ABSTRACT

BOLDEA, OTILIA. Estimation and Inference in Unstable Nonlinear Least Squares Models. (Under the direction of Dr. Alastair R. Hall).

In this thesis, we extend Bai and Perron's (1998, Econometrica, pp. 47-78) method for detecting multiple breaks to nonlinear models. To that end, we consider an unstable univariate nonlinear least squares (NLS) model with a limited number of parameter shifts occurring at unknown dates. In our framework, the break-dates are simultaneously estimated with the parameters via minimization of the residual sum of squares. Using nonlinear asymptotic theory, we derive the asymptotic distributions of both break-point and parameter estimates and propose several instability tests. We also present simulation results that validate our procedure. Our method is useful for estimating and testing nonlinear macroeconomic models with multiple unknown breaks. As an empirical illustration, we explore the relationship between our model and smooth transition models in the context of a US interest rate reaction function. Unlike previous studies, our model can nest nonlinearities and breaks. We provide evidence for at least two breaks while allowing for smooth transition within each regime, before and after a break.

Estimation and Inference in Unstable Nonlinear Least Squares Models

by
Otilia Boldea

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Economics

Raleigh, North Carolina

2008

APPROVED BY:

Dr. Alastair R. Hall
Chair of Advisory Committee

Dr. David A. Dickey

Dr. Denis Pelletier

Dr. John J. Seater

Dedication

To my birth mother, Aurelia, even though she is no longer with us, and to my father, Ion, for his dedication

Biography

Otilia Boldea was born in Timisoara, Romania, on April 25, 1979. She received a Bachelor in Finance from the West University of Timisoara in 2001. In parallel, starting 1999, she attended University of Saarland, Germany, where she received a Masters in Business and Administration in 2002. She joined the Department of Economics for pursuing a Doctorate Degree in January 2003.

Acknowledgements

I would like to express my sincere gratitude to Dr. Alastair Hall, my academic advisor. Without his patience, support, guidance, technical and editorial advice, the completion of this endeavor would not have been possible. I would like to thank him also for inspiring me and for teaching me how to present and write articulately.

I would like to acknowledge the tremendous support of Dr. Bibhuti Bhattacharyya and Dr. David Dickey, as well as their professional enthusiasm that gave me wings during this journey through time series econometrics. Dr. Atsushi Inoue I thank for his valuable comments and for always being available for questions and sudden drop-ins into his office.

I cannot forget Dr. John Seater, whom I want to thank from the bottom of my heart for his valuable support not only with my thesis, but throughout my doctoral years. I really started to like economics only when I took his first class. I really started to understand it by his second class. For his collaboration, friendliness and professional, career and life advice I consider him as a second father to me. Had it not been for him, I would not have loved my department and my work so much.

I would like to thank my good friends Xenia, Berna and Eren for their steady moral support throughout this journey. I would also like to deeply thank my dear friend Vangelis for being there for me at the writing stage of my thesis and for his help with the computer codes.

I am grateful to my mother, Ana-Maria, for her dedication and kindness that helped me through hard moments, and my elder sister, Monica. My little sister, Ioana, receives my deepest thanks, too, for being my good friend throughout. My uncle, Bujor, and my cousin, Serban, have been of great help to me during these years.

Last, but not least, I would like to express my deepest gratitude to my father for

believing that I can do this and much more. He has been my one true example in life. Had it not been for his moral, professional and financial support, I would not be the person I am today, nor would I have come so far.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The issue of structural instability has received much attention from economists in the last few decades. Because of events such as the Great Depression, oil price shocks, the Great Moderation, changes in policy regimes, the assumption that models are stable may fail. Over long time spans, a stable model might not be the appropriate tool to capture the features of economic decisions. When a regime change occurs, the decision model used by economic agents before the change may no longer be valid after the change. Instability can then be thought of as an abrupt or constant change in the whole model, as a smooth transition from one model to another, maybe of different functional form or as a switch back-and-forth between models.

A special case of instability is that of parameter change, without the entire model specification changing. One way economists came to think about parameter change originates in Lucas' critique - Lucas (1976). Policy shifts affect the expectations of aggregate variables; when decisions are made based on statistical models that embed expectations, the latter have already changed, causing the statistical relationships to break down. An empirical study by Bai (1997) shows that the parameter that quantifies the response of the interest rate to fluctuations in the discount rate changes at the same time the Fed's operating procedures change, confirming Lucas' argument. Another related strand of literature dealing with parameter change originates in the debate of unit roots versus trends in macroeconomic time series. Nelson and Plosser (1982) found that the secular component of several macroeconomic time series is a unit root. This created unrest among business cycle theorists, who interpreted the results as evidence that random shocks have a permanent effect on the economy. In a seminal paper, Perron (1989) argued that the effects are temporary and represented the shocks as breaks in underlying deterministic trends. He found that if a sudden change in parameter is present (following and followed by a period of parameter constancy), but unaccounted for, it would yield to erroneous non-rejection of the unit root hypothesis for 11 out of the 14 time series considered by Nelson and Plosser (1982). Subsequently, his work was criticized by *inter alia* Zivot and Andrews (1992), who pointed out that the choice of change-point in Perron (1989) was influenced by prior examination of the data, invalidating the distributional properties of the unit root tests used. This critique equally applied to empirical studies using the Chow (1960) test, an F-test of stability in linear stationary models when the break-point is known. All arguments were in favor of viewing the break-point as endogenous.

For the statistical literature, this was not news. Pioneering work of Quandt (1958) dealt with estimating an unknown break in parameters of a linear model by means of maximum likelihood. He also proposed an F-test for equality of parameters Quandt (1960). Extensions to his work are collected in statistical literature surveys by Zacks (1983), Krishnaiah and Miao (1988), Bhattacharya (1994) and Csörgö and Horváth (1997).

The econometric literature on structural change evolved a great deal in the last three decades - see surveys by Dufour and Ghysels (1996); Banerjee and Urga (2005). It can be grouped in various ways: with focus on inference or estimation, considering linear or non-linear, univariate or multivariate, parametric or nonparametric, classical or Bayesian models, with stationary or non-stationary regressors, with single break or multiple breaks, known or unknown, with instability in parameters or of a more general kind, with focus on *a posteriori* detection of breaks or prediction. Unless otherwise specified, we will focus on sudden shifts in parameters as the main source of instability.

Regarding inference for one known break, the earliest tests for structural stability can be dated back to Chow (1960). Several authors, including Toyoda (1974) and Phillips and McCabe (1983), have demonstrated that the power of Chow's F-test is sensitive to heteroskedasticity. Since the variance of error may change at the same time with the parameter vector, they proposed modifying the test such that the errors have variance σ_1^2 before and σ_2^2 after the break. An easier and more general approach is the MacKinnon (1989) test, because it is robust to heteroskedasticity of unknown form. Lo and Newey (1985) extended Chow's test to linear simultaneous equations. Andrews and Fair (1988) provided Wald, Lagrange Multiplier (LM) and Likelihood Ratio (LR)-tests for a known break in a variety of nonlinear models, including nonlinear least squares (NLS), maximum likelihood (ML), and Maximum (M) estimators.

Since Quandt (1960), various tests deal with the issue of one unknown break. Brown, Durbin, and Evans (1975) emphasize the use of graphical methods in decisions based on Quandt's test, a test whose distribution was unknown at that time. When the regressors are independent of the errors, they propose using recursive least-squares (LS) residuals to compute cumulative sum (CUSUM) and CUSUM-of-squares tests. This test is generalized to dynamic models by Krämer, Ploberger, and Alt (1988). The CUSUM of squares test based on recursive residuals was shown to equally apply to ordinary least-squares (OLS) residuals by Harrison and McCabe (1979). Ploberger and Krämer (1992) show that the CUSUM test can also be performed in case of dependent and heteroskedastic errors. A drawback of the CUSUM test is that it lacks power when testing for stability of a model's intercept but not of other parameters (see Maddala and Kim, 1998, pp. 393). As a remedy, Ploberger, Krämer, and Kontrus (1989) propose using a test based on successive OLS parameter estimates, test that was initially suggested by Sen (1980) for slope parameters. Simulations studies show that none of the variants of the CUSUM test or the fluctuation test is superior to the other.

All the above tests consider unknown changes in mean parameters. Several financial series have constant mean, but the variance (volatility) changes. It is tempting to think of squared financial returns as ARMA processes and apply CUSUM tests such as the ones mentioned above. However, they are more suitably expressed as generalized autoregressive conditional heteroskedasticy models (GARCH) and the resemblance between ARMA and GARCH is deceiving. Moreover, Carrasco and Chen (2000) show that GARCH processes are β-mixing, requiring weaker conditions than the tests mentioned above can handle. Among others, Kokoszka and Leipus (1998, 2000) provide CUSUM-type tests for linear models under these weaker conditions. Least-Squares (LS) type tests that apply for nonlinear models, as well as for weak mixing conditions, are provided by Lavielle and Moulines (2000). Andreou and Ghysels (2002) provide extensions of both Kokoszka and Leipus's and Lavielle and Moulines's tests to high frequency data and multivariate volatility models.

Andrews (1993) redirects attention to Quandt's (1960) LR test. He derives its limiting distribution and proposes other sequential tests, such as supremum (sup) Wald and LM, valid for more general classes of nonlinear models than the ones Quandt (1960) considered. The latter tests implicitly use an estimate of the unknown change-point, at the value where the sequence of test statistics is maximized. Andrews and Ploberger (1994) suggest using weighting schemes for test statistics sequences, assigning equal or different probabilities to each break-point candidate. The tests are known today as average (ave) and exponential (exp) tests. Using Andrews's (1994) interpretation, if the weighting densities of the ave and exp tests are viewed as prior distributions over the space of local alternatives, then the optimal test statistics can be viewed as Bayesian posterior odds ratios. Although all tests proposed by Andrews and Ploberger (1994) are locally optimal tests, they are only valid for ML methods. Sowell (1996) extends these tests to GMM and introduces new weighting densities to direct the power of the tests to a specific location in the sample. Hansen (2000) points out that the marginal distribution of the regressors may experience parameter changes at the same time the break occurs, invalidating the asymptotic properties of the sup, ave and exp Wald, LM and LR tests. He suggests using a fixed regressor bootstrap to solve the size problems of these tests. Ongoing work by Kim and Perron (2007b) assesses the relative asymptotic efficiency of all tests described above. They find that sup and exp Wald tests are superior to all other tests when using an approximate Bahadur slope criterion for comparison.[1] They also argue for globally rather than locally optimal tests.

All tests presented have power against one-time parameter variation of unknown form. How-

[1]For details on the Bahadur slope as a criterion for comparing tests, see Bahadur (1960).

ever, it is plausible that, under the alternative, parameters follow a certain predefined process, such as a random walk. Nyblom and Mäkeläinen (1983) derive locally most powerful sup F-tests against this alternative. Nyblom (1989) provides locally most powerful tests against parameter variation in the form of martingales. This specification comprises single jumps at unknown times and random walks. Another possibility is that the parameter switches back and forth between regimes. Goldfeld and Quandt (1973) introduced a Markov Switching Regression (MSR). Each period, there are transition probabilities assigned to each regime, and the transition is gradual. The time of the switch may be known or hidden and there is a large body of literature dealing with MSR estimation and inference (see Wang, 2003, pp. 82-98). An even smoother transition between regimes can be modeled by means of smooth transition autoregressive processes, usually denoted STAR (see Granger and Teräsvirta, 1993, pp. 39-41; Fan and Yao, 2005, pp. 125-142) or, alternatively, in a Bayesian framework (see Broemeling and Tsurumi, 1987, pp. 136-189; Maddala and Kim, 1998, pp. 402-406). STAR became a popular modeling tool for economic time series due to its ability to explain nonlinearity and regime change by means of state variables - see van Dijk, Teräsvirta, and Franses (2002).

From parameter variation, the focus shifted to more general model misspecification with papers such as Ghysels and Hall (1990) and Ghysels, Guay, and Hall (1998). They propose out-of-sample predictive tests for GMM, assessing whether estimates in one sub-sample are useful for predicting the second. Their asymptotic results suggest that the predictive test can be decomposed into a parameter variation test and test of the validity of over-identifying restrictions in one of the sub-samples. This test works if both sub-samples are asymptotically large. This caveat is overcome by Dufour, Ghysels, and Hall's (1994) more general predictive tests and Andrews's (2003) end-of-sample tests. The disadvantage of the predictive tests is that they compute the OIR test in only one sub-sample. Within a GMM framework, Hall and Sen (1999) show that it is desirable to use the OIR test in both subsamples, in conjunction with stability tests, rather than the predictive tests alone.

More often that not, neglected instability causes series to appear to exhibit long-memory.[2] It is therefore desirable to have tests of unit roots or long memory against breaks as well as stability tests with nonstationary regressors under both the null and the alternative hypotheses. Papers for tests of unit root against a break in trend include *interalia* Banerjee, Lumsdaine, and Stock (1992), Zivot and Andrews (1992), Chu and White (1992) and Hansen (2000). Sequential CUSUM tests of change in mean against a unit root are introduced by Aue, Horváth, and Horváth (2006). Perron (1989) proposes modified Dickey-Fuller (DF) tests for unit roots when a break in trend is present under both the null and alternative hypothesis. These tests are carried under the assumption of a known break. Work in progress by Kim and Perron (2007a) develops unit root tests under the assumption of an endogenous break. However, not all economic processes are unit roots. In fact, several financial series are suitably modeled as near unit root processes.[3] As is true for unit roots, near unit roots can be easily confused with structural change, too. For example, Diebold and Inoue (2001) and Granger and Hyung (2004) show how a stochastic regime switching process, respectively a linear process with breaks can mimic the properties of long memory processes. As shown by Wright (1998) and Krämer and Sibbertsen (2002), the sup Wald and CUSUM tests in unstable polynomial regressions, respectively in linear models, can easily mistake long memory for structural change when standard

[2] For example, see Hillebrand (2005).
[3] They are also denoted as long memory processes or fractionally integrated processes.

critical values are employed. Both authors derive the distribution of these tests under the null hypothesis of no structural change. Most of these tests are of long memory against structural change (typically against switching regimes). In contrast, Lazarová (2005) derives a test for structural change that allows the regressors and errors to be fractionally integrated time series under both the null and alternative hypothesis. There is still an ongoing debate about instability and long memory detection.

As shown by the numerous strands of literature mentioned above, the papers on testing for one unknown break are quite numerous. In contrast, the issue of estimating at least one unknown break has received considerable less attention. Within linear parametric classical models, work by Quandt (1958) shows that picking the supremum over the maximum likelihood function for each potential break yields a consistent estimate of the true change point. Bai (1994, 1995) obtains the same results for least squares and least absolute deviation methods. Furthermore, he derives the finite sample and the limiting distribution of the change-point under different assumptions about the magnitude of parameter shifts.

Although most tests and estimation methods allow for only one unknown change, there is always the possibility of multiple breaks. This motivated researchers to develop methods for estimating their location when unknown. Yao and Au (1989) show that multiple unknown shifts in mean can be dated and estimated consistently by OLS the same way one shift can. Subsequently, Bai (1997) proposed a sequential OLS method to estimate multiple breaks one at a time. Perhaps the most comprehensive study in linear parametric models is that of Bai and Perron (1998). For general classes of heteroskedastic dynamic unstable linear time series models, they show that minimizing OLS over all possible partitions of the sample yields consistent estimates of the break points and estimates. They provide sup F-tests for selecting the number of breaks. They also propose a sequential procedure for selecting the number of break points in the sample based on various F-statistics for parameter constancy. The methods are extended to multivariate models by Perron and Qu (2006), restricted univariate models by Perron and Qu (2007), and two stage least squares models by Hall, Han, and Boldea (2007). Another method of estimation for general linear time series models is minimum description length (see Davis, Lee and Rodriguez-Yam, 2006). This method is in fact a selection criterion in which the change-point locations, the parameters and the number of break points are simultaneously determined by minimizing the length of a coded regressor.[4]

Estimating the break-points can be done in a classical or Bayesian framework. In a Bayesian model, Barry and Hartigan (1993) assume that there is an underlying sequence of parameters partitioned into contiguous blocks of equal parameter values, and that the beginning of each block is a change-point. They provide approximation techniques, based on Markov sampling, that greatly reduces computational time in large samples. Building on this technique, Chib (1998) specifies a latent discrete variable indicative of the regime from which a particular observation has been drawn. Each regime is modeled as a discrete-time discrete-state Markov process with transition probabilities constrained in such a way that the state variable can either stay in the current value or jump to the next higher value. The author points out that different change-point models can be viewed as nonnested, and propose using Bayes factors (ratio of marginal likelihoods) to choose between competing models. As in classical models,

[4] For details on this procedure, see Rissanen (1989).

the model above can predict the new value for the parameters. However, unless a model for when the next break will occur is added, the break cannot be predicted and this criticism applies equally to classical models. Not that all the models described above are appropriate for significant discrete jumps. For continuous changes, nonparametric estimation methods can be employed. The interested reader is referred to Brodsky and Darkovsky (1993) for details.

We are interested in a discrete, small number of historical macroeconomic changes. It is often the case that macroeconomists prefer parametric models, hence we will remain in a classical parametric setting. As can be noted from above, all papers within this class of models focus on linear estimation and inference. If the model is nonlinear (in parameters), to our knowledge, only tests for a single break or at most two breaks, without estimation results, are employed. But several economic models are nonlinear and may experience multiple shifts. Our goal is to provide a comprehensive treatment of estimation and inference issues in a class of nonlinear models with multiple unknown breaks that can be estimated via nonlinear least squares (NLS). In the spirit of Bai and Perron's (1998) paper, we propose obtaining estimates of multiple unknown breaks and parameters by minimizing the sum of squared residuals over all possible change-points. Using nonlinear asymptotics and empirical process theory, we show that the resulting estimates are consistent and we derive their asymptotic distributions. Like in linear models, we show that the distributions of the parameter estimates are the same as if the break-points were known. Under certain assumptions, the change-point estimators follow a two-sided random walk, and their asymptotic distribution is a two-sided Gaussian stochastic process. We propose sup F-tests to detect the number of change points when unknown and provide simulations that validate our procedures. Since our procedure nests nonlinearities and breaks, we explore the relationship between the (nonlinear) STAR models and our model in the context of a US interest rate function with one or two state variables driving the transitions. We find that, in addition to smooth transition between regimes, a nonlinearity caused by three month differences in the interest rate hitting a certain threshold, there is evidence of instability for at least two locations in our sample. Our model seems to fit the data quite well.

The rest of the thesis is organized as follows. Chapter 2 presents a brief review of linear asymptotics in unstable models and nonlinear asymptotic theory in stable models. Chapter 3 presents our model with the main assumptions maintained throughout the thesis. It shows consistency of the break point estimates and parameters. Chapter 4 derives the asymptotic distributions of parameter estimates. Chapter 5 is devoted to the distributional properties of the break point estimators under different magnitudes of parameter shifts. Stability tests are proposed in Chapter 6. Simulation results are the subject of Chapter 7. In Chapter 8 we explore the relationship between our model and STAR models, and we provide our empirical results in connection to the US interest rate reaction function. Finally, Chapter 9 provides conclusions and directions for future research.

Chapter 2

Stable and Unstable Least Squares Models: A Review

This chapter is devoted to reviewing the asymptotics of linear and nonlinear models. By means of a very simple linear model, we show in Section 1 how things can go wrong when a breakpoint is ignored. Section 2 deals with OLS asymptotics in the presence of multiple structural changes. Section 3 reviews the NLS asymptotic theory in stable models. These two frameworks are combined in the next chapter, where we introduce NLS models with multiple structural changes.

2.1 Neglecting a Parameter Shift in OLS Models

Consider for now the following simple univariate linear data generation process, characterized by one parameter change:

$$y_t = \begin{cases} x_t'\,\theta_1^0 + u_t & t = 1, \ldots, k \\ x_t'\,\theta_2^0 + u_t & t = k+1, \ldots, T \end{cases} \tag{2.1}$$

Here y_t is the dependent variable, x_t is a $p \times 1$ stochastic regressor, θ_1^0, θ_2^0 are the $p \times 1$ parameter vectors, T is the sample size and $1 < k < T$ is an integer denoting the change-point. From now on, we will drop the *0*-superscripts only to denote parameter values different from the true ones. We can rewrite the data generation process in matrix form. To that end, let X be the matrix whose t^{th} row is x_t', Y the $T \times 1$ vector whose t^{th} element is y_t, and similarly for U. Also let X_1 and X_2 be the matrices that vertically stack all x_t' for $t = 1, \ldots, k$, respectively $t = k+1, \ldots, T$. Additionally, let $\bar{X}_2 = [O, X_2']'$, where O is a null matrix with p rows and k columns, and $\delta^0 = \theta_2^0 - \theta_1^0 \neq 0$. Then the true model in (2.1) becomes:

$$Y = X\theta_1^0 + \bar{X}_2\delta^0 + u \tag{2.2}$$

One of the usual classical linear model assumptions is:

Assumption A 2.1 (**Dependence and Memory of Processes**).
(i) The errors u_t are independent and identically distributed (i.i.d) with mean 0 and variance σ^2.
(ii) Morever, they are independent of the regressors, i.e. $u_t \perp x_s$ for all integers t, s.

Add the following fixed break-fraction assumption:

Assumption A 2.2 (**Fixed Break Fraction**). *Denote by $[\cdot]$ the smallest integer part. Then $k = [T\lambda]$.*

This assumption allows both sub-samples to grow proportionately with the sample size. We also impose the following identification assumption:

Assumption A 2.3 (**Sub-Sample Identification**). *$T^{-1}X_i'X_i \xrightarrow{p} Q_i$, some positive definite matrices of constants, for $i = 1, 2$.*

Under assumptions 2.1-2.3, the OLS estimators of θ_1^0, θ_2^0 from the two separate sub-samples yield consistent estimators of the corresponding parameter values. Suppose the researcher is not aware of the parameter shift. Instead, he / she will estimate a single parameter θ from the model:

$$Y = X\theta + U. \tag{2.3}$$

Unaware of the shift, he / she will use the full-sample OLS estimation while imposing the usual identification assumption[1]:

Assumption A 2.4 (**Full-Sample Identification**). $T^{-1}X'X \xrightarrow{p} Q$, *a positive definite matrix of constants.*

Note that Assumption 2.4 is implied by 2.3, but the researcher is unaware of the shift, so he imposes 2.4. Given (2.1), the full-sample OLS estimator is:

$$\begin{aligned} \hat{\theta}_T &= \left(X'X\right)^{-1} X'Y = \theta_1^0 + \left(X'X\right)^{-1} X'\bar{X}_2\, \delta^0 + \left(X'X\right)^{-1} X'U \\ \hat{\theta}_T &= \theta_1^0 + \left(X'X\right)^{-1} X_2{}'X_2\, \delta^0 + \left(X'X\right)^{-1} X'U \end{aligned} \tag{2.4}$$

Given Assumptions 2.1-2.3, $\hat{\theta}$ will be a biased and inconsistent estimator of θ_1^0, since:

$$\begin{aligned} \text{plim } E\ [\hat{\theta}_T - \theta^0\,] &= \text{plim }\ [\left(T^{-1}X'X\right)^{-1} (T^{-1}X_2{}'X_2)]\, \delta^0 \\ &= \text{plim }\ \left(T^{-1}X'X\right)^{-1} \text{plim } (T^{-1}X_2{}'X_2)\delta^0 \\ &= Q^{-1}Q_2\delta^0 \neq 0 \end{aligned} \tag{2.5}$$

Note that $Q^{-1}Q_2\delta^0 \neq 0$ because $Q^{-1}Q_2$ is invertible and $\delta^0 \neq 0$. Similar arguments hold for θ_2^0. Note that the bias depends on the magnitude of the shift δ^0. Hence, awareness of shifts, their locations in the sample and their magnitude is crucial for obtaining good estimates.

