[\hat{k}_1 + 1, T]$ and test both subsamples for constancy of their parameters. If the sup F-test is not rejected in either subsample, then the model has only one break (the break estimated in Step 2) and the testing procedure ends. However, if the sup F-test is rejected in only one subsample, then there are at most two break points in the model; go on to Step 4. Likewise, if the sup F-test is rejected in both subsamples, then there are at least three break points in the model and the process goes on to Step 4.

Step 4: Estimate a break point (as done in Step 2) in any of the subsamples for which the sup F-test is rejected. This break point will again be the location at which $S_T(k)$ is minimised within the respective subsamples, $[1, \hat{k}_1]$ or/and $[\hat{k}_1 + 1, T]$. Note that if any break fraction is estimated in either of the subsamples, it will have the same asymptotic properties as if $[1, k_0^{(1)}]$ and $[k_0^{(1)} + 1, T]$ were used since the first break fraction estimator $\hat{\lambda}_1$ is consistent for the true break fraction, $\lambda_0^{(1)}$, as shown in Proposition 3 on page 66.

Step 5: Split the subsamples again based on the location of any identified break point and continue the testing and estimation cycle until the sup F-test fails to reject stability in any of the $m + 1$ subsamples. A non rejection of the test implies the subsamples are stable and hence the estimated number of significant break points in the model is one less than the number of stable subsamples.

Bai (1997a) showed this combined procedure yields consistent estimates of the true number of break points in a model. Note, however, that Bai (1997a) assumes there is at least one break in the model and so omits the first step above. Nonetheless, we include Step 1 in our estimation procedure because the estimation of the first break point is fundamental to all further sequentially estimated break points in the model. Also, the sup F-test is designed for OLS estimations and hence would not be suitable within our GMM framework.

[2]They assume the critical value and size of the sup F-test are based on the asymptotic distribution as given in Table I in Bai and Perron (1998).

6.2.1 GMM Hypothesis Tests for Stability

The various GMM-based test statistics existing in the literature are usually in the form of a *supremum*, an average or an exponential of a sequence of some test statistic. For example, Andrews and Fair (1988), Andrews (1993), Andrews and Ploberger (1994) and Gagliardini et al. (2005). Others include Ghysels and Hall (1990), Sowell (1996) and more recently, Elliot and Müller (2014). We discuss two tests specifically developed for models estimated using GMM.

In the approach taken by Hall and Sen (1999), the GMM population moment condition is decomposed into two orthogonal parts. The first, known as the identifying restrictions, are imposed during estimation, hence they are directly linked to the constancy of the parameters. The other, known as the overidentifying restrictions, are associated with the structural stability of the model. They propose hypothesis tests for both components separately which is useful in determining the source of the instability since a violation of a specific null would indicate instability in either the parameters or in the structure of the econometric model itself.

Still within the GMM framework, Ghysels et al. (1997) propose a Predictive test for structural change with a single unknown break point. The process involves evaluating the moment conditions of $T-k$ subset of observations at the parameter estimates of the observations in k. If there is no structural change, then the moment conditions predicted for the $T-k$ observations should be statistically insignificant and converge in probability to zero.

All these tests are valid in their own different terms, however, the interest of this research is on parameter variation, hence we employ the hypothesis test for parameter instability at an unknown location in the sample as proposed by Andrews (1993). Though the test is developed for nonlinear models, Augustine-Ohwo (2012) adapts it to the following simple linear model,

$$y_t = x_t^{'}\theta_t + u_t,$$

where y_t is the dependent variable, x_t is the endogenous regressor, θ_t is the parameter vector and u_t is the error term. Define z_t to be the instruments satisfying $E[z_t u_t(\theta_0)] = 0$, where θ_0 is the true parameter vector. We adopt this model and provide an overview of the hypothesis test below.

6.2.2 Andrews (1993) Tests for Parameter Instability

We start with the hypotheses of interest. The null is given as,

$$H_0 : \theta_t = \theta_0 \quad for \;\; t = 1, \ldots, T \;\; and \;\; \theta_0 = (\theta_0^{(1)'}, \theta_0^{(2)'})' \tag{6.3}$$

while the alternative hypothesis for an unknown break point is presented as,

$$H_1(\Lambda) = \bigcup_{\lambda \in \Lambda} H_{1,T}(\lambda), \tag{6.4}$$

where

$$H_{1,T}(\lambda) : \theta_t = \begin{cases} \theta_1(\lambda) & for \;\; t = 1, \ldots, [\lambda T] \\ \theta_2(\lambda) & for \;\; t = [\lambda T] + 1, \ldots, T \quad and \; \theta_1(\lambda) \neq \theta_2(\lambda). \end{cases} \tag{6.5}$$

Note that under the null hypothesis of parameter constancy, we have that the parameters, $\theta_0^{(1)} = \theta_0^{(2)}$ and the break point λ, is present only under the alternative hypothesis. The GMM parameters are obtained in the usual way by minimising the GMM criterion, $Q_T(\theta(\lambda); \lambda)$, as given in Appendix A.1 on page 37, that is,

$$\hat{\theta}_T(\lambda) = argmin_{\Theta(\lambda)} Q_T(\theta(\lambda); \lambda).$$

Although the parameters used in our estimations in previous chapters were based on the first step GMM estimators, in these hypothesis tests however, the second step GMM parameters are used to construct the test statistics. The main difference between the first and second step GMM estimators lies in the weighting matrix used in their construction. Recall the first step GMM estimator[3] is obtained by minimising the GMM criterion using a preliminary weighting matrix - usually the identity matrix or the instruments cross product matrix. On the other hand, to obtain the second step GMM estimator, an optimal weighting matrix is required. Hansen (1982) provides the choice of this optimal weighting matrix as the inverse of the variance obtained from the first step GMM estimators, denoted as Ω^{-1} where $\Omega = lim_{T \to \infty} Var[T^{-1/2} \sum_1^T z_t u_t]$, a finite positive definite matrix. The uniform convergence useful in obtaining the limiting properties of the estimator as stated in Assumption 4 on page 27 is given as $sup_{\lambda \in \Lambda} \| Var[T^{-1/2} \sum_1^{[\lambda T]} z_t u_t(\theta_0)] - \lambda \Omega \| \xrightarrow{p} 0$. These second step GMM estimators are consistent and efficient (Hansen, 1982) and Augustine-Ohwo (2012) establish these second step GMM parameters obtained from the

[3] Also see page 29.

linear model adopted in this study have the same asymptotic properties as those given in Andrews (1993).

The Test Statistics

Andrews (1993) considers three types of test statistics based on the Wald, the Lagrange Multiplier and the Difference-type test statistics, which we denote as $Wald_T(\lambda)$, $LM_T(\lambda)$ and $\ddot{D}_T(\lambda)$ respectively. The $Wald_T(\lambda)$ and $LM_T(\lambda)$ are both constructed as quadratic forms. $Wald_T(\lambda)$ is based on the difference between the unrestricted estimated value of the restrictions and their value under the null, that is,

$$Wald_T(\lambda) = T\big(\hat{\theta}_1(\lambda) - \hat{\theta}_2(\lambda)\big)' V_W(\lambda)^{-1}\big(\hat{\theta}_1(\lambda) - \hat{\theta}_2(\lambda)\big), \tag{6.6}$$

where $\hat{\theta}_1(\lambda)$ and $\hat{\theta}_2(\lambda)$ are the unrestricted estimators and $V_W(\lambda)$ is the variance-covariance matrix of these unrestricted estimators. $LM_T(\lambda)$ is based on the vector of first order conditions from the minimisation of the unrestricted estimator evaluated at the restricted estimator, that is,

$$LM_T(\lambda) = T\bar{\kappa}' V_{LM}(\lambda)^{-1}\bar{\kappa}, \tag{6.7}$$

where $\bar{\kappa}$ is the Lagrange Multiplier given in Appendix A.2 on page 38 and $V_{LM}(\lambda)$ is the variance-covariance matrix of $\hat{\kappa}$. $\ddot{D}_T(\lambda)$, on the other hand, is given as the sample size T multiplied by the difference between the GMM minimands evaluated at the restricted and unrestricted parameters, that is,

$$\ddot{D}_T(\lambda) = T\big(Q_T(\bar{\theta}_1, \bar{\theta}_1, \lambda) - Q_T(\hat{\theta}_1, \hat{\theta}_1, \lambda)\big), \tag{6.8}$$

where $Q_T(\cdot)$ are the GMM minimands given in (2.3) on page 28. We use $\ddot{D}_T(\lambda)$ here to differentiate the Difference test statistic from $D_T(\lambda)$ used in previous chapters to denote the test statistic when using the first step GMM estimators.

As stated in Andrews (1993), these test statistics are also applicable to statistics of the form $g(\{Wald_T(\lambda); \lambda \in \Lambda\})$, where g is an arbitrary continuous function; popular choices of g in the literature are the *supremum*, average and exponential functions. Andrews (1993) requires continuity of the criterion which is satisfied by the statistics given in (6.6) to (6.8). Also, compactness of the parameter space is necessary as earlier fulfilled on page 29. Other conditions given in Assumption 1(a) to 1(h) in Andrews (1993) are analogous to those used in this linear model as laid out in Assumptions 1 to 5 on page 27 and

Assumptions 6 to 8 on page 29. Additionally, the choice of the asymptotically optimal weight matrix, Ω^{-1} is similar.

To obtain the test statistic, we follow a similar process to the break point estimation procedure outlined on page 30 where we use the *supremum* as the choice of our continuous function. Thus, a sequence of $LM_T(\lambda)$, $Wald_T(\lambda)$ and $\ddot{D}_T(\lambda)$ as given in (6.6) to (6.8) are obtained for all $\lambda \in \Lambda$ and the maximum of these sequences obtained. These maxima, as given below, are the statistics used to test the null in (6.3) against the alternative of an unknown break point given in (6.4). They are presented as,

$$sup_{\lambda\in\Lambda} Wald_T(\lambda), \tag{6.9}$$

$$sup_{\lambda\in\Lambda} LM_T(\lambda), \tag{6.10}$$

$$sup_{\lambda\in\Lambda} \ddot{D}_T(\lambda). \tag{6.11}$$

These three tests are asymptotically equivalent as established in Andrews (1993) and Augustine-Ohwo (2012) for the linear case. They also have identical limit distributions which are functions of independent Brownian motions. As established in Theorem 3 in Andrews (1993),

$$sup_{\lambda\in\Lambda} Wald_T(\lambda) \xrightarrow{d} sup_{\lambda\in\Lambda}\{\lambda(1-\lambda)\}^{-1}\{(B_q(\lambda) - \lambda B_q(1))'(B_q(\lambda) - \lambda B_q(1))\},$$

where $B_q(\lambda)$ is a q-dimensional Brownian motion. Hence, the critical values[4] provided in Table 1 in Andrews (1993) are also suitable for our purpose. Furthermore, since $\hat{\lambda}_1$ is consistent for $\lambda_0^{(1)}$, then the limit distribution of the test statistic in the subsamples, say $[1, k_0^{(1)}]$ would be similar to that of the whole sample $[1, T]$ and so the critical values can also be used.

Parameter stability is rejected for high values of $sup_{\lambda\in\Lambda} Wald_T(\lambda)$, $sup_{\lambda\in\Lambda} LM_T(\lambda)$ and $sup_{\lambda\in\Lambda} \ddot{D}_T(\lambda)$; or more specifically, if the values are greater than the relevant critical value given in Table 1 in Andrews (1993).

6.3 Finite Sample Performance

This section presents results from Monte Carlo simulations conducted to assess the performance of the test outlined in the previous section. In this simulation exercise, our focus

[4]The critical values in Andrews (1993) cover the standard significant levels (1%, 5% and 10%) and the whole range of $\Lambda = [\lambda_0,\ 1-\lambda_0]$ for $\lambda_0 = 0.05, \ldots, 0.50$.

is on the ability of the test statistics to correctly detect a true break point in the model, hence, no results of the estimated break point is presented rather, the number of rejections of the test statistic is analysed. We search for a maximum of three breaks as outlined in the procedure below.

6.3.1 Monte Carlo Simulation Process

The estimation procedure follows these five steps:

Step 1: Test for parameter instability using the whole sample. We denote the whole sample used in this first step as S_0.

Step 2: If the test statistic is greater than the critical value, reject parameter stability and go to Step 3. The number of rejections at this stage are noted against S_0 in the results presented. Otherwise, if stability is not rejected, then there is no break point in the model and a zero rejection is recorded against S_0.

Step 3: Estimate a break point as the location of the *supremum* of $D_T(\lambda)$ using the whole sample $[1, T]$. As highlighted above, this break point is not presented, but this step is required in order to continue the estimation process.

Step 4: Based on the location of the estimated break point obtained in Step 3, split the sample into two subsamples. Denote the subsample before and after this break point as S_1 and S_2 respectively.

Step 5: Test for parameter stability in both subsamples and again reject stability in any subsample where the test statistic is greater than the critical value. The number of rejections (or zero rejections) are noted against the respective subsamples.

This process is repeated a thousand times and the number of rejections obtained for S_0, S_1 and S_2 are reported. These rejections determine the number of break points estimated in each model. Note that if the process terminates in Step 2 then a zero rejection is recorded for S_1 and S_2 as well.

6.3.2 The Monte Carlo Simulation Results

The results of the simulation for all three test statistics, $sup_{\lambda\in\Lambda}\ddot{D}_T(\lambda)$, $sup_{\lambda\in\Lambda}Wald_T(\lambda)$ and $sup_{\lambda\in\Lambda}LM_T(\lambda)$ are presented in this section. The results are organised into two broad categories. The first category presents the rejection frequencies of the test statistics in each sample and subsample, while the second category presents the number of estimated

breaks selected by the three test statistics. The results cover all the six different models discussed in Chapter 5.

Since the second step GMM estimators are used in the test, there are two possible ways to estimate $\ddot{D}_T(\lambda)$ depending on which first step GMM estimator used. When the inverse of the variance of the unrestricted GMM estimator is used as the weighting matrix in the second step, we denote the test statistic as $sup_{\lambda\in\Lambda}\ddot{D}_T^U$.

Similarly, when the inverse of the variance of the restricted GMM estimators are used, we denote the test statistic as $sup_{\lambda\in\Lambda}\ddot{D}_T^R$.

The Data Generating Process (DGP)

The models examined here are constructed similar to those used in the Monte Carlo simulations presented in Chapter 5. The results cover three sample sizes of 140, 300 and 600 at the three conventional statistical levels of significance 1%, 5% and 10%. A total of 1000 replications were performed in all cases. We present results using a trimming of $\epsilon = 0.15$, that is, $\Lambda = [0.15, 0.85]$ as well as when $\epsilon = 0.20$, that is, $\Lambda = [0.20, 0.80]$. Other general details about the DGP are given on page 129.

Category 1: Rejection Frequencies of the Test Statistics

The critical values used here are as given in Table 1 in Andrews (1993). When $\epsilon = 0.15$, the values are 15.56, 11.72 and 10.00 at 1%, 5% and 10% level of significance, respectively. The counterpart when $\epsilon = 0.20$ are 15.09, 11.26 and 9.54 respectively. The number of times the test statistics are greater than these critical values for the one thousand simulations are reported in the tables.

Results for Models with a Stable Jacobian Equation

For all models in this section, the JE is constructed identical to that in (5.1) on page 129. That is,

$$x_t = [1, z_t^{'}]\delta_0 + v_t, \quad t = 1, 2, \ldots, T,$$

with specific values of the JE parameters, $\delta_0 = (1, 0.5, 0.5, 0.5, 0.5)^{'}$.

For the SE on the other hand, three types of models are considered: models with no break, one break and two breaks in their SE. We present the DGP and the results from these three

types of models below.

I. Model with No Break in the SE or JE

The SE which is identical to (6.1) on page 167 is given as,

$$y_t = [1, x_t]'\theta_0 + u_t, \quad t = 1, 2, \ldots, T, \quad where \;\; \theta_0 = [1, 0.1]'.$$

Since there is no break in the model, we would expect the null hypothesis of a break in the model to be rejected at about 1%, 5% and 10% of the time at the 1%, 5% and 10% significance levels respectively. This is equivalent to the size of the test. In this model, the focus should be on the results obtained when the whole sample is used, that is S_0. As seen in the results displayed for $\epsilon = 0.15$ across all the test statistics in Table 6.1, the rejection frequencies converge towards the size of the test as the sample size gets larger. A similar trend is observed when $\epsilon = 0.20$ as seen in Table 6.2. Notice a reduction in the rejection frequencies in Table 6.2 as against those displayed in Table 6.1. This indicates the results are improved across all test statistics since there are potentially more observations available for estimation in each partition when $\epsilon = 0.20$ than when $\epsilon = 0.15$.

II. Model with One Break in the SE only

The SE used here is similar to that in (5.3) on page 135. It is given as,

$$y_t = [1, x_t]'\theta_0^{(i)} + u_t, \quad for \;\; t = 1, 2, \ldots, T \; and \; i = 1, 2,$$

where $\theta_0^{(1)} = (1, 0.1)'$, $\theta_0^{(2)} = (-1, -0.1)'$ and the true break fraction, $\lambda_0 = 0.5$.
As seen in the results of the rejection frequencies reported in Tables 6.3 and 6.4 for $\epsilon = 0.15$ and $\epsilon = 0.20$ respectively, all test statistics strongly reject stability in the whole model (S_0) with probability 1 for all sample sizes, signifying there is at least one break in the model which has to be estimated. After this break is estimated and the sample split as detailed in the procedure, the test is carried out on the two subsamples, S_1 and S_2. Based on the results of the estimated break point presented in the Monte Carlo simulations given in Chapter 5, S_1 and S_2 should have similar number of observations and as they both contain no breaks, we expect the rejection frequencies to be comparable which is evident in both tables.

III. Model with Two Breaks in the SE only

The DGP of the SE in a model with two breaks in its SE is given as

$$y_t = [1, x_t]'\theta_0^{(i)} + u_t, \quad for\ t = 1, 2, \ldots, T\ and\ i = 1, 2, 3,$$

with true parameter values, $(\theta_0^{(1)}, \theta_0^{(2)}, \theta_0^{(3)}) = \left([1, 0.1]', [-1; -5.1]', [1; -4.1]'\right)$ and true break fractions, $\lambda_0^{(1)} = 0.3$ and $\lambda_0^{(2)} = 0.70$. Notice a bigger magnitude of shift is imposed at the first break point so that $D_T(\lambda_0^{(1)}) > D_T(\lambda_0^{(2)})$. In this way, $\lambda_0^{(1)}$ is the dominant break fraction and will be identified first, consistent with the theory laid out in Chapter 3 as well as the results of the Monte Carlo simulations in Chapter 5. Thus, after estimating the first break point, we expect the test to reject stability in only S_2 because it is the subsample within which the second break point is located.

As displayed in Tables 6.5 and 6.6, for all three sample sizes, the test statistics exhibit strong power against the null of stability in the whole sample, S_0, as well as in S_2 where the second true break point lies. Although there is no break in S_1, the rejection frequencies are relatively high especially with the smallest sample size of 140. This is understandable since $\lambda_0^{(1)}$ is at 0.3, leaving very few observations available for estimations in that subsample. However, these rejection frequencies reduce drastically, across all test statistics, as the sample size increases.

Results for Models with an Unstable Jacobian Equation

For all models presented in this section, the JE is constructed similar to (5.7) on page 154. That is,

$$\begin{aligned} x_t &= [1, z_t]'\delta_0^{(1)} + v_t, \quad for\ \ t = 1, 2, \ldots, [T\pi_0] \\ &= [1, z_t]'\delta_0^{(2)} + v_t, \quad for\ \ t = [T\pi_0] + 1, \ldots, T, \end{aligned}$$

where $(\delta_0^{(1)}, \delta_0^{(2)}) = ([1, 0.8, 0.8, 0.8, 0.8]', [0.1, 0.2, 0.2, 0.2, 0.2]')$.

Three types of models are examined in this section. The first is one with a single break only in the JE while the SE remains stable. The second and third have a break in both the JE and SE. In the second, the breaks occur at the same location while they occur at different locations in the third. As in Chapter 4, the former is referred to as the Coincidental Break and the latter as the Separate Break.

I. Model with One Break in the JE only

In this model, the break in the JE is placed midway in the sample, that is, $\pi_0 = 0.5$. Conversely, as there is no break in the SE, we use an identical model as that used in the case of a model with no breaks given above, that is,

$$y_t = [1, x_t]'\theta_0 + u_t, \quad for \ \ t = 1, 2, \ldots, T, \quad where \ \ \theta_0 = [1, 0.1]'.$$

As established in the theory laid out in Section 4.1 as well as the results of the Monte Carlo simulations presented in Chapter 5, an unstable JE does not confound estimations of a true break in the SE, if it exists. Hence, we expect the results from this model with a break only in the JE to be similar to those obtained above for a model with no break in the SE. Focusing again on S_0, the results shown in Tables 6.7 and 6.8 are comparable to those seen earlier in Tables 6.1 and 6.2, albeit with slightly higher rejection values. A similar pattern is noticed across all test statistics although $sup_{\lambda\in\Lambda}\ddot{D}_T^U$ and $sup_{\lambda\in\Lambda}\ddot{D}_T^R$ recorded higher rejection values than $sup_{\lambda\in\Lambda}Wald_T$ and $sup_{\lambda\in\Lambda}LM_T$ for all sample sizes.

II. Model with One Coincidental Break in the SE and JE

To construct this model, the JE given above is combined with the SE from the model with one break discussed earlier. The break points are set at $\lambda_0 = \pi_0 = 0.5$. From the theory set out in Chapter 4 and the Monte Carlo simulations presented in Chapter 5, any rejection in this model is attributable to the break in the SE and hence it is expected that the results from this model would be similar to those obtained from the model with one break in the SE presented earlier.

As shown in Tables 6.9 and 6.10, all the test statistics reject stability in S_0 with probability 1, across all sample sizes. Although there is no break in S_1 and S_2, the rejection frequencies are relatively high for the small sample size of 140 but this quickly reduces as the sample size increases. Similarities between these results and those in the model with one break in the SE only presented in Tables 6.3 and 6.4 are clearly noticeable.

III. Model with One Separate Break in the SE and JE

The model used here is identical to the Coincidental break model except that in this case, the break fractions occur at different locations. Firstly, we examine the case when $\lambda_0 = 0.7$ and $\pi_0 = 0.5$ as displayed in Tables 6.11 and 6.12. There is again a strong rejection of stability in S_0 across all the test statistics, indicating the presence of a break in the model.

Although there is no other break in the two subsamples, the rejection frequencies in S_2 are much higher than those in S_1. This is because the location of the break in the SE at 0.7 implies that S_2 is a much smaller subsample than S_1, with fewer observations available for estimations. For example, when the sample size of 140 is split at 0.7, there would potentially be about 42 observations in S_2. With a trimming of $\epsilon = 0.15$ or $\epsilon = 0.20$, there would be less than 10 observations available for estimations in a partition which can lead to spurious rejections of stability.

Secondly, we also examine the case when $\lambda_0 = 0.5$ and $\pi_0 = 0.7$ as presented in Tables 6.13 and 6.14 for $\epsilon = 0.15$ and $\epsilon = 0.20$ respectively. As expected, the results are similar to those in the model with one break in the SE only given in Tables 6.3 and 6.4 since the location of the break in the SE are identical at 0.5. As seen, stability in S_0 is strongly rejected with probability 1 across all test statistics while S_1 and S_2 have comparable rejection frequencies since they both have no breaks and have about the same number of observations. Again, estimations using $\epsilon = 0.20$ produce lower rejection frequencies than those obtained when $\epsilon = 0.15$ is used because there are potentially more observations available for estimations in each partition. For example, using $sup_{\lambda \in \Lambda} \ddot{D}_T^R$, the rejection frequency at the 1% significance level is 6.8% as against 4.5% for a sample of size 600.

Table 6.1: Relative rejection frequencies of the test statistics - Model has no break in SE or JE. $\Lambda = [0.15, 0.85]$

Sample Used	T	$sup_{\lambda\in\Lambda} Wald_T$			$sup_{\lambda\in\Lambda} LM_T$			$sup_{\lambda\in\Lambda} \ddot{D}_T^U$			$sup_{\lambda\in\Lambda} \ddot{D}_T^R$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
S_0	140	0.113	0.225	0.329	0.103	0.217	0.322	0.213	0.399	0.475	0.204	0.367	0.466
	300	0.035	0.116	0.190	0.037	0.111	0.202	0.088	0.227	0.298	0.076	0.199	0.286
	600	0.028	0.084	0.145	0.016	0.071	0.120	0.039	0.092	0.152	0.021	0.062	0.115
S_1	140	0.011	0.041	0.094	0.015	0.026	0.064	0.032	0.089	0.156	0.023	0.078	0.133
	300	0.010	0.051	0.087	0.010	0.022	0.047	0.012	0.065	0.113	0.011	0.051	0.105
	600	0.005	0.015	0.037	0.003	0.014	0.030	0.007	0.024	0.055	0.002	0.019	0.040
S_2	140	0.012	0.044	0.090	0.012	0.026	0.069	0.028	0.079	0.152	0.021	0.085	0.134
	300	0.006	0.034	0.080	0.006	0.020	0.046	0.013	0.055	0.106	0.011	0.059	0.098
	600	0.002	0.016	0.040	0.002	0.011	0.023	0.002	0.018	0.044	0.002	0.011	0.032

Rejection frequency of $sup_{\lambda\in\Lambda} A_T$ is derived as the proportion of the number of times $sup_{\lambda\in\Lambda} A_T >$critical value.

The critical values used are 15.56, 11.72 and 10.00 for 1%, 5% and 10% significance levels respectively.

S_0 uses the full sample; S_1 and S_2 are the subsamples before & after the first estimated break respectively.

The SE parameters are $\theta_0 = (1, 0.1)'$ and the JE parameters are $\delta_0 = (1, 0.5, 0.5, 0.5, 0.5)'$.

Table 6.2: Relative rejection frequencies of the test statistics - Model has no break in SE or JE. $\Lambda = [0.20, 0.80]$

Sample	T	$sup_{\lambda\in\Lambda} Wald_T$			$sup_{\lambda\in\Lambda} LM_T$			$sup_{\lambda\in\Lambda} \ddot{D}_T^U$			$sup_{\lambda\in\Lambda} \ddot{D}_T^R$		
Used		1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
	140	0.060	0.167	0.244	0.052	0.149	0.239	0.112	0.256	0.376	0.123	0.236	0.354
S_0	300	0.027	0.094	0.171	0.022	0.093	0.180	0.074	0.195	0.270	0.057	0.127	0.244
	600	0.020	0.079	0.131	0.015	0.062	0.114	0.036	0.087	0.143	0.017	0.053	0.102
	140	0.020	0.078	0.130	0.026	0.085	0.152	0.042	0.097	0.171	0.053	0.127	0.153
S_1	300	0.003	0.028	0.058	0.006	0.026	0.073	0.015	0.035	0.076	0.021	0.047	0.081
	600	0.002	0.009	0.028	0.002	0.010	0.032	0.003	0.017	0.037	0.002	0.017	0.022
	140	0.017	0.070	0.123	0.022	0.088	0.163	0.037	0.091	0.158	0.061	0.119	0.147
S_2	300	0.001	0.022	0.051	0.005	0.027	0.085	0.018	0.031	0.074	0.027	0.038	0.078
	600	0.001	0.001	0.027	0.001	0.009	0.033	0.001	0.011	0.035	0.001	0.011	0.022

The critical values used are 15.09, 11.26 and 9.54 for 1%, 5% and 10% significance levels respectively.

Other details are similar to those in Table 6.1.

Table 6.3: Relative rejection frequencies of the test statistics - Model with one break in the SE. $\Lambda = [0.15, 0.85]$

Sample	T	$sup_{\lambda\in\Lambda} Wald_T$			$sup_{\lambda\in\Lambda} LM_T$			$sup_{\lambda\in\Lambda} \ddot{D}_T^U$			$sup_{\lambda\in\Lambda} \ddot{D}_T^R$		
Used		1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
	140	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
S_0	300	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	600	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	140	0.297	0.476	0.594	0.303	0.482	0.580	0.568	0.730	0.801	0.549	0.712	0.794
S_1	300	0.109	0.246	0.332	0.098	0.228	0.318	0.213	0.395	0.495	0.187	0.348	0.470
	600	0.063	0.150	0.231	0.030	0.118	0.210	0.093	0.206	0.306	0.067	0.180	0.279
	140	0.323	0.500	0.603	0.320	0.503	0.613	0.593	0.732	0.811	0.554	0.719	0.787
S_2	300	0.134	0.279	0.379	0.103	0.229	0.341	0.252	0.396	0.518	0.210	0.371	0.477
	600	0.059	0.160	0.233	0.033	0.123	0.190	0.093	0.204	0.284	0.072	0.183	0.275

The SE parameters are $(\theta_0^{(1)}, \theta_0^{(2)}) = \left([1, 0.1]', [-1, -0.1]'\right)$.

The break point in the SE is at $\lambda_0 = 0.5$.

Table 6.4: Relative rejection frequencies of the test statistics - Model with one break in the SE. $\Lambda = [0.20, 0.80]$

Sample	T	$sup_{\lambda\in\Lambda} Wald_T$			$sup_{\lambda\in\Lambda} LM_T$			$sup_{\lambda\in\Lambda} \ddot{D}_T^U$			$sup_{\lambda\in\Lambda} \ddot{D}_T^R$		
Used		1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
	140	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
S_0	300	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	600	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	140	0.187	0.348	0.473	0.193	0.344	0.455	0.406	0.573	0.657	0.383	0.560	0.659
S_1	300	0.066	0.183	0.270	0.054	0.158	0.252	0.122	0.286	0.380	0.115	0.268	0.373
	600	0.044	0.114	0.193	0.027	0.100	0.166	0.061	0.152	0.246	0.047	0.137	0.224
	140	0.203	0.359	0.480	0.198	0.361	0.469	0.405	0.561	0.655	0.355	0.562	0.655
S_2	300	0.095	0.198	0.292	0.067	0.184	0.273	0.158	0.285	0.399	0.130	0.277	0.387
	600	0.041	0.117	0.181	0.028	0.092	0.164	0.059	0.144	0.210	0.050	0.124	0.213

The SE parameters are $(\theta_0^{(1)}, \theta_0^{(2)}) = \left([1, 0.1]', [-1, -0.1]'\right)$.