2.2 Asymptotics in Unstable OLS Models

Bai and Perron (1998) provide a comprehensive treatment of estimation in linear regression

[1] See Wooldridge (2006), Ch. 2-3.

models. They also consider partial structural change (meaning only a subset of parameter change from one regime to another), but for simplicity we are going to present their model in pure structural change form (when all parameters shift at the same date). Assume that the data generation process is:

$$y_t = x_t' \theta_i^0 + u_t \qquad t = [T_i^0 + 1, T_{i+1}^0] \qquad i = 0, 1, \ldots m \tag{2.6}$$

where $T_0^0 = 0$ and $T_{m+1}^0 = T$ by convention. Here y_t is the dependent variable, x_t $(p \times 1)$ are the regressors, θ_{i+1}^0 $(p \times 1)$ are coefficients that change at dates T_i^0, and m is a known finite positive integer. The location of the break-points are unknown to the researcher, while m is known.[2] Denote by $\bar{T}^m \equiv (T_0 = 1, T_1, \ldots, T_m, T_{m+1} = T)$ any m-partition of the interval $[1, T]$. To simplify the notation, we will stack the column vectors θ_{i+1}^0 and θ_{i+1} into two corresponding $p \times (m+1)$ matrices, θ^0 and θ. For a given sample partition and given parameter values θ, denote by $S_T(\bar{T}^m, \theta)$ the sum of squares. The sub-sample OLS estimator for a given partition is:

$$\hat{\theta}_T(\bar{T}^m) = \underset{\theta(\bar{T}^m)}{\operatorname{argmin}} \ S_T(\bar{T}^m, \theta(\bar{T}^m)\,) \tag{2.7}$$

Searching over all possible partitions yields the estimates $\hat{T} = (1, \hat{T}_1, \ldots, \hat{T}_m, T)$ for change-points and $\hat{\theta}_T = (\hat{\theta}_{T,1}, \ldots, \hat{\theta}_{T,m+1})$ for parameters:

$$\hat{T} = \underset{\bar{T}^m}{\operatorname{argmin}}\, S_T\ (\ \bar{T}^m, \hat{\theta}_T(\bar{T}^m)\ ) \ \text{ and } \hat{\theta}_T = \hat{\theta}_T\ (\hat{T}) \tag{2.8}$$

Since in Chapter 3 we propose a similar estimation procedure for nonlinear models, of special interest are the assumptions under which the estimators above are consistent. Let X_i be the vector that stacks the row vectors x_t' when $t \in I_i^0$, for $I_i^0 = \{T_{i-1}^0 + 1, \ldots, T_i^0\}$ and $i = 1, \ldots, m+1$. Assign $\Delta_i^0 = T_i^0 - T_{i-1}^0$. Using this notation, Bai and Perron (1998) impose the following assumption on the data generation process:

[2] Or else can be inferred with appropriate testing procedures - see Bai and Perron (1998).

Assumption A 2.5 (**Dependence and Memory of Processes**).

(i) With $\{\mathcal{F}^i_{-\infty}, i = 1, 2, \ldots\}$ *a sequence of increasing* σ*-fields,* $\{u_i, \mathcal{F}^i_{-\infty}\}$ *form a* $\mathcal{L}^r$*-mixingale sequence with* $r = 4 + \gamma$ *, for some* $\gamma > 0$.[3]

(ii) The regressors are independent of the errors, i.e. $u_t \perp x_s$ *for all integers* t, s.

Assumption 2.5(i) is more general than its counterpart 2.1(i) because it allows for substantial correlation and heterogeneity in the errors, as can be seen from the McLeish's (1975) definition:

Definition 2.1 ($\mathcal{L}^{\mathbf{r}}$ **Mixingale**). Let $\{v_t\}$ be a sequence of random vectors defined on a probability space $(\Omega, \mathcal{F}, P)$ and let $\mathcal{F}^b_a = \sigma\,(v_t; a \leq t \leq b)$ be the Borel σ-algebra of events generated by $v_a, \ldots, v_b$. Denote the $\mathcal{L}^r$-norm of v_t by $E^{1/r}\|v_t\|^r = \|v_t\|_{\,r}$, where $\|\cdot\|$ without a subscript denotes the Euclidean vector norm. Then $\{v_i, \mathcal{F}^i_{-\infty}\}$ forms an $\mathcal{L}^r$ mixingale of size $-a$ if there exist sequences of non-negative constants $\{c_i, i \geq 1\}$ and $\{\phi_j, j \geq 0\}$ such that $\phi_j = O(j^{-a})$ with $a > 0$, and for all $i, j \geq 0$,

(i) $\|\, v_i \mid \mathcal{F}^{i-j}_{-\infty} \,\|_{\,r} \;\leq c_i\, \phi_j;$

(ii) $\|\, v_i - E(v_i \mid \mathcal{F}^{i+j}_{-\infty}) \,\|_{\,r} \;\leq c_i\, \phi_{j+1};$

(iii) $\max_i c_i \;\leq K < \infty$ for some $K > 0$;

(iv) $\sum_{j=1}^{\infty} j^{1+\kappa}\phi_j \;< \infty$ for some $\kappa > 0$.

A mixingale is the asymptotic analogue of a martingale. When v_i is $\mathcal{F}^i_{-\infty}$-measurable, meaning that it does not depend on future values, (ii) holds automatically. Condition (i) is esentially a memory condition, because the order of ϕ_j indicates the rate of memory decay. Condition (iv) is a summability condition. Examples of $\mathcal{L}^1$-mixingales include: martingale difference sequences, ARMA processes, strong and uniform mixing sequences that are uniformly $\mathcal{L}^r$-bounded for some $r > 1$ (i.e. $\sup_t E|u_t|^r < \infty$).[4] Note that part (ii) of assumption 2.5

[3] For properties of mixingales, see McLeish (1975) and Andrews (1988).

[4] For details, see McLeish (1975) and Gallant and White (1988), pp. 25.

is rather restrictive because it excludes the presence of lagged dependent variables in the regressors. Bai and Perron (1998) relax this assumption at the expense of assuming $\{u_t\}$ is a martingale difference and $\{x_t\}$ are non-trending.

Having these in mind, Bai and Perron (1998) consider the following additional assumptions:

Assumption A 2.6 (**Identification**).

(i) $X_i X_i'/\Delta_i^0 \xrightarrow{p} Q_i$*, some random positive definite matrices not necessarily the same for all* $i = 1, \ldots, m+1$.

(ii) There exists an $l_0 > 0$ *such that for all* $l > l_0$*, the minimum eigenvalues of* $A_{il} = l^{-1} \sum_{T_i^0+1}^{T_i^0+l} x_t x_t'$ *and* $A_{il}^* = l^{-1} \sum_{T_i^0-l}^{T_i^0} x_t x_t'$ *are bounded away from* 0.

(iii) The matrix $B_{il} = \sum_k^l x_t x_t'$ *is invertible for* $l - k \geq p$.

Assumption A 2.7 (**Fixed Break Fractions**). *Assume* $T_i^0 = [T\lambda_i^0]$*, where* $0 < \lambda_1^0 < \ldots < \lambda_m^0 < 1$.

Assumption 2.6 is specific to change-point models. Part (i) is the generalization to m breaks of 2.3. Part (ii) requires that there are enough observations near the true break-point so that they can be identified. Part (iii) is imposed because the change-points are estimated by a global least-squares search. In other words, there need to be at least p observations in each segment so that the $p \times 1$ parameters are identified. Assumption 2.7, like its counterpart 2.2, ensures that the break-fractions are fixed, hence allowing the break points to be asymptotically distinct. Alternatively, one can allow the magnitude of shifts to converge to zero as the sample size increases.[5]

Let $\hat{T}_i = [T\hat{\lambda}_i]$, $\bar{X} = \text{diag}\,(X_1, \ldots, X_{m+1})$, $E(UU') = \Omega$, and $V = \text{diag}\,(Q_1, \ldots, Q_{m+1})$. Hence $T^{-1}\bar{X}'\bar{X} \xrightarrow{p} V$. Additionally, let $\Phi = \text{plim}\ T^{-1}\bar{X}'\Omega\bar{X}$. Given this notation, Bai and Perron (1998) show:

Theorem 2.1. *Under Assumptions 2.5-2.7,*

(i) For each $i = 1, \ldots m$*, the break-fraction estimates are consistent:* $\hat{\lambda}_i \xrightarrow{p} \lambda_i^0$*. Also, their rate of*

[5] See Billingsley (1968), Ch. 4.

convergence is T*, i.e. for every* $\epsilon > 0$*, there is a* $C > 0$ *such that* $\limsup P(|T(\hat{\lambda}_i - \lambda_i^0)| > C) < \epsilon$*;*
(ii) The parameter estimates are asymptotically normal: $T^{1/2}\,(\hat{\theta}_T - \theta^0) \xrightarrow{p} \mathcal{N}(0, V^{-1}\Phi V^{-1})$.

Consider part (i) first. We will briefly describe the proof technique, as it it similar to the one we employ in Ch. 3. To that end, let $\hat{u}_t$ be the estimated residuals. The key step in proving consistency of break-fraction estimates is letting $d_t = \hat{u}_t - u_t$. This way, the minimized sum of squared residuals can be written as:

$$T^{-1}\sum_{t=1}^{T}\hat{u}_t^2 \;=\; T^{-1}\sum_{t=1}^{T}u_t^2 + T^{-1}\sum_{t=1}^{T}d_t^2 + 2T^{-1}\sum_{t=1}^{T}d_t u_t \tag{2.9}$$

On the other hand, the definition of minimization implies:

$$T^{-1}\sum_{t=1}^{T}\hat{u}_t^2 \le T^{-1}\sum_{t=1}^{T}u_t^2, \tag{2.10}$$

Using (2.9) and (2.10), it follows that

$$T^{-1}\sum_{t=1}^{T}d_t^2 + 2T^{-1}\sum_{t=1}^{T}d_t u_t \le 0. \tag{2.11}$$

Bai and Perron (1998) show that the second term on the LHS of (2.11) is $o_p(1)$, implying that $T^{-1}\sum_{t=1}^{T} d_t^2$ is $o_p(1)$. They also show that if a change-point is not consistently estimated, then $T^{-1}\sum_{t=1}^{T} d_t^2$ is not $o_p(1)$. By contradiction, consistency follows. We will delay talking about the rate of convergence proof until Ch. 3.

Now consider part (ii) of Theorem 2.1. Note that the distribution of the parameter estimates are the same as if the change-points were known. If the errors are serially uncorrelated and homoskedastic with variance σ^2, then $\Phi = \sigma^2 V$. Hence, (ii) simplifies to:

$$T^{1/2}\,(\hat{\theta}_T - \theta^0) \xrightarrow{p} \mathcal{N}(0, \sigma^2 V^{-1}). \tag{2.12}$$

Behind Theorem 2.1 are the general assumptions and proof techniques for OLS with multiple unknown breaks. For NLS models, we will provide similar theorems in Ch 3. To do so, we will briefly present the NLS asymptotics, which is a little different since it involves models that are nonlinear in parameters.

2.3 Asymptotics in Stable NLS Models

A nonlinear univariate data generation process can often be written as:

$$y_t = f(x_t\ , \theta^0) + u_t \tag{2.13}$$

where $\theta^0 \in \Theta$ is the true known $p \times 1$ parameter vector, the function $f : \mathbb{R}^q \times \Theta \to \mathbb{R}$ is a known measurable function on $\mathbb{R}$ for each $\theta \in \Theta$, and the sample size is T. For simplicity, denote $f_t(\theta) = f(x_t, \theta)$ and $S_T(\theta) = \sum_{t=1}^{T} u_t^2(\theta)$, where $u_t(\theta) = y_t - f_t(\theta)$.

Given that $E\left[u_t\ f_t(\theta^0)\right] = 0$, an appropriate estimation method for the above model is NLS:

$$\hat{\theta}_T = \underset{\theta\in\Theta}{\text{argmin}}\ S_T(\theta) \tag{2.14}$$

Gallant (1989) derives a simple analogy between OLS and NLS. Let F be the $T \times p$ matrix with rows $F_t'(\theta^0)$, where $F_t(\theta) = \partial f_t(\theta)/\partial\theta$. Then:

$$\begin{aligned} \hat{\theta}_{OLS} &= \ (X'X)^{-1}X'Y \\ \hat{\theta}_{NLS} &= \ (F'F)^{-1}F'Y + o_p(T^{-1/2}). \end{aligned} \tag{2.15}$$

Hence, we could ignore the $o_p(T^{-1/2})$ term, regard F as the regressor and proceed as we did for linear models. However, things are not that simple, because the above approximation does not hold unless sufficient regularity conditions are imposed.

The regularity assumptions leading to consistency and asymptotic normality of the NLS estimator are different than for OLS because of three reasons. First, the proofs are based on using Taylor approximations rather than analytical formulae for parameters and minimands. Secondly, because the model (2.13) involves a nonlinear function, certain smoothness conditions are required. Thirdly, this function depends on parameters, so that uniform bounds are often useful.

Besides boundedness and smoothness conditions, assumptions on the data generation process are needed. The early paper of Jennrich (1969) derives the asymptotic properties of NLS estimators under fixed regressors and i.i.d. errors. Hannan (1971) extends these results to time-series data by allowing for stationary and ergodic errors. Hansen (1982) includes results for nonlinear models with jointly strictly stationary and ergodic regressors and errors. Bierens (1982) assumes i.i.d. errors, but allows explanatory variables to be uniform mixing processes. Finally, a treatment of NLS under general mixing assumptions on both errors and regressors is provided by White and Domowitz (1984). The results above are combined and extended in the books of Gallant and White (1988) and Gallant (1989). They provide a unified treatment of general nonlinear models (including NLS).

In this review of NLS models, we are going to follow closely White and Domowitz (1984) and Gallant and White (1988). The focus will be on joint strong mixing processes for errors and regressors. This allows for fairly general models, in which both errors and regressors may be serially correlated and heteroskedastic, as can be seen from the definition of strong (or α) mixing processes:

Definition 2.2 (**α-Mixing Process**)**.** Let $\{v_t\}$ be a sequence of random vectors defined on a probability space $(\Omega, \mathcal{F}, P)$ and let $\mathcal{F}_a^b = \sigma\left(v_t; a \leq t \leq b\right)$ be the Borel σ-algebra of events generated by $v_a, \ldots, v_b$. Define:

$$\alpha(m) = \sup_n \sup_{F \in \mathcal{F}_{-\infty}^n, G \in \mathcal{F}_{n+m}^\infty} |P(F\,G) - P(F)\,P(G)|, \tag{2.16}$$

a measure of dependence between events separated by m time periods. If $\alpha(m) = O(m^{-a})$ for some $a > 0$, $\{v_t\}$ is called α-mixing of size $-a$.

The order of a mixing process indicates how fast the dependence dies out in the limit. Examples of α-mixing processes are martingale differences and, under suitable conditions, stationary Markov chains, GARCH processes, finite ARMA processes, but also processes with memories longer than the latter.[6]

Hence, we assume:

Assumption A 2.8 (Dependence and Memory of Processes).
(i) Let $v_t = \{x_t, u_t\}$. Then v_t is an α-mixing process of size $-2r/(r-1)$, with $r > 1$.
(ii) $E[u_t\, f_t(\theta)] = 0$ for all θ, t.

Part (ii) is a weaker assumption than its linear counterpart 2.5(ii), only requiring that the regressors and errors are uncorrelated. This orthogonality assumption is the main reason θ^0 can be estimated by NLS. It rules out the possibility that u_t is serially correlated at the same time with x_t containing lagged dependent variables.[7] If we focus only on u_t, part (i) is comparable to Assumption 2.5(i) because there is a close relationship between mixing and mixingale processes, and the first can be transformed into the second under certain conditions.[8] Mixing processes have the advantage that measurable functions of them are themselves mixing. Hence, the regression function $u_t f_t(\theta)$ and its derivative $u_t F_t(\theta)$, will be mixing of the same size[9] for each θ, or uniformly in θ if α can be bounded uniformly in θ.

Next, we provide smoothness conditions on the regression function:

Assumption A 2.9 (Smoothness). *$f_t(\cdot)$ is twice continuously differentiable on $\theta \in \Theta$, uniformly in t.*

This assumption allows for the use of mean value expansions in deriving asymptotic normality of estimates, but can be replaced with the weaker assumption that $f_t(\cdot)$ is continuous on $\theta \in \Theta$ almost surely (a.s.), uniformly in t, for consistency purposes. Since we deal with

[6] The autocovariances of a finite lag ARMA model die out exponentially fast, while the autocovariances of an α-mixing process of size $-a$ dissappear at the slower rate m^a.
[7] In that case, the appropriate estimation method is GMM.
[8] See McLeish (1975) and Ch.4, Section 4.2.
[9] See White and Domowitz (1984), Lemma 2.1.

nonlinear models, uniform convergence in θ of certain quantities is needed. Hence, we impose, as in White and Domowitz (1984), uniform boundedness assumptions. To that end, define:

Definition 2.3 (**Uniformly $\mathcal{L}^s$-Integrable Functions**). The measurable function $B_t(v_t)$, $B_t : \mathbb{R}_+^{q+1} \to \mathbb{R}$ is $\mathcal{L}^s$-integrable if $\|B_t(v_t)\|_s \leq \Delta < \infty$ for some constant $\Delta > 0$ and all t.

Definition 2.4 (**Uniformly $\mathcal{L}^s$-Bounded Functions**). Let $q_t(v_t, \theta)$ be measurable functions for each θ in Θ, satisfying the smoothness assumption 2.9. Suppose there exist uniformly s-integrable functions B_t such that $\|q_t(v_t, \theta)\| \leq B_t(v_t)$ for all θ in Θ. Then the sequence $\{q_t\}$ is said to be uniformly $\mathcal{L}^s$ bounded.

With these definitions in mind, we can adapt the boundedness assumptions in White and Domowitz (1984) as follows:

Assumption A 2.10 (**Boundedness**).

(i) Θ *is a compact subset of* $\mathbb{R}^p$.

(ii) $\{u_t\}$ *is uniformly* $2r + 2\delta$*-integrable, for some* $\delta > 0$*;*

(iii) Either (a) $\{u_t f_t(\theta)\}$ *is uniformly* $\mathcal{L}^{2r+2\delta}$ *bounded or (b)* $E(u_t|x_t) = 0$ *and* $\{u_t f_t(\theta^0)\}$ *is uniformly* $\mathcal{L}^{2r+2\delta}$ *bounded;*

(iv) $\{u_t F_{t,i}(\theta)\}$ *and* $\{F_{t,i}(\theta)\}$ *are uniformly* $\mathcal{L}^{2r+2\delta}$ *bounded, for* $i = 1, \ldots, p$*;*

(v) For $i, j = 1, \ldots p$, $\{u_t \dfrac{\partial^2 f_t(\theta)}{\partial\theta_i \partial\theta_j}\}$ *are uniformly* $\mathcal{L}^{2r+2\delta}$ *bounded.*

The first three conditions restrict the moments of $u_t(\theta)$ and ensure the existence of the expected value of the minimand function, of its limit and of the limit of its derivative. Uniform boundednes is also needed to ensure uniform convergence in θ. In other words, let $S(\theta) = \lim\ E\left[S_T(\theta)\right] = \lim\ T^{-1} \sum_{t=1}^T E\left[u_t^2(\theta)\right]$. Then:

$$S_T(\theta) \xrightarrow{p} S(\theta), \text{ uniformly in } \theta. \tag{2.17}$$

The other conditions, (iv) and (v), are needed for asymptotic normality of $\hat{\theta}_T$, but not for consistency. We postpone their discussion to Ch. 3.

As with linear models, we still need an identification assumption:

Assumption A 2.11 (**Identification**).

(i) $S(\theta)$ has a unique global minimum at θ^0.

(ii) Let $A_{T,a}(\theta) = Var\ T^{-1/2} \sum_{a+1}^{a+T} u_t F_t(\theta)$. Then, for each $\theta \in \Theta$, $A_{T,a}(\theta) \xrightarrow{p} A(\theta)$, a positive definite matrix, uniformly in a.

(iii) Let $D_T(\theta) = T^{-1} \sum_{t=1}^{T} F_t(\theta) F_t(\theta)'$. Then, for each $\theta \in \Theta$, $D_T(\theta) \xrightarrow{p} D(\theta)$, a positive definite matrix not depending on T.

Note that $A(\cdot)$ and $D(\cdot)$ are both well defined by the boundedness assumptions 2.10(iii) and (iv). The third condition above, with θ replaced by θ^0, is the nonlinear equivalent of the identification assumption 2.6, because $F_t(\theta^0)$ are nonlinear equivalents of x_t in a linear model. The other two are specific to the nonlinear model we considered. More exactly, condition (i) is necessary because it ensures that the minimum over $S_T(\theta)$ is asymptotically identifiable and achieved at the true parameter value. Together with (2.17), it implies that $S_T(\hat{\theta}) = \min_\theta S_T(\theta) \xrightarrow{p} \min_\theta S(\theta) = S(\theta^0)$. By a 'subsequence of subsequence' argument, $\hat{\theta}_T \xrightarrow{p} \theta^0$. Condition (ii), with $a = 0$, ensures that the long-run variance $A(\theta)$ exists. The uniform convergence in a is a statement about the rate of memory decay of processes dependent on $\{\ldots, u_t F_t(\theta), \ldots\}$. Its relevance will be seen in Ch. 4.

We can now summarize the main NLS asymptotic results:

Theorem 2.2.

(i) Under 2.8, 2.9, 2.10(i)-(iii) and A2.11(i), $\hat{\theta}_T \xrightarrow{p} \theta^0$.

(ii) Under 2.8- 2.11, $T^{1/2}\,(\hat{\theta}_T - \theta^0) \xrightarrow{d} \mathcal{N}(0, D^{-1} A D^{-1})$, where $D = D(\theta^0)$ and $A = A(\theta^0)$.

These results are similar to the ones of Theorem 2.1. As for linear models, when the errors are i.i.d and independent of the regressors, $A = \sigma^2 D = \sigma^2 \text{ plim } T^{-1} \sum_{t=1}^{T} F_t(\theta^0) F_t(\theta^0)'$, hence:

$$T^{1/2}\,(\hat{\theta}_T - \theta^0) \xrightarrow{d} \mathcal{N}(0, \sigma^2 D^{-1}), \tag{2.18}$$

which is indeed the nonlinear equivalent of (2.12).

In this chapter, we contrasted OLS and NLS theory. In subsection 2.2, we saw how the linear framework of LS can be extended to multiple breaks. Next, we consider the extension of nonlinear LS to multiple breaks. In doing so, we will exploit some of the connections between linear and nonlinear LS theory. This is the subject of Ch. 3.

Chapter 3

Break-Fraction Estimates in Unstable NLS Models

In the previous chapter, we provided a review of asymptotic theory in OLS models with or without multiple unknown change points and NLS models with no change. Several econometric models are nonlinear in parameters, and they too may be subject to parameter variation the same way linear models are. Moreover, our model can nest nonlinearities and breaks while several current literature strands consider them as competing specifications. Here, we extend the present econometric methodology to NLS models with multiple unknown changes. To our knowledge, this hasn't been done before.

In Section 1, we lay out the model and the main assumptions that we will use throughout the dissertation. In Section 2, by means of a simple LS procedure, we show that we can consistently locate the unknown break dates in nonlinear models. In Section 3, we show that the break-fractions are T-rate convergent. Upon this result hinge all asymptotic properties laid out in Ch. 4.

3.1 The Unstable NLS Model and Assumptions

Consider a univariate nonlinear data generation process with m unknown change-points:

$$y_t = f(x_t, \theta^0_{i+1}) + u_t \qquad t = [T^0_i + 1, T^0_{i+1}] \qquad i = 0, 1, \dots m \tag{3.1}$$

where $T_0^0 = 0$ and $T_{m+1}^0 = T$ by convention. Here y_t is the dependent variable, x_t $(q \times 1)$ are the regressors, θ_{i+1}^0 $(p \times 1)$ are parameters that change at dates T_i^0, the function $f : \mathbb{R}^q \times \Theta \to \mathbb{R}$ is a known measurable function on $\mathbb{R}$ for each $\theta \in \Theta$, and T is the sample size. As for linear models, we consider m as a known finite positive integer, but we allow for the break dates to be unknown to the researcher. For simplicity, let $f_t(\theta) = f(x_t, \theta)$ and denote by $\bar{T}^m \equiv (T_0 = 1, T_1, \ldots, T_m, T_{m+1} = T)$ any m-partition of the interval $[1, T]$. To further simplify the notation, we will stack the column vectors θ_{i+1}^0 and θ_{i+1} into two corresponding $p \times (m+1)$ matrices, θ^0 and θ. For a given sample partition and given parameter values θ, denote by $S_T(\bar{T}^m, \theta)$ the sum of squares. Given that $E\left[u_t f_t(\theta)\right] = 0$, the estimation can be carried out by NLS within each partition. The sub-sample NLS estimator for a given partition is:

$$\hat{\theta}_T(\bar{T}^m) = \underset{\theta(\bar{T}^m)}{\operatorname{argmin}} \; S_T(\bar{T}^m, \theta(\bar{T}^m)\,) \tag{3.2}$$

As Bai and Perron (1998), we consider searching over all possible partitions for obtaining the break-point estimates. Hence, the estimates $\hat{T} = (1, \hat{T}_1, \ldots, \hat{T}_m, T)$ for change-points and $\hat{\theta}_T = (\hat{\theta}_{T,1}, \ldots, \hat{\theta}_{T,m+1})$ for parameters are obtained as follows:

$$\hat{T} = \underset{\bar{T}^m}{\operatorname{argmin}} \, S_T\left(\,\bar{T}^m, \hat{\theta}_T(\bar{T}^m)\,\right) \text{ and } \hat{\theta}_T = \hat{\theta}_T\,(\hat{T}) \tag{3.3}$$

For clarity of expositions, we will drop the T-subscripts on parameter estimates.