The break point in the SE is at $\lambda_0 = 0.5$.

Table 6.5: Relative rejection frequencies of the test statistics - Model with two breaks in the SE. $\Lambda = [0.15, 0.85]$

Sample	T	$sup_{\lambda\in\Lambda} Wald_T$			$sup_{\lambda\in\Lambda} LM_T$			$sup_{\lambda\in\Lambda} \ddot{D}_T^U$			$sup_{\lambda\in\Lambda} \ddot{D}_T^R$		
Used		1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
	140	1.000	1.000	1.000	0.995	0.990	1.000	1.000	1.000	1.000	0.998	0.999	1.000
S_0	300	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	600	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	140	0.606	0.774	0.831	0.608	0.742	0.821	0.882	0.927	0.944	0.845	0.922	0.937
S_1	300	0.217	0.397	0.506	0.252	0.430	0.530	0.468	0.631	0.717	0.422	0.602	0.703
	600	0.087	0.212	0.304	0.097	0.207	0.304	0.153	0.315	0.419	0.128	0.292	0.397
	140	1.000	1.000	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	0.999	1.000
S_2	300	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	600	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

The SE parameters are $(\theta_0^{(1)}, \theta_0^{(2)}, \theta_0^{(3)}) = \left([1, 0.1]', [-1; -5.1]', [1; -4.1]'\right)$.

The break points in the SE are at $\lambda_0^{(1)} = 0.3$ and $\lambda_0^{(2)} = 0.7$.

Notice a bigger magnitude of shift is imposed at the first break point so that $D_T(\lambda_0^{(1)}) > D_T(\lambda_0^{(2)})$.

Table 6.6: Relative rejection frequencies of the test statistics - Model with two breaks in the SE. $\Lambda = [0.20, 0.80]$

Sample	T	$sup_{\lambda\in\Lambda} Wald_T$			$sup_{\lambda\in\Lambda} LM_T$			$sup_{\lambda\in\Lambda} \ddot{D}_T^U$			$sup_{\lambda\in\Lambda} \ddot{D}_T^R$		
Used		1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
	140	1.000	1.000	1.000	0.996	0.999	1.000	1.000	1.000	1.000	0.998	1.000	1.000
S_0	300	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	600	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	140	0.403	0.590	0.702	0.462	0.602	0.706	0.770	0.862	0.910	0.705	0.834	0.885
S_1	300	0.149	0.297	0.403	0.169	0.314	0.413	0.284	0.477	0.571	0.257	0.443	0.564
	600	0.057	0.159	0.257	0.065	0.162	0.251	0.096	0.226	0.330	0.085	0.212	0.309
	140	1.000	1.000	1.000	0.996	0.999	1.000	1.000	1.000	1.000	0.998	1.000	1.000
S_2	300	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	600	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

The SE parameters are $(\theta_0^{(1)}, \theta_0^{(2)}, \theta_0^{(3)}) = \left([1, 0.1]', [-1; -5.1]', [1; -4.1]'\right)$.

The break points in the SE are at $\lambda_0^{(1)} = 0.3$ and $\lambda_0^{(2)} = 0.7$.

Details are similar to those in Table 6.5.

Table 6.7: Relative rejection frequencies of the test statistics - Model with one break in the JE only. $\Lambda = [0.15, 0.85]$

Sample	T	$sup_{\lambda\in\Lambda} Wald_T$			$sup_{\lambda\in\Lambda} LM_T$			$sup_{\lambda\in\Lambda} \ddot{D}_T^U$			$sup_{\lambda\in\Lambda} \ddot{D}_T^R$		
Used		1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
	140	0.143	0.299	0.399	0.147	0.269	0.383	0.286	0.456	0.510	0.267	0.473	0.516
S_0	300	0.061	0.153	0.234	0.083	0.183	0.269	0.159	0.294	0.371	0.146	0.249	0.378
	600	0.036	0.113	0.181	0.033	0.098	0.171	0.063	0.142	0.219	0.049	0.126	0.224
	140	0.022	0.083	0.144	0.049	0.103	0.186	0.099	0.118	0.219	0.074	0.109	0.189
S_1	300	0.010	0.038	0.080	0.017	0.053	0.118	0.076	0.093	0.109	0.036	0.080	0.113
	600	0.002	0.021	0.047	0.005	0.018	0.043	0.013	0.033	0.065	0.012	0.041	0.062
	140	0.021	0.076	0.117	0.056	0.121	0.178	0.112	0.134	0.226	0.082	0.097	0.171
S_2	300	0.015	0.052	0.099	0.013	0.047	0.110	0.097	0.125	0.189	0.041	0.088	0.124
	600	0.001	0.016	0.043	0.004	0.016	0.051	0.016	0.043	0.072	0.013	0.053	0.081

The SE parameters are $\theta_0 = (1, 0.1)'$.

The JE coefficients are $(\delta_0^{(1)}, \delta_0^{(2)}) = ([1, 0.8, 0.8, 0.8, 0.8]', [0.1, 0.2, 0.2, 0.2, 0.2]')$.

The break in the JE is at $\pi_0 = 0.5$.

Table 6.8: Relative rejection frequencies of the test statistics - Model with one break in the JE only. $\Lambda = [0.20, 0.80]$

Sample	T	$sup_{\lambda\in\Lambda} Wald_T$			$sup_{\lambda\in\Lambda} LM_T$			$sup_{\lambda\in\Lambda} \ddot{D}_T^U$			$sup_{\lambda\in\Lambda} \ddot{D}_T^R$		
Used		1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
	140	0.098	0.224	0.314	0.112	0.257	0.307	0.231	0.303	0.416	0.221	0.313	0.463
S_0	300	0.046	0.125	0.200	0.048	0.133	0.218	0.143	0.275	0.367	0.106	0.227	0.347
	600	0.029	0.089	0.148	0.031	0.087	0.141	0.052	0.138	0.205	0.041	0.101	0.212
	140	0.023	0.072	0.138	0.038	0.089	0.153	0.079	0.095	0.197	0.081	0.103	0.213
S_1	300	0.003	0.018	0.055	0.012	0.033	0.097	0.059	0.073	0.096	0.048	0.089	0.181
	600	0.000	0.009	0.028	0.001	0.006	0.023	0.002	0.011	0.034	0.001	0.019	0.037
	140	0.032	0.108	0.173	0.043	0.119	0.169	0.088	0.105	0.215	0.087	0.097	0.198
S_2	300	0.007	0.031	0.059	0.008	0.025	0.090	0.063	0.076	0.083	0.041	0.068	0.192
	600	0.002	0.010	0.027	0.001	0.006	0.022	0.002	0.018	0.044	0.001	0.011	0.039

The SE parameters are $\theta_0 = (1, 0.1)'$.

The JE coefficients are $(\delta_0^{(1)}, \delta_0^{(2)}) = ([1, 0.8, 0.8, 0.8, 0.8]', [0.1, 0.2, 0.2, 0.2, 0.2]')$.

The break in the JE is at $\pi_0 = 0.5$.

Table 6.9: Relative rejection frequencies of the test statistics - Model with Coincidental break in the SE and JE. $\Lambda = [0.15, 0.85]$

Sample Used	T	$sup_{\lambda\in\Lambda} Wald_T$			$sup_{\lambda\in\Lambda} LM_T$			$sup_{\lambda\in\Lambda} \ddot{D}_T^U$			$sup_{\lambda\in\Lambda} \ddot{D}_T^R$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
S_0	140	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	300	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	600	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
S_1	140	0.336	0.519	0.615	0.297	0.480	0.582	0.599	0.738	0.812	0.560	0.708	0.787
	300	0.124	0.262	0.357	0.119	0.262	0.365	0.221	0.404	0.516	0.186	0.351	0.467
	600	0.057	0.150	0.235	0.036	0.118	0.224	0.088	0.214	0.324	0.065	0.183	0.280
S_2	140	0.195	0.368	0.488	0.331	0.509	0.607	0.486	0.664	0.739	0.497	0.680	0.767
	300	0.073	0.204	0.305	0.125	0.293	0.406	0.188	0.343	0.450	0.198	0.376	0.484
	600	0.050	0.129	0.199	0.059	0.156	0.228	0.067	0.177	0.256	0.076	0.188	0.279

The SE parameters are $(\theta_0^{(1)}, \theta_0^{(2)}) = \left([1, 0.1]', [-1, -0.1]'\right)$.

The JE coefficients are $(\delta_0^{(1)}, \delta_0^{(2)}) = \left([1, 0.8, 0.8, 0.8, 0.8]', [0.1, 0.2, 0.2, 0.2, 0.2]'\right)$.

The coincidental break is at $\lambda_0 = \pi_0 = 0.5$.

Table 6.10: Relative rejection frequencies of the test statistics - Model with Coincidental break in the SE and JE. $\Lambda = [0.20, 0.80]$

Sample Used	T	$sup_{\lambda\in\Lambda} Wald_T$			$sup_{\lambda\in\Lambda} LM_T$			$sup_{\lambda\in\Lambda} \ddot{D}_T^U$			$sup_{\lambda\in\Lambda} \ddot{D}_T^R$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
S_0	140	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	300	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	600	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
S_1	140	0.218	0.395	0.489	0.178	0.345	0.464	0.416	0.581	0.666	0.371	0.552	0.646
	300	0.072	0.198	0.284	0.070	0.178	0.271	0.129	0.292	0.402	0.115	0.270	0.370
	600	0.040	0.122	0.195	0.031	0.108	0.194	0.063	0.160	0.262	0.044	0.143	0.229
S_2	140	0.126	0.264	0.367	0.207	0.373	0.491	0.332	0.483	0.578	0.351	0.531	0.643
	300	0.051	0.148	0.234	0.080	0.217	0.328	0.113	0.238	0.335	0.128	0.287	0.383
	600	0.033	0.089	0.156	0.044	0.129	0.194	0.043	0.115	0.193	0.054	0.139	0.229

The SE parameters are $(\theta_0^{(1)}, \theta_0^{(2)}) = \left([1, 0.1]', [-1, -0.1]'\right)$.

The JE coefficients are $(\delta_0^{(1)}, \delta_0^{(2)}) = ([1, 0.8, 0.8, 0.8, 0.8]', [0.1, 0.2, 0.2, 0.2, 0.2]')$.

The coincidental break is at $\lambda_0 = \pi_0 = 0.5$.

Details are as given in Table 6.9.

Table 6.11: Relative rejection frequencies of the test statistics - Model with Separate breaks in SE and JE. $\Lambda = [0.15, 0.85]$

Sample	T	$sup_{\lambda\in\Lambda} Wald_T$			$sup_{\lambda\in\Lambda} LM_T$			$sup_{\lambda\in\Lambda} \ddot{D}_T^U$			$sup_{\lambda\in\Lambda} \ddot{D}_T^R$		
Used		1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
	140	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
S_0	300	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	600	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	140	0.284	0.464	0.562	0.226	0.393	0.515	0.464	0.624	0.712	0.404	0.566	0.664
S_1	300	0.120	0.246	0.342	0.064	0.163	0.260	0.189	0.331	0.424	0.126	0.278	0.388
	600	0.060	0.145	0.231	0.020	0.088	0.159	0.079	0.176	0.274	0.054	0.139	0.226
	140	0.457	0.618	0.695	0.551	0.733	0.811	0.828	0.912	0.932	0.830	0.917	0.947
S_2	300	0.161	0.308	0.425	0.236	0.426	0.545	0.368	0.537	0.642	0.390	0.584	0.684
	600	0.058	0.156	0.242	0.087	0.222	0.320	0.121	0.258	0.368	0.140	0.295	0.407

The SE and JE parameters are identical to those in Table 6.9.

The break in SE is at $\lambda_0 = 0.7$ while that in the JE is at $\pi_0 = 0.5$.

Table 6.12: Relative rejection frequencies of the test statistics - Model with Separate breaks in SE and JE. $\Lambda = [0.20, 0.80]$

Sample	T	$sup_{\lambda\in\Lambda} Wald_T$			$sup_{\lambda\in\Lambda} LM_T$			$sup_{\lambda\in\Lambda} \ddot{D}_T^U$			$sup_{\lambda\in\Lambda} \ddot{D}_T^R$		
Used		1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
	140	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
S_0	300	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	600	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	140	0.173	0.348	0.471	0.134	0.271	0.390	0.335	0.491	0.594	0.263	0.424	0.531
S_1	300	0.083	0.190	0.271	0.036	0.137	0.221	0.132	0.235	0.333	0.075	0.193	0.293
	600	0.048	0.119	0.203	0.023	0.075	0.150	0.055	0.139	0.227	0.041	0.110	0.193
	140	0.298	0.474	0.591	0.383	0.570	0.669	0.642	0.783	0.837	0.694	0.814	0.856
S_2	300	0.093	0.218	0.319	0.149	0.315	0.427	0.225	0.386	0.497	0.267	0.442	0.547
	600	0.032	0.111	0.186	0.053	0.158	0.245	0.065	0.171	0.276	0.093	0.214	0.320

The SE and JE parameters are identical to those in Table 6.9.

The break in SE is at $\lambda_0 = 0.7$ while that in the JE is at $\pi_0 = 0.5$.

Table 6.13: Relative rejection frequencies of the test statistics - Model with Separate breaks in SE and JE. $\Lambda = [0.15, 0.85]$

Sample	T	$sup_{\lambda\in\Lambda} Wald_T$			$sup_{\lambda\in\Lambda} LM_T$			$sup_{\lambda\in\Lambda} \ddot{D}_T^U$			$sup_{\lambda\in\Lambda} \ddot{D}_T^R$		
Used		1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
	140	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
S_0	300	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	600	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	140	0.330	0.522	0.620	0.279	0.453	0.564	0.596	0.737	0.808	0.547	0.712	0.790
S_1	300	0.124	0.257	0.354	0.090	0.225	0.316	0.222	0.408	0.511	0.182	0.352	0.459
	600	0.059	0.154	0.239	0.029	0.113	0.214	0.090	0.216	0.326	0.065	0.186	0.279
	140	0.357	0.525	0.622	0.329	0.500	0.600	0.603	0.753	0.820	0.535	0.711	0.792
S_2	300	0.145	0.296	0.411	0.109	0.227	0.329	0.262	0.414	0.519	0.202	0.370	0.479
	600	0.074	0.166	0.242	0.041	0.102	0.175	0.109	0.210	0.308	0.068	0.170	0.253

The SE and JE parameters are identical to those in Table 6.9.

The break in SE is at $\lambda_0 = 0.5$ while that in the JE is at $\pi_0 = 0.7$.

Table 6.14: Relative rejection frequencies of the test statistics - Model with Separate breaks in SE and JE. $\Lambda = [0.20, 0.80]$

Sample	T	$sup_{\lambda\in\Lambda} Wald_T$			$sup_{\lambda\in\Lambda} LM_T$			$sup_{\lambda\in\Lambda} \ddot{D}_T^U$			$sup_{\lambda\in\Lambda} \ddot{D}_T^R$		
Used		1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
	140	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
S_0	300	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	600	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	140	0.208	0.398	0.496	0.167	0.329	0.453	0.417	0.588	0.672	0.366	0.549	0.644
S_1	300	0.068	0.192	0.283	0.048	0.156	0.248	0.125	0.290	0.396	0.119	0.264	0.364
	600	0.041	0.124	0.198	0.028	0.099	0.177	0.063	0.161	0.264	0.047	0.142	0.232
	140	0.245	0.399	0.505	0.207	0.367	0.480	0.428	0.579	0.677	0.373	0.536	0.649
S_2	300	0.099	0.216	0.312	0.070	0.170	0.268	0.167	0.312	0.412	0.138	0.284	0.376
	600	0.050	0.130	0.206	0.032	0.083	0.147	0.072	0.158	0.235	0.045	0.117	0.206

The SE and JE parameters are identical to those in Table 6.9.

The break in SE is at $\lambda_0 = 0.5$ while that in the JE is at $\pi_0 = 0.7$.

Category 2: Estimated Number of Breaks

This section presents the results of the estimated number of break points in each model, as established by each of the four test statistics used. The estimated break point is denoted as $\hat{m}$. Based on the estimation procedure outlined in Section 6.3.1, a maximum of 3 break points can be estimated in any model; one using the whole sample S_0 and one each from the subsamples S_1 and S_2. The models and critical values presented here are identical to those used in Category 2.

Results for Models with a Stable Jacobian Equation

As in the previous category, this section also examines models with no break, one break and two breaks in their SE.

I. Model with No Break in the SE or JE

As the model has no breaks, it is expected that for all test statistics, the number of break points estimated, $\hat{m}$, should be zero for majority of the simulations. This is indeed the case as shown in Tables 6.15 and 6.16 for $\epsilon = 0.15$ and $\epsilon = 0.20$ respectively. Of the 1,000 repetitions when $\epsilon = 0.15$, $sup_{\lambda\in\Lambda}Wald_T$ and $sup_{\lambda\in\Lambda}LM_T$ estimate zero breaks 96.5% and 96.3% of the time respectively, at 1% significance level for a sample size of 300. This increased slightly to 97.3% and 97.8% respectively when $\epsilon = 0.20$. However, one break point was estimated 5.1%, 2.3% and 1.3% of the time for sample sizes 140, 300 and 600 respectively using $sup_{\lambda\in\Lambda}\ddot{D}_T^R$ when $\epsilon = 0.20$.

II. Model with One Break in the SE only

The results for this model with one break are presented in Tables 6.17 and 6.18. At 1% significance level for a sample size of 140 when using $sup_{\lambda\in\Lambda}Wald_T$, one break point is estimated only 47.9% of the time; this drops to 27.3% and 16% at the 5% and 10% significance levels respectively. Using $sup_{\lambda\in\Lambda}\ddot{D}_T^U$ and $sup_{\lambda\in\Lambda}\ddot{D}_T^R$ for the same sample size, one break point is estimated only 17.2% and 19.9% times respectively. Also, at 10% significance level for a sample size of 300, two break points are estimated more often across all the test statistics. These results are consistent with the high rejection frequencies in Category 1 presented earlier in Tables 6.3 and 6.4. However, the results improve for the bigger sample size of 600.

III. Model with Two Breaks in the SE only

As seen in the results displayed in Tables 6.19 and 6.20, there is a strong rejection of no break and one break across all test statistics. The results from the sample with a size of 140 is again distorted as three break points are estimated more often across all significance levels and all test statistics. This is attributable to the few observations in S_1 which led to spurious rejections of stability as seen in Tables 6.5 and 6.6. In contrast, the model with a bigger sample size of 600 estimates correctly two break points more often across all test statistics.

Results for Models with an Unstable Jacobian Equation

As in Category 1, the same three models with an Unstable JE are considered here: a model with a break only in the JE and models with a Coincidental and Separate break in the JE and SE.

I. Model with One Break in the JE only

With the break only in the JE, the results are expected to be similar to those obtained from a model with no breaks presented earlier. Indeed, this is the case as seen from Tables 6.21 and 6.22. For a sample size of 300 at 1% significance level, $sup_{\lambda\in\Lambda}Wald_T$ and $sup_{\lambda\in\Lambda}LM_T$ estimate zero break 95.4% and 95.2% of the time respectively when $\epsilon = 0.20$. Again, $sup_{\lambda\in\Lambda}\ddot{D}_T^U$ and $sup_{\lambda\in\Lambda}\ddot{D}_T^R$ are comparatively lower at 85.7% and 89.4% respectively. Using $sup_{\lambda\in\Lambda}\ddot{D}_T^R$ at 1% significance level, only a small fraction estimated one break in the model at 5.6%, 3.3% and 1.4% for sample size 140, 300 and 600 respectively.

II. Model with One Coincidental Break in the SE and JE

The results for this model are displayed in Tables 6.23 and 6.24 for $\epsilon = 0.15$ and $\epsilon = 0.20$ respectively. From the theory, this model should yield similar results to the model with a single break in the SE discussed above in Tables 6.17 and 6.18. The results obtained from the sample of size 140 is again distorted as it only estimates one break point most of the time using $sup_{\lambda\in\Lambda}Wald_T$ and $sup_{\lambda\in\Lambda}LM_T$ at 1% significance level. For both $sup_{\lambda\in\Lambda}\ddot{D}_T^U$ and $sup_{\lambda\in\Lambda}\ddot{D}_T^R$, three breaks are estimated at both 5% and 10% significance levels for a sample size of 140. This implies that in addition to the true break in S_0, other breaks

were obtained in S_1 and S_2. However, for the large sample size, a single break is correctly estimated across all test statistics.

III. Model with One Separate Break in the SE and JE

In the first model displayed in Tables 6.25 and 6.26, the break in the SE lies at 0.7 while that in the JE lies at 0.5. Across all sample sizes, there is a substantial estimation of two break points, particularly for sample sizes 140 and 300. This is reasonable considering the location of the break in the SE; as discussed earlier, the observations available for estimation in S_2 are few, giving rise to spurious rejections of stability in that subsample. When the location of the break points are swapped, that is, when the break in the SE, $\lambda_0 = 0.5$, and that in the JE, $\pi_0 = 0.7$, then there is an improvement in the estimation of one break point across the test statistics, particularly for the large sample size, as seen in Tables 6.27 and 6.28. For example, using $sup_{\lambda \in \Lambda} \ddot{D}_T^R$ at 1% significance level for sample size 600, 87% estimate one break when $\lambda_0 = 0.7$ as against 91% when $\lambda_0 = 0.5$.

Table 6.15: Relative frequencies of the estimated number of breaks - Model has no break in the SE or JE. $\Lambda = [0.15, 0.85]$

Sign	Breaks	$sup_{\lambda\in\Lambda} Wald_T$			$sup_{\lambda\in\Lambda} LM_T$			$sup_{\lambda\in\Lambda} \ddot{D}_T^U$			$sup_{\lambda\in\Lambda} \ddot{D}_T^R$		
Level	Est.	140	300	600	140	300	600	140	300	600	140	300	600
1%	$\hat{m}=0$	0.887	0.965	0.972	0.897	0.963	0.984	0.787	0.912	0.961	0.796	0.924	0.979
	$\hat{m}=1$	0.091	0.021	0.022	0.088	0.026	0.011	0.149	0.055	0.026	0.138	0.040	0.015
	$\hat{m}=2$	0.021	0.012	0.005	0.012	0.009	0.005	0.042	0.021	0.012	0.039	0.025	0.005
	$\hat{m}=3$	0.001	0.002	0.001	0.003	0.002	0.000	0.022	0.012	0.001	0.027	0.011	0.001
5%	$\hat{m}=0$	0.775	0.884	0.916	0.783	0.889	0.929	0.601	0.773	0.908	0.633	0.801	0.938
	$\hat{m}=1$	0.149	0.042	0.055	0.147	0.066	0.043	0.189	0.070	0.045	0.207	0.111	0.039
	$\hat{m}=2$	0.067	0.063	0.027	0.056	0.035	0.022	0.165	0.124	0.037	0.136	0.075	0.016
	$\hat{m}=3$	0.009	0.011	0.002	0.014	0.010	0.006	0.045	0.033	0.010	0.024	0.013	0.007
10%	$\hat{m}=0$	0.671	0.810	0.855	0.678	0.798	0.880	0.525	0.702	0.848	0.534	0.714	0.885
	$\hat{m}=1$	0.177	0.052	0.077	0.167	0.101	0.062	0.205	0.102	0.063	0.263	0.176	0.082
	$\hat{m}=2$	0.120	0.109	0.059	0.129	0.083	0.053	0.194	0.145	0.072	0.166	0.084	0.023
	$\hat{m}=3$	0.032	0.029	0.009	0.026	0.018	0.005	0.076	0.051	0.017	0.037	0.026	0.010

$\hat{m} = j$, denotes j breaks were estimated in 1,000 simulations. Other details of the model are given in Table 6.1.

Table 6.16: Relative frequencies of the estimated number of breaks - Model has no break in the SE or JE. $\Lambda = [0.20, 0.80]$

Sign	Breaks	$sup_{\lambda\in\Lambda} Wald_T$			$sup_{\lambda\in\Lambda} LM_T$			$sup_{\lambda\in\Lambda} \ddot{D}_T^U$			$sup_{\lambda\in\Lambda} \ddot{D}_T^R$		
Level	Est.	140	300	600	140	300	600	140	300	600	140	300	600
1%	$\hat{m}=0$	0.940	0.973	0.980	0.948	0.978	0.985	0.888	0.926	0.964	0.877	0.943	0.983
	$\hat{m}=1$	0.026	0.023	0.017	0.040	0.019	0.012	0.047	0.042	0.021	0.051	0.023	0.013
	$\hat{m}=2$	0.031	0.004	0.003	0.010	0.003	0.003	0.043	0.032	0.015	0.044	0.034	0.004
	$\hat{m}=3$	0.003	0.000	0.000	0.002	0.000	0.000	0.022	0.000	0.000	0.028	0.000	0.000
5%	$\hat{m}=0$	0.833	0.906	0.921	0.851	0.907	0.938	0.744	0.805	0.913	0.764	0.873	0.947
	$\hat{m}=1$	0.038	0.051	0.061	0.036	0.063	0.044	0.029	0.057	0.050	0.038	0.027	0.036
	$\hat{m}=2$	0.110	0.036	0.017	0.098	0.025	0.017	0.167	0.090	0.036	0.152	0.076	0.016
	$\hat{m}=3$	0.019	0.007	0.001	0.015	0.005	0.001	0.060	0.048	0.001	0.046	0.024	0.001
10%	$\hat{m}=0$	0.756	0.829	0.869	0.761	0.820	0.886	0.624	0.730	0.857	0.646	0.756	0.898
	$\hat{m}=1$	0.047	0.082	0.082	0.068	0.092	0.076	0.119	0.084	0.071	0.093	0.084	0.053
	$\hat{m}=2$	0.141	0.069	0.043	0.107	0.060	0.033	0.160	0.132	0.057	0.186	0.097	0.039
	$\hat{m}=3$	0.056	0.020	0.006	0.064	0.028	0.005	0.097	0.054	0.015	0.075	0.063	0.010

Details of the model are given in Table 6.2.

Table 6.17: Relative frequencies of the estimated number of breaks - Model has one break in the SE. $\Lambda = [0.15, 0.85]$

Sign	Breaks	$sup_{\lambda\in\Lambda} Wald_T$			$sup_{\lambda\in\Lambda} LM_T$			$sup_{\lambda\in\Lambda} \ddot{D}_T^U$			$sup_{\lambda\in\Lambda} \ddot{D}_T^R$		
Level	Est.	140	300	600	140	300	600	140	300	600	140	300	600
1%	$\hat{m}=0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{m}=1$	0.479	0.767	0.881	0.476	0.810	0.939	0.172	0.578	0.821	0.199	0.645	0.866
	$\hat{m}=2$	0.422	0.223	0.116	0.425	0.179	0.059	0.495	0.379	0.172	0.499	0.313	0.129
	$\hat{m}=3$	0.099	0.010	0.003	0.099	0.011	0.002	0.333	0.043	0.007	0.302	0.042	0.005
5%	$\hat{m}=0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{m}=1$	0.273	0.535	0.707	0.238	0.593	0.770	0.065	0.363	0.632	0.083	0.395	0.667
	$\hat{m}=2$	0.478	0.405	0.276	0.539	0.357	0.219	0.408	0.483	0.326	0.403	0.491	0.303
	$\hat{m}=3$	0.249	0.060	0.017	0.223	0.050	0.011	0.527	0.154	0.042	0.514	0.114	0.030
10%	$\hat{m}=0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{m}=1$	0.160	0.403	0.587	0.152	0.436	0.643	0.031	0.228	0.504	0.045	0.261	0.519
	$\hat{m}=2$	0.483	0.483	0.362	0.503	0.469	0.314	0.326	0.531	0.402	0.329	0.531	0.408
	$\hat{m}=3$	0.357	0.114	0.051	0.345	0.095	0.043	0.643	0.241	0.094	0.626	0.208	0.073

Details of the model are given in Table 6.3.

Table 6.18: Relative frequencies of the estimated number of breaks - Model has one break in the SE. $\Lambda = [0.20, 0.80]$

Sign	Breaks	$sup_{\lambda\in\Lambda} Wald_T$			$sup_{\lambda\in\Lambda} LM_T$			$sup_{\lambda\in\Lambda} \ddot{D}_T^U$			$sup_{\lambda\in\Lambda} \ddot{D}_T^R$		
Level	Est.	140	300	600	140	300	600	140	300	600	140	300	600
1%	$\hat{m}=0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{m}=1$	0.647	0.843	0.917	0.651	0.884	0.946	0.353	0.734	0.882	0.386	0.765	0.905
	$\hat{m}=2$	0.316	0.153	0.081	0.307	0.111	0.053	0.483	0.252	0.116	0.490	0.225	0.093
	$\hat{m}=3$	0.037	0.004	0.002	0.042	0.005	0.001	0.164	0.014	0.002	0.124	0.010	0.002
5%	$\hat{m}=0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{m}=1$	0.423	0.650	0.777	0.408	0.678	0.818	0.185	0.497	0.728	0.191	0.519	0.756
	$\hat{m}=2$	0.447	0.319	0.215	0.479	0.302	0.172	0.496	0.435	0.248	0.496	0.417	0.227
	$\hat{m}=3$	0.130	0.031	0.008	0.113	0.020	0.010	0.319	0.068	0.024	0.313	0.064	0.017
10%	$\hat{m}=0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{m}=1$	0.274	0.506	0.657	0.276	0.533	0.704	0.119	0.359	0.597	0.112	0.378	0.613
	$\hat{m}=2$	0.499	0.426	0.312	0.524	0.409	0.262	0.450	0.503	0.350	0.462	0.484	0.337
	$\hat{m}=3$	0.227	0.068	0.031	0.200	0.058	0.034	0.431	0.138	0.053	0.426	0.138	0.050

Details of the model are given in Table 6.4.