One of the main tasks of this chapter is to lay out the assumptions under which the parameter estimates are asymptotically normal. This task involves combining the main nonlinear LS assumptions with conditions that are specific to linear change-point models. As in Ch. 2, we start with assumptions about dependency and memory of the regressors and errors:

Assumption A 3.1 (Dependence and Memory of Processes).

(i) Let $v_t = (x_t, u_t)$. Then $\{v_t\}$ is a strictly stationary α-mixing process of size $-2s/(s-2) - \delta$,

where $s > 2$, $\delta > 0$[1]*;*

(ii) The errors are uncorrelated with the regression function, i.e. $E[u_t f_t(\theta)] = 0$ *for all* θ, t.

This is a slightly stronger α-mixing process than that assumed for stable NLS theory in Ch. 2, because the size of the mixing coefficients is larger in absolute value. It is also more restrictive than the mixingale assumption for unstable OLS (because mixingales include near epoch-dependent processes that are $\mathcal{L}^r$-bounded, with $r > 1$, more general than α-mixing processes).

Additionally, we imposed strict stationarity:

Definition 3.1 (**Strict Stationarity**[2]). A time series $\{v_t\}_{-\infty}^{+\infty}$ is strictly stationary if $\{v_1, \ldots, v_n\}$ and $\{v_{1+h}, \ldots, v_{n+h}\}$ have the same distribution for any integers n, h.

Assumption 3.1 is rather strong because it doesn't allow the mean or variance of v_t to change. It was recently criticized by Hansen (2000), who shows that at the same time a parameter change, changes in the variance and/or marginal distribution of regressors and/or errors may occur, and one may not be able to distinguish between the two. This is a valid argument, and for linear models weak stationarity with existing second moments often suffices. However, if we focus on a nonlinear relationship, we need to look beyond the second moments, hence justifying the strict stationarity assumption. Examples of models with strictly stationary strong mixing sequences include appropriately redefined threshold autoregressive models (TARs), ARCH and GARCH under some conditions and models underlying ergodic Markov chains. While the strict stationarity mixing condition is not easily verifiable for a given process, it can be done by means of techniques extensively described in Fan and Yao (2003), Ch. 4.[3]

The next assumption imposes smoothness of the regression function:

Assumption A 3.2 (**Smoothness**). *The function* $f_t(\cdot)$ *is twice continuously differentiable in* θ.

[1]For definition of an α-mixing process, see Ch. 2, Section 2.3

[2]Fan and Yao (2003), pp. 30.

[3] Perhaps their most useful result is the following: A strictly stationary process $\{v_t\}_{-\infty}^{+\infty}$ is α-mixing if and only if its positive half $\{v_t\}_1^{+\infty}$ is α-mixing.

Note that unlike Assumption 2.10, we do not impose this condition uniformly in t. Gallant and White (1988) notes that uniformity in t is very strong because it restricts the number of lags present in $f_t(\theta)$ to a finite number, which is not always realistic.

Next, we impose some bounds on functions of the data. For that, denote, as in Ch. 2, $F_t(\theta) = \partial f_t(\theta)/\partial\theta$. We assume:

Assumption B (Boundedness).

(i) Θ *is a compact subset of* $\mathbb{R}^p$.

(ii) $\{u_t\}$ *is uniformly* $2s$*-integrable;*

(iii) $\{u_t f_t(\theta)\}$ *is uniformly* $\mathcal{L}^{2s}$ *bounded;*

(iv) $\{u_t F_t(\theta)\}$ *and* $\{F_t(\theta)\}$*, are uniformly* $\mathcal{L}^{2s}$ *bounded;*

(v) For $i, j = 1, \ldots p$, $\{u_t \dfrac{\partial^2 f_t(\theta)}{\partial\theta_i\partial\theta_j}\}$ *are uniformly* $\mathcal{L}^s$ *bounded.*

We denote this assumption with the letter B instead of A because we are going to actually use a modified version of it. First, note that the moments $2r + 2\delta$ in Assumption 2.10 are only imposed for existence of a moment higher than the second, because there $r > 1, \delta > 0$. Here, they are replaced by the stronger assumption of existence of more than the forth moments, because $s > 2$. Second, note that if $v_t = \{x_t, u_t\}$ is strictly stationary, then any measurable function of v_t is strictly stationary too. Hence $\{u_t\}$, $\{f_t(\theta)\}$, $\{F_t(\theta)\}$, $\{u_t f_t(\theta)\}$ and $\{u_t F_t(\theta)\}$ are strictly stationary for each θ. According to Definition 2.3, (ii) above implies $E|u_t|^s < \infty$. Also, because of strict stationarity, uniform $\mathcal{L}^s$ boundedness simply implies that functions of v_t and θ possess s (hence greater than 2) moments bounded uniformly in θ (without the bounds being uniform in t). Hence, we can simplify Assumption 3.3 as follows:

Assumption A 3.3 (Boundedness).

(i) Θ *is a compact subset of* $\mathbb{R}^p$*;*

(ii) $E|u_t|^{2s} < \infty$*;*

(iii) $\sup_\theta E|u_t f_t(\theta)|^{2s} < \infty$*;*

(iv) $\sup_\theta E\|u_t F_t(\theta)\|^{2s} < \infty$ *and* $\sup_\theta E|F_t(\theta)|^{2s} < \infty$*;*

(v) For $i, j = 1, \dots p$, $\sup_\theta E\left|u_t \dfrac{\partial^2 f_t(\theta)}{\partial\theta_i \partial\theta_j}\right|^s < \infty$.

We also slightly modify Assumption 2.11. To that end, let $S(\theta) = \operatorname{plim} S_T(\bar{T}_0, \theta\ )$, where $\bar{T}_0$ is the true m-partition of the $[1, T]$ interval. Then:

Assumption A 3.4 (**Identification**).

(i) $S(\theta)$ has a unique global minimum at θ^0;

(ii) Let $A_i(\theta_i^0) = Var\ T^{-1/2}\sum_{t=T_{i-1}^0+1}^{T_i^0} u_t F_t(\theta_i^0)$, for $i = 1, \dots, m+1$. Then $A_i(\theta_i^0) \xrightarrow{p} \bar{A}_i$, a positive definite matrix not depending on T.

(iii) Let $D_T(\theta) = T^{-1}\sum_{t=1}^{[Tr]} F_t(\theta)F_t(\theta)'$. Then, uniformly in $\theta \times r$, $D_T(\theta) \xrightarrow{p} rD(\theta)$, a positive definite matrix not depending on T;

(iv) $E[f_t(\theta_i^0)] \neq E[f_t(\theta_{i+1}^0)]$, for each $i = 1, 2, \dots, m$.

The first three conditions are similar to Assumption 2.11. Part (iv) above ensures that the shifts are identifiable. As we will see later, the above renders any identification assumption similar to the unstable linear model superfluous. The only assumption we will preserve is:

Assumption A 3.5 (**Fixed Break Fractions**).

$T_i^0 = [T\lambda_i^0]$, where $0 < \lambda_1^0 < \dots < \lambda_m^0 < 1$.

Having these assumptions in mind, we can derive the asymptotic properties of our estimators.

3.2 Consistency of Break-Fraction Estimates

By break-fraction estimates, we will mean $\hat{\lambda}_i$ such that $\hat{T}_i = [T\hat{\lambda}_i]$, for $i = 1, \dots, m$. We will show their consistency, as summarized below:

Theorem 3.1. *Under Assumptions 3.1-A3.5, for each $i = 1, \dots, m$, $\hat{\lambda}_i \xrightarrow{p} \lambda_i^0$.*

We first outline the steps of the proof. As in Bai and Perron (1998), define: $\hat{u}_t = y_t - f_t(\hat{\theta}_{k+1})$, for $t \in [\hat{T}_k + 1, \hat{T}_{k+1}]$ and $d_t = \hat{u}_t - u_t = f_t(\theta_{j+1}^0) - f_t(\hat{\theta}_{k+1})$, for $t \in I_j^0 = [T_j^0 + 1, T_{j+1}^0] \cap \hat{I}_k =$

$[\hat{T}_k+1, \hat{T}_{k+1}]$, where $k, j = 0, 1, \ldots, m$. By similar arguments to Section 2.2,

$$T^{-1}\sum_{t=1}^{T} d_t^2 + 2T^{-1}\sum_{t=1}^{T} d_t u_t \leq 0. \tag{3.4}$$

As in Bai and Perron (1998), we will show that:

Lemma 3.1. $T^{-1}\sum_{t=1}^{T} d_t u_t \xrightarrow{p} 0$.

Lemma 3.1, together with equation (3.4), imply that $T^{-1}\sum_{t=1}^{T} d_t^2 \leq 0$ with large probability for large T. Since $T^{-1}\sum_{t=1}^{T} d_t^2 \geq 0$ by construction, it can only be true that $T^{-1}\sum_{t=1}^{T} d_t^2 \xrightarrow{p} 0$. We will show that if at least one break-fraction is not consistently estimated, $T^{-1}\sum_{t=1}^{T} d_t^2 \stackrel{p}{\nrightarrow} 0$:

Lemma 3.2. *If* $\hat{\lambda}_j \stackrel{p}{\nrightarrow} \lambda_j^0$ *for some* j*, then* $\limsup\, P\left[\, T^{-1}\sum_{t=1}^{T} d_t^2 > C \,\right] > \epsilon$*, for some* $C > 0, \epsilon > 0$.

This yields a contradiction, ensuring that all break-fractions are consistently estimated. Hence, for proving Theorem 3.1, it suffices to prove Lemma 3.1 and 3.2.

Proof of Lemma 3.1:

If we let i, respectively $\hat{i}$ subscripts denote summing over intervals I_i^0, respectively $\hat{I}_j$, while $\psi_t(\theta) = u_t f_t(\theta)$, then:

$$T^{-1}\sum_{t=1}^{T} u_t d_t \;=\; T^{-1}\sum_{i=0}^{m}\sum_{i}\psi_t(\theta_{i+1}^0) - T^{-1}\sum_{i=0}^{m}\sum_{\hat{i}}\psi_t(\hat{\theta}_{i+1}) \tag{3.5}$$

First note that:

$$\begin{aligned}
\left|T^{-1}\sum_{i=0}^{m}\sum_{\hat{i}}\psi_t(\hat{\theta}_{i+1})\right| &\leq \sum_{i=0}^{m} T^{-1}\left|\sum_{\hat{i}}\psi_t(\hat{\theta}_{i+1})\right| \leq \sum_{i=1}^{m+1} T^{-1}\left|\sum_{t=1}^{\hat{T}_i}\psi_t(\hat{\theta}_i) - \sum_{t=1}^{\hat{T}_{i-1}}\psi_t(\hat{\theta}_i)\right| \\
&\leq T^{-1}\sum_{i=1}^{m+1}\left|\sum_{t=1}^{[T\hat{\lambda}_i]}\psi_t(\hat{\theta}_i)\right| + T^{-1}\sum_{i=1}^{m+1}\left|\sum_{t=1}^{[T\hat{\lambda}_{i-1}]}\psi_t(\hat{\theta}_i)\right| \\
&\leq 2(m+1)\left|\sup_{\theta\times r} T^{-1}\sum_{t=1}^{[Tr]}\psi_t(\theta)\right|
\end{aligned}$$

By similar arguments,

$$T^{-1}\sum_{i=0}^{m}\left|\sum_{i}\psi_t(\theta^0_{i+1})\right| \leq 2(m+1)\left|\sup_{\theta\times r} T^{-1}\sum_{t=1}^{[Tr]}\psi_t(\theta)\right|$$

Hence,

$$T^{-1}\left|\sum_{t=1}^{T}u_t d_t\right| \leq 4(m+1)\left|\sup_{\theta\times r} T^{-1}\sum_{t=1}^{[Tr]}\psi_t(\theta)\right| \tag{3.6}$$

The novelty in this approach is taking the supremum over $\theta \times r$ rather than the supremum over θ and then over r or vice versa. This is a necessary step because $\hat{\theta}_i$ depends on the estimated partition. To show that $T^{-1}\sum_{t=1}^{[Tr]}\psi_t(\theta) \xrightarrow{p} 0$ uniformly in $\theta \times r$, we use the following proposition in Caner (2007), with β-mixing replaced by α-mixing (since β-mixing implies α-mixing of the same size). To that end, define the norm of a matrix as $\|A\| = [tr(A'A)]^{1/2}$. Then:

Proposition 3.1 (**Caner, 2007**). *If:*

(i) $\{v_t\}$ *is a strictly stationary* α*-mixing process of size* $-a$*, with* $a > 2 + 4/\delta$*, for some* $\delta > 0$*;*

(ii) $h_t(\theta) = h(v_t, \theta)$ *is a vector valued function such that* $\sup_\theta E\|h_t(\theta)\|^{2+\delta} < \infty$*;*

(iii) $\|h_t(\theta_1) - h_t(\theta_2)\| \leq B_t\|\theta_1 - \theta_2\|$*, with* $E\|B_t\|^{2+\delta} < \infty$*, then:*

$$T^{-1}\sum_{t=1}^{[Tr]}[(h_t(\theta) - E[h_t(\theta)])] = O_p(T^{-1/2}) = o_p(1) \text{ uniformly in } \theta \times r.$$

Note the typo in Caner (2007): in his Assumption AS1, $|\cdot|$ should be replaced by $\|\cdot\|$, or else his Lemma A1 should be a univariate result (while he proves a multivariate one). We will show that the conditions of Proposition 3.1 are implied by our assumptions. Since $s > 2$, define δ such that $s = 2+\delta$. Then $2s/(s-2)+\delta = 2+4/(s-2)+\delta = 2+4/\delta+\delta > 2+4/\delta$. Hence, by the mixing assumption, $\{v_t\}$ is also a strictly stationary α-mixing process of size $a > 2 + 4/\delta$. Now let $h_t(\cdot) = \psi_t(\theta) = u_t f_t(\theta)$. Since $s = 2 + \delta$, by A3.3(iii), $\sup_\theta E|\psi_t(\theta)|^{2+\delta} < \infty$. By the smoothness assumption A3.2, $\psi_t(\theta)$ is continuously differentiable in θ, hence we can apply the mean value theorem (MVT). Therefore, $|\psi_t(\theta_1) - \psi_t(\theta_2)| = \|u_t F_t(\bar{\theta})\| \times \|\theta_1 - \theta_2\|$, where $\bar{\theta} = \bar{\lambda}\theta_1 + (1-\bar{\lambda})\theta_2$, for some $\bar{\lambda} \in [0,1]$. Now by A3.3(ii) and A3.3(iv), for $s = 2+\delta$, we have $\sup_\theta E\|u_t F_t(\theta)\|^s < \infty$ by A3.3(iv). Hence, all conditions of Proposition 3.1 are satisfied, so:

$$T^{-1}\sum_{t=1}^{[Tr]} \psi_t(\theta) = o_p(1) \text{ uniformly in } \theta \times r. \tag{3.7}$$

Combining (3.7) with (3.6), we have:

$$T^{-1}\sum_{t=1}^{T} u_t d_t \xrightarrow{p} 0,$$

which completes the proof of Lemma 3.1. □

For Lemma 3.2, we will only need to prove uniform convergence in θ of certain sums, and not in $\theta \times r$. To that end, we use the following ULLN, which is a specialization of the ULLN in Gallant and White (1988), pp. 34[4]

Proposition 3.2 (ULLN). *Let there be a probability space $(\Omega, \mathcal{F}, P)$ and a compact set $\Theta \in \mathbf{R}^p$. If:*

(i) $h_t(\cdot)$ is a continuous function for each t[5];

(ii) $\sup_\theta E\|h_t(\theta)\|^s < \infty$;

(iii) $\{h_t(\theta)\}$ is a strictly stationary α-mixing process of size $-2s/(s-2)$,

Then $T^{-1}\sum_{t=1}^{T} h_t(\theta) \xrightarrow{p} H(\theta) = lim T^{-1}\sum_{t=1}^{T} E\left[h_t(\theta)\right]$, uniformly in θ.

Proof of Lemma 3.2:

Consider $\eta > 0$ such that $[T\eta]$ is an integer. Let $1*$ and $2*$ denote summing over the ordered sets $I_1(\eta) = \{\ [T\lambda_j^0] - T\eta + 1, \ldots, [T\lambda_j^0]\ \}$, respectively $I_2(\eta) = \{\ [T\lambda_j^0] + 1, \ldots, [T\lambda_j^0] + T\eta\ \}$.

[4] More exactly, we replace the assumption that $\{h_t(\theta)\}$ is NED in Gallant and White (1988) with the stronger condition (iii).

[5] This assumption is slightly stronger than the assumption that q_t is almost sure Lipschitz-$\mathcal{L}_1$ in Gallant and White (1988).

If $\hat{\lambda}_j \overset{p}{\nrightarrow} \lambda_j^0$ for at least one j, then with positive probability, there is a mismatch between the true and the estimated partitions. This mismatch occurs for at least two neighboring intervals. To see this, consider the interior of the interval $\hat{I}_{k-1} = [\hat{T}_{k-1}+1, \hat{T}_k]$ (to avoid confusion with true values, we will replace the subscript j by k for estimated values). Then, a true break T_j^0 must occur in this interval with positive probability if one break-point is not consistently estimated. Consequently, there is an η such that with positive probability, $\hat{\theta}_k$ will be estimating θ_j^0 on $I_1(\eta) \in \hat{I}_{k-1}$, but θ_{j+1}^0 on $I_2(\eta) \in \hat{I}_{k-1}$. Hence, with positive probability greater than $\epsilon > 0$,

$$\begin{aligned} T^{-1}\sum_{t=1}^{T} d_t^2 &\geq T^{-1}\sum_{1^*}[f_t(\hat{\theta}_k) - f_t(\theta_j^0)]^2 + T^{-1}\sum_{2^*}[f_t(\hat{\theta}_k) - f_t(\theta_{j+1}^0)]^2 \\ &\geq \inf_{\theta}\left[T^{-1}\sum_{1^*}[f_t(\theta) - f_t(\theta_j^0)]^{2^*} + T^{-1}\sum_{2}[f_t(\theta) - f_t(\theta_{j+1}^0)]^2\right] \end{aligned}$$

Let $d_t(\theta_A, \theta_B) = f_t(\theta_A) - f_t(\theta_B)$, where $\theta_A, \theta_B \in \Theta$. Also, denote $H_{T,i}(\theta) = T^{-1}\sum_{i*} d_t^2(\theta, \theta_{j-1+i}^0)$, for $i = 1, 2$, and $H_T(\theta) = \sum_{i=1,2} H_{T,i}(\theta)$. With this notation ,we can rewrite the last inequality as:

$$T^{-1}\sum_{t=1}^{T} d_t^2 \geq \inf_{\theta} H_T(\theta) \tag{3.8}$$

If we can show that $\inf_\theta H_T(\theta) > C$, for some $C > 0$, with probability one, then $T^{-1}\sum_{t=1}^{T} d_t^2 > C$ with probability $> \epsilon$. Then:

$$\limsup\ P\left[T^{-1}\sum_{t=1}^{T} d_t^2 > C\right] > \epsilon,$$

which establishes Lemma 3.2. Hence, we are left with showing $\inf_\theta H_T(\theta) > C$, which amounts to proving uniform convergence in θ of $H_T(\theta)$ to a positive quantity. We will prove this by means of Proposition 3.2 for $H_{t,i}(\theta)$. Let $h_{t,i}(\theta) = d_t^2(\theta, \theta_{j-1+i}^0)$, $i = 1, 2$. Now $h_{t,i}(\theta)$ is a measurable function of the underlying process v_t. According to A3.1, v_t is a strictly stationary process, hence any measurable function of it is strictly stationary. Moreover, since v_t is α-mixing, any measurable function of it is α-mixing too, of the same size.[6] We are going to state this in the following more general proposition:

Proposition 3.3. *Let $h_t(\theta) = h(v_t, \theta)$, where h is a measurable function. If $\{v_t\}$ is a strictly stationary α-mixing process of size $-a$, then $\{h_t(\theta)\}$ is a strictly stationary α-mixing process of the same size.*

By the above proposition and A3.1, $h_{t,i}(\theta)$ is a strictly stationary α-mixing process of size $-2s/(s-2) - \delta$. Hence, it is also of size $-2s/(s-2)$, because $\delta > 0$. Moreover, $h_{t,i}(\cdot)$ is a continuous function of θ for each t, by the smoothness assumption A3.2. As for the boundedness condition (ii) in ULLN, note that:

$$h_{t,i}(\theta) = d_t^2(\theta, \theta_{j-1+i}^0) = \left[\theta - \theta_{j-1+i}^0\right]' F_t(\bar{\theta}_i) F_t(\bar{\theta}_i)' \left[\theta - \theta_{j-1+i}^0\right]$$

Here $\bar{\theta}_i, i = 1, 2$ are obtained from a mean value theorem. Then:

$$E|h_{t,i}(\theta)|^s \leq \|\theta - \theta_{j-1+i}^0\|^{2s} E\|F_t(\bar{\theta}_i) F_t(\bar{\theta}_i)'\|^s,$$

where the matrix norm was defined as $\|A\| = [\mathrm{tr}(A'A)]^{1/2}$. Since $\theta, \theta_{j-1+i} \in \Theta$, a compact set by A3.4(i), it follows that $\|\theta - \theta_{j+i}^0\|^{2s}$ is bounded uniformly in θ. On the other hand, $\sup_\theta E|F_{t,i}(\theta)|^s$ bounded by A 3.3(iv). If we denote by $F_{t,l}(\bar{\theta}_i)$ the l^{th} element of $F_t(\bar{\theta}_i)$,

$$\begin{aligned} E\|F_t(\bar{\theta}_i) F_t(\bar{\theta}_i)'\|^s &= E\left[\ \mathrm{tr}\left(F_t(\bar{\theta}_i) F_t(\bar{\theta}_i)'\right)^2\right]^{s/2} = E\left[\sum_{l=1}^{p} F_{t,l}^4(\bar{\theta}_i)\right]^{s/2} \\ &\leq E\left[\sum_{l=1}^{p} F_{t,l}^2(\bar{\theta}_i)\right]^s = E\left[\ \mathrm{tr}\left(F_t(\bar{\theta}_i) F_t(\bar{\theta}_i)'\right)\right]^{(1/2)\,2s} \\ &= E\|F_t(\bar{\theta}_i)\|^{2s} < \infty. \end{aligned} \tag{3.9}$$

[6] See White and Domowitz (1984), Lemma 2.1.

It follows that:

$$E|h_{t,i}(\theta)|^s < \infty.$$

Hence, all conditions in ULLN are satisfied, and we can apply it to $H_{T,i}(\theta) = \sum_{i*} h_{t,i}(\theta)$:

$$H_{T,i}(\theta) \xrightarrow{p} H_i(\theta) = \lim T^{-1} \sum_{i*} E\left[h_{t,i}(\theta)\right] \text{ uniformly in } \theta. \tag{3.10}$$

Since sums of uniform convergent processes are themselves uniform convergent, we have:

$$H_T(\theta) = \sum_{i=1,2} H_{T,i} \xrightarrow{p} H(\theta) = \sum_{i=1,2} H_i(\theta) \text{ uniformly in } \theta. \tag{3.11}$$

It remains to show that $\inf_\theta H(\theta) > 0$.

$$\begin{aligned}
H(\theta) &= \lim T^{-1} \sum_{1*} E d_t^2(\theta, \theta_j^0) + \lim T^{-1} \sum_{2*} E d_t^2(\theta, \theta_{j+1}^0) \\
&= \eta \left\{ \, E\, d_t^2(\theta, \theta_j^0) + E\, d_t^2(\theta, \theta_{j+1}^0) \, \right\} \\
&\geq \frac{\eta}{2}\, E\left[f_t(\theta_j^0) - f_t(\theta_{j+1}^0)\right]^2 \\
&\geq \frac{\eta}{2} \left\{ Var\left[f_t(\theta_j^0) - f_t(\theta_{j+1}^0)\right] + \; E^2\left[f_t(\theta_j^0) - f_t(\theta_{j+1}^0)\right] \right\} \\
&\geq \frac{\eta}{2}\, E^2\left[f_t(\theta_j^0) - f_t(\theta_{j+1}^0)\right] \\
&= \frac{\eta}{2} \left\{ \, E\left[f_t(\theta_j^0)\right] - E\left[f_t(\theta_{j+1}^0)\right] \, \right\}^2 > C
\end{aligned}$$

In the inequality above, line 2 follows from strict stationarity of x_t, hence of $d_t^2(\theta, \theta_{j+i}^0)$, for $i = 1, 2$. The first inequality follows from the fact that $a^2 + b^2 \geq (a-b)^2/2$, and the last one from Assumption A3.4(iv). Hence $\inf_\theta H(\theta) > C$. If we substitute this into (3.8), then , with probability greater than ϵ,

$$T^{-1} \sum_{t=1}^{T} d_t^2 \geq \inf_\theta H_T(\theta) > C.$$

□

3.3 Convergence Rates of Break-Fraction Estimates

In the previous section, we showed consistency of break-fraction estimates. This section is devoted to showing that the break-fractions are T-rate consistent.