Table 6.19: Relative frequencies of the estimated number of breaks - Model has two breaks in the SE. $\Lambda = [0.15, 0.85]$

Sign	Breaks	$sup_{\lambda\in\Lambda} Wald_T$			$sup_{\lambda\in\Lambda} LM_T$			$sup_{\lambda\in\Lambda} \ddot{D}_T^U$			$sup_{\lambda\in\Lambda} \ddot{D}_T^R$		
Level	Est.	140	300	600	140	300	600	140	300	600	140	300	600
1%	$\hat{m}=0$	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{m}=1$	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{m}=2$	0.394	0.783	0.913	0.386	0.748	0.903	0.118	0.532	0.847	0.153	0.578	0.872
	$\hat{m}=3$	0.606	0.217	0.087	0.608	0.252	0.097	0.882	0.468	0.153	0.845	0.422	0.128
5%	$\hat{m}=0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{m}=1$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{m}=2$	0.226	0.603	0.788	0.257	0.570	0.793	0.073	0.369	0.685	0.077	0.398	0.708
	$\hat{m}=3$	0.774	0.397	0.212	0.742	0.430	0.207	0.927	0.631	0.315	0.922	0.602	0.292
10%	$\hat{m}=0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{m}=1$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{m}=2$	0.169	0.494	0.696	0.179	0.470	0.703	0.056	0.283	0.581	0.063	0.297	0.603
	$\hat{m}=3$	0.831	0.506	0.304	0.821	0.530	0.297	0.944	0.717	0.419	0.937	0.703	0.397

Details of the model are given in Table 6.5.

Table 6.20: Relative frequencies of the estimated number of breaks - Model has two breaks in the SE. $\Lambda = [0.20, 0.80]$

Sign	Breaks	$sup_{\lambda\in\Lambda} Wald_T$			$sup_{\lambda\in\Lambda} LM_T$			$sup_{\lambda\in\Lambda} \ddot{D}_T^U$			$sup_{\lambda\in\Lambda} \ddot{D}_T^R$		
Level	Est.	140	300	600	140	300	600	140	300	600	140	300	600
1%	$\hat{m}=0$	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000
	$\hat{m}=1$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{m}=2$	0.597	0.851	0.943	0.534	0.831	0.935	0.230	0.716	0.904	0.293	0.743	0.915
	$\hat{m}=3$	0.403	0.149	0.057	0.462	0.169	0.065	0.770	0.284	0.096	0.705	0.257	0.085
5%	$\hat{m}=0$	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{m}=1$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{m}=2$	0.410	0.703	0.841	0.397	0.686	0.838	0.138	0.523	0.774	0.166	0.557	0.788
	$\hat{m}=3$	0.590	0.297	0.159	0.602	0.314	0.162	0.862	0.477	0.226	0.834	0.443	0.212
10%	$\hat{m}=0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{m}=1$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{m}=2$	0.298	0.597	0.743	0.294	0.587	0.749	0.090	0.429	0.670	0.115	0.436	0.691
	$\hat{m}=3$	0.702	0.403	0.257	0.706	0.413	0.251	0.910	0.571	0.330	0.885	0.564	0.309

Details of the model are given in Table 6.6.

Table 6.21: Relative frequencies of the estimated number of breaks - Model has one break in the JE only. $\Lambda = [0.15, 0.85]$

Sign	Breaks	$sup_{\lambda\in\Lambda} Wald_T$			$sup_{\lambda\in\Lambda} LM_T$			$sup_{\lambda\in\Lambda} \ddot{D}_T^U$			$sup_{\lambda\in\Lambda} \ddot{D}_T^R$		
Level	Est.	140	300	600	140	300	600	140	300	600	140	300	600
1%	$\hat{m}=0$	0.857	0.939	0.964	0.853	0.917	0.967	0.714	0.841	0.937	0.733	0.854	0.951
	$\hat{m}=1$	0.104	0.040	0.033	0.044	0.058	0.021	0.090	0.048	0.028	0.184	0.089	0.031
	$\hat{m}=2$	0.035	0.017	0.003	0.098	0.022	0.012	0.169	0.091	0.032	0.064	0.048	0.013
	$\hat{m}=3$	0.004	0.004	0.000	0.005	0.003	0.000	0.027	0.020	0.003	0.019	0.009	0.005
5%	$\hat{m}=0$	0.701	0.847	0.887	0.731	0.817	0.902	0.544	0.706	0.858	0.527	0.751	0.874
	$\hat{m}=1$	0.164	0.076	0.078	0.081	0.049	0.032	0.138	0.029	0.023	0.164	0.031	0.021
	$\hat{m}=2$	0.111	0.064	0.003	0.131	0.117	0.062	0.231	0.197	0.066	0.213	0.151	0.069
	$\hat{m}=3$	0.024	0.013	0.002	0.057	0.017	0.004	0.087	0.068	0.053	0.096	0.067	0.036
10%	$\hat{m}=0$	0.601	0.766	0.819	0.617	0.731	0.829	0.490	0.629	0.781	0.484	0.622	0.776
	$\hat{m}=1$	0.189	0.081	0.101	0.095	0.099	0.073	0.118	0.069	0.034	0.084	0.069	0.039
	$\hat{m}=2$	0.159	0.127	0.070	0.217	0.141	0.079	0.281	0.201	0.106	0.319	0.213	0.107
	$\hat{m}=3$	0.051	0.026	0.010	0.071	0.029	0.019	0.111	0.101	0.079	0.113	0.096	0.078

Details of the model are given in Table 6.7.

Table 6.22: Relative frequencies of the estimated number of breaks - Model has one break in the JE only. $\Lambda = [0.20, 0.80]$

Sign	Breaks	$sup_{\lambda\in\Lambda} Wald_T$			$sup_{\lambda\in\Lambda} LM_T$			$sup_{\lambda\in\Lambda} \ddot{D}_T^U$			$sup_{\lambda\in\Lambda} \ddot{D}_T^R$		
Level	Est.	140	300	600	140	300	600	140	300	600	140	300	600
1%	$\hat{m}=0$	0.902	0.954	0.971	0.888	0.952	0.969	0.769	0.857	0.948	0.779	0.894	0.959
	$\hat{m}=1$	0.049	0.037	0.027	0.035	0.025	0.017	0.061	0.043	0.023	0.056	0.033	0.014
	$\hat{m}=2$	0.043	0.008	0.002	0.057	0.019	0.014	0.147	0.083	0.027	0.135	0.052	0.022
	$\hat{m}=3$	0.006	0.001	0.000	0.020	0.004	0.000	0.023	0.017	0.002	0.030	0.021	0.005
5%	$\hat{m}=0$	0.776	0.875	0.911	0.743	0.867	0.913	0.697	0.725	0.862	0.687	0.773	0.899
	$\hat{m}=1$	0.077	0.083	0.072	0.070	0.029	0.027	0.037	0.031	0.019	0.067	0.043	0.017
	$\hat{m}=2$	0.114	0.035	0.015	0.134	0.093	0.057	0.201	0.188	0.081	0.203	0.153	0.059
	$\hat{m}=3$	0.033	0.007	0.002	0.053	0.011	0.003	0.065	0.056	0.038	0.043	0.031	0.025
10%	$\hat{m}=0$	0.686	0.800	0.852	0.693	0.782	0.859	0.584	0.633	0.795	0.537	0.653	0.788
	$\hat{m}=1$	0.071	0.103	0.099	0.065	0.087	0.061	0.067	0.063	0.025	0.070	0.052	0.027
	$\hat{m}=2$	0.175	0.080	0.043	0.181	0.111	0.068	0.258	0.216	0.108	0.289	0.208	0.122
	$\hat{m}=3$	0.068	0.017	0.006	0.061	0.020	0.012	0.091	0.088	0.072	0.104	0.087	0.063

Details of the model are given in Table 6.8.

Table 6.23: Relative frequencies of the estimated number of breaks - Model has Coincidental breaks in the SE and JE. $\Lambda = [0.15, 0.85]$

Sign	Breaks	$sup_{\lambda\in\Lambda} Wald_T$			$sup_{\lambda\in\Lambda} LM_T$			$sup_{\lambda\in\Lambda} \ddot{D}_T^U$			$sup_{\lambda\in\Lambda} \ddot{D}_T^R$		
Level	Est.	140	300	600	140	300	600	140	300	600	140	300	600
1%	$\hat{m}=0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{m}=1$	0.538	0.807	0.894	0.462	0.772	0.908	0.201	0.627	0.848	0.226	0.647	0.863
	$\hat{m}=2$	0.393	0.189	0.105	0.448	0.212	0.089	0.513	0.337	0.149	0.491	0.322	0.133
	$\hat{m}=3$	0.069	0.004	0.001	0.090	0.016	0.003	0.286	0.036	0.003	0.283	0.031	0.004
5%	$\hat{m}=0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{m}=1$	0.297	0.571	0.733	0.239	0.518	0.744	0.085	0.571	0.643	0.088	0.571	0.659
	$\hat{m}=2$	0.519	0.392	0.255	0.533	0.409	0.238	0.428	0.507	0.323	0.436	0.503	0.311
	$\hat{m}=3$	0.184	0.037	0.012	0.228	0.073	0.018	0.487	0.120	0.034	0.476	0.112	0.030
10%	$\hat{m}=0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{m}=1$	0.196	0.437	0.606	0.162	0.372	0.598	0.046	0.204	0.500	0.046	0.255	0.508
	$\hat{m}=2$	0.505	0.464	0.354	0.487	0.485	0.352	0.357	0.552	0.420	0.354	0.539	0.425
	$\hat{m}=3$	0.299	0.099	0.040	0.351	0.143	0.050	0.597	0.207	0.080	0.600	0.206	0.067

Details of the model are given in Table 6.9.

Table 6.24: Relative frequencies of the estimated number of breaks - Model has Coincidental breaks in the SE and JE. $\Lambda = [0.20, 0.80]$

Sign	Breaks	$sup_{\lambda\in\Lambda} Wald_T$			$sup_{\lambda\in\Lambda} LM_T$			$sup_{\lambda\in\Lambda} \ddot{D}_T^U$			$sup_{\lambda\in\Lambda} \ddot{D}_T^R$		
Level	Est.	140	300	600	140	300	600	140	300	600	140	300	600
1%	$\hat{m}=0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{m}=1$	0.668	0.877	0.927	0.647	0.856	0.927	0.389	0.770	0.895	0.407	0.767	0.904
	$\hat{m}=2$	0.280	0.123	0.073	0.321	0.138	0.071	0.474	0.218	0.104	0.464	0.223	0.094
	$\hat{m}=3$	0.032	0.000	0.000	0.032	0.006	0.002	0.137	0.012	0.001	0.129	0.010	0.002
5%	$\hat{m}=0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{m}=1$	0.438	0.678	0.800	0.397	0.643	0.777	0.220	0.532	0.740	0.212	0.508	0.734
	$\hat{m}=2$	0.465	0.298	0.189	0.488	0.319	0.209	0.496	0.406	0.245	0.493	0.427	0.250
	$\hat{m}=3$	0.097	0.024	0.011	0.115	0.038	0.014	0.284	0.062	0.015	0.295	0.065	0.016
10%	$\hat{m}=0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{m}=1$	0.314	0.536	0.673	0.264	0.478	0.654	0.138	0.386	0.596	0.125	0.378	0.593
	$\hat{m}=2$	0.516	0.410	0.303	0.517	0.445	0.304	0.480	0.491	0.353	0.461	0.491	0.356
	$\hat{m}=3$	0.170	0.054	0.024	0.219	0.077	0.042	0.382	0.123	0.051	0.414	0.131	0.051

Details of the model are given in Table 6.10.

Table 6.25: Relative frequencies of the estimated number of breaks - Model has Separate breaks in SE and JE. $\Lambda = [0.15, 0.85]$

Sign	Breaks	$sup_{\lambda\in\Lambda} Wald_T$			$sup_{\lambda\in\Lambda} LM_T$			$sup_{\lambda\in\Lambda} \ddot{D}_T^U$			$sup_{\lambda\in\Lambda} \ddot{D}_T^R$		
Level	Est.	140	300	600	140	300	600	140	300	600	140	300	600
1%	$\hat{m}=0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{m}=1$	0.380	0.738	0.884	0.359	0.722	0.896	0.091	0.506	0.812	0.092	0.533	0.812
	$\hat{m}=2$	0.499	0.243	0.114	0.505	0.256	0.101	0.526	0.431	0.176	0.582	0.418	0.182
	$\hat{m}=3$	0.121	0.019	0.002	0.136	0.022	0.003	0.383	0.063	0.012	0.326	0.049	0.006
5%	$\hat{m}=0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{m}=1$	0.193	0.517	0.716	0.172	0.481	0.709	0.030	0.322	0.609	0.032	0.306	0.607
	$\hat{m}=2$	0.532	0.412	0.267	0.530	0.449	0.272	0.404	0.488	0.348	0.453	0.526	0.352
	$\hat{m}=3$	0.275	0.071	0.017	0.298	0.070	0.019	0.566	0.190	0.043	0.515	0.168	0.041
10%	$\hat{m}=0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{m}=1$	0.118	0.377	0.587	0.094	0.341	0.571	0.018	0.202	0.458	0.018	0.193	0.471
	$\hat{m}=2$	0.507	0.479	0.353	0.486	0.513	0.379	0.319	0.530	0.442	0.353	0.542	0.425
	$\hat{m}=3$	0.375	0.144	0.060	0.420	0.146	0.050	0.663	0.268	0.100	0.629	0.265	0.104

Details of the model are given in Table 6.11.

Table 6.26: Relative frequencies of the estimated number of breaks - Model has Separate breaks in SE and JE. $\Lambda = [0.20, 0.80]$

Sign	Breaks	$sup_{\lambda\in\Lambda} Wald_T$			$sup_{\lambda\in\Lambda} LM_T$			$sup_{\lambda\in\Lambda} \ddot{D}_T^U$			$sup_{\lambda\in\Lambda} \ddot{D}_T^R$		
Level	Est.	140	300	600	140	300	600	140	300	600	140	300	600
1%	$\hat{m}=0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{m}=1$	0.589	0.833	0.922	0.544	0.820	0.924	0.243	0.665	0.886	0.217	0.674	0.870
	$\hat{m}=2$	0.351	0.158	0.076	0.395	0.175	0.076	0.537	0.313	0.108	0.609	0.310	0.126
	$\hat{m}=3$	0.060	0.009	0.002	0.061	0.005	0.000	0.220	0.022	0.006	0.174	0.016	0.004
5%	$\hat{m}=0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{m}=1$	0.357	0.630	0.784	0.326	0.590	0.778	0.112	0.469	0.711	0.102	0.454	0.699
	$\hat{m}=2$	0.464	0.332	0.202	0.507	0.368	0.211	0.502	0.441	0.268	0.558	0.457	0.278
	$\hat{m}=3$	0.179	0.038	0.014	0.167	0.042	0.011	0.386	0.090	0.021	0.340	0.089	0.023
10%	$\hat{m}=0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{m}=1$	0.217	0.497	0.651	0.203	0.446	0.635	0.063	0.339	0.549	0.059	0.316	0.558
	$\hat{m}=2$	0.504	0.416	0.309	0.535	0.460	0.335	0.443	0.492	0.399	0.495	0.528	0.371
	$\hat{m}=3$	0.279	0.087	0.040	0.262	0.094	0.030	0.494	0.169	0.052	0.446	0.156	0.071

Details of the model are given in Table 6.12.

Table 6.27: Relative frequencies of the estimated number of breaks - Model has Separate breaks in SE and JE. $\Lambda = [0.15, 0.85]$

Sign	Breaks	$sup_{\lambda\in\Lambda} Wald_T$			$sup_{\lambda\in\Lambda} LM_T$			$sup_{\lambda\in\Lambda} \ddot{D}_T^U$			$sup_{\lambda\in\Lambda} \ddot{D}_T^R$		
Level	Est.	140	300	600	140	300	600	140	300	600	140	300	600
1%	$\hat{m}=0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{m}=1$	0.434	0.744	0.873	0.485	0.809	0.931	0.156	0.572	0.810	0.218	0.650	0.870
	$\hat{m}=2$	0.445	0.243	0.121	0.422	0.183	0.068	0.489	0.372	0.181	0.482	0.316	0.127
	$\hat{m}=3$	0.121	0.013	0.006	0.093	0.008	0.001	0.355	0.056	0.009	0.300	0.034	0.003
5%	$\hat{m}=0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{m}=1$	0.229	0.520	0.708	0.259	0.599	0.795	0.063	0.336	0.621	0.088	0.385	0.676
	$\hat{m}=2$	0.495	0.407	0.264	0.529	0.350	0.195	0.384	0.506	0.332	0.401	0.508	0.292
	$\hat{m}=3$	0.276	0.073	0.028	0.212	0.051	0.010	0.553	0.158	0.047	0.511	0.107	0.032
10%	$\hat{m}=0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{m}=1$	0.141	0.368	0.580	0.158	0.450	0.644	0.036	0.215	0.469	0.040	0.261	0.532
	$\hat{m}=2$	0.476	0.499	0.359	0.520	0.455	0.323	0.300	0.540	0.428	0.338	0.540	0.404
	$\hat{m}=3$	0.383	0.133	0.061	0.322	0.095	0.033	0.664	0.245	0.103	0.622	0.199	0.064

Details of the model are given in Table 6.13.

Table 6.28: Relative frequencies of the estimated number of breaks - Model has Separate breaks in SE and JE. $\Lambda = [0.20, 0.80]$

Sign	Breaks	$sup_{\lambda\in\Lambda} Wald_T$			$sup_{\lambda\in\Lambda} LM_T$			$sup_{\lambda\in\Lambda} \ddot{D}_T^U$			$sup_{\lambda\in\Lambda} \ddot{D}_T^R$		
Level	Est.	140	300	600	140	300	600	140	300	600	140	300	600
1%	$\hat{m}=0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{m}=1$	0.597	0.839	0.912	0.661	0.886	0.940	0.327	0.727	0.870	0.401	0.756	0.909
	$\hat{m}=2$	0.353	0.155	0.085	0.304	0.110	0.060	0.501	0.254	0.125	0.459	0.231	0.090
	$\hat{m}=3$	0.050	0.006	0.003	0.035	0.004	0.000	0.172	0.019	0.005	0.140	0.013	0.001
5%	$\hat{m}=0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{m}=1$	0.356	0.631	0.765	0.418	0.697	0.827	0.180	0.484	0.702	0.207	0.520	0.756
	$\hat{m}=2$	0.491	0.330	0.216	0.468	0.280	0.164	0.473	0.430	0.277	0.501	0.412	0.229
	$\hat{m}=3$	0.153	0.039	0.019	0.114	0.023	0.009	0.347	0.086	0.021	0.292	0.068	0.015
10%	$\hat{m}=0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{m}=1$	0.238	0.483	0.632	0.272	0.544	0.699	0.106	0.342	0.564	0.124	0.389	0.608
	$\hat{m}=2$	0.523	0.439	0.332	0.523	0.396	0.278	0.439	0.508	0.373	0.459	0.482	0.346
	$\hat{m}=3$	0.239	0.078	0.036	0.205	0.060	0.023	0.455	0.150	0.063	0.417	0.129	0.046

Details of the model are given in Table 6.14.

6.4 Conclusion

This chapter deals with the issue of determining the true number of break points in linear models using the method proposed in Bai (1997a). This approach combines hypothesis testing of parameter stability with the sequential estimation of the break point. The chapter also presents a review of some hypothesis tests for stability that exist in literature. Although Bai (1997a) propose the sup$F$-test which is suitable for OLS estimations, in this study however, we apply the three GMM-based *supremum*-type tests provided in Andrews (1993) which are based on the Wald, Lagrange Multiplier (LM) and Difference-type tests.

Through a series of Monte Carlo simulations, the finite sample performance of the three test statistics in detecting the true number of breaks in the model are examined. From the results obtained, this combined procedure generally yields the correct number of break points across all three test statistics for samples of sizes 300 and 600. The results for a sample size of 140 were a bit distorted and tended to overestimate the true number of breaks in the models mainly because the observations used in the estimations were too few, leading to spurious rejections of the test statistics. The way the test statistic is constructed by splitting the sample into smaller subsamples can result in few observations in a partition. Additionally, when the break point is close to the end points of the sample, the number of observations in a partition can be very few.

On the basis of the results obtained, the Wald and LM type tests performed comparably better than the Difference type test statistic. However, in Appendix E.1 on page 216, we present the results of the Monte Carlo simulation carried out at the 5% significance level to determine the size of the test when the sample size is increased to 1000. The results seen in Table E.1 show a similar convergence to 0.05 across all three test statistics when the whole sample S_0 is used. Consequently, as shown in Table E.2, the procedure correctly estimates no break in the model 94.7% and 94.9% of the time when using $sup_{\lambda\in\Lambda}Wald_T$ and $sup_{\lambda\in\Lambda}\ddot{D}_T^R$ respectively. Both $sup_{\lambda\in\Lambda}LM_T$ and $sup_{\lambda\in\Lambda}\ddot{D}_T^U$ indicate there is no break 94.8% of the time.

In the next chapter, we explore this combined testing and estimating procedure in an empirical setting using the New Keynesian Philips Curve used in Hall et al. (2012).

Appendix E

E.1 Monte Carlo simulation for sample size T=1000

Table E.1: Relative rejection frequencies of the test statistics - Model has no break in SE or JE. Sample Size T=1000.

	$sup_{\lambda\in\Lambda} Wald_T$			$sup_{\lambda\in\Lambda} LM_T$			$sup_{\lambda\in\Lambda} \ddot{D}_T^U$			$sup_{\lambda\in\Lambda} \ddot{D}_T^R$		
	S_0	S_1	S_2	S_0	S_1	S_2	S_0	S_1	S_2	S_0	S_1	S_2
$\Lambda = [0.15, 0.85]$	0.054	0.004	0.004	0.053	0.002	0.003	0.054	0.003	0.004	0.052	0.003	0.003
$\Lambda = [0.20, 0.80]$	0.053	0.001	0.002	0.052	0.001	0.001	0.052	0.002	0.001	0.051	0.001	0.000

Rejection frequency of $sup_{\lambda\in\Lambda} A_T$ is derived as the proportion of the number of times $sup_{\lambda\in\Lambda} A_T >$critical value.
The critical value at 5% significance level is 11.72.
S_0 uses the full sample; S_1 and S_2 are the subsamples before & after the first estimated break respectively.
The SE parameters are $\theta_0 = (1, 0.1)'$ and the JE parameters are $\delta_0 = (1, 0.5, 0.5, 0.5, 0.5)'$.

Table E.2: Relative frequencies of estimated number of breaks - Model has no break in the SE or JE. Sample Size T=1000.

Breaks	$sup_{\lambda\in\Lambda} Wald_T$		$sup_{\lambda\in\Lambda} LM_T$		$sup_{\lambda\in\Lambda} \ddot{D}_T^U$		$sup_{\lambda\in\Lambda} \ddot{D}_T^R$	
Est.	$\epsilon = 0.15$	$\epsilon = 0.20$	$\epsilon = 0.15$	$\epsilon = 0.20$	$\epsilon = 0.15$	$\epsilon = 0.20$	$\epsilon = 0.15$	$\epsilon = 0.20$
$\hat{m} = 0$	0.946	0.947	0.947	0.948	0.946	0.948	0.948	0.949
$\hat{m} = 1$	0.045	0.042	0.044	0.039	0.041	0.041	0.042	0.043
$\hat{m} = 2$	0.009	0.011	0.009	0.013	0.013	0.011	0.010	0.008
$\hat{m} = 3$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

$\hat{m} = j$, denotes j breaks were estimated in 1,000 simulations.

When $\epsilon = 0.15$, then $\Lambda = [0.15, 0.85]$ and when $\epsilon = 0.20$, $\Lambda = [0.20, 0.80]$.

Chapter 7

Empirical Application

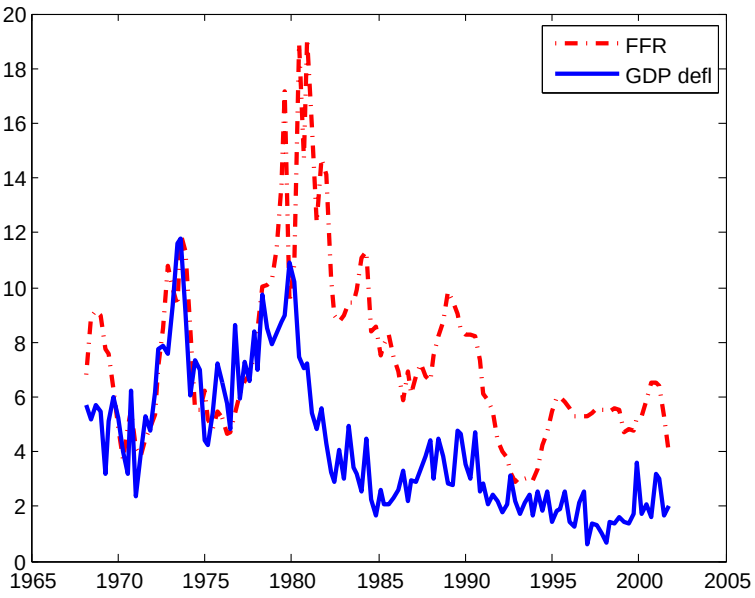
In the previous chapter, we presented a method of estimating the true number of break points in a linear model using simulated data in a series of Monte Carlo experiments. This was conducted by combining Andrews (1993) test for parameter stability and our proposed GMM break point estimation method. The results indicated that our proposed break point estimation technique was consistent for the true break points in the model. In this chapter, we investigate the performance of this combined testing and estimation method using real data. We adopt the New Keynesian Phillips Curve (NKPC) model used in Hall et al. (2012) since they found two of the endogenous regressors had unstable JE when estimating using Two Stage Least Squares (2SLS). It thus forms a good reference with which our proposed break point estimation approach can be compared.

The NKPC model, which is used by monetary policy makers as a good approximation of inflation dynamics, is based mainly on previous (backward-looking) and expected (forward-looking) inflation rates. The existing literature holds varied outlooks on the dominance of either of these inflation rates in determining the existing inflation rate. Although the focus of this research is not to form an opinion thereof, Gali and Gertler (1999), Rudd and Whelan (2005) and Gali et al. (2005) form noteworthy references on this issue.

Furthermore, there is an ongoing debate in the literature on the preferred approach to use when estimating the NKPC model - the traditional single equation estimation versus the system estimation methods. In the latter, additional equations for the interest rates and output gap are included when estimating the NKPC. Beyer et al. (2008) provides a stimulating discussion of these two estimation options.

A strong assumption when using the NKPC model is that it is structurally stable. It is on this basis that monetary policy makers take vital decisions. However, changes in

Figure 7.1: GDP Deflator Inflation and the Federal Funds Rate



monetary policy regimes could cause changes in economic variables which could lead to readjustments of the relationships between these variables in the NKPC model. For example, Figure 7.1 shows the annualised U.S. quarterly gross domestic product (GDP) deflator inflation rate and the US Federal Funds Rate (FFR) over the period 1968:3 to 2001:4. The GDP deflator inflation rate is used as a measure of inflation while the FFR is the interest rate which banks can lend to other institutions overnight. Figure 7.1 shows a reduction in the inflation rate (smooth line) towards the end of the 1970s by The Federal Reserve under it's chairman, Paul Volcker. The rate was further reduced by his successor, Alan Greenspan, in the late 1980s into the 1990s. The FFR (dash line) shows a similar pattern to the inflation rate throughout the period. A comparable structure is exhibited by the three month Treasury Bill rate (not shown). This behaviour indicates changes to monetary policy instruments occur in response to changes in inflation. The NKPC is thus subject to shifts and as it is a vital tool in monetary policy, it is essential to test for, and estimate any break points existing in the model.

The chapter is outlined as follows. Section 7.1 presents the model and its data. Section 7.2 and 7.3 discuss the break point estimation using 2SLS and GMM approaches respectively. The results are discussed alongside the two estimation approaches. Section 7.4 concludes.

7.1 The Model and Data

The NKPC model we adopt in this paper is identical to the model in Hall et al. (2012). It is referred to as an extended NKPC model introduced in Zhang et al. (2008) and Zhang et al. (2009). The presence of lagged values of the changes in inflation differentiates this NKPC model from the traditional NKPC models in the literature. The lags are included to account for any serial correlation in the errors. They state their extended model is sufficient to describe the empirical NKPC of the U.S. for the period under review. The model is given as

$$inf_t = c_0 + \alpha_f inf_{t+1|t} + \alpha_b inf_{t-1} + \alpha_{og} og_t + \sum_{i=1}^{3} \alpha_i \Delta inf_{t-i} + u_t, \tag{7.1}$$

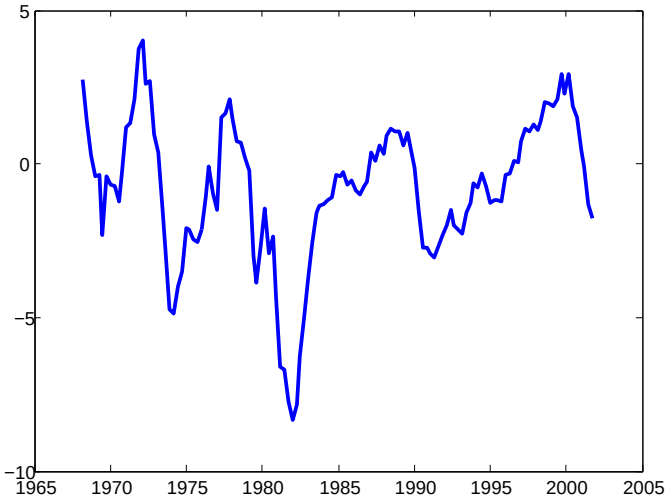
where inf_t is the measure of inflation obtained from the annualised quarterly growth rate of the GDP deflator obtained from the Federal Reserve Bank of St. Louis, c_0 is the core or underlying (steady-state) inflation rate, og_t is the output gap obtained from the estimates of real potential GDP as published by the Congressional Budget Office[1], $inf_{t+1|t}$ is the Greenbook one quarter ahead forecast of inflation[2] prepared at time t within the US Federal Reserve Bank of Philadelphia and Δinf_t is the lagged value of the change in inflation, obtained as $\Delta inf_t = inf_t - inf_{t-1}$. In the literature, α_f and α_b are referred to as the coefficients of forward- and backward-looking inflation respectively.