Since the sum of squared residuals involves differences between sums over $\hat{I}_i$ and I_i^0, it is important that these sums do not become too large. A T-rate of convergence of the break-fractions $\hat{\lambda}_i$ implies that $|\hat{T}_i - T_i^0| < C$ with large probability for some finite $C > 0$. Hence, $\hat{I}_i \setminus I_i^0$ cannot contain more than a bounded number of terms. If these terms in turn are small enough, then the sum of squared residuals converges to the true sum of squares, a prerequisite of good asymptotic properties of NLS estimators.

The rate of convergence for break-fraction estimates can be summarized as follows:

Theorem 3.2.

Under A3.1-A3.5, for any $\epsilon > 0$, there is a $C > 0$ such that: $\limsup P(|T(\hat{\lambda}_i - \lambda_i^0)| > C\ ) < \epsilon$.

Proof of Theorem 3.2.

Step 1.

As in Bai and Perron (1998), without loss of generality, we assume only three breaks. We will focus on proving Theorem 4.1 for $\hat{\lambda}_2$; the analyses for $\hat{\lambda}_1$ and $\hat{\lambda}_3$ are similar. For any $\epsilon > 0$, define:

$$V_\epsilon = \{(T_1, T_2, T_3) : |\ T_i - T_i^0\ | \leq \epsilon T\ (i = 1, 2, 3)\} \tag{3.12}$$

Since $\hat{\lambda}_i \xrightarrow{p} \lambda_i^0$, $\lim P\{|\ \hat{T}_i - T_i^0\ | \leq \epsilon T\} = 1$, hence $\lim P\{(\hat{T}_1, \hat{T}_2, \hat{T}_3) \in V_\epsilon\} = 1$. It follows that we need to examine only the behavior of those breakpoints contained in V_ϵ. Consider, without loss of generality, the case $\hat{T}_2 < T_2^0$.[7] For $C > 0$, define:

$$V_\epsilon(C) = \{(T_1, T_2, T_3) : |\ T_i - T_i^0\ | \leq \epsilon T\ (i = 1, 2, 3);\ T_2^0 - T_2 > C\} \tag{3.13}$$

Note that $V_\epsilon(C) \subset V_\epsilon$. We will show that the probability that the break-points are contained

[7] Bai and Perron (1998) note that the case $\hat{T}_2 \geq T_2^0$ can be handled by a symmetric argument.

in $V_\epsilon(C)$ is very small. By similar arguments, the same results will hold for sets of the type $V_\epsilon(C) \in V_\epsilon$, redefined such that $T_2 - T_2^0 > C$ and $|T_i - T_i^0| > C$, where $i = 1, 3$. Hence, with large probability, $|\hat{T}_i - T_i^0| < C$, for $i = 1, 2, 3$, which is exactly the content of Theorem 3.2. So, for proving the latter theorem, it suffices to show that the break-points will not be contained in $V_\epsilon(C)$ with large probability.

To that end, denote by $S_T(T_1, T_2, T_3)$ the minimized sum of squared residuals for a given 3-break-partition $(1, T_1, T_2, T_3, T)$ of the sample interval. By definition of minimized sum of squared residuals, $S_T(\hat{T}_1, \hat{T}_2, \hat{T}_3) \leq S_T(\hat{T}_1, T_2^0, \hat{T}_3)$. Let $\Delta_2 = T_2 - T_2^0$. We will show that for any $\eta > 0$, we can pick ϵ and C such that:

$$P\left\{\min_{V_\epsilon(C)} \frac{S_T(T_1, T_2, T_3) - S_T(T_1, T_2^0, T_3)}{\Delta_2} \leq 0\right\} < \eta, \text{ for } T \geq T(\eta). \tag{3.14}$$

Equation (3.14) implies that for large T, with probability smaller than η, the break-point estimates pertaining to $V_\epsilon(C)$ satisfy $S_T(\hat{T}_1, \hat{T}_2, \hat{T}_3) \leq S_T(\hat{T}_1, T_2^0, \hat{T}_3)$, or that with a probability larger or equal to $1 - \eta$, $S_T(\hat{T}_1, \hat{T}_2, \hat{T}_3) > S_T(\hat{T}_1, T_2^0, \hat{T}_3)$, contradicting the sum of squares minimization definition. This contradiction implies that $\hat{T}_2$ cannot belong to $V_\epsilon(C)$. Hence, (3.14) is all we need to show so that Theorem 3.2 is verified.

From now on we will implicitly work on $V_\epsilon(C)$, unless indicated otherwise. Define $SSR_1 = S_T(T_1, T_2, T_3)$, $SSR_2 = S_T(T_1, T_2^0, T_3)$ and introduce $SSR_3 = S_T(T_1, T_2, T_2^0, T_3)$. Then:

$$\begin{aligned} S_T(T_1, T_2, T_3) - S_T(T_1, T_2^0, T_3) &= SSR_1 - SSR_2 \\ &= (SSR_1 - SSR_3) - (SSR_2 - SSR_3). \end{aligned}$$

This approach helps us carry out the analysis in terms of two problems involving a single structural change: the first imposing an additional break at T_2^0 between T_2 and T_3, and the second introducing an additional break at T_2 between T_1 and T_2^0. Let $(\theta_1^*, \theta_2^*, \theta_3^{**}, \theta_4^*)$, $(\theta_1^*, \theta_2^{**}, \theta_3^*, \theta_4^*)$ and $(\theta_1^*, \theta_2^*, \theta_2^\delta, \theta_3^*, \theta_4^*)$ be the NLS parameter estimates based on partitions $(1, T_1, T_2, T_3, T)$ respectively $(1, T_1, T_2^0, T_3, T)$ and $(1, T_1, T_2, T_2^0, T_3, T)$. Note that $\theta_2^*, \theta_2^\delta, \theta_2^{**}$ are all estimating θ_2^0, while $\theta_3^*, \theta_3^{**}$ are both estimators of θ_3^0.

In the light of proving (3.14), our goal is to locate the dominating terms in $SSR_1 - SSR_2$ and show that we can pick ϵ and C such they are positive with large probability for large T. To that end, let $V_\epsilon(C)$ be the domain on which some quantity $q_T(\cdot)$ is defined. We will denote by $q_T \sim O_p(T^b)$, if for some $b \in \mathbb{R}$ and any $\eta > 0$, we have:

$$P(|q_T| > T^b) < \eta \text{ for } T \geq T(\eta), \tag{3.15}$$

where T as defined here is large but not necessarily infinite. Note that the statement above depends on the C and ϵ we pick. We will write $q_T \sim O_p^+(T^b)$ if the equation above holds and plim q_T is positive (or positive definite for matrices). Also, we will denote $q_T \sim O_p(T^b) + a_T$, if $q_T - a_T \sim O_p(T^b)$ for some a_T, and $q_T \sim O_p^+(T^b) + a_T$, if $q_T - a_T \sim O_p^+(T^b)$. Similarly, let $q_T \sim o_p(1)$ mean that on $V_\epsilon(C)$, for any $\eta > 0$,

$$P(|q_T| > 0 < \eta) \text{ for } T \geq T(\eta). \tag{3.16}$$

Also denote $q_T \sim o_p(1) + a_T$ if $q_T - a_T \sim o_p(1)$. Under this notation, equation (3.14) is equivalent to:

$$\frac{SSR_1 - SSR_2}{\Delta_2} \sim O_p^+(1) \tag{3.17}$$

because then the probability that $SSR_1 - SSR_2$ is negative is small. So, for proving Theorem 3.2, a proof of (3.17) suffices.

Step 2:

To simplify the notation, let $I_1 = [1, T_1]$, $I_2 = [T_1 + 1, T_2]$, $I_2^\Delta = [T_2 + 1, T_2^0]$, $I_3 = [T_2^0 + 1, T_3]$,

$I_4 = [T_3+1, T]$. Recall that $\Delta_2 = T_2^0 - T_2 > C$. Consider $SSR_1 - SSR_3$ first:

$$
\begin{aligned}
&\frac{SSR_1 - SSR_3}{T_2^0 - T_2} \\
&= \frac{1}{\Delta_2}\{ \sum_{I_1} u_t^2(\theta_1^*) + \sum_{I_2} u_t(\theta_2^*) + \sum_{I_2^\Delta} u_t^2(\theta_3^{**}) + \sum_{I_3} u_t^2(\theta_3^{**}) + \sum_{I_4} u_t^2(\theta_4^*) \; \} \\
&- \frac{1}{\Delta_2}\{ \sum_{I_1} u_t^2(\theta_1^*) + \sum_{I_2} u_t(\theta_2^*) + \sum_{I_2^\Delta} u_t^2(\theta_2^\delta) + \sum_{I_3} u_t^2(\theta_3^*) + \sum_{I_4} u_t^2(\theta_4^*) \; \} \\
&= \frac{1}{\Delta_2}\sum_{I_2^\Delta}[u_t^2(\theta_3^{**}) - u_t^2(\theta_2^\delta)] + \frac{1}{\Delta_2}\sum_{I_3}[u_t^2(\theta_3^{**}) - u_t^2(\theta_3^*)] \\
&= D_1 + D_2.
\end{aligned}
$$

Heuristically speaking, D_1 involves a "mismatch" in estimators, because θ_3^{**} is estimating θ_3^0, while θ_2^δ is estimating θ_2^0. This "mismatch" is not present in D_2, because θ_3^{**} and θ_3^* are both estimating θ_3^0. Hence, D_1 should be dominating D_2 for a large enough $\Delta_2 > C$. We will prove this below.

For $i = 1, \ldots, 4$, in an interval where θ_i^0 is the true parameter value, and $\theta \in \Theta$,

$$
\begin{aligned}
u_t^2(\theta) - u_t^2 &= [y_t - f_t(\theta)]^2 - u_t^2 \\
&= [u_t + f_t(\theta_i^0) - f_t(\theta)]^2 - u_t^2 \\
&= [u_t + f_t(\theta_i^0) - f_t(\theta) - u_t]\,[u_t + f_t(\theta_i^0) - f_t(\theta) + u_t] \\
&= [f_t(\theta_i^0) - f_t(\theta)]\,[f_t(\theta_i^0) - f_t(\theta) + 2u_t] \\
&= [f_t(\theta) - f_t(\theta_i^0)]^2 - 2u_t[f_t(\theta) - f_t(\theta_i^0)] \\
&= d_t^2(\theta, \theta_i^0) - 2u_t\, d_t(\theta, \theta_i^0)
\end{aligned}
$$

Note that the true parameter value on I_2^Δ, the interval on which D_1 is defined, is θ_2^0. Then for any $\theta_A, \theta_B \in \Theta$ and $t \in I_2^\Delta$, $u_t^2(\theta_A) - u_t^2(\theta_B) = d_t^2(\theta_A, \theta_2^0) - d_t^2(\theta_B, \theta_2^0) - 2u_t\, d_t(\theta_A, \theta_B)$.

According to the above, we can write:

$$
\begin{aligned}
D_1 &= \frac{1}{\Delta_2} \sum_{I_2^\Delta} [u_t^2(\theta_3^{**}) - u_t^2(\theta_2^\delta)] \qquad (3.18) \\
&= \frac{1}{\Delta_2} \sum_{I_2^\Delta} d_t^2(\theta_3^{**}, \theta_2^0) - \frac{1}{\Delta_2} \sum_{I_2^\Delta} d_t^2(\theta_2^\delta, \theta_2^0) + \frac{2}{\Delta_2} \sum_{I_2^\Delta} u_t \, d_t(\theta_2^\delta, \theta_3^{**}) \\
&= D_{1,1} - D_{1,2} + D_{1,3}.
\end{aligned}
$$

We will find the order of each of the terms above. In Ch. 3, Proof of Lemma 3.2, we have shown that strictly stationary processes such as $\{d_t^2(\theta, \theta_2^0)\}$ satisfy the ULLN. In other words, if we pick C large enough such that the interval I_2^Δ is large enough, then, uniformly in θ,

$$
H_T(\theta) = \frac{1}{\Delta_2} \sum_{I_2^\Delta} d_t^2(\theta, \theta_2^0) - \lim E\left[d_t^2(\theta, \theta_2^0)\right] \sim o_p(1) \qquad (3.19)
$$

Hence, as long as we pick C large enough,

$$
D_{1,1} = \frac{1}{\Delta_2} \sum_{I_2^\Delta} d_t^2(\theta_3^{**}, \theta_2^0) - \text{ plim}\left[d_t^2(\theta_3^{**}, \theta_2^0)\right] \sim o_p(1) \qquad (3.20)
$$

$$
D_{1,2} = \frac{1}{\Delta_2} \sum_{I_2^\Delta} d_t^2(\theta_2^\delta, \theta_2^0) - \text{ plim}\left[d_t^2(\theta_2^\delta, \theta_2^0)\right] \sim o_p(1) \qquad (3.21)
$$

We need to find the probability limits above. For that, we need to consider the properties of θ_3^{**}, an estimator of θ_3^0. It is obtained by minimizing the sum of squared residuals on the interval $I_3 = [T_2^0, T_3]$. This differs from the true interval $I_3^0 \equiv [T_2^0, T_3^0]$ by ϵT terms. If we estimated θ_3^0 on I_3^0, we would obtain the usual NLS estimator, within $o_p(1)$ of the true value θ_3^0.

Case 1 $T_3 \leq T_3^0$. Then $I_3 \subset I_3^0$, and θ_3^{**} is simply a sub-sample estimator of θ_3^0. As long as ϵ is small enough, the sub-sample is large enough to ensure that $\theta_3^{**} - \theta_3^0 \sim O_p(T^{-1/2})$.

Case 2 $T_3 > T_3^0$. Let $I_3^\epsilon = [T_3^0 + 1, T_3]$. Then:

$$\theta_3^{**} = \underset{\theta}{\operatorname{argmin}}[\frac{1}{\Delta_2}\sum_{I_3^0} u_t^2(\theta) + \frac{1}{\Delta_2}\sum_{I_3^\epsilon} u_t(\theta)^2]$$

Let l_3 be the length of the interval I_3^ϵ. Because $\{u_t^2(\theta)\}$ is strictly stationary, α-mixing by A3.1(i) process, a continuous function by A3.2 and uniformly bounded by A3.3(ii), by ULLN, we have:

$$0 \leq \frac{1}{\Delta_2}\sum_{I_3^\epsilon} u_t(\theta)^2 \leq \frac{l_3}{C}E[u_t^2(\theta)] + o_p(1)$$

Since $l_3 \leq T\epsilon$, if ϵ is small enough compared to C, then l_3/C is close to 0. Hence,

$$\min_\theta[\frac{1}{\Delta_2}\sum_{I_3^0} u_t^2(\theta)] \leq \min_\theta[\frac{1}{\Delta_2}\sum_{I_3^0} u_t^2(\theta) + \frac{1}{\Delta_2}\sum_{I_3^\epsilon} u_t^2(\theta)]$$
$$\leq \min_\theta[\frac{1}{\Delta_2}\sum_{I_3^0} u_t^2(\theta)] + o_p(1)$$

So,

$$\min_\theta[\frac{1}{\Delta_2}\sum_{I_3^0} u_t^2(\theta) + \frac{1}{\Delta_2}\sum_{I_3^\epsilon} u_t(\theta)^2] \leq \min_\theta[\frac{1}{\Delta_2}\sum_{I_3^0} u_t^2(\theta)] + o_p(1)$$

Hence, θ_3^{**} is asymptotically equivalent to the NLS estimator on I_3^0:

$$\theta_3^{**} \sim o_p(1) + \underset{\theta}{\operatorname{argmin}}[\frac{1}{\Delta_2}\sum_{I_3^0} u_t^2(\theta)] \sim \theta_3^0 + O_p(T^{-1/2})$$

So, in both cases, if we pick ϵ small enough, the misspecification of the estimation interval will be small, hence so will be the deviation of θ_3^{**} from the properties of the usual NLS estimator. Hence, we can pick ϵ small enough so that:

$$\theta_3^{**} - \theta_3^0 \sim O_p(T^{-1/2}) = o_p(1) \tag{3.22}$$

So $f_t(\theta_3^{**}) - f_t(\theta_3^0) \sim o_p(1)$. Also, $\theta_3^{**} - \theta_2^0 = (\theta_3^{**} - \theta_3^0) + (\theta_3^0 - \theta_2^0) \sim o_p(1) + (\theta_3^0 - \theta_2^0)$. Hence:

$$\begin{aligned} d_t^2(\theta_3^{**}, \theta_2^0) &= \left[f_t(\theta_3^{**}) - f_t(\theta_3^0)\right]^2 = \left[\left\{f_t(\theta_3^{**}) - f_t(\theta_2^0)\right\} - \left\{f_t(\theta_3^0) - f_t(\theta_2^0)\right\}\right]^2 \\ &= \left[f_t(\theta_3^{**}) - f_t(\theta_3^0)\right]^2 + \left[f_t(\theta_3^0) - f_t(\theta_2^0)\right]^2 \\ &- 2\left[f_t(\theta_3^{**}) - f_t(\theta_3^0)\right]\left[f_t(\theta_3^0) - f_t(\theta_2^0)\right] \end{aligned}$$

We can write:

$$d_t^2(\theta_3^{**}, \theta_2^0) \sim o_p(1) + E\left[f_t(\theta_3^0) - f_t(\theta_2^0)\right]^2 + o_p(1),$$

or plim $d_t^2(\theta_3^{**}, \theta_2^0) = E\left[f_t(\theta_3^0) - f_t(\theta_2^0)\right]^2 = \mathrm{Var}\left[f_t(\theta_3^0) - f_t(\theta_2^0)\right] + \{\ E\left[f_t(\theta_3^0)\right] - E\left[f_t(\theta_2^0)\right]\ \}^2 > 0$. The last inequality follows from A3.4(iv). Combining this with (3.20) yields:

$$D_{1,1} = \frac{1}{\Delta_2} \sum_{I_2^\Delta} d_t^2(\theta_3^{**}, \theta_2^0) \ \sim O_p^+(1). \tag{3.23}$$

We will later see that this is the only positive dominating term we are looking for in $SSR_1 - SSR_2$.

For analyzing $D_{1,2}$, we need to take a close look at the properties of θ_2^δ. The properties of θ_2^δ, an estimator of θ_2^0, depend on the length of the sample interval, $I_2^\Delta = [T_2, T_2^0]$. Since its length is $C <$ length $I_2^\Delta < \epsilon T$, for any large T, we can choose ϵ such that $I_2^\Delta \subset [T_1^0, T_2^0]$. In this case, θ_2^δ will be a sub-sample NLS estimator of θ_2^0. If we pick C big enough, the sub-sample is large enough to ensure that:

$$\theta_2^\delta - \theta_2^0 \sim o_p(1).$$

Hence $d_t^2(\theta_2^\delta, \theta_2^0) = \left[f_t(\theta_2^\delta) - f_t(\theta_2^0)\right]^2 \sim o_p(1)$. Combining this with (3.21) yields:

$$D_{1,2} = \frac{1}{\Delta_2} \sum_{I_2^\Delta} d_t^2(\theta_2^\delta, \theta_2^0) \ \sim o_p(1). \tag{3.24}$$

For analyzing $D_{1,3} = \frac{2}{\Delta_2} \sum_{I_2^\Delta} u_t\ d_t(\theta_2^\delta, \theta_3^{**}) = \frac{2}{\Delta_2} \sum_{I_2^\Delta} u_t\ \left[f_t(\theta_2^\delta) - f_t(\theta_3^{**})\right]$, we need uniform convergence of infinite sums of $\{u_t f_t(\theta)\}$. But we have already shown in the proof of Lemma 3.1 that, uniformly in $\theta \times r$,

$$T^{-1} \sum_{t=1}^{Tr} u_t f_t(\theta) = o_p(1)$$

As a special case, if C is big enough, we have:

$$D_{1,3} = \frac{2}{\Delta_2} \sum_{I_2^{\Delta}} u_t \; f_t(\theta_2^{\delta}) - \frac{2}{\Delta_2} \sum_{I_2^{\Delta}} u_t \; f_t(\theta_3^{**}) \sim o_p(1) + o_p(1) = o_p(1) \tag{3.25}$$

From (3.23) - (3.25), it follows that for large C and small ϵ,

$$D_1 = D_{1,1} - D_{1,2} + D_{1,3} \sim O_p^+(1) - o_p(1) + o_p(1) = O_p^+(1) \tag{3.26}$$

For deriving the order of D_2:

$$D_2 = \frac{1}{\Delta_2} \sum_{I_3} \left[u_t^2(\theta_3^{**}) - u_t^2(\theta_3^{*}) \right] \tag{3.27}$$

in a similar fashion to D_1, we have to consider two cases that indicate different true parameter values in interval I_3. To that end, note that $I_3 = [T_2^0 + 1, T_3]$. If $T_3 \leq T_3^0$, then $I_3 \subset I_3^0 = [T_2^0 + 1, T_3^0]$, and the true parameter value on I_3 is θ_3^0. Otherwise, $I_3 = I_3^0 \cup [T_3^0 + 1, T_3]$, where the true parameter value is θ_3^0 on I_3^0, but θ_4^0 on $I_3^{\epsilon} = [T_3^0 + 1, T_3]$.

Case 1: $T_3 \leq T_3^0$. Here, D_2 can be rewritten as:

$$\begin{aligned} D_2 &= \frac{1}{\Delta_2} \sum_{I_3} \left[u_t^2(\theta_3^{**}) - u_t^2(\theta_3^{*}) \right] & (3.28) \\ &= \frac{1}{\Delta_2} \sum_{I_3} d_t^2(\theta_3^{**}, \theta_3^0) \; - \; \frac{1}{\Delta_2} \sum_{I_3} d_t^2(\theta_3^{*}, \theta_3^0) \; + \; \frac{2}{\Delta_2} \sum_{I_3} u_t \; d_t(\theta_3^{*}, \theta_3^{**}) \\ &= D_{2,1} - D_{2,2} + D_{2,3} \end{aligned}$$

By (3.22), for small enough ϵ, $\theta_3^{**} - \theta_3^0 \sim O_p(T^{-1/2})$. Similarly, $\theta_3^{*} - \theta_3^0 \sim O_p(T^{-1/2})$. Let $\breve{\theta}_3 = T^{1/2}\left[\theta_3^{**} - \theta_3^0\right] \sim O_p(1)$. The analyses of $D_{2,1}$ and $D_{2,2}$ will be slightly different than for $D_{1,1}$ and $D_{1,2}$ given that we are summing over a different interval. Since we will use MVT repeatedly, we will standardize its notation. For each t, consider θ_i^k and θ_j^s, where $i, j \in \{1, 2, \ldots, m+1\}$, and $k \in \{0, *, **, \delta\}$. We will denote by $\tilde{\theta}_{t,i,j}^{k,s} = \tilde{\lambda}_{t,i,j}^{k,s}\, \theta_i^k + [1 - \tilde{\lambda}_{t,i,j}^{k,s}]\, \theta_j^s$, where $\tilde{\lambda}_{t,i,j}^{k,s} \in (0,1)$, such that $d_t(\theta_i^k, \theta_j^s) = [\theta_i^k - \theta_j^s]' F_t(\tilde{\theta}_{t,i,j}^{k,s})$. For example, an expansion of $f_t(\theta_3^{**})$ around $f_t(\theta_3^0)$ yields $\tilde{\theta}_{t,3,3}^{**,0}$.

Under this notation,

$$
\begin{aligned}
D_{2,1} &= \breve{\theta}_3{}' \left[\frac{1}{T\Delta_2} \sum_{I_3} F_t(\tilde{\theta}_{t,3,3}^{**,0}) F_t(\tilde{\theta}_{t,3,3}^{**,0})' \right] \breve{\theta}_3 \\
&\leq \|\breve{\theta}_3\|^2 \left[\frac{T_3 - T_2^0}{T\Delta_2} \right] \|F_t(\tilde{\theta}_{t,3,3}^{**,0}) F_t(\tilde{\theta}_{t,3,3}^{**,0})'\| \\
&\leq \|\breve{\theta}_3\|^2 \left[\frac{\lambda_3^0 - \lambda_2^0}{C} \right] \sup_{\theta,t} \|F_t(\theta) F_t(\theta)'\|
\end{aligned}
$$

By (3.9), $\sup_{\theta,t} \|F_t(\theta)F_t(\theta)'\|$ is bounded a.s. Hence:

$$
D_{2,1} \sim O_p(1) \frac{1}{C} O_p(1) \;=\; \frac{1}{C} O_p(1) \tag{3.29}
$$

By similar arguments,

$$
D_{2,2} \sim \frac{1}{C} O_p(1) \tag{3.30}
$$

Now:

$$
\begin{aligned}
D_{2,3} &= T^{1/2}[\theta_3^* - \theta_3^{**}]' \left[\frac{2}{T\Delta_2} \sum_{I_3} u_t\, F_t(\tilde{\theta}_{t,3,3}^{*,**}) \right] \\
&\sim O_p(1) \frac{2}{\Delta_2}\, T^{-1/2} \sum_{I_3} u_t\, F_t(\tilde{\theta}_{t,3,3}^{*,**})
\end{aligned}
$$

Recall that the length of I_3 is $T_3 - T_2^0$. Let $T_3 = [T\lambda_3]$, for $\lambda_3 \in (0,1)$, and $\Psi_t(\theta) = u_t F_t(\theta)$.

Then:

$$
\begin{aligned}
\|T^{-1/2}\sum_{I_3}\Psi_t(\tilde{\theta}^{*,**}_{t,3,3})\| &= T^{-1/2}\|\sum_{t=[T\lambda_2^0]}^{[T\lambda_3]}\Psi_t(\tilde{\theta}^{*,**}_{t,3,3})\| \\
&\leq T^{-1/2}\left\|\sum_{t=1}^{[T\lambda_3]}\Psi_t(\tilde{\theta}^{*,**}_{t,3,3}) - \sum_{t=1}^{[T\lambda_2^0]}\Psi_t(\tilde{\theta}^{*,**}_{t,3,3})\right\| \\
&\leq T^{-1/2}\left\|\sum_{t=1}^{[T\lambda_3]}\Psi_t(\tilde{\theta}^{*,**}_{t,3,3})\right\| + T^{-1/2}\left\|\sum_{t=1}^{[T\lambda_2^0]}\Psi_t(\tilde{\theta}^{*,**}_{t,3,3})\right\| \\
&\leq 2\sup_r\, T^{-1/2}\sum_{t=1}^{[Tr]}\sup_\theta\|\Psi_t(\theta)\|
\end{aligned}
$$

Now $\{\sup_\theta\|\Psi_t(\theta)\|\}$ exists a.s. by A3.3(iv). It is a strictly stationary α-mixing process of size $-2s/(s-2)-\delta$, because $\{v_t\} = \{(x_t, u_t)\}$ is, by A3.1 (i). Also, $\sup_\theta\|\Psi_t(\theta)\|^s = |u_t|^s\sup_\theta\|F_t(\theta)\|^s$ exists a.s., because of the bounds assumed on u_t and F_t in A3.3(ii) and (iv). Hence, $E\sup_\theta\|\Psi_t(\theta)\| < \infty$. If we set $s = 2+\delta$, then the assumptions of Proposition 3.1 are satisfied, hence:

$$\sup_r\, T^{-1/2}\sum_{t=1}^{[Tr]}\sup_\theta\|\Psi_t(\theta)\| = O_p(1). \tag{3.31}$$

Hence,

$$T^{-1/2}\sum_{I_3}\Psi_t(\tilde{\theta}^{*,**}_{t,3,3}) = O_p(1), \tag{3.32}$$

allowing us to write:

$$
\begin{aligned}
D_{2,3} &\sim O_p(1)\frac{2}{\Delta_2}\, T^{-1/2}\sum_{I_3}\Psi_t(\tilde{\theta}^{*,**}_{t,3,3}) \\
&\leq\sim O_p(1)\frac{1}{C}O_p(1) \sim \frac{1}{C}O_p(1).
\end{aligned}
$$

Hence, the order of D_2 is:

$$D_2 \;=\; D_{2,1} - D_{2,2} + D_{2,3} \sim \frac{1}{C}O_p(1) - \frac{1}{C}O_p(1) + \frac{1}{C}O_p(1) \;=\; \frac{1}{C}O_p(1) \tag{3.33}$$

Case 2: $T_3 > T_3^0$. Here, D_2 can be rewritten as:

$$D_2 = \frac{1}{\Delta_2} \sum_{I_3^0} \left[u_t^2(\theta_3^{**}) - u_t^2(\theta_3^*) \right] \; + \; \frac{1}{\Delta_2} \sum_{I_3^\epsilon} \left[u_t^2(\theta_3^{**}) - u_t^2(\theta_3^*) \right] \tag{3.34}$$
$$\sim \frac{1}{C} O_p(1) + D_2^*,$$

because $\frac{1}{\Delta_2} \sum_{I_3^0} \left[u_t^2(\theta_3^{**}) - u_t^2(\theta_3^*) \right] \sim \frac{1}{C} O_p(1)$ by similar arguments as for case 1 above. The difference is the presence of D_2^*. For $t \in I_3^\epsilon$,

$$\begin{aligned} u_t^2(\theta_3^{**}) - u_t^2(\theta_3^*) &= [y_t - f_t(\theta_3^{**})]^2 \; - \; [y_t - f_t(\theta_3^*)]^2 \\ &= [f_t(\theta_3^*) - f_t(\theta_3^{**})] \; \{2u_t + 2f_t(\theta_4^0) - [f_t(\theta_3^*) + f_t(\theta_3^{**})]\} \\ &= -d_t^2(\theta_3^*, \theta_3^{**}) \; + \; 2d_t(\theta_3^*, \theta_3^{**}) \, f_t(\theta_4^0) \; + \; 2u_t \, d_t(\theta_3^*, \theta_3^{**}) \end{aligned}$$

Hence,

$$\begin{aligned} D_2^* &= -\frac{1}{\Delta_2} \sum_{I_3^\epsilon} d_t^2(\theta_3^*, \theta_3^{**}) + \frac{2}{\Delta_2} \sum_{I_3^\epsilon} d_t(\theta_3^*, \theta_3^{**}) \, f_t(\theta_4^0) + \frac{2}{\Delta_2} \sum_{I_3^\epsilon} u_t \, d_t(\theta_3^*, \theta_3^{**}) \\ &= -D_{2,1}^* + D_{2,2}^* + D_{2,3}^* \end{aligned}$$

By similar arguments as for $D_{2,1}$ in case 1, $D_{2,1}^*$ can be written as:

$$\begin{aligned} D_{2,1}^* &= T^{1/2} \, [\theta_3^* - \theta_3^{**}]' \; \left[\frac{1}{T\Delta_2} \sum_{I_3^\epsilon} F_t(\tilde{\theta}_{t,3,3}^{*,**}) F_t(\tilde{\theta}_{t,3,3}^{*,**})' \right] \; T^{1/2} \, [\theta_3^* - \theta_3^{**}] \\ &\sim O_p(1) \; \frac{\epsilon}{C} \; \sup_{\theta,\, t} \|F_t(\theta) F_t(\theta)'\| \sim \frac{\epsilon}{C} O_p(1) \end{aligned}$$

On the other hand, $D_{2,2}^*$ is:

$$D_{2,2}^* = T^{1/2} \, [\theta_3^* - \theta_3^{**}]' \; \left[\frac{1}{T^{1/2}\Delta_2} \sum_{I_3^\epsilon} F_t(\tilde{\theta}_{t,3,3}^{*,**}) \; f_t(\theta_4^0) \right]$$

On the other hand, we can show that:

$$\|\frac{1}{T^{1/2}\Delta_2}\sum_{I_3^\epsilon} F_t(\tilde{\theta}^{*,**}_{t,3,3}) f_t(\theta_4^0)\| \leq \frac{l_3}{T^{1/2}C}\sup_{\theta,t}\|F_t(\theta)\|\,|f_t(\theta_4^0)|$$

Now, $F_t(\theta)$ is bounded a.s. by A3.3(iv). On the other hand, $u_t f_t(\theta)$ and u_t are bounded a.s. by A3.3(ii) and (iii). Hence $f_t(\theta)$ is bounded a.s. Also, we can make ϵ small enough compared to C such that the number of terms l_3 in I_3^ϵ is small. Hence:

$$D^*_{2,2} \sim O_p(1)\frac{1}{C}O_p(1) = \frac{1}{C}O_p(1)$$

By a similar analysis to $D_{2,3}$ in case 1, we have:

$$D^*_{2,3} \sim \frac{1}{C}O_p(1)$$

Hence, for ϵ small enough, the order of D_2^* is:

$$D_2^* = -D^*_{2,1} + D^*_{2,2} + D^*_{2,3} \sim -\frac{\epsilon}{C}O_p(1) + \frac{1}{C}O_p(1) + \frac{1}{C}O_p(1) \sim \frac{1}{C}O_p(1)$$

Finally, the order of D_2 in case 2, when ϵ is small, is:

$$D_2 \sim \frac{1}{C}O_p(1) + \frac{1}{C}O_p(1) = \frac{1}{C}O_p(1) \tag{3.35}$$

From (3.33) and (3.35), we conclude that in both cases above,

$$D_2 \sim \frac{1}{C}O_p(1) \tag{3.36}$$

This result, together with (3.26), determine the order of $SSR_1 - SSR_3$. When picking a small

ϵ and a large C,

$$\frac{SSR_1 - SSR_3}{T_2^0 - T_2} = D_1 + D_2 \sim O_p^+(1) + \frac{1}{C} O_p(1) = O_p^+(1) \tag{3.37}$$

Step 3:

In this step we will show that the order of $SSR_2 - SSR_3$ is smaller than the order of $SSR_1 - SSR_3$. To that end, note:

$$\begin{aligned}
&\frac{SSR_2 - SSR_3}{T_2^0 - T_2} \\
&= \frac{1}{\Delta_2}\{ \sum_{I_1} u_t^2(\theta_1^*) + \sum_{I_2} u_t(\theta_2^{**}) + \sum_{I_2^\Delta} u_t^2(\theta_2^{**}) + \sum_{I_3} u_t^2(\theta_3^*) + \sum_{I_4} u_t^2(\theta_4^*) \} \\
&- \frac{1}{\Delta_2}\{ \sum_{I_1} u_t^2(\theta_1^*) + \sum_{I_2} u_t(\theta_2^*) + \sum_{I_2^\Delta} u_t^2(\theta_2^\delta) + \sum_{I_3} u_t^2(\theta_3^*) + \sum_{I_4} u_t^2(\theta_4^*) \} \\
&= \frac{1}{\Delta_2} \sum_{I_2} [u_t^2(\theta_2^{**}) - u_t^2(\theta_2^*)] + \frac{1}{\Delta_2} \sum_{I_2^\Delta} [u_t^2(\theta_2^{**}) - u_t^2(\theta_2^\delta)] \\
&= D_3 + D_4.
\end{aligned}$$

Noticing that both θ_2^{**} and θ_2^* are estimators of θ_2^0, it is immediately apparent that D_3 looks exactly like D_2. Hence, by similar arguments, for small ϵ,

$$D_3 \sim \frac{1}{C} O_p(1). \tag{3.38}$$

On the other hand, D_4 is not very similar to D_1, because it does not involve a "mismatch" between estimators (i.e. both θ_2^{**} and θ_2^δ are estimating θ_2^0). We can write D_4 as:

$$\begin{aligned} D_4 &= \frac{1}{\Delta_2}\sum_{I_2^\Delta}[u_t^2(\theta_2^{**}) - u_t^2(\theta_2^\delta)] \\ &= \frac{1}{\Delta_2}\sum_{I_2^\Delta} d_t^2(\theta_2^{**}, \theta_2^0) - \frac{1}{\Delta_2}\sum_{I_2^\Delta} d_t^2(\theta_2^\delta, \theta_2^0) + \frac{1}{\Delta_2}\sum_{I_2^\Delta} u_t d_t(\theta_2^\delta, \theta_2^{**}) \\ &= D_{4,1} - D_{4,2} + D_{4,3} \end{aligned}$$

Note that, for a large enough C and small enough ϵ, $\theta_2^{**} - \theta_2^0$ and $\theta_2^\delta - \theta_2^0$ are $\sim o_p(1)$. Applying ULLN in a similar fashion as for $D_{1,2}$, we get:

$$D_{4,1} \sim o_p(1) \text{ and } D_{4,2} \sim o_p(1)$$

By similar arguments as for $D_{1,3}$, we have:

$$D_{4,3} \sim o_p(1)$$

It follows that:

$$D_4 = D_{4,1} - D_{4,2} + D_{4,3} \sim o_p(1) \tag{3.39}$$

Combining (3.38) with (3.39) yields:

$$\frac{SSR_2 - SSR_3}{\Delta_2} = D_3 + D_4 \sim \frac{1}{C}O_p(1) + o_p(1) = \frac{1}{C}O_p(1), \tag{3.40}$$

if we pick C large enough and ϵ small enough.

Hence, we can combine (3.37) in Step 2 with (3.40) in Step 3 to get:

$$\begin{aligned} \frac{SSR_1 - SSR_2}{\Delta_2} &= \frac{SSR_1 - SSR_3}{\Delta_2} - \frac{SSR_2 - SSR_3}{\Delta_2} \\ &\sim O_p^+(1) - \frac{1}{C}O_p(1) = O_p^+(1), \end{aligned}$$

provided that C is large enough and ϵ small enough. This is in fact (3.17), hence the proof of Theorem 3.2 is complete. $\square$

Chapter 4

Asymptotic Distribution of Parameter Estimates

In the previous chapter, we have found the convergence rate of the break-fraction estimates. To derive the asymptotic distributions of the NLS parameter estimates, we will need to show that they can be written in a similar fashion to (2.15). Section 1 shows that the parameter estimates are consistent. Section 2 derives their asymptotic distribution, which is the same as if the change-points were known. It also discusses specific cases and the estimation of the covariance matrix in finite samples.

4.1 Consistency of Parameter Estimates

In this section, we are going to show that the parameter estimates are consistent. This is stated in the following theorem:

Theorem 4.1. *Under A3.1-A3.5, for $i = 1, \ldots, m+1$, we have $\hat{\theta}_i \xrightarrow{p} \theta_i^0$.*

To prove the theorem above, we will first resort to showing an additional lemma. As a matter of notation, consider some partition of the interval $[1, T]$, denoted $(1, T_1, \ldots, T_m, T)$. Let $S_{T,I_i}(\theta) = T^{-1} \sum_{t=T_{i-1}}^{T_i} u_t^2(\theta)$ be the partial sum of squares in interval $I_i = [T_{i-1}+1, T_i]$, for $i = 1, \ldots, m+1$. Hence, we will denote by $S_{T,I_i^0}(\theta)$ and $S_{T,\hat{I}_i}(\theta)$ the sum of squares in interval $I_i^0 = [T_{i-1}+1, T_i]$, respectively $\hat{I} = [\hat{T}_{i-1}+1, \hat{T}_i]$. Since θ_i^0 is the true value in interval

I_i^0,

$$S_{T,I_i^0}(\theta) = T^{-1}\sum_i u_t^2(\theta) = T^{-1}\sum_i u_t^2 + T^{-1}\sum_i d_t^2(\theta,\theta_i^0) + 2T^{-1}\sum_i u_t\, d_t(\theta,\theta_i^0)$$

The sum over i has been defined in the proof of Lemma 3.1, and means summing over I_i^0. Let $S_{I_i^0}(\theta) = \operatorname{plim} T^{-1}\sum_i u_t^2 + \operatorname{plim} T^{-1}\sum_i d_t^2(\theta,\theta_i^0)$. Since $\{u_t\}$ is a strictly stationary α-mixing process of size $-2s/(s-2)-\delta$, and $E|u_t|^s < \infty$, this process obeys the ULLN in chapter 3 (note that this is a special case since u_t is not a function of θ, hence uniformity in θ is not needed). Let $E[u_t^2] = \sigma^2$ and $\Delta_i^0 = \lambda_i^0 - \lambda_{i-1}^0$, for $i = 1, \dots, m+1$. Hence,

$$T^{-1}\sum_i u_t^2 \xrightarrow{p} \Delta_i^0\, \sigma^2 \tag{4.1}$$

Similarly, we have seen that $\{d_t^2(\theta,\theta_i^0)\}$ obeys the ULLN. Let $h(\theta) = E\left[d_t^2(\theta,\theta_i^0)\right]$. Then:

$$T^{-1}\sum_i d_t^2(\theta,\theta_i^0) \xrightarrow{p} \Delta_i^0\, h(\theta) \tag{4.2}$$

Hence:

$$T^{-1}S_{I_i^0}(\theta) = \operatorname{plim} T^{-1}\sum_i u_t^2 + \operatorname{plim} T^{-1}\sum_i d_t^2(\theta,\theta_i^0) = \Delta_i^0\left[\sigma^2 + h(\theta)\right] \tag{4.3}$$

Having this in mind, we will show the following lemma:

Lemma 4.1. *$S_{T,I_i^0}(\theta) \xrightarrow{p} S_{I_i^0}(\theta)$, uniformly in θ.*

Proof of Lemma 4.1: Because of (4.3), all we need to prove is that $T^{-1}\sum_i u_t\, d_t(\theta,\theta_i^0) \xrightarrow{p} 0$, uniformly in θ. This is an exercise similar to the proof of Lemma 3.1. Recall that $\psi_t(\theta) = u_t f_t(\theta)$. Then, by Proposition 3.1, $\sum_{t=1}^{[Tr]} \psi_t(\theta) \xrightarrow{p} 0$ uniformly in $\theta \times r$. Hence, for fixed $r = \lambda_i^0$, respectively $r = \lambda_{i-1}^0$, we have, uniformly in θ,

$$T^{-1}\sum_{t=1}^{[T\lambda_i^0]} \psi_t(\theta) \xrightarrow{p} 0 \quad \text{and } T^{-1}\sum_{t=1}^{[T\lambda_{i-1}^0]} \psi_t(\theta) \xrightarrow{p} 0$$

So, uniformly in θ,

$$T^{-1} \sum_{i} \psi_t(\theta) = T^{-1} \sum_{t=1}^{[T\lambda_i^0]} \psi_t(\theta) - T^{-1} \sum_{t=1}^{[T\lambda_{i-1}^0]} \psi_t(\theta) \xrightarrow{p} 0 \tag{4.4}$$

Hence, $T^{-1} \sum_i \psi_t(\theta_i^0) \xrightarrow{p} 0$, so $T^{-1} \sum_i u_t \, d_t(\theta, \theta_i^0) \xrightarrow{p} 0$, uniformly in θ, completing the proof of Lemma 4.1. □

We are now in place to prove Theorem 4.1.

Proof of Theorem 4.1. We are first going to show that $S_{T,\hat{I}_i}(\theta) - S_{T,I_i^0}(\theta) = o_p(1)$, uniformly in $\theta \in \Theta$. Let $\hat{I}_i \nabla I_i^0 = (\hat{I}_i \backslash I_i^0) \cup (I_i^0 \backslash \hat{I}_i)$, and define as indicator function $\iota_i : \hat{I}_i \nabla I_i^0 \to \{-1, 1\}$, where $\iota_i(t) = \iota_{i,t} = 1$, if $t \in \hat{I}_i \setminus I_i^0$, and $\iota_{i,t} = -1$, if $t \in I_i^0 \setminus \hat{I}_i$. Then we can write:

$$\begin{aligned} S_{T,\hat{I}_i}(\theta) - S_{T,I_i^0}(\theta) &= T^{-1} \sum_{\hat{i}} u_t^2(\theta) - T^{-1} \sum_{i} u_t^2(\theta) \\ &= \sum_{\hat{I}_i \nabla I_i^0} \iota_{i,t} \left[T^{-1} u_t^2\right] + \sum_{\hat{I}_i \nabla I_i^0} \iota_{i,t} \left[T^{-1} d_t^2(\theta, \theta_i^0)\right] + \\ &+ \sum_{\hat{I}_i \nabla I_i^0} \iota_{i,t} \left[T^{-1} u_t \, d_t(\theta, \theta_i^0)\right] \end{aligned}$$

By Theorem 3.2, there can be no more than $2C$ integer values contained in $\hat{I}_i \nabla I_i^0$. Hence, the sums above involve a bounded number of terms. These terms have the property that $\operatorname{plim} T^{-1} u_t^2 = \lim T^{-1} \sigma^2 = 0$, $\operatorname{plim} T^{-1} d_t^2(\theta, \theta_i^0) = \lim T^{-1} h_t(\theta) = 0$, uniformly in θ, and $\operatorname{plim} T^{-1} u_t \, d_t(\theta, \theta_i^0) = T^{-1} 0 = 0$, uniformly in θ. Hence, $S_{T,\hat{I}_i}(\theta) - S_{T,I_i^0}(\theta)$ involves sums of a bounded number of terms that are $o_p(1)$ uniformly in θ. Hence, uniformly in θ,

$$S_{T,\hat{I}_i}(\theta) - S_{T,I_i^0}(\theta) = o_p(1)$$

Combining the above with Lemma 4.1 yields:

$$S_{T,\hat{I}_i}(\theta) \xrightarrow{p} S_{I_i^0}(\theta), \text{ uniformly in } \theta \in \Theta \tag{4.5}$$

Now Θ is compact by Assumption A3.3(i). It follows that $\{\hat{\theta}_i = \hat{\theta}_{i,T}\}_{T=1}^{\infty}$ has at least one limit point $\vartheta_i \in \Theta$. Moreover, there exists at least one subsequence of $\{\hat{\theta}_{i,T}\}$, call it $\{\hat{\theta}_{i,T_k^i}\}$, such that $\hat{\theta}_{i,T_k^i} \xrightarrow{p} \vartheta_i$ as $k \to \infty$. By uniform convergence of $T^{-1}S_{T,\hat{I}_i}(\theta)$ - see (4.5) - it follows that:

$$T^{-1}S_{T,\hat{I}_i}(\hat{\theta}_{i,T_k^i}) \xrightarrow{p} S_{I_i^0}(\vartheta_i) \leq S_{I_i^0}(\theta_i^0), \tag{4.6}$$

where the inequality follows from the fact that $T^{-1}S_{T,\hat{I}_i}(\hat{\theta}_{i,T_k^i})$ is the minimized sum of squared residuals. But θ^0 is the unique minimizer of $S(\theta)$ by A3.4(i), hence θ_i^0 are the unique minimizers of $S_{I_i^0}(\theta)$. This result and equation (4.6) imply that $\vartheta_i = \theta_i^0$. Hence, we have one and only one limit point of $\hat{\theta}_{i,T}$ in the compact space Θ, namely θ_i^0. It follows that every subsequence of $\hat{\theta}_{i,T}$ must converge to the same limit θ_i^0. Hence $\hat{\theta}_i \xrightarrow{p} \theta_i^0$ for any $i = 1, \ldots, m+1$, completing the consistency proof. □

4.2 Asymptotic Normality of Parameter Estimates

Theorems 3.2 and 4.1 allow us to show that the parameter estimates are asymptotically normal. Moreover, we will show in this section that their distribution is the same as if the break-points were known, a common result in the literature.

To define this result formally, recall that plim $A_T(\theta) =$ Var $[T^{-1/2}\sum_{t=1}^{T} u_t\, F_t(\theta)] = A(\theta)$, exists and is positive definite by Assumption A3.4(ii) in Chapter 3. For $i = 1, \ldots, m+1$, $\bar{A}_i =$ plim Var $[T^{-1/2}\sum_i u_t\, F_t(\theta_i^0)]$[1] Let $\bar{D}^{-1}$ be the $(m+1)p \times (m+1)p$ block-diagonal matrix whose $i - i^{th}$ $p \times p$ block is $\bar{D}_i^{-1} = [\Delta_i D(\theta_i^0)]^{-1}$, with $D(\theta)$ the positive definite matrix defined in Assumption **??**(iii). Moreover, let $\Phi_i = \bar{D}_i^{-1}\bar{A}_i\bar{D}_i^{-1}$ and $\Phi = \text{diag}[\Phi_1, \ldots, \Phi_{m+1}$.

[1] $A_i(\theta_i^0)$ exists because of Assumption 3.4(ii).

Given this notation, we can state the asymptotic normality result as follows:

Theorem 4.2.

Under A3.1-A3.5, $T^{1/2}(\hat{\theta} - \theta^0) \overset{d}{\to} \mathcal{N}\,(0, \Phi)$.

This is a similar result to Theorem 2.2. Theorem 4.2 implies that the estimates in different sub-samples are asymptotically independent.

The proof of Theorem 4.2 follows the lines of proof found in Gallant (1989), ch. 4, making use of the MVT. Consider a mean value expansion of $T^{1/2}\,\partial S_{T,\hat{I}_i}/\partial\theta$ around θ_i^0, where we denote by $\bar{\theta}_i = \bar{\lambda}_i\hat{\theta}_i + (1-\bar{\lambda}_i)\theta_i^0$, with $\bar{\lambda}_i \in (0,1)$:

$$T^{1/2}\frac{\partial S_{T,\hat{I}_i}}{\partial\theta}\bigg|_{\hat{\theta}_i} = T^{1/2}\,\frac{\partial S_{T,\hat{I}_i}}{\partial\theta}\bigg|_{\theta_i^0} + \frac{\partial^2 S_{T,\hat{I}_i}}{\partial\theta\partial\theta'}\bigg|_{\bar{\theta}_i} T^{1/2}(\hat{\theta}_i - \theta_i^0) \tag{4.7}$$

Note that $\dfrac{\partial S_{T,\hat{I}_i}}{\partial\theta}\bigg|_{\hat{\theta}_i} = 0$, because our minimization problem in fact states that $\hat{\theta}_i = \text{argmin}_\theta\, S_{T,\hat{I}_i}(\theta)$. So, provided that $-\dfrac{\partial^2 S_{T,\hat{I}_i}}{\partial\theta\partial\theta'}\bigg|_{\bar{\theta}_i}$ is invertible[2], (4.7) can be rewritten as:

$$T^{1/2}(\hat{\theta}_i - \theta_i^0) \;=\; \left[\frac{\partial^2 S_{T,\hat{I}_i}}{\partial\theta\partial\theta'}\bigg|_{\bar{\theta}_i}\right]^{-1}\;\left[-\frac{\partial T^{1/2} S_{T,\hat{I}_i}}{\partial\theta}\bigg|_{\theta_i^0}\right] \tag{4.8}$$

For proving Theorem 4.2, we are going to use (4.8) in conjunction with two lemmas. The first one is:

Lemma 4.2. $-T^{1/2}\dfrac{\partial S_{T,I_i^0}}{\partial\theta}\bigg|_{\theta_i^0} \overset{d}{\to} \mathcal{N}\,\left(0, 4\bar{A}_i\right)$.