The output gap, og_t, is used in the model as a measurement of the marginal cost of firms in the economy and its behaviour over the period under review is displayed in Figure 7.2. The fall in the mid-1970s and early 1980s is evident in the plots.

[1] Zhang et al. (2009) also use the Hodrick-Prescott filter as a proxy for output gap.

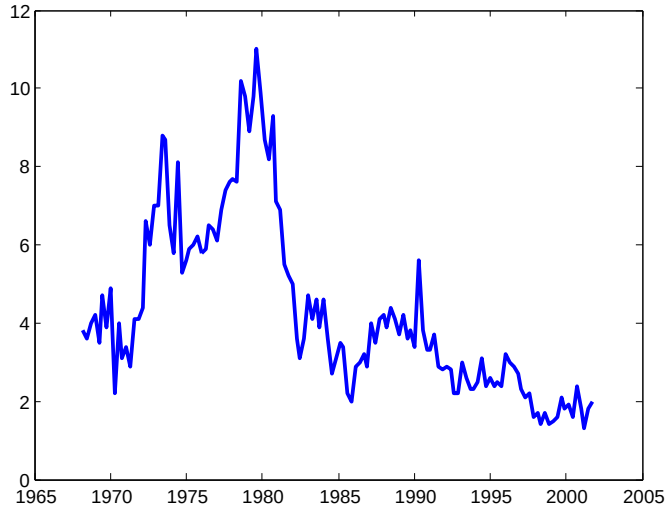
[2] Zhang et al. (2008) also employ three other measures of inflation forecasts prepared by the Survey of Professional Forecasters and the Michigan Survey.

Figure 7.2: Output gap (real potential GDP)



The Greenbook inflation forecasts, $inf_{t+1|t}$, is displayed in Figure 7.3. Although it is less volatile than the GDP deflator inflation rate, inf_t, seen earlier in Figure 7.1, it however exhibits a similar behaviour. Notice $inf_{t+1|t}$ tends to lag inf_t.

Figure 7.3: Greenbook One-quarter ahead forecasts



As highlighted in Zhang et al. (2009), since data used for the Greenbook inflation forecast is collected during the quarter, it may be influenced by current economic and financial activities, and hence, it can be classified as endogenous. Similarly, since demand shocks

may affect both the contemporaneous noise u_t and the current period real variable measured by the output gap, then the output gap can also be treated as endogenous[3]. As a result, Instrumental Variables (IV) need to be used in the estimations. For these IV to be suitable, they would have to be correlated with both $inf_{t+1|t}$ and og_t, but uncorrelated with the other variables in the NKPC given in (7.1). Similar to Hall et al. (2012), the instrument set z_t used in our analysis comprises of eight instruments: one lag each of the unemployment rate, short term interest rate (in this case, the three-month Treasury Bill rate) and the growth rate of the money aggregate M2, as well as the five exogenous variables in (7.1), that is, c_0, inf_{t-1} and $\sum_{i=1}^{3} \Delta inf_{t-i}$. All the data used in this empirical analysis are quarterly U.S. data spanning 1968.3 to 2001.4.

7.2 Estimating Break Points Using 2SLS

This section outlines the process of estimating break points using 2SLS estimation approach (within the scope of this NKPC model), as proposed in Hall et al. (2012).

Step 1: Test for break points in the Reduced Form[4] (RF) of the endogenous regressors given as,

$$inf_{t+1|t} = z_t^{'}\omega_1 + v_t \tag{7.2}$$

$$og_t = z_t^{'}\omega_2 + v_t \tag{7.3}$$

In this step, the commonly used method of testing for multiple break points given in Bai and Perron (1998) is used. Hall et al. (2012) allow for a maximum of 5 breaks and set $\epsilon = 0.1$, that is, the range of $\Lambda = [0.1, 0.9]$ and hence they search for breaks using the central 80% of the sample. Their results, as set out in Table 10 of Hall et al. (2012), clearly indicate a rejection of parameter stability for both $inf_{t+1|t}$ and og_t.

Step 2: Since there is evidence of parameter variation from Step 1, they adopt a sequential estimation strategy to determine the location of the break points. Two break points were estimated for $inf_{t+1|t}$ at 1975:2 and 1981:1, while only one break point was found at 1975:2 for og_t. In their strategy, it is the combination of all the break points obtained in the RF that is important, hence, they conclude there are two break points in the RF at $\{$1975:2, 1981:1$\}$.

[3]Additionally, Zhang et al. (2009) confirm orthogonality of $inf_{t+1|t}$ and og_t with u_t is rejected at conventional significance levels.

[4]This is analogous to the Jacobian Equation (JE) when using our GMM method.

Step 3: Conditioning on these two break points, the NKPC model given in (7.1) is split into three subsamples, each of which can now be said to have a stable RF. We define those subsamples as Sub1 for the period 1968:3 - 1975:2, Sub2 for the period 1975:3 - 1981:1 and Sub3 for the period 1981:2 - 2001:4.

Step 4: The three subsamples are then tested for any additional break points in the NKPC, which were not obtained in the RF estimations in Step 2. As Sub1 and Sub2 are quite small with less than 30 observations each, they test for only one break in each of them. For Sub3, a maximum of two breaks are allowed. The test results indicate no further breaks are in the NKPC model.

Step 5: Using the F-test and Wald test for their fixed break tests, they examine the NKPC model to confirm if the breaks obtained in the RF in Step 2 are also present. Their results indicate the two breaks are also present in the NKPC model[5].

With these findings, they conclude that the NKPC model given in (7.1) is unstable over the period examined having two break points at 1975:2 and 1981:1.

7.3 Estimating Break Points Using GMM

In this section, we present the estimation procedure as well as the results obtained when the GMM method proposed in this study is applied to an identical NKPC model. It is assumed that this model has two breaks in the Jacobian Equations (JE) of the endogenous regressors, $inf_{t+1|t}$ and og_t, as established in the previous section by Hall et al. (2012). The same data and set of instruments are maintained and a similar testing and estimation process as laid out in Chapter 6 is followed.

The results presented cover the three test statistics: the Wald, $sup_{\lambda\in\Lambda}Wald_T$; the Lagrange Multiplier, $sup_{\lambda\in\Lambda}LM_T$ and the two Difference-type test statistics, $sup_{\lambda\in\Lambda}\ddot{D}_T^U$ and $sup_{\lambda\in\Lambda}\ddot{D}_T^R$. Similar results were obtained for both $\epsilon = 0.15$ and $\epsilon = 0.20$, hence we report the results for only the former. As given in Table 1 in Andrews (1993), the critical values for a model with seven parameters are 26.23, 21.84 and 19.69 for 1%, 5% and 10% significance levels respectively. The steps and results (see Table 7.1) are discussed below.

Step 1: The whole sample is tested for evidence of parameter variation using Andrews (1993) discussed in the previous chapter. A test statistic of 41.80, 18.57, 41.77 and 20.81 was obtained for $sup_{\lambda\in\Lambda}Wald_T$, $sup_{\lambda\in\Lambda}LM_T$, $sup_{\lambda\in\Lambda}\ddot{D}_T^U$ and $sup_{\lambda\in\Lambda}\ddot{D}_T^R$, respectively.

[5]These types of breaks are similar to the Coincidental Breaks scenario under our GMM estimation approach.

There is a clear rejection of parameter stability for $sup_{\lambda\in\Lambda}Wald_T$ and $sup_{\lambda\in\Lambda}\ddot{D}_T^U$ at all levels of significance, indicating the presence of at least one break in the model. Using $sup_{\lambda\in\Lambda}\ddot{D}_T^R$, stability is rejected only at 10% significance level while no break point is detected in $sup_{\lambda\in\Lambda}LM_T$ at all levels.

Step 2: The Sequential Estimation Method proposed in this research as outlined in Section 3.2 is now applied to the whole sample, $[1, T]$, to obtain the first break fraction. The break fraction is estimated at the location of the *suprema* of the test statistics. The first break fraction occurs at $\lambda = 0.21$ for $sup_{\lambda\in\Lambda}Wald_T$ at all significance levels and $sup_{\lambda\in\Lambda}\ddot{D}_T^U$ and $sup_{\lambda\in\Lambda}\ddot{D}_T^R$ at 10% significance level. This corresponds to the first break date $\hat{k}_1$ in 1975:2. Note this first estimated break point is at an identical location to that obtained in Hall et al. (2012).

Step 3: The whole sample is now split into two subsamples $[1, \hat{k}_1]$ and $[\hat{k}_1, T]$ and Andrews (1993) test is repeated on them to determine their stability. However, from the location of $[\hat{k}_1, T]$, there are too few observations to carry out this test in the first subsample; in fact, there are less than thirty observations pre-1975:2. Considering the 15% trimming, the eight instruments as well as the seven coefficients being estimated, then Andrews (1993) test would suffer size distortions[6] and spurious rejections of stability will be recorded. Hence, only the second subsample is tested for an additional break point.

With a test statistic of 20.73, 24.70 and 21.69 obtained for $sup_{\lambda\in\Lambda}Wald_T$, $sup_{\lambda\in\Lambda}\ddot{D}_T^U$ and $sup_{\lambda\in\Lambda}\ddot{D}_T^R$, respectively, there is evidence the NKPC is statistically unstable in the post 1975 period at 10% significance level.

Step 4: The break point estimation procedure is applied to this second subsample to obtain the second break which occurs in 1988:3 for $sup_{\lambda\in\Lambda}Wald_T$ and 1981:1 for both $sup_{\lambda\in\Lambda}\ddot{D}_T^U$ and $sup_{\lambda\in\Lambda}\ddot{D}_T^R$. Again, the location of this second break in 1981:1 is identical to that in Hall et al. (2012). Thus, applying our proposed GMM estimation method directly to the NKPC model using the Difference tests yield similar break points at 1975:2 and 1981:1 as found in Hall et al. (2012). Using the Wald test also produced evidence of two breaks in the model, though one is at a different location to those estimated in the Difference tests. On the other hand, the LM test showed the NKPC model was stable over the period reviewed.

[6]This was also observed in the simulation experiment carried out for the small sample size in Chapter 6.

Table 7.1: Break points estimated for the NKPC model.

Sign level	$sup_{\lambda\in\Lambda} Wald_T$	$sup_{\lambda\in\Lambda} LM_T$	$sup_{\lambda\in\Lambda} \ddot{D}_T^U$	$sup_{\lambda\in\Lambda} \ddot{D}_T^R$
1%	1975:2	No Break	1975:2	No Break
5%	1975:2	No Break	1975:2	No Break
	No Break	No Break	1981:1	No Break
10%	1975:2	No Break	1975:2	1975:2
	1988:3	No Break	1981:1	1981:1

Estimation results for $\Lambda = [0.15, 0.85]$. Similar results obtained for $\epsilon = 0.20$.
Test stat of 41.80 and 18.57 were obtained for $sup_{\lambda\in\Lambda} Wald_T$ and $sup_{\lambda\in\Lambda} LM_T$ resp.
Test stat of 41.77 and 20.81 were obtained for $sup_{\lambda\in\Lambda} \ddot{D}_T^U$ and $sup_{\lambda\in\Lambda} \ddot{D}_T^R$ resp.
Critical values are 26.23, 21.84 and 19.69 for 1%, 5% and 10% s.l. resp.

7.4 Conclusion

This chapter investigates the performance of our proposed GMM break point estimation technique using the New Keynesian Phillips Curve model in Hall et al. (2012). The NKPC model uses U.S. data spanning from the third quarter of 1968 to the first quarter of 2001. Hall et al. (2012) identify two break points in the endogenous regressor, output gap and one break point in the inflation forecast. In their approach using 2SLS, the breaks in the JE are estimated before any additional breaks in the NKPC model. Using our GMM approach, however, the breaks in the NKPC model are estimated directly without the need to pre-estimate the JE. Similar to Hall et al. (2012), we also find the NKPC to be unstable over the period. Specifically, the results indicate there are two break points at 1975:2 and 1981:1 and potentially a third at 1988:3. Identical break locations of the first two breaks were found in Hall et al. (2012) using 2SLS. This indicates that in the presence of an Unstable JE, our proposed method is able to detect break points in the main SE, if they exist. Furthermore, the results from this empirical application support the theoretical analysis and Monte Carlo simulations carried out in previous chapters.

Chapter 8

Conclusion

This research provides a GMM-based method of estimating break points which occur at unknown locations in linear econometric models with endogenous regressors. When regressors are endogenous, that is when the regressors are correlated with the errors, OLS yields inconsistent estimators and hence common break point estimation tests[1] in the literature are inappropriate. For models with endogenous regressors, it is usual to use Instrumental Variables, of which Two Stage Least Squares (2SLS) and Generalised Method of Moments (GMM) are the two common choices in the literature. For the purpose of this research, we refer to the equation which sets out the relationship between the endogenous regressor and its instruments as the Jacobian Equation (JE). This JE can be stable or unstable and we examine both types of Jacobian structures in this research.

When the JE is stable, Hall et al. (2012) obtain inconsistent break fraction estimators using an approach that minimises the GMM objective function over all possible break partitions. They attribute the inconsistencies of the break fraction estimator to the structure of the GMM minimand which is essentially a square of sums; this allows the effects of any misspecification from the model to be offset within the minimand. Furthermore, Ghysels and Hall (1990) establish that the $J - test$ which is usually used in GMM literature as a diagnostic for model stability has no power in cases of parameter instability. The test is essentially a test of the overidentifying restrictions which are orthogonal to the identifying restrictions that are actually used in parameter estimation within the GMM moment conditions.

For an Unstable JE on the other hand, Hall et al. (2012) obtain consistent break fraction estimators using 2SLS by first identifying any break points in the JE before searching

[1] See Bai (1994a), Bai (1994b), Bai (1997a), Bai (1997b) and Bai and Perron (1998).

for additional break points in the Structural Equation (SE). They show this break fraction estimator is consistent and Boldea et al. (2012) go further to establish the limiting distributions of these estimators.

To our knowledge, there is no study conducted within the GMM framework to address the issue of estimating break points in models with endogenous regressors. Our study sets out to fill this gap. This study reveals that our GMM-based approach yields consistent break fraction estimators in linear models with endogenous regressors. Our proposed method focuses on parameter variation and it is based on the Wald, Lagrange Multiplier and Difference type test statistics. Additionally, we show that the results hold irrespective of a Stable or Unstable JE. We give a summary of the main findings below.

Chapters 2 and 3 examine models with one break and multiple breaks in the SE, respectively, assuming the JE is stable. Our proposed estimation strategy was discussed and the test statistic and break fraction estimators were examined. We present the theoretical analysis using the Difference type test while the Sequential Estimation Method is used in estimating the multiple breaks. The asymptotic properties of the break fraction estimators obtained from both models were established, specifically, the consistency and rate T-convergence were proved. Consequently, our GMM break point estimator has similar properties to those that have been shown for break point estimators in other settings such as a linear regression model with exogenous regressors estimated via OLS. Therefore, our approach of using a statistic that is centred on parameter variation offers an improvement to the GMM approach in Hall et al. (2012).