Proof of Lemma 4.2.

Recall that $S_{T,I_i^0}(\theta_i^0) = \sum_i u_t^2(\theta_i^0)$. Now:

$$\frac{\partial u_t^2(\theta)}{\partial\theta} = \frac{\partial[y_t - f_t(\theta)]^2}{\partial\theta} = -2u_t(\theta)\,F_t(\theta) \tag{4.9}$$

[2]This assumption can be dispensed of using generalized inverses.

Hence,

$$-\,T^{1/2}\frac{\partial S_{T,I_i^0}}{\partial \theta}\Big|_{\theta_i^0} \;=\; 2T^{-1/2}\sum_i u_t\,F_t(\theta_i^0). \tag{4.10}$$

To find the distribution of $2T^{-1/2}\sum_i u_t\,F_t(\theta_i^0) \;=\; 2T^{-1/2}\sum_i \Psi_t(\theta_i^0)$, we need a central limit theorem (CLT). To do so, we are going to use another version of Proposition 3.1, adapted again from Caner (2007)[3]:

Proposition 4.1 (CLT). *If:*

(i) $\{v_t\}$ is a strictly stationary α-mixing process of size $-a$, with $a > 2 + 4/\delta$, for some $\delta > 0$;

(ii) $E[v_t] = 0$ and $E\|v_t\|^{2+\delta} < \infty$, then:

$$T^{-1/2}\sum_{t=1}^{T} v_t \xrightarrow{d} \mathcal{N}\,(\,0, H\,)$$

where $H = \operatorname{plim} Var[T^{-1/2}\sum_{t=1}^{T} v_t]$ exists and is a positive definite matrix not depending on T.

In our case, $v_t = \Psi_t(\theta_i^0)$. Hence, $H = \bar{A}_i$ in our notation, and it exists by A3.4(ii). As before, let $s = 2+\delta$. Then condition (i) above is satisfied by A3.1(i). By interchanging differentiation and integration, we have:

$$E\left[\Psi_t(\theta)\right] = E\left[\frac{\partial}{\partial\theta}\psi_t(\theta)\right] = \int \frac{\partial}{\partial\theta}\psi_t(\theta)dP = \frac{\partial}{\partial\theta}\int \psi_t(\theta)dP = \frac{\partial}{\partial\theta}\,0 = 0$$

So, $E\left[\Psi_t(\theta_i^0)\right] = 0$. On the other hand, $E\|\Psi_t(\theta_i^0)\|^s < \infty$ by A3.3(iv). Hence, the conditions of Proposition 4.1 are satisfied, so:

$$-T^{1/2}\,\frac{\partial S_{T,I_i^0}}{\partial\theta}\Big|_{\theta_i^0} \;=\; -2T^{-1/2}\sum_i \Psi_t(\theta_i^0) \xrightarrow{d} \mathcal{N}\,(\,0, 4\bar{A}_i\,)$$

[3] We assumed $\theta = \theta_i^0$, so that uniform convergence and the Lipschitz condition are unnecessary. Then the limiting process Caner (2007) found, a Kiefer process defined on (θ, r), reduces to a Brownian bridge in r. Moreover, since the expectation of the series is 0, it reduces to a Brownian motion. Finally, because $r = 1$, the limit further reduces to a normal distribution.

This completes the proof of Lemma 4.2. □

In addition to Lemma 4.2, we need another lemma to deal with functions of the second derivative of the minimand, present in (4.8). This lemma is stated below:

Lemma 4.3. $\frac{\partial^2 S_{T,I_i^0}}{\partial\theta\partial\theta'}\Big|_{\theta_i^0} \xrightarrow{p} 2\bar{D}_i$

Proof of Lemma 4.3.

Note that:

$$\frac{\partial^2 u_t^2(\theta)}{\partial\theta\partial\theta'} = \frac{\partial}{\partial\theta}\left[-2u_t(\theta)F_t(\theta)\right] = 2F_t(\theta)F_t(\theta)' - 2u_t(\theta)\frac{\partial^2 f_t(\theta)}{\partial\theta\partial\theta'}$$

Hence,

$$\frac{\partial^2 S_{T,I_i^0}}{\partial\theta\partial\theta'}\Big|_{\theta_i^0} \;=\; 2T^{-1}\sum_i F_t(\theta_i^0)F_t(\theta_i^0)' - 2T^{-1}\sum_i u_t\frac{\partial^2 f_t(\theta)}{\partial\theta\partial\theta'}\Big|_{\theta_i^0} \tag{4.11}$$

By A3.4(iii), $2T^{-1}\sum_i F_t(\theta_i^0)F_t(\theta_i^0)' \xrightarrow{p} 2\bar{D}_i$. Hence, to validate Lemma 4.3, we are left with proving:

$$2T^{-1}\sum_i u_t\frac{\partial^2 f_t(\theta)}{\partial\theta\partial\theta'}\Big|_{\theta_i^0} \xrightarrow{p} 0 \tag{4.12}$$

We are going to prove the stronger result:

$$2T^{-1}\sum_i u_t\frac{\partial^2 f_t(\theta)}{\partial\theta\partial\theta'}\Big|_{\theta} \xrightarrow{p} 0,\ \text{uniformly in } \theta \tag{4.13}$$

For that, we are going to employ the ULLN. By interchanging differentiation and integration, $E\left[u_t\frac{\partial^2 f_t(\theta)}{\partial\theta\partial\theta'}\Big|_{\theta}\right] = 0$. Now, by A3.1 (i), $\{u_t\frac{\partial^2 f_t(\theta)}{\partial\theta\partial\theta'}\}$ is a strictly stationary α-mixing process of size $-2s/(s-2)-\delta$. Moreover, it is a continuous function for each t by the smoothness

Assumption A3.2. The boundedness condition in ULLN is satisfied, i.e.

$$\begin{aligned}
\sup_\theta E\left\|u_t\frac{\partial^2 f_t(\theta)}{\partial\theta\partial\theta'}\right\|^s &= \sup_\theta E\left\{tr\left[u_t\frac{\partial^2 f_t(\theta)}{\partial\theta\partial\theta'}\right]^2\right\}^{s/2} \\
&= \sup_\theta E\left\{\sum_{k=1}^p\left[u_t\frac{\partial^2 f_t(\theta)}{\partial\theta_k\partial\theta_k'}\right]^2\right\}^{s/2} \\
&\le \sup_\theta E\left\{\sum_{k=1}^p\left|u_t\frac{\partial^2 f_t(\theta)}{\partial\theta_k\partial\theta_k'}\right|\right\}^s \\
&\le \left\{\sum_{k=1}^p\left[\sup_\theta\ E\left|u_t\frac{\partial^2 f_t(\theta)}{\partial\theta_k\partial\theta_k'}\right|^s\right]^{1/s}\right\}^s < \infty,
\end{aligned} \tag{4.14}$$

where the second inequality follows from Minkowski's inequality, and the third from A3.3(v). So, we can apply the ULLN in Chapter 3 to obtain:

$$2T^{-1}\sum_i u_t\frac{\partial^2 f_t(\theta)}{\partial\theta\partial\theta'}\Big|_\theta \xrightarrow{p} 0,\ \text{uniformly in } \theta,$$

which is the desired result stated in (4.13) and ensures that (4.12) holds, hence Lemma 4.3 holds. □

Having Lemma 4.2 and 4.3 in mind, we can now prove Theorem 4.2.

Proof of Theorem 4.2.

On the RHS of (4.8), we have two terms that we need to analyze. We are going to show:

$$\frac{\partial^2 S_{T,\hat{I}_i}}{\partial\theta\partial\theta'}\Big|_{\bar\theta_i} - \frac{\partial^2 S_{T,I_i^0}}{\partial\theta\partial\theta'}\Big|_{\theta_i^0} = o_p(1) \tag{4.15}$$

Combining (4.15) with Lemma 4.3, we get:

$$\frac{\partial^2 S_{T,\hat{I}_i}}{\partial\theta\partial\theta'}\Big|_{\bar\theta_i} \xrightarrow{p} \bar{D}_i \tag{4.16}$$

On the other hand, we will show that:

$$-T^{1/2}\frac{\partial S_{T,\hat{I}_i}}{\partial\theta}\Big|_{\theta_i^0}+T^{1/2}\frac{\partial S_{T,I_i^0}}{\partial\theta}\Big|_{\theta_i^0}=o_p(1) \tag{4.17}$$

By equation (4.17) and Lemma 4.2,

$$-T^{1/2}\frac{\partial S_{T,\hat{I}_i}}{\partial\theta}\Big|_{\theta_i^0}\xrightarrow{d}\mathcal{N}\left(0,4\bar{A}_i\right) \tag{4.18}$$

Then (4.8), together with (4.16) and (4.18) and Slutsky's Theorem imply:

$$T^{1/2}(\hat{\theta}_i-\theta_i^0)\xrightarrow{d}\mathcal{N}\left(0,(\bar{D}_i)^{-1}\bar{A}_i(\bar{D}_i)^{-1}\right),$$

Hence, if $\hat{\theta}_i$ and θ_j exhibit the correlation pattern described in Theorem 4.2, then Theorem 4.2 holds.

So we will show Theorem 4.2 in three steps: step 1 involves proving (4.15), step 2 involves showing (4.17), and step 3 shows that the estimates in different sub-samples are asymptotically independent.

Step 1.

$$\frac{\partial^2 S_{T,\hat{I}_i}}{\partial\theta\partial\theta'}\Big|_{\bar{\theta}_i}-\frac{\partial^2 S_{T,I_i^0}}{\partial\theta\partial\theta'}\Big|_{\bar{\theta}_i}=\sum_{\hat{I}_i\nabla I_i^0}\iota_{t,i}\left[T^{-1}F_t(\bar{\theta}_i)F_t(\bar{\theta}_i)'\right] \tag{4.19}$$
$$-\sum_{\hat{I}_i\nabla I_i^0}\iota_{t,i}\left[T^{-1}u_t\frac{\partial^2 f_t(\theta)}{\partial\theta\partial\theta'}\Big|_{\bar{\theta}_i}\right]$$

Note that both sums on the RHS of (4.19) involve a finite number of terms by Theorem 3.2. The terms of the form $T^{-1}F_t(\bar{\theta}_i)F_t(\bar{\theta}_i)'$ are $o_p(1)$ by a similar argument to (3.9). On the other hand, $T^{-1}u_t\frac{\partial^2 f_t(\theta)}{\partial\theta\partial\theta'}\Big|_{\bar{\theta}_i}$ are $o_p(1)$ because $\sup_\theta E\left\|u_t\frac{\partial^2 f_t(\theta)}{\partial\theta\partial\theta'}\right\|<\infty$ by (4.14). Hence, the RHS of

(4.19) involves a finite number of $o_p(1)$ terms, so:

$$\frac{\partial^2 S_{T,\hat{I}_i}}{\partial\theta\partial\theta'}\Big|_{\bar{\theta}_i} - \frac{\partial^2 S_{T,I_i^0}}{\partial\theta\partial\theta'}\Big|_{\bar{\theta}_i} = o_p(1) \tag{4.20}$$

By consistency of parameter estimates, stated in Theorem 4.1, $\hat{\theta}_i - \theta_i^0 = o_p(1)$, implying that $\bar{\theta}_i - \theta_i^0 = o_p(1)$. Since $\dfrac{\partial^2 S_{T,I_i^0}}{\partial\theta\partial\theta'}$ is a continuous function of θ by the smoothness assumption A3.2, it follows that:

$$\frac{\partial^2 S_{T,I_i^0}}{\partial\theta\partial\theta'}\Big|_{\bar{\theta}_i} - \frac{\partial^2 S_{T,I_i^0}}{\partial\theta\partial\theta'}\Big|_{\theta_i^0} = o_p(1). \tag{4.21}$$

Combining (4.20) with (4.21), we get:

$$\frac{\partial^2 S_{T,\hat{I}_i}}{\partial\theta\partial\theta'}\Big|_{\bar{\theta}_i} - \frac{\partial^2 S_{T,I_i^0}}{\partial\theta\partial\theta'}\Big|_{\theta_i^0} = o_p(1),$$

which is exactly (4.15).

Step 2.

Note that:

$$-T^{1/2}\frac{\partial S_{T,\hat{I}_i}}{\partial\theta}\Big|_{\theta_i^0} + T^{1/2}\frac{\partial S_{T,I_i^0}}{\partial\theta}\Big|_{\theta_i^0} = 2\sum_{\hat{I}_i \nabla I_i^0} \iota_{t,i}\left[T^{-1/2}\Psi_t(\theta_i^0)\right] \tag{4.22}$$

The RHS of (4.22) involves, by Theorem 3.2, a finite number of terms that are $o_p(1)$ because $E\|\Psi_t(\theta_i^0)\| < \infty$ by A3.3(iv). Hence,

$$-T^{1/2}\frac{\partial S_{T,\hat{I}_i}}{\partial\theta}\Big|_{\theta_i^0} + T^{1/2}\frac{\partial S_{T,I_i^0}}{\partial\theta}\Big|_{\theta_i^0} = o_p(1),$$

which is exactly (4.22).

Step 3.

To find the correlation between $\hat{\theta}_i$ and $\hat{\theta}_j$, where $i, j = 1, \ldots, m+1$ and $i \neq j$, we are going to derive approximation formulae similar to (2.15). To that end, we substitute (4.15) and (4.17) into (4.8) to get:

$$
\begin{aligned}
& T^{1/2}[\hat{\theta}_i - \theta_i^0] \\
= & \left[\frac{\partial^2 S_{T,I_i^0}}{\partial\theta\partial\theta'}\Big|_{\theta_i^0} \right]^{-1} \left[-T^{1/2} \frac{\partial S_{T,I_i^0}}{\partial\theta}\Big|_{\theta_i^0} \right] + o_p(1) \\
= & \left[T^{-1} \sum_i F_t(\theta_i^0) F_t(\theta_i^0) - T^{-1} \sum_i u_t \frac{\partial^2 f_t(\theta)}{\partial\theta\partial\theta'}\Big|_{\theta_i^0} \right]^{-1} \left[T^{-1/2} \sum_i u_t F_t(\theta_i^0) \right] + o_p(1)
\end{aligned}
$$

Now, $T^{-1} \sum_i u_t \frac{\partial^2 f_t \theta}{\partial\theta\partial\theta'}\Big|_{\theta_i^0} = o_p(1)$ by (4.13). Hence:

$$
\begin{aligned}
& T^{1/2}[\hat{\theta}_i - \theta_i^0] \\
= & [T^{-1} \sum_i F_t(\theta_i^0) F_t(\theta_i^0) - o_p(1)]^{-1} \, [T^{-1/2} \sum_i u_t F_t(\theta_i^0)] + o_p(1) \\
= & [T^{-1} \sum_i F_t(\theta_i^0) F_t(\theta_i^0)]^{-1} \, [T^{-1/2} \sum_i u_t F_t(\theta_i^0)] - o_p(1) \, [T^{-1/2} \sum_i u_t F_t(\theta_i^0)] + o_p(1) \\
= & [T^{-1} \sum_i F_t(\theta_i^0) F_t(\theta_i^0)]^{-1} \, [T^{-1/2} \sum_i u_t F_t(\theta_i^0)] - o_p(1) \, O_p(1) + o_p(1) \\
= & [T^{-1} \sum_i F_t(\theta_i^0) F_t(\theta_i^0)]^{-1} \, [T^{-1/2} \sum_i u_t F_t(\theta_i^0)] + o_p(1)
\end{aligned}
$$

Let $\bar{F}_i$ be the $[T_i^0 - T_{i-1}^0] \times p$ matrix with rows $F_t(\theta_i^0)'$, and $\bar{U}_i$ be the $[T_i^0 - T_{i-1}^0] \times 1$ vector with elements u_t. Then we can write:

$$
T^{1/2}[\hat{\theta}_i - \theta_i^0] = (T^{-1} \bar{F}_i' \, \bar{F}_i)^{-1} \, T^{-1/2} \bar{F}_i' \, \bar{U}_i + o_p(1) \tag{4.23}
$$

Hence, for $i, j = 1, \ldots, m+1$ and $i \neq j$,

$$
\begin{aligned}
&\operatorname{plim} Cov(T^{1/2}[\hat{\theta}_i - \theta_i^0], T^{1/2}[\hat{\theta}_j - \theta_j^0]) \\
&= \operatorname{plim}\{(T^{-1}\bar{F}_i'\,\bar{F}_i)^{-1}\; T^{-1}\,[\bar{F}_i'\,\bar{U}_i\bar{U}_j\bar{F}_j]\;(T^{-1}\bar{F}_j'\,\bar{F}_j)^{-1}\} \\
&= \operatorname{plim}(T^{-1}\bar{F}_i'\,\bar{F}_i)^{-1}\operatorname{plim}\left[T^{-1}\bar{F}_i'\,\bar{U}_i\bar{U}_j\bar{F}_j\right]\operatorname{plim}(T^{-1}\bar{F}_j'\,\bar{F}_j)^{-1} \\
&= [\Delta_i^0 D(\theta_i^0)]^{-1}\{\operatorname{plim}\,[T^{-1/2}\sum_{t\in I_i^0} u_t F_t(\theta_i^0)]\;[T^{-1/2}\sum_{s\in I_j^0} u_s F_s(\theta_j^0)']\}\,[\Delta_j^0 D(\theta_j^0)]^{-1} \\
&\equiv \bar{D}_i \bar{A}_{i,j} \bar{D}_i,
\end{aligned}
$$

where:

$$
\bar{A}_{i,j} = \lim\, E\,[T^{-1/2}\sum_{t\in I_i^0} u_t F_t(\theta_i^0)]\;[T^{-1/2}\sum_{s\in I_j^0} u_s F_s(\theta_j^0)']
$$

for $i, j = 1, \ldots, m+1$, $i \neq j$. Notice that $\{u_t F_t(\theta)\}$ is an α-mixing process. We can write:

$$
\begin{aligned}
T^{-1/2}\sum_{t\in I_i^0} u_t F_t(\theta_i^0) \;&\in \mathcal{F}_{[T\lambda_{i-1}^0]+1}^{[T\lambda_i^0]} \;\subset \mathcal{F}_{-\infty}^{[T\lambda_i^0]} \\
T^{-1/2}\sum_{t\in I_j^0} u_t F_t(\theta_j^0) \;&\in \mathcal{F}_{[T\lambda_{j-1}^0]+1}^{[T\lambda_j^0]} \;\subset \mathcal{F}_{[T\lambda_{j-1}^0]+1}^{\infty}
\end{aligned}
$$

For non-adjacent segments I_i^0 and I_j^0, i.e. $j-i \geq 2$, the two σ-algebras $\mathcal{F}_{-\infty}^{[T\lambda_i^0]}$ and $\mathcal{F}_{[T\lambda_{j-1}^0]+1}^{\infty}$ are at least $[T\lambda_{j-1}]-[T\lambda_i]$ observations apart. As $T \to \infty$, hence $[T\lambda_{j-1}]-[T\lambda_i] \to \infty$, by definition of an α-mixing process, the dependence between the two σ-algebras above dies out. Hence, for $|i-j| \geq 2$, $T^{-1/2}\sum_{t\in I_i^0} u_t F_t(\theta_i^0)$ and $T^{-1/2}\sum_{t\in I_j^0} u_t F_t(\theta_j^0)$ are asymptotically independent, so $\bar{A}_{i,j} = O_{p\times p}$. For adjacent segments $|i-j| = 1$, say for $i = j-1$, let $\varrho_{|k-l|} = u_k u_l F_k(\theta_i^0) F_l(\theta_j^0)'$. Then:

$$
\bar{A}_{i,j} = \operatorname{plim} T^{-1}\Big(\varrho_1 + 2\varrho_2 + 3\varrho_2 + \ldots\Big)
$$

Note that, because of the α-mixing assumption, as $|k-l| \to \infty$, $E[\varrho_{|k-l|}] \to 0$. The first terms in the equation above are non-zero, but the T^{-1} scaling ensures that they disappear in the limit. Hence, $\bar{A}_{i,i+1} = O_{p\times p}$ and $\bar{A}_{i,i-1} = O_{p\times p}$, hence the adjacent segments (and estimates) are asymptotically independent.

Putting together Steps 1-3, the proof of Theorem 4.2 is complete. □

A natural concern for the practitioner is finding positive definite estimates for the long-run covariance matrix Φ so that he/she can construct confidence intervals. Some estimates are readily available from the Proof of Theorem 4.2. For example, to estimate the long run variance of θ_i^0, we need to estimates the i^{th} $(p \times p)$ diagonal block of Φ, namely $\bar{D_i}^{-1}\bar{A}_i\bar{D_i}^{-1}$. Since $\hat{\lambda}_i - \lambda_i^0 \xrightarrow{p} 0$, $\hat{\Delta}_i = \hat{\lambda}_i - \hat{\lambda}_{i-1} \xrightarrow{p} \Delta_i^0$. Also, $\hat{\theta}_i - \theta_i^0 \xrightarrow{p} 0$ and $D(\theta)$ is a continuous function of θ, so we can use $\hat{D}_T(\hat{\theta}_i) = T^{-1}\sum_{t=[T\hat{\lambda}_{i-1}+1]}^{[T\hat{\lambda}_i]} F_t(\hat{\theta}_i)F_t(\hat{\theta}_i)'$:

$$\hat{D}_T(\hat{\theta}_i) - \Delta_i D(\theta_i^0) = T^{-1}\sum_{t\in\hat{I}_i} F_t(\hat{\theta}_i)F_t(\hat{\theta}_i)' - \operatorname{plim} T^{-1}\sum_{t\in I_i^0} F_t(\theta_i^0)F_t(\theta_i^0)' = o_p(1),$$

by similar arguments as in the proof of Theorem 4.2. Also, $\hat{D}_T(\hat{\theta}_i)$ is positive definite because it is the sum of positive definite matrices $F_t(\hat{\theta}_i)F_t(\hat{\theta}_i)'$. In the case of $\bar{A}_i$, we can use heteroskedasticity autocorrelation matrix estimators (HACs), which are positive definite too. Denote them by $\hat{A}_{T,i}$. Also let $\hat{D}_{T,i} = \hat{D}_T(\hat{\theta}_i)$. Then we can use the following long-run estimator of the variance of θ_i^0:

$$\hat{\Phi}_{T,i} \; = \; \hat{D}_{T,i}^{-1}\hat{A}_{T,i}\hat{D}_{T,i}^{-1} \xrightarrow{p} \bar{D_i}^{-1}\bar{A}_i\bar{D_i}^{-1}$$

This way, the diagonal blocks of the variance-covariance matrix estimators are positive definite and consistent, ensuring that the estimating variance-covariance matrix $\hat{\Phi}_T$, with $i-j^{th}$ block $\hat{\Phi}_{T,i}$, is itself positive definite and consistent:

$$\hat{\Phi}_T \xrightarrow{p} \Phi \tag{4.24}$$

Note that assumption A3.1 implies that the errors are identically distributed, but not independent of each other. So, another case of interest is stated in the following stronger condition:

Assumption A 4.1 (Dependence and Memory of Processes).
(i) Let $\{x_t, u_t\}$ *be a strictly stationary* α*-mixing process of size* $-2s/(s-2) - \delta^*$*, with* $s > 2$ *and* $\delta^* > 0$*. Moreover,* $\{v_t\}_{t \in I_i^0} \perp \{v_t\}_{t \in I_j^0}$*, for* $|i - j| = 1$*;*
(ii) The errors are conditionally homoskedastic and uncorrelated, i.e. $u_t \perp x_k$ *for all* t, k*, so that* $E[u_t \mid x_k] = 0$*,* $E[u_t^2 \mid x_k] = \sigma^2$ *and* $E[u_t u_k \mid x_i x_j] = 0$*, for all* t, k, i, j*.*

With this assumption replacing A3.1, the variance-covariance matrix Φ admits further simplification:

$$\begin{aligned}
\bar{A}_i &= \text{plim Var}\,[T^{-1/2} \sum_i u_t F_t(\theta_i^0)] = \lim T^{-1} \sum_i E[u_t^2 F_t(\theta_i^0 F_t(\theta_i^0)'] \\
&= \lim T^{-1} \sum_i E[u_t^2 \mid x_t]\; E[F_t(\theta_i^0 F_t(\theta_i^0)'] \\
&= \sigma^2 \lim T^{-1} \sum_i E[F_t(\theta_i^0 F_t(\theta_i^0)'] = \sigma^2 (F_i' F_i)^{-1} = \sigma^2 \bar{D}_i.
\end{aligned}$$

Then the variance of θ_i is $\sigma^2 \bar{D}_i^{-1} \bar{D}_i \bar{D}_i^{-1} = \sigma^2 \bar{D}_i^{-1}$, hence the variance-covariance matrix is:

$$\Phi = \sigma^2 \text{ diag}[\bar{D}_i^{-1}] = \sigma^2 \bar{D}^{-1} = \sigma^2 \text{ diag } [(F_1' F_1)^{-1}, \ldots, (F_{m+1}' F_{m+1})^{-1}]$$

This is stated in the following corollary:

Corrolary 1 to Theorem 4.2. *Under A4.1 and A3.2-A3.5,*

$$T^{-1/2}(\hat{\theta}_T - \theta_0) \xrightarrow{d} \mathcal{N}\left(\, 0, \sigma^2 \bar{D}^{-1} \,\right).$$

In this chapter, we showed how to construct confidence intervals for parameter estimates. What we have not discussed is construction of confidence intervals for the break-fraction estimates. This is the subject of the next chapter.

Chapter 5

Asymptotic Distribution of Break-Point Estimators

Given the rate of convergence of the break-points and parameter estimates in Ch 3. and Ch. 4, a natural next step is to consider their asymptotic distributions. Bai (1994, 1995, 1997) derives it for linear least-squares methods. Hall and Han (2005) extend it to linear IV (instrument variables) models. All these papers consider two cases : the magnitude of shifts in regression parameters can be fixed or shrinking at a certain rate.

For fixed magnitudes of shifts, the distribution of break-point estimators depends on the underlying distribution of the regressors and errors $\{f_t(\theta_1^0), \ldots, f_t(\theta_{m+1}^0), u_t\}_{-\infty}^{+\infty}$. In this case, an analytical solution for the limiting distribution is hard to obtain. However, in linear models, under strict stationarity of regressors and errors and independence among the two, Bai (1995, 1997) finds it to be a two-sided random walk with stochastic drifts.

An alternative approach is to allow for small magnitudes of shifts that shrink to zero as the sample size increases. In this case, the limiting distributions of the break-point estimators are invariant to the underlying distribution of $\{f_t(\theta_1^0), \ldots, f_t(\theta_{m+1}^0), u_t\}_{-\infty}^{+\infty}$ and they remain valid for moderate shifts. Assuming that regressors and errors are jointly second order stationary, Bai (1995, 1997) finds the limiting distribution to be a functional of a two-sided Brownian motion defined on $[0, \infty]$. This is the limiting process for the two-sided random walk with stochastic drifts mentioned above.

Consider the framework of one unknown break. For both cases - fixed and shrinking shifts, the proofs in Bai (1994, 1997) and Hall and Han (2005) all perform a first step in which the sum of squared residuals (for each partition, evaluated at the LS parameter estimates) is shown to be a linear transformation of the Wald-test statistic of the equality of parameters. With this transformation, the break-point estimator is the argmax of a Wald statistic over all possible breaks, whose distribution is easy to derive.

This transformation does not hold for nonlinear models in finite samples, so we took the alternative path of deriving the asymptotic distributions by examining directly the behavior of the objective function. Our approach is somewhat similar to Bai (1995), where no transformation can be performed on the least-absolute deviation (LAD) minimand.

5.1 Fixed Magnitudes of Shifts

We will only consider the case of one unknown break ($m = 1$). Provided that adjacent segments are asymptotically distinct, the proof is easily generalizable to m unknown breaks, as long as m is fixed.

Consider the following NLS model with one unknown break at date k:

$$y_t = \begin{cases} f(x_t, \theta_1^0) + u_t & t = 1, \dots, k \\ f(x_t, \theta_2^0) + u_t & t = k+1, \dots, T. \end{cases}$$

The true break occurs at time k_0. An implicit assumption in Chapters 3 and 4 was that the parameter shifts are constant:

Assumption A 5.1. $\delta = \theta_2^0 - \theta_1^0$*, a fixed number.*

Denote by $S_T(k, \theta_1, \theta_2)$ the sum of squared residuals evaluated at a potential break-point $1 \leq k \leq T$. Also, let $S_T(k) = \min_{\theta_1, \theta_2} S_T(k, \theta_1, \theta_2)$. Recall that the break-point estimator is obtained by minimizing $S_T(k)$:

$$\hat{k} = \underset{1\le k\le T}{\text{argmin}}\, S_T(k) = \underset{1\le k\le T}{\text{argmin}}\ \underset{\theta_1,\theta_2}{\text{argmin}}\ S_T(k,\theta_1,\theta_2)$$

Since $S_T(k_0,\theta_1^0,\theta_2^0) = \sum_{t=1}^{T} u_t^2$ is a random quantity that does not depend on k, θ_1 or θ_2, we can subtract it from the minimand without affecting its distribution:

$$\hat{k} = \underset{1\le k\le T}{\text{argmin}}\ \underset{\theta_1,\theta_2}{\text{argmin}}\ T\left[S_T(k,\theta_1,\theta_2) - S_T(k_0,\theta_1^0,\theta_2^0)\right] \tag{5.1}$$

As long as:

Assumption A 5.2. *$sup_t \mid u_t \mid < \infty$,*

$S_T(k_0,\theta_1^0,\theta_2^0)$ will be finite in finite samples. We can relax this assumption to $sup_t|u_t| < \infty$ almost surely, which is then equivalent to $E\ \sup_t|u_t| < \infty$, already embedded in Assumption A3.3 (ii). We will consider A5.1 instead of A3.3 (ii) only for simplicity of exposition. Since the argmin operation is taking prior to deriving the limiting distribution of the break-point estimator, (5.1) is a valid way to rewrite the minimization problem. Denote $V(k,\theta_1,\theta_2) \equiv T\left[S_T(k,\theta_1,\theta_2) - S_T(k_0,\theta_1^0,\theta_2^0)\right]$. We will provide a large sample approximation to it. The approximation is given by the following theorem:

Theorem 5.1. *Under Assumptions A3.1 or A4.1 and A3.2-A3.5, A5.1 defined for $m = 1$, we have:*

$$[\hat{k} - k_0] - \underset{n\in C}{\text{argmax}}\, W^*(n) \xrightarrow{p} 0,$$

where $C = [-N, N]$ is a compact subset of $\mathbb{R}$, N is fixed, and $W^*(n)$ is a double-sided stochastic *process defined as follows:*

$$W^*(n) = \begin{cases} W_1^*(n) = -\sum_{t=n+1}^{0} d_t^2(\theta_2^0, \theta_1^0) + 2\sum_{t=n+1}^{0} u_t d_t(\theta_2^0, \theta_1^0) & \textit{if } n < 0 \\ 0 & \textit{if } n = 0 \\ W_2^*(n) = -\sum_{t=1}^{n} d_t^2(\theta_2^0, \theta_1^0)^2 - 2\sum_{t=1}^{n} u_t d_t(\theta_2^0, \theta_1^0) & \textit{if } n > 0. \end{cases}$$

To compare this result to linear models, consider the MVT expansion of $f_t(\theta_2^0)$ around $f_t(\theta_1^0)$.

Since we will use MVT repeatedly, we will standardize its notation further. Denote by $\bar{\theta}_{t,i,j} = \bar{\lambda}_{t,i,j}\,\hat{\theta}_i + [1 - \bar{\lambda}_{t,i,j}]\,\theta_j^0$, with $\bar{\lambda}_{t,i,j} \in (0,1)$ and $i,j \in \{1,2,\ldots,m+1\}$, such that $d_t(\hat{\theta}_i, \theta_j^0) = [\hat{\theta}_i - \theta_j^0]' F_t(\bar{\theta}_{t,i,j})$.

Consider now the MVT expansion $d_t(\theta_2^0, \theta_1^0) = [\theta_2^0 - \theta_1^0]' F_t(\bar{\theta}_{t,1,2}^{0,0}) = \delta' F_t(\bar{\theta}_{t,1,2}^{0,0})$. Then:

$$W^*(n) = \begin{cases} W_1^*(n) = -\delta' \sum_{t=n+1}^{0} F_t(\bar{\theta}_{t,1,2}^{0,0}) F_t(\bar{\theta}_{t,1,2}^{0,0})'\delta + 2\delta' \sum_{t=n+1}^{0} u_t F_t(\bar{\theta}_{t,1,2}^{0,0}) & \text{if } n < 0 \\ 0 & \text{if } n = 0 \\ W_2^*(n) = -\delta' \sum_{t=1}^{n} F_t(\bar{\theta}_{t,1,2}^{0,0}) F_t(\bar{\theta}_{t,1,2}^{0,0})'\delta - 2\delta' \sum_{t=1}^{n} u_t F_t(\bar{\theta}_{t,1,2}^{0,0}) & \text{if } n > 0. \end{cases}$$

The expression above looks similar to the linear result in Bai (1997), where the regressors z_t replace $F_t(\bar{\theta}_{t,1,2}^{0,0})$. As noted by Bai (1995), continuity of $f_t(\theta)$ guarantees that $W^*(n)$ has a continuous distribution. By definition of continuous distributions, we have:

(i) $P\left[W_j^*(n_1) = W_j^*(n_2)\right] = 0$ for $j = 1,2$ and $n_1 \neq n_2$;

(ii) $P\left[W_1^*(n_1) = W_2^*(n_2)\right] = 0$ for all n_1, n_2.

This implies that the maximum of $W^*(n)$ is unique almost surely (by definition of almost sureness, the probability of an undesired event - such as $W_j^*(n_1) = W_j^*(n_2)$ or $W_1^*(n_1) = W_2^*(n_2)$ - that would not yield unique maxima, is 0). This uniqueness allows us to invoke the continuous mapping theorem (CMT) for the "argmax" functional - see Kim and Pollard (1990). This theorem asserts that the " argmax " is a continuous functional on a set of functions with compact domain and unique maximum. Since $n \in C$, a compact set, and $W^*(n)$ has a unique maximum, we can use CMT to express the distribution of $\hat{k}$ in the way Theorem 5.1 states it.

Proof of Theorem 5.1.

The distribution of $\hat{k}$ depends on the distribution of $\text{argmin}_{\theta_1,\theta_2} V_T(k,\theta_1,\theta_2)$:

$$\hat{k} = \underset{k}{\text{argmin}} \underset{\theta_1,\theta_2}{\text{argmin}} V_T(k,\theta_1,\theta_2).$$

The proof will be in three steps, summarized below:

Step 1. The key to deriving an approximating asymptotic distribution of $V_T(k,\hat{\theta}_1(k),\hat{\theta}_2(k))$ $= \text{argmin}_{\theta_1,\theta_2} V_T(k,\theta_1,\theta_2)$ lies in knowing the convergence rates of $\hat{k}$ and $\hat{\theta}_i(k)$. Then the minimization problem is defined over a neighborhood of (k,θ_1,θ_2).

Step 2. Given Step 1, we split the objective function into three parts. Assume $k < k_0$; the proof for $k \geq k_0$ is similar.

$$\begin{aligned} V_T(k,\hat{\theta}_1(k),\hat{\theta}_2(k)) &= \sum_{t=1}^{k}[u_t^2(\hat{\theta}_1(k)) - u_t^2(\theta_1^0)] + \sum_{t=k+1}^{k_0}[u_t^2(\hat{\theta}_2(k)) - u_t^2(\theta_1^0)] + \\ &+ \sum_{t=k_0+1}^{T}[u_t^2(\hat{\theta}_2(k)) - u_t^2(\theta_2^0)] \\ &= \Sigma_1 + \Sigma_2 + \Sigma_3. \end{aligned}$$

Step 3. Given the rate of convergence of parameter and break-fraction estimates, one can show that Σ_1 and Σ_3 have asymptotic distributions that are independent of k. We are left with Σ_2, representing magnitude of the mismatch between the objective function evaluated at $[k,\hat{\theta}_1(k),\hat{\theta}_2(k)]$ and at the true parameter values. For large samples, this mismatch yields an approximation to the distribution of the objective function in a neighborhood of the true values. Then Σ_2 governs the distribution of $\hat{k}$. The complete proof follows.

Case 1 : $k < k_0$.

Let $v = k_0 - k$, $0 < v \leq N$. The latter is a restatement of Theorem 3.2, indicating that the difference between the true and estimated number of break-points is a fixed number, and does not diverge to ∞ as the sample size increases. Given that the break-fractions are T-rate convergent, we know by Theorem 4.2:

$$\breve{\theta}_i(v) = T^{1/2}(\hat{\theta}_i(k) - \theta_i^0) \Rightarrow \mathcal{N}\ (\ 0, \Phi_i), \tag{5.2}$$

for $i = 1, 2$, where λ_1^0 indicates the true break-fraction, $\lambda_2^0 = 1 - \lambda_1^0$, and $\Phi_i = \bar{D}_i^{-1}\bar{A}_i\bar{D}_i^{-1}$ is a variance not depending on v. We will start by analyzing Σ_1:

$$\begin{aligned}\Sigma_1 &= \sum_{t=1}^{k}[u_t^2(\hat{\theta}_1(k)) - u_t^2] &(5.3)\\ &= \sum_{t=1}^{k} d_t^2(\hat{\theta}_1(k), \theta_1^0) - 2\sum_{t=1}^{k} u_t\ d_t(\hat{\theta}_1(k), \theta_1^0).\end{aligned}$$

Since $k = k_o - v$, by MVT, we get:

$$\begin{aligned}\Sigma_1 &= \breve{\theta}_1(v)'\left[T^{-1}\sum_{t=1}^{k} F_t(\bar{\theta}_{t,1,1})F_t(\bar{\theta}_{t,1,1})'\right]\breve{\theta}_1(v) - 2\breve{\theta}_1(v)'\left[T^{-1/2}\sum_{t=1}^{k} u_t F_t(\bar{\theta}_{t,1,1})\right]\\ &= \Sigma_{1,1} + \Sigma_{1,2}.\end{aligned}$$

Start with $\Sigma_{1,1}$. We know from A3.4(iii) that $T^{-1}\sum_{t=1}^{k_0} F_t(\theta)F_t(\theta)' \xrightarrow{p} \lambda_1^0 D(\theta) = \bar{D}_1$, uniformly in θ. Since $F_t(\theta)F_t(\theta)'$ is uniformly bounded and $\bar{\theta}_{t,1,1} \xrightarrow{p} \theta_1^0$,

$$T^{-1}\sum_{t=1}^{k_0} F_t(\bar{\theta}_{t,1,1})F_t(\bar{\theta}_{t,1,1})' = \bar{D}_1 + o_p(1). \tag{5.4}$$

Given that $\mid v \mid \leq N$, $\sum_{t=k+1}^{k_0} T^{-1}F_t(\theta)F_t(\theta)'$ involves v (hence a finite number of) terms that are $o_p(1)$ uniformly in θ, by A3.4(ii). Hence, uniformly in v, we have:

$$T^{-1} \sum_{t=k+1}^{k_0} F_t(\bar{\theta}_{t,1,1}) F_t(\bar{\theta}_{t,1,1})' = o_p(1). \tag{5.5}$$

Putting together (5.4) and (5.5), we can write, uniformly in v:

$$\begin{aligned} T^{-1} \sum_{t=1}^{k} F_t(\bar{\theta}_{t,1,1}) F_t(\bar{\theta}_{t,1,1}) &= T^{-1} \sum_{t=1}^{k_0} F_t(\bar{\theta}_{t,1,1}) F_t(\bar{\theta}_{t,1,1})' - T^{-1} \sum_{t=k+1}^{k_0} F_t(\bar{\theta}_{t,1,1}) F_t(\bar{\theta}_{t,1,1})' \\ &= \bar{D}_1 + o_p(1). \end{aligned}$$

From the equation above and (5.2), we have:

$$\begin{aligned} \Sigma_{1,1} &= \breve{\theta}_1(v)' \left[T^{-1} \sum_{t=1}^{k} F_t(\bar{\theta}_{t,1,1}) F_t(\bar{\theta}'_{t,1,1}) \right] \breve{\theta}_1(v) \\ &= \mathcal{N}(0, \Phi_1)' \times \bar{D}_1 \times \mathcal{N}(0, \Phi_1) + o_p(1) \\ &= \mathfrak{D}_1^* + o_p(1), \end{aligned} \tag{5.6}$$

where the $o_p(1)$ term is uniform in v, and $\mathfrak{D}_1^*$ is a distribution not depending on v. Note that if A3.1 is replaced by A4.1, then $\Phi_1 = \sigma^2 \bar{D}_1^{-1}$, hence $\mathfrak{D}_1^* = \sigma^2 \chi_p^2$. Hence, under both A3.1 and A4.1, while $\Sigma_{1,1}$ depends on v in finite samples, its asymptotic distribution does not.

Consider now $\Sigma_{1,2}$. The part in square brackets of this term can be written as:

$$T^{-1/2} \sum_{t=1}^{k} u_t F_t(\bar{\theta}_{t,1,1}) = T^{-1/2} \sum_{t=1}^{k_0} u_t F_t(\bar{\theta}_{t,1,1}) - \sum_{t=k+1}^{k_0} T^{-1/2} u_t F_t(\bar{\theta}_{t,1,1}) \tag{5.7}$$

Since $\bar{\theta}_{t,1,1} \xrightarrow{p} \theta_1^0$, we have, by similar arguments as before,

$$T^{-1/2} \sum_{t=1}^{k_0} u_t F_t(\bar{\theta}_{t,1,1}) - T^{-1/2} \sum_{t=1}^{k_0} u_t F_t(\theta_1^0) = o_p(1) \tag{5.8}$$

Recall from Lemma 4.3, that:

$$\frac{\partial S_{T,I_1^0}(\theta)}{\partial\theta}\Big|_{\theta_1^0} = 2T^{-1/2}\sum_{t=1}^{k_0} u_t F_t(\theta_1^0) \xrightarrow{d} \mathcal{N}\left(0, 4\bar{A}_1\right)$$

Hence,

$$T^{-1/2}\sum_{t=1}^{k_0} u_t F_t(\theta_1^0) \xrightarrow{d} \mathcal{N}\left(0, \bar{A}_1\right) \tag{5.9}$$

From (5.8) and (5.9), it follows that:

$$T^{-1/2}\sum_{t=1}^{k_0} u_t F_t(\bar{\theta}_{t,1,1}) \xrightarrow{d} \mathcal{N}\left(0, \bar{A}_1\right) \tag{5.10}$$

On the other hand, $\sum_{t=k+1}^{k_0} T^{-1/2} u_t F_t(\theta)$ involves v, a bounded number of terms that are $o_p(1)$ uniformly in v. The latter holds because $E[u_t F_t(\theta)] = 0$, as shown before. It follows that, uniformly in v:

$$\sum_{t=k+1}^{k_0} T^{-1/2} u_t F_t(\theta) = o_p(1) \tag{5.11}$$

Substituting (5.10) and (5.11) into (5.7), we get that, uniformly in v,:

$$\begin{aligned} T^{-1/2}\sum_{t=1}^{k} u_t F_t(\bar{\theta}_{t,1,1}) &= T^{-1/2}\sum_{t=1}^{k_0} u_t F_t(\bar{\theta}_{t,1,1}) - \sum_{t=k+1}^{k_0} T^{-1/2} u_t F_t(\bar{\theta}_{t,1,1}) \\ &= \mathcal{N}\left(0, \bar{A}_1\right) + o_p(1) \end{aligned} \tag{5.12}$$

Then, the distribution of $\Sigma_{1,2}$ is:

$$\begin{aligned} \Sigma_{1,2} &= -2\breve{\theta}_1(v)' \left[T^{-1/2}\sum_{t=1}^{k} u_t F_t(\bar{\theta}_{t,1,1})\right] \\ &= -2\mathcal{N}\left(0, \Phi_1\right)' \times \mathcal{N}\left(0, \bar{A}_1\right) + o_p(1) \\ &= \mathfrak{D}_1^{**} + o_p(1), \end{aligned} \tag{5.13}$$

where $\mathfrak{D}_1^{**} = -2\mathcal{N}(0, \Phi_1)' \times \mathcal{N}(0, A_1)$ is a distribution not depending on v. Again, under A4.1, $\mathfrak{D}_1^{**} = -2\mathcal{N}(0, \sigma^2 \bar{D}_1^{-1})' \times \mathcal{N}(0, \sigma^2 \bar{D}_1) = -2\sigma^2 \chi_p^2$. So, under both A3.1 and A4.1, while $\Sigma_{1,2}$ might depend on v in finite samples, its asymptotic distribution does not.

Using (5.6) and (5.13), we can write:

$$\Sigma_1 = \Sigma_{1,1} + \Sigma_{1,2} = \mathfrak{D}_1^* - \mathfrak{D}_1^{**} + o_p(1) = \mathfrak{D}_1 + o_p(1) \tag{5.14}$$

Note that, under A4.1, $\mathfrak{D}_1 = \sigma^2 \chi_p^2 - 2\sigma^2 \chi_p^2 = -\sigma^2 \chi_p^2$.

Next, consider Σ_3.

$$\Sigma_3 = \sum_{t=k_0+1}^{T} [u_t^2(\hat{\theta}_2(k)) - u_t^2(\theta_2^0)] \tag{5.15}$$

By similar arguments as for Σ_1,

$$\Sigma_3 = \mathfrak{D}_2 + o_p(1), \tag{5.16}$$

where $\mathfrak{D}_2$ is a distribution not depending on v, and $\mathfrak{D}_2 = -\sigma^2 \chi_p^2$ under A4.1. By combining the results for Σ_1 and Σ_3 in (5.14), respectively (5.16), with the expression for the minimand, we have uniformly in $v = k_0 - k,\ 0 < v \leq N$,

$$V_T(k, \hat{\theta}_1(k), \hat{\theta}_2(k)) = \mathfrak{D}_1 + \mathfrak{D}_2 + o_p(1) + \Sigma_2 = \mathfrak{D} + o_p(1) \tag{5.17}$$

Under A4.1, the above simplifies to:

$$V_T(k, \hat{\theta}_1(k), \hat{\theta}_2(k)) = -\sigma^2 \chi_{2p}^2 + o_p(1) + \Sigma_2 \tag{5.18}$$

The latter is true because the χ_p^2 quantities derived from Σ_1 and Σ_3 are independent, because the segments are asymptotically independent as shown in the proof of Theorem 4.2.