Chapter 4 examines the break fraction estimators obtained from models with an Unstable JE. Two types of models were explored. The first was a model with a break point in the JE only while the second had an additional break point in the SE. We showed that a break in the JE did not confound the estimations of a break in the SE, as our estimators were still consistent for the true break. Our estimation approach is applied directly to the SE, irrespective of a break in the JE or not. This is unlike the 2SLS method in Hall et al. (2012) where the break points in the JE had to be pre-estimated before those in the SE. This potentially saves computing time and the researcher can focus on the SE, which is the main interest. Some theoretical analyses still needs to be carried out to conclusively prove the asymptotic properties of the break fraction estimators obtained when the Jacobian is unstable, nonetheless, the extensive research carried out, as well as the results of the Monte Carlo simulation results presented support the conjecture made.

Chapter 5 presents results from a series of Monte Carlo simulations carried out to estimate the break points in both the Stable and Unstable Jacobian models. In the simulations,

we assume the number of break points are known and only their locations need to be estimated. Five different sample sizes were used in the estimations and we present a few simulation results from all the three test statistics used, though more focus is on the Difference type test on which the theoretical analysis is based. The results show all three tests behave similarly in finite samples and hence, indicates that the theoretical analysis carried out in this research may extend to them as well. It is worth noting that the results of the simulations support the theoretical analysis carried out in Chapters 2 to 4.

Chapter 6 provides a method of testing the significance of a break point, hence, we assume the true number of break points in the model is unknown. From the way the Sequential Estimation Method is constructed, there will always be a break point estimated which does not necessarily mean it is a genuine break. This, therefore, demands the derivation of a rule that terminates the process by determining if the break estimated is large enough to be a true break. This is done through hypothesis testing. We discuss some of the hypothesis tests existing in the literature in this chapter. For our purpose, we adopt the Andrews (1993) test for parameter instability and show that combining it with our estimation procedure yields the true number of break points in the model. However, in small sample sizes, the tests frequently overestimated the true number of breaks. A similar overestimation was found when the break points were close to the ends of the sample.

Lastly, Chapter 7 provides an empirical application of our proposed methodology. We test for parameter stability and estimate break points in the New Keynesian Phillips Curve (NKPC) model similar to that used in Hall et al. (2012). This model is suitable because Hall et al. (2012) established the presence of break points in the reduced form equations of the two endogenous regressors - output gap and inflation forecast. Hence, this model is a good reference to examine the performance of our proposed GMM estimation method. In our case, we do not need to estimate the Unstable JE prior to the SE as done in Hall et al. (2012), rather, the estimation method is applied directly to the NKPC model. Based on our results, the Difference tests yields identical break dates to those found in Hall et al. (2012). This result verifies our theoretical conclusions in an empirical setting. Thus, using our method in the presence of an Unstable JE produces consistent break point estimators of the main NKPC model and hence, provides an alternative break point estimation technique when researchers are faced with unstable models with endogenous regressors.

In our estimation approach proposed in this study, we assume the break points in the parameters in the SE occur at the same time. A similar assumption is made for the parameters in the JE. This is popularly referred to in the literature as a pure structural change. With some restrictions placed on these parameters, the analyses provided in this research

could be extended to the partial structural change case where only some of the parameters are allowed to change. This would potentially allow more observations for estimations.

Although we establish the consistency and rate T- convergence of the break fractions under the fixed break framework in the Stable Jacobian in Chapters 2 and 3, we do not provide the limiting distributions in the proofs. In the existing OLS literature on break point estimation, the limiting distributions obtained under the fixed break approach is dependent on the data, making it not useful for inference. To develop a useful approximation, it is necessary to resort to the shrinking break framework where the magnitude of parameter change converges to zero as the sample size grows. This is commonly used in the literature and results in a distribution that is invariant to the underlying distribution of the data. Although this framework is designed for small shifts, the resulting distribution is used as an approximation for moderate and large shifts. Additionally, while the preliminary studies carried out as well as the Monte Carlo simulation results are encouraging, the asymptotic properties of the break fraction estimators obtained from the Unstable Jacobian models have to be determined within the shrinking breaks approach as the fixed break framework proved intractable. Going the shrinking breaks route will circumvent the problems encountered with the fixed break approach, as well as establish a limiting distribution theory that can be useful as a basis for inference. This is an interesting area that will merit future research.

Bibliography

Altissimo, F. and Corradi, V. (2003). Strong rules for detecting the number of breaks in a time series, *Journal of Econometrics* **117**(2): 207 – 244.

Andreou, E. and Ghysels, E. (2002). Detecting multiple breaks in financial market volatility dynamics, *Journal of Applied Econometrics* **17**(5): 579–600.

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

Andrews, D. and Fair, R. (1988). Inference in Nonlinear Econometric Models with Structural Change, *The Review of Economic Studies* **55**: 615–640.

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

Antoshin, S., Berg, A. and Souto, M. (2008). Testing for Structural Breaks in Small Samples, *IMF Working Paper,* http://www.imf.org/external/pubs/ft/wp/2008/wp0875.pdf .

Augustine-Ohwo, O. (2012). *Testing for Parameter Variation at a Single Unknown Breakpoint*, Master's thesis, School of Social Sciences, University of Manchester.

Bai, J. (1994a). Estimation of Structural Change Based On Wald Type Statistics, *Working Paper, Department of Economics, MIT, Cambridge, Massachusetts* pp. 94–96.

Bai, J. (1994b). Least Squares Estimation of a Shift in Linear Processes, *Journal of Time Series Analysis* **15**(5): 453–472.

Bai, J. (1997a). Estimating Multiple Breaks One at a Time, *Econometric Theory* **13**(3): 315–352.

Bai, J. (1997b). Estimation of a change point in multiple regression models, *Review of Economics and Statistics* **79**(4): 551 – 563.

Bai, J. (1999). Likelihood ratio tests for multiple structural changes, *Journal of Econometrics* **91**(2): 299 – 323.

Bai, J. and Perron, P. (1998). Estimating and Testing Linear Models with Multiple Structural Changes, *Econometrica* **66**: 47–78.

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

Bai, J. and Perron, P. (2006). *Multiple Structural Change Models: A Simulation Analysis*, Cambridge University Press, New York. in D. Corbae, S. N. Durlauf and B. E. Hansen (eds) pages 212–237.

Banerjee, A., Lazarova, S. and Urga, G. (2002). Bootstrapping Sequential Tests for Multiple Structural Breaks. Available at https://www.researchgate.net/publication/2591068_Bootstrapping_Sequential_Tests_For_Multiple_Structural_Breaks.

Barry, D. and Hartigan, J. A. (1993). A Bayesian Analysis for Change Point Problems, *Journal of the American Statistical Association* **88**(421): 309–319.

Beyer, A., Farmer, R. E. A., Henry, J. and Marcellino, M. (2008). Factor Analysis in a Model with Rational Expectations, *Econometrics Journal* **11**: 271–286.

Bhattacharya, P. (1987). Maximum likelihood estimation of a change-point in the distribution of independent random variables: General multiparameter case, *Journal of Multivariate Analysis* **23**(2): 183 – 208.

Boldea, O. and Hall, A. (2013). Estimation and Inference in Unstable Nonlinear Least Squares Models, *Journal of Econometrics* **172**(1).

Boldea, O., Hall, A. and Han, S. (2012). Asymptotic Distribution Theory for Break Point Estimators in Models Estimated via 2sls, *Econometric Reviews* **31**: 1–33.

Brown, R., Durbin, J. and Evans, J. (1975). Techniques for Testing the Constancy of Regression Relationsips over Time, *Journal of the Royal Statistical Society* **37**(2): 149–192.

Caceres, C., Corbacho, A. and Medina, L. (2010). Structural Breaks in Fiscal Performance: Did Fiscal Responsibility Laws have anything to do with Them?, *IMF Working Paper,* https://www.imf.org/external/pubs/ft/wp/2010/wp10248.pdf .

Chernoff, H. and Zacks, S. (1964). Estimating the current mean of a normal distribution which is subjected to changes in time, *The Annals of Mathematical Statistics* **35**(3): 999–1018.

Chong, T. T.-L. (2001). Estimating the Locations and Number of Change points by the Sample-splitting Method, *Statistical Papers* **42**: 53–79.

Chow, G. (1960). Tests of Equality between sets of Coefficients in two linear regressions, *Econometrica* **28**(3): 591–605.

Diebold, F. and Chen, C. (1996). Testing structural stability with endogenous breakpoint a size comparison of analytic and bootstrap procedures, *Journal of Econometrics* **70**(1): 221–241.

Dunn, K. B. and Singleton, K. J. (1986). Modeling the term structure of interest rates under nonseparable utility and durability of goods, *Journal of Financial Economics* **17**: 27 – 55.

Elliot, G. and Müller, U. K. (2014). Pre and Post Break Parameter Inference, *Journal of Econometrics* **180**: 141–157.

Erdman, C. and Emerson, J. W. (2008). A fast Bayesian change point analysis for the segmentation of microarray data, *Bioinformatics* **24**(19): 2143–2148. Available at http://bioinformatics.oxfordjournals.org/content/24/19/2143.full.pdf.

Gagliardini, P., Trojani, F. and Urga, G. (2005). Robust GMM Tests for Structural Breaks, *Journal of Econometrics* **129**: 139–182.

Gali, J. and Gertler, M. (1999). Inflation Dynamics: A Structural Econometric Analysis, *Journal of Monetary Economics* **44**(2): 195–222.

Gali, J., Gertler, M. and López-Salido, D. (2005). Robustness of the estimates of the hybrid new keynesian phillips curve, *Journal of Monetary Economics* **52**(6): 1107–1118.

Garcia, R. and Perron, P. (1996). An Analysis Of The Real Iinterest Rate Under Regime Shifts, *The Review of Economics and Statistics* **78**(1): 111–125.

Ghysels, E., Guay, A. and Hall, A. (1997). Predictive Tests for Structural Change with Unknown Breakpoint, *Journal of Econometrics* **82**: 209–233.

Ghysels, E. and Hall, A. (1990). Are consumption-based intertemporal capital asset pricing models structural?, *Journal of Econometrics* **45**: 121–139.

Hájek, J. and Rényi, A. (1955). Generalisation of an Inequality of Kolmogorov, *Acta Mathematica Academiae Scientiarum Hungarica* **6**: 281–283.

Hall, A. (2005). *Generalized Method of Moments*, Oxford University Press, New York.

Hall, A., Han, S. and Boldea, O. (2012). Inference Regarding Multiple Structural Changes in Linear Models with Endogenous Regressors, *Journal of Econometrics* **170**: 281–302.

Hall, A., Osborn, D. and Sakkas, N. (2013). Inference On Structural Breaks Using Information Criteria. University of Bath. Available at http://opus.bath.ac.uk/36583/5/Published_version.pdf.

Hall, A. R., Osborn, D. R. and Sakkas, N. (2015). Structural Break Inference Using Information Criteria in Models Estimated by Two-Stage Least Squares, *Journal of Time Series Analysis* **36**(5): 741–762.

Hall, A. and Sen, A. (1999). Structural Stability Testing in Models Estimated by Generalized Method of Moments, *Journal of Business and Economic Statistics* **17**(3): pp. 335–348.

Hansen, B. (1992). Testing for Parameter Instability in Linear Models, *Journal of Policy Modeling* **14**(4): 517–533.

Hansen, B. E. (2001). The new econometrics of structural change: Dating breaks in u.s. labour productivity, *Journal of Economic Perspectives* **15**(4): 117–128.

Hansen, L. (1982). Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica* **50**(4).

Hansen, L. and Singleton, K. (1982). Generalized instrumental variables estimation of nonlinear rational expectations models, *Econometrica* **50**: 1269–1286.

Hinkley, D. V. and Hinkley, E. A. (1970). Inference about the change-point in a sequence of binomial variables, *Biometrika* **57**(3): 477–488.

Jouini, J. and Boutahar, M. (2005). Evidence On Structural Changes in U.S Time Series, *Economic Modelling* **22**: 391–422.

Li, H. and Müller, U. K. (2009). Valid inference in partially unstable generalized method of moments models, *Review of Economic Studies* **76**(1): 343–365.

Loschi, R. H., Moura, C. R. and Iglesias, P. L. (2005). Bayesian analysis for change points in the volatility of latin american emerging markets, *Journal of Data Science* **3**: 101–122.

Lucas, R. (1976). Econometric policy evaluation: A critique, *Carnegie-Rochester Conference Series on Public Policy* **1**(1): 19–46.

Nyblom, J. (1989). Testing for the Constancy of Parameters Over Time, *Journal of the American Statistical Association* **84**(405): pp. 223–230.

Perron, P. and Qu, Z. (2006). Estimating restricted structural change models, *Journal of Econometrics* **134**(2): 373 – 399.

Perron, P. and Yamamoto, Y. (2015). Using OLS to Estimate and Test for Structural Changes in Models with Endogenous Regressors, *Journal of Applied Econometrics* **30**(1).

Quandt, R. E. (1958). The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes, *Journal of The American Statistical Association* **53**: 873–880.

Quandt, R. E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes, *Journal of The American Statistical Association* **55**(290): 324–330.

Rudd, J. and Whelan, K. (2005). New Tests of the New-Keynesian Phillips Curve, *Journal of Monetary Economics* **52**(6): 1167–1181.

Sowell, F. (1996). Optimal tests for parameter instability in the generalized method of moments framework, *Econometrica* **64**(5): 1085–1107.

Verbeek, M. (2008). *A Guide to Modern Econometrics*, third edition edn, John Wiley and Sons, Ltd, Chichester, West Sussex, England.

Wang, Z. (2007). *A Study on the Hájek-Rényi Inequality and its Applications*, PhD thesis, Faculty of Graduate Studies and Research, University of Regina, Saskatchewan, Canada, available at: URL http://stat.math.uregina.ca/∼kozdron/Research/Students/Wang.pdf.

White, H. (2001). *Asymptotic Theory for Econometricians*, Emerald Group Publishing Limited, Great Britain.

Wooldridge, J. (2009). *Introductory Econometrics*, fourth edition edn, South Western, Cengage Learning, Ohio.

Wright, P. G. (1928). *The tariff on animal and vegetable oils*, The Macmillan company, New York.

Yao, Y.-C. (1988). Estimating the Number of Change-points via Schwarz' Criterion, *Statistics and Probability Letters* **6**: 181–189.

Zhang, C., Osborn, D. and Kim, D. (2008). The New Keynesian Phillips Curve: from sticky inflation to sticky prices, *Journal of Money, Credit and Banking* **40**(4): 667699.

Zhang, C., Osborn, D. and Kim, D. (2009). Observed Inflation Forecasts and the New Keynesian Phillips Curve, *Oxford Bulletin of Economics and Statistics* **71**(3): 375398.

Zhang, N. R. and Siegmund, D. O. (2007). A modified bayes information criterion with applications to the analysis of comparative genomic hybridization data, *Biometrics* **63**(1): 22–32.