Now turn to Σ_2, the term that approximates the distribution of the break-point estimator in finite samples.

$$\Sigma_2 = \sum_{t=k+1}^{k_0} [u_t^2(\hat{\theta}_2(k)) - u_t^2(\theta_1^0)] \tag{5.19}$$
$$= \sum_{t=k+1}^{k_0} d_t^2(\hat{\theta}_2(k), \theta_1^0) - 2 \sum_{t=k+1}^{k_0} u_t \, d_t(\hat{\theta}_2(k), \theta_1^0)$$

$$\Sigma_2 = \sum_{t=k+1}^{k_0} [d_t(\hat{\theta}_2(k), \theta_2^0) + d_t(\theta_2^0, \theta_1^0)]^2 - 2 \sum_{t=k+1}^{k_0} u_t \, d_t(\hat{\theta}_2(k), \theta_1^0)$$
$$= \sum_{t=k+1}^{k_0} d_t^2(\hat{\theta}_2(k), \theta_2^0) + \sum_{t=k+1}^{k_0} d_t^2(\theta_2^0, \theta_1^0) +$$
$$+ 2 \sum_{t=k+1}^{k_0} d_t(\hat{\theta}_2(k), \theta_2^0) \, d_t(\theta_2^0, \theta_1^0) - 2 \sum_{t=k+1}^{k_0} u_t \, d_t(\hat{\theta}_2(k), \theta_2^0) - 2 \sum_{t=k+1}^{k_0} u_t \, d_t(\theta_2^0, \theta_1^0)$$
$$= \Sigma_{2,1} + \Sigma_{2,2} + \Sigma_{2,3} + \Sigma_{2,4} + \Sigma_{2,5}$$

Consider $\Sigma_{2,1}$:

$$\Sigma_{2,1} = \sum_{t=k+1}^{k_0} d_t(\hat{\theta}_2(k), \theta_2^0) \tag{5.20}$$
$$= \breve{\theta}_2(v)' \left[\sum_{t=k+1}^{k_0} T^{-1} F_t(\bar{\theta}_{t,2,2}) F_t(\bar{\theta}_{t,2,2})' \right] \breve{\theta}_2(v) \tag{5.21}$$

Since $F_t(\theta)F_t(\theta)'$ is uniformly bounded in θ by A3.4 (ii), $\sum_{t=k+1}^{k_0} T^{-1} F_t(\theta) F_t(\theta)'$ involves a finite number of terms v that are $o_p(1)$ uniformly in θ . Hence, uniformly in v:

$$\sum_{t=k+1}^{k_0} T^{-1} F_t(\bar{\theta}_{t,2,2}) F_t(\bar{\theta}_{t,2,2})' = o_p(1)$$

Since $\breve{\theta}_2 = O_p(1)$ uniformly in v, we have:

$$\Sigma_{2,1} = \breve{\theta}_2(v)' \left[\sum_{t=k+1}^{k_0} T^{-1} F_t(\bar{\theta}_{t,2,2}) F_t(\bar{\theta}_{t,2,2})' \right] \breve{\theta}_2(v) = O_p(1) \times o_p(1) \times O_p(1) = o_p(1)$$

Now turn to $\Sigma_{2,2}$:

$$\Sigma_{2,2} = \sum_{t=k+1}^{k_0} d_t^2(\theta_2^0, \theta_1^0) = \sum_{t=k_0-v+1}^{k_0} d_t^2(\theta_2^0, \theta_1^0) = \sum_{t=-v+1}^{0} d_t^2(\theta_2^0, \theta_1^0),$$

where the latter follows from the strict stationarity assumption on $\{x_t\}$ imposed in A3.1. Note that this is exactly the drift term present in the definition of $W_1^*(n)$, because if $n = -v$, then $n < 0$ and $\mid n \mid \leq N$. So, we have:

$$\Sigma_{2,2} = \sum_{t=n+1}^{0} d_t^2(\theta_2^0, \theta_1^0), \; n < 0.$$

Next, consider $\Sigma_{2,3}$:

$$\begin{aligned} \Sigma_{2,3} &= 2 \sum_{t=k+1}^{k_0} d_t(\hat{\theta}_2(k), \theta_2^0) \, d_t(\theta_2^0, \theta_1^0) \\ &= 2\breve{\theta}_2(v)' \left[\sum_{t=k+1}^{k_0} T^{-1/2} F_t(\bar{\theta}_{t,2,2}) F_t(\tilde{\theta}_{t,2,1}^{0,0})' \right] \delta \end{aligned}$$

Now $F_t(\bar{\theta}_{t,2,2}) F_t(\tilde{\theta}_{t,2,1}^{0,0})'$ is bounded, by A3.3(iv). Hence, uniformly in v,

$$\sum_{t=k+1}^{k_0} T^{-1/2} F_t(\bar{\theta}_{t,2,2}) F_t(\tilde{\theta}_{t,2,1}^{0,0})' = o_p(1)$$

By Assumption A7, $\delta = O(1)$. Then $\Sigma_{2,3}$ is, uniformly in v,

$$\Sigma_{2,3} = 2\breve{\theta}_2(v)' \left[\sum_{t=k+1}^{k_0} T^{-1/2} F_t(\bar{\theta}_{t,2,2}) F_t(\breve{\theta}_{t,2,1}^{0,0})' \right] \delta = O_p(1) \times o_p(1) \times O(1) = o_p(1)$$

We are left with $\Sigma_{2,4}$ and $\Sigma_{2,5}$.

$$\Sigma_{2,4} = -2 \sum_{t=k+1}^{k_0} u_t \, d_t(\hat{\theta}_2(k), \theta_2^0) = -2\breve{\theta}_2(v) \left[T^{-1/2} \sum_{t=k+1}^{k_0} u_t F_t(\bar{\theta}_{t,2,2}) \right] \tag{5.22}$$

The quantity $\sum_{t=k+1}^{k_0} T^{-1/2} u_t F_t(\theta)$ involves a bounded number of terms that are $o_p(T^{-1/2})$, because $E\left[u_t F_t(\theta)\right] = 0$. It follows that:

$$T^{-1/2} \sum_{t=k+1}^{k_0} u_t F_t(\bar{\theta}_{t,2,2}) = o_p(T^{-1/2}). \tag{5.23}$$

Using (5.23) into (5.22), uniformly in v:

$$\Sigma_{2,4} = -2\breve{\theta}_2(v) \left[T^{-1/2} \sum_{t=k+1}^{k_0} u_t F_t(\bar{\theta}_{t,2,2}) \right] = -O_p(1) \times o_p(T^{-1/2}) = o_p(T^{-1/2}).$$

We also have:

$$\begin{aligned}
\Sigma_{2,5} &= -2 \sum_{t=k+1}^{k_0} u_t \, d_t(\theta_2^0(k), \theta_1^0) = -2 \sum_{t=-v+1}^{0} u_t \, d_t(\theta_2^0(k), \theta_1^0) \\
&= -2 \sum_{t=n+1}^{0} u_t \, d_t(\theta_2^0(k), \theta_1^0),
\end{aligned}$$

where $n = -v < 0$. The latter follows from strict stationarity of $\{u_t f_t(\theta)\}$. $\Sigma_{2,5}$ is the second term appearing in the definition of $W_1^*(n)$. Since $\Sigma_2 = \Sigma_{2,1} + \Sigma_{2,2} + \Sigma_{2,3} + \Sigma_{2,4} + \Sigma_{2,5}$, we have, uniformly in v:

$$\begin{aligned}
\Sigma_2 &= o_p(1) + \sum_{t=n+1}^{0} d_t^2(\theta_2^0, \theta_1^0) + o_p(1) + o_p(T^{-1/2}) - 2 \sum_{t=n+1}^{0} u_t \, d_t(\theta_2^0, \theta_1^0) \\
&= \sum_{t=n+1}^{0} d_t^2(\theta_2^0, \theta_1^0) - 2 \sum_{t=n+1}^{0} u_t \, d_t(\theta_2^0, \theta_1^0) + o_p(1) \qquad (5.24)
\end{aligned}$$

Equations (5.17) and (5.24) imply that the minimand is, uniformly in $k - k_0 = n$, where

$-N \leq n < 0$:

$$V(k, \hat{\theta}_1, \hat{\theta}_2) = \mathfrak{D} + \sum_{t=n+1}^{0} d_t^2(\theta_2^0, \theta_1^0) - 2 \sum_{t=n+1}^{0} u_t \, d_t(\theta_2^0, \theta_1^0) + o_p(1)$$

If follows that:

$$\hat{k} = \underset{k=k_0+n,\, -N\leq n<0}{\text{argmin}} V(k, \hat{\theta}_1, \hat{\theta}_2) = \underset{k=k_0+n,\, -N\leq n<0}{\text{argmax}} -V(k, \hat{\theta}_1, \hat{\theta}_2)$$

Hence,

$$\begin{aligned}
\hat{k} - k_0 &= \underset{-N\leq n<0}{\text{argmax}} \; V(k, \hat{\theta}_1, \hat{\theta}_2) \\
&= \underset{-N<n<0}{\text{argmax}} \left[\mathfrak{D} - \sum_{t=n+1}^{0} d_t^2(\theta_2^0, \theta_1^0) + 2 \sum_{t=n+1}^{0} u_t \, d_t(\theta_2^0, \theta_1^0) + o_p(1) \right] \\
&= \underset{-N\leq n<0}{\text{argmax}} \left[W_1^*(n) + o_p(1) + \mathfrak{D} \right].
\end{aligned}$$

Note that the term $\mathfrak{D}$ is independent of n. Hence we can write, as a finite sample approximation for the distribution of $\hat{k}$, when $\hat{k} < k_0$:

$$\hat{k} - k_0 = \underset{-N\leq n<0}{\text{argmax}} \, W_1^*(n) \tag{5.25}$$

Case 2: $k > k_0$.

This case can be handled in a similar fashion to obtain:

$$\hat{k} - k_0 = \underset{0<n\leq N}{\text{argmax}} \, W_2^*(n) \tag{5.26}$$

Case 3: $k = k_0$.

Then $\hat{k} = k_0$, in which case $V(k, \hat{\theta}_1, \hat{\theta}_2)$ is just the difference between the NLS estimated and true sum of squared residuals. Standard NLS asymptotic theory (described in Section 2.3) shows that the difference converges in probability to zero. Hence we can write $W^*(n=0)=0$ as a finite sample approximation. We obtain:

$$\left[\hat{k}-k_0\right] - \underset{|n|\leq N}{\operatorname{argmax}} W^*(n) \xrightarrow{p} 0,$$

completing the proof of Theorem 5.1. □

Let us briefly discuss the properties of the approximating distribution $W^*(n)$. It usually depends on the distribution of $\{f_t(\theta), u_t\}$. For $n<0$, as n grows unbounded, $W_1^*(n)$ converges to $-\infty$ because it has a negative stochastic drift. This can be seen by noting that, when $n<0$,

$$\Delta W_1^*(n) = W_1^*(n) - W_1^*(n-1) = -d_n^2(\theta_2^0, \theta_1^0) + 2u_n d_n(\theta_2^0, \theta_1^0)$$

$$E\Delta W_1^*(n) = -E[d_n^2(\theta_2^0, \theta_1^0)] = -E^2[d_n(\theta_2^0, \theta_1^0)] - Var[d_n(\theta_2^0, \theta_1^0)] < 0.$$

Secondly, $W_2^*(n)$ converges to $-\infty$ because it has a negative stochastic drift, since $-E\Delta W_2^*(n) < 0$. Hence, the process $W^*(n)$ achieves its unique maximum near zero, as expected since $\hat{k}-k_0 \xrightarrow{p} 0$.[1] So, we can write the following:

Corrolary 1 to Theorem 5.1. *Under A3.1 or A4.1 and A3.2-A3.4, A5.1 , $\hat{k}-k_0-\operatorname{argmax}_{n\in\mathbb{Z}} W^*(n) \xrightarrow{p} 0$, where $\mathbb{Z}$ is the set of all integers.*

This corollary simply follows from the fact that $W^*(\cdot)$ has a unique maximum.

We would eventually like to obtain a limiting distribution rather than a large sample approximation to it. If this limiting distribution doesn't depend on the data, then we can easily construct confidence intervals for the break-fractions. We will show that this is indeed the case given A4.1:

[1] As noted by Bai(1997), if $\{v_t\} = \{(u_t, x_t)\}$ is an independent process, then $W^*(n)$ is a two-sided random walk with stochastic drifts.

Corrolary 2 to Theorem 5.1. *Given A3.2-A5.2,*

$$\frac{a^*}{\sigma^4}[\hat{k} - k_0] \xrightarrow{d} \underset{n}{\operatorname{argmax}}\, Z(n) = [\; W(n) - 0.5|n| \;]$$

where $W(n) = W_1(-n), n < 0$, $W(n) = W_2(n), n \geq 0$ *and* $W_1(n), W_2(n)$ *are two independent standard scalar Gaussian processes defined on* $[0, \infty]$, *and* $a^* = E[d_t(\theta_1^0, \theta_2^0)]^2$.

Note that the quantity a^*/σ^4 is just a scaling factor, where a^* depends on δ. It can be dropped but we included it to symbolize the fact the break-fractions' distributions may depend on how large the shift is. Its significance will become clear in the next chapter.

Proof of Corollary 2 to Theorem 5.1.

Case 1. Consider first the case $n < 0$. Then:

$$W_1^*(n) = -\sum_{t=n+1}^{0} d_t^2(\theta_2^0, \theta_1^0) \;+\; 2\sum_{t=n+1}^{0} u_t\, d_t(\theta_2^0, \theta_1^0)$$

Since $\{d_t^2(\theta_2^0, \theta_1^0)\}$ is strictly stationary, it has constant expectation for each t. Denote $E[d_t^2(\theta_2^0, \theta_1^0)] = a^*$. So,

$$E[\sum_{t=n+1}^{0} d_t^2(\theta_2^0, \theta_1^0)] = \sum_{t=n+1}^{0} E[d_t^2(\theta_2^0, \theta_1^0)] = |n|\, a^*$$

On the other hand, $2\sum_{t=n+1}^{0} u_t\, d_t(\theta_2^0, \theta_1^0)$ is a process with independent increments that have expectation $E[u_t\, d_t(\theta_2^0, \theta_1^0)] = 0$ and:

$$\begin{aligned}
\text{Var}\;[u_t\, d_t(\theta_2^0, \theta_1^0)] &= \text{Var}\;[\psi_t(\theta_1^0) - \psi_t(\theta_1^0)] \\
&= E[\psi_t^2(\theta_1^0)] + E[\psi_t(\theta_2^0)] - 2E[\psi_t(\theta_1^0)\psi_t(\theta_2^0)] \\
&= E[u_t^2|x_t]\; E[f_t(\theta_1^0) - f_t(\theta_2^0)]^2 = \sigma^2 a^*.
\end{aligned}$$

At this point, we need a functional central limit theorem (FCLT). We are going to adapt it again from Caner (2007):

Proposition 5.1 (FCLT). *If:*

(i) $\{v_t\}$ *is a strictly stationary* α*-mixing process of size* $-a$*, with* $a > 2 + 4/\delta$*, for some* $\delta > 0$*;*

(ii) $h_t(\theta) = h(v_t, \theta)$ *is a vector valued function such that* $E[h_t(\theta)] = 0$ *and* $\sup_\theta E\|h_t(\theta)\|^{2+\delta} < \infty$*;*

(iii) $\|h_t(\theta_1) - h_t(\theta_2)\| \le H_t\|\theta_1 - \theta_2\|$*, with* $E\|H_t\|^{2+\delta} < \infty$*, then:*

$$T^{-1}\sum_{t=1}^{[Tr]} [h_t(\theta)] \xrightarrow{d} \{E[h_t(\theta)]^2\}^{1/2} B_p(r),$$

where $B_p(r)$ *is a* $p \times 1$ *Brownian motion defined on* $r \in (0, 1)$*.*

We have already shown that quantities of the type $\psi_t(\theta)$ satisfy the conditions above, since they are identical to the ones in Proposition 3.1. Hence, $h_t(\theta)$, here replaced by $2u_t\, d_t(\theta_2^0, \theta_1^0)$, satisfy these conditions too. By the FCLT,

$$T^{-1}\sum_{t=1}^{[Tr]} 2\left[u_t d_t(\theta_1^0, \theta_2^0)\right] \xrightarrow{d} 2(a^*)^{1/2}\sigma B_1(r).$$

Let r be such that $n = -[Tr]$. Also, $TB_1(r) = B_1(Tr) = W_1(Tr)$. Note that for Tr an integer, $B_1(Tr) = W_1(-n)$. Hence:

$$\sum_{t=n+1}^{0} 2\left[u_t d_t(\theta_1^0, \theta_2^0)\right] \xrightarrow{d} 2(a^*)^{1/2}\sigma W_1(-n),$$

so:

$$W_1^* \xrightarrow{d} 2(a^*)^{1/2}\sigma W_1(-n) - a^*|n|$$

If we perform the change in variable $n = bn'$, it can be shown that $\text{argmax}_{n'} B(n') = b^{-1}\,\text{argmax}_n B(n)$ (see Bai (1997), Proof of Proposition 3). Let $b = a^*/2\sigma^2$. Then:

$$\begin{aligned} \hat{k} - k_0 &= \frac{2\sigma^2}{a*} \underset{n\leq 0}{\text{argmax}} \left[W_1(-\sigma^4 n) - 0.5|n|\sigma^2\right] + o_p(1) \qquad (5.27)\\ &= \frac{\sigma^4}{a*} \underset{n\leq 0}{\text{argmax}}\; o_p(1) + [W_1(-n) - 0.5|n|\,] \equiv \frac{\sigma^4}{a*} \underset{n\leq 0}{\text{argmax}}\, [o_p(1) + Z_1(n)] \end{aligned}$$

The properties of the latter distribution are described in Bai (1997).

Case 2. By a similar argument, when $n \geq 0$,

$$\hat{k} - k_0 = \frac{\sigma^4}{a*} \underset{n\leq 0}{\text{argmax}}\, [o_p(1) + Z_2(n)] = o_p(1) + [W_2(n) - 0.5|n|\,], \qquad (5.28)$$

where $W_2(n)$ is a Gaussian stochastic process defined on $[0, \infty]$. Combining (5.27) with (5.28) yields:

$$\frac{a*}{\sigma^4}[\hat{k} - k_0] \xrightarrow{d} \underset{n}{\operatorname{argmax}}\, Z(n),$$

which is the desired result in Corollary 2 to Theorem 5.1. □

5.2 Shrinking Magnitudes of Shifts

Consider our initial regression model with m unknown shifts, defined in (3.1). Now instead of A5.1, consider A5.3, which imposes parameter shifts that are not fixed, but shrinking at a certain rate v_T:

Assumption A 5.3. *For $i = 1, \ldots, m$, $\delta_{i,T} = \delta_i\, v_T = \theta^0_{i+1,T} - \theta^0_{i,T}$, where δ_i is a fixed $p \times 1$ vector and $\{v_T\}$ is a scalar series such that $v_T \to 0$ and $T^{1/2-\nu} v_T^2 \to \infty$ as $T \to \infty$, for some $\nu \in \left[0, \frac{1}{2}\right)$.*

This assumption is sufficient for obtaining pivotal statistics. In other words, it ensures that the asymptotic distributions of the change-point estimates do not depend on the underlying distributions of $\{u_t, f_t(\theta)\}^{+\infty}_{t=-\infty}$. The order of v_T says that the parameter shifts are decaying slow enough to permit good asymptotic behavior of our estimates. Similar assumptions are *interalia* in $T^{1/2-\nu} v_T \to \infty$, for $\nu \in \left(0, \frac{1}{2}\right)$ in Bai and Perron (1998) and $T^{1/2} v_T/(logT)^2 \to \infty$ in Qu and Perron (2007). Our assumption is more restrictive because it allows only shifts of order $T^{-1/4}$ or larger, compared to an order arbitrarily close to $T^{-1/2}$ maintained in the above-mentioned papers.

Under shrinking magnitudes of shift, all properties of break-fraction estimates need to be re-derived. The consistency proof reveals that our stronger requirement is a consequence of the lack of finite sample closed-form expressions for the NLS parameter estimates.

5.2.1 Consistency of Break-Fraction Estimates

Even though the magnitude of the shifts decreases, the change-point estimates remain consistent, as stated below:

Theorem 5.2.

Under Assumptions A3.1-A3.5 and A5.3, $\hat{\lambda}_i \xrightarrow{p} \lambda_i^0$*, for* $i = 1, \ldots, m$.

The proof of Theorem 5.2 is similar to that of Theorem 3.1, but modifications are required in light of the possibility that $T^{-1}\sum_{t=1}^{T} d_t^2 \xrightarrow{p} 0$ even if a break-fraction is not consistently estimated.

Example. *To illustrate this possibility, consider for a moment the following single break model:*

$$y_{t,T} = \begin{cases} f_t(\theta_{1,T}^0) & t = 1, \ldots, k_0 \\ f_t(\theta_{2,T}^0) & t = k_0 + 1, \ldots, T. \end{cases}$$

Assume the change-fraction estimate has a fixed bias $\lambda_1^0 - \hat{\lambda}_1 = c$*, where* $c \in (0,1)$*, and* cT *is an integer, but the NLS-estimates are consistent, hence* $\hat{\theta}_i - \theta_{i,T}^0 = o_p(1)$ *for* $i = 1, 2$*. Assumption A5.3 implies that* $\theta_{1,T}^0 - \theta_{2,T}^0 = o_p(1)$ *and hence* $\hat{\theta}_2 - \theta_{1,T}^0 = o_p(1)$*. We will drop the* T *subscripts for convenience. Then:*

$$\begin{aligned} & T^{-1}\sum_{t=1}^{T} d_t^2 \\ &= T^{-1}\sum_{t=1}^{[T\hat{\lambda}]} d_t^2(\hat{\theta}_1, \theta_1^0) \,+\, T^{-1}\sum_{t=[T\hat{\lambda}]+1}^{[T\lambda_0]} d_t^2(\hat{\theta}_2, \theta_1^0) \,+\, T^{-1}\sum_{t=[T\lambda_0]+1}^{T} d_t^2(\hat{\theta}_2, \theta_2^0) \\ &= B_1 + B_2 + B_3. \end{aligned}$$

Applying MVT to B_1, we obtain:

$$B_1 = [\hat{\theta}_1 - \theta_1^0]' \, T^{-1} \sum_{t=1}^{[T\hat{\lambda}]} F_t(\bar{\theta}_{t,1,1}) F_t(\bar{\theta}_{t,1,1})' \, [\hat{\theta}_1 - \theta_1^0]$$
$$= o_p(1) \times O_p(1) \times o_p(1) = o_p(1).$$

Similarly, $B_3 = o_p(1)$. For B_2, we get:

$$B_2 = [\hat{\theta}_2 - \theta_1^0]' \, T^{-1} \sum_{t=[T\hat{\lambda}]+1}^{[T\lambda_0]} F_t(\bar{\theta}_{t,2,1}) F_t(\bar{\theta}_{t,2,1})' \, [\hat{\theta}_2 - \theta_1^0].$$

Now $T^{-1} \sum_{t=[T\hat{\lambda}]+1}^{k_0} F_t(\theta) F_t(\theta)'$ contains an infinite number of terms Tc. Hence, by A3.4 (ii), $T^{-1} \sum_{t=[T\hat{\lambda}]+1}^{[T\lambda_0]} F_t(\theta) F_t(\theta)' \xrightarrow{p} cW(\theta)$, uniformly in θ. Then $\hat{\theta}_2 - \theta_1^0 = o_p(1)$ ensures that $\bar{\theta}_{t,2,1} - \theta_1^0 = o_p(1)$, hence:

$$T^{-1} \sum_{t=[T\hat{\lambda}]+1}^{[T\lambda_0]} F_t(\bar{\theta}_{t,2,1}) F_t(\bar{\theta}_{t,2,1})' \xrightarrow{p} cD(\theta_1^0) = O_p(1).$$

The above implies:

$$B_2 = [\hat{\theta}_2 - \theta_1^0]' \, T^{-1} \sum_{t=[T\hat{\lambda}]+1}^{[T\lambda_0]} F_t(\bar{\theta}_{t,2,1}) F_t(\bar{\theta}_{t,2,1})' \, [\hat{\theta}_2 - \theta_1^0]$$
$$= o_p(1) \times O_p(1) \times o_p(1) = o_p(1).$$

Hence, even if the break-fraction estimate $\hat{\lambda}$ is inconsistent, we still get:

$$T^{-1} \sum_{t=1}^{T} d_t^2 = B_1 + B_2 + B_3 = o_p(1) + o_p(1) + o_p(1) = o_p(1),$$

showing that the consistency proof needs to be rethought in light of Assumption A5.3.

Proof of Theorem 5.2.

Step 1. We will start with inequality (3.4), which is deduced independently of the magnitude of shifts.

$$T^{-1}\sum_{t=1}^{T} d_t^2 + 2T^{-1}\sum_{t=1}^{T} d_t u_t \leq 0.$$

Independently of the magnitude of shifts, Proposition 3.1 holds, implying that:

$$\sum_{t=1}^{[Tr]} u_t f_t(\theta) = O_p(T^{1/2}) \text{ uniformly in } \theta \times r, \tag{5.29}$$

as we have shown in the Proof of Lemma 3.1.

Step 2.

We slightly modify Lemma 3.1 to the Lemma below :

Lemma 5.1. *Under A3.1-A3.5 and A5.3, we have:* $\sum_{t=1}^{T} u_t d_t = O_p(T^{1/2+\nu})$*, uniformly over the space of all partitions and parameters* $(T_1, \ldots, T_m) \times \theta$*.*

Step 3.

We replace Lemma 3.2 with the Lemma below:

Lemma 5.2. *Consider the framework of A3.1-A3.5 and A5.3. If* $\hat{\lambda}_j \overset{p}{\nrightarrow} \lambda_j^0$ *for some j, then limsup* $P\left[\sum_{t=1}^{T} d_t^2 > TCv_t^2\right] > \epsilon$*, for some* $C > 0, \epsilon > 0$.

From Lemma 5.2, with positive probability greater than $\epsilon > 0$:

$$\sum_{t=1}^{T} d_t^2 > C(Tv_T^2). \tag{5.30}$$

Combining Lemma 5.2 with(5.30), we obtain:

$$\begin{aligned}\sum_{t=1}^{T} d_t^2 + 2\sum_{t=1}^{T} d_t u_t &> C(Tv_T^2) + 2O_p(T^{1/2+\nu}) \qquad (5.31)\\ &= O_p(T^{1/2+\nu})\left[O_p(T^{1/2-\nu}v_T^2) + 2\right] \to \infty,\end{aligned}$$

by A5.3, which ensures $T^{1/2-\nu}v_T^2 \to \infty$ as long as $v_T \to 0$ and $\nu \in [0, \frac{1}{2})$. But (5.31) contradicts (3.4). Hence, Lemma 5.2 is contradicted, implying that all break-fractions must be consistently estimated. We are left with proving Lemma 5.1 and 5.2. □

Proof of Lemma 5.1:

By Proposition 3.1, we have that $\sup_{\theta\times r}\sum_{t=1}^{Tr} u_t f_t(\theta) = O_p(T^{1/2}) \leq O_p(T^{1/2+\nu})$, for any $\nu \geq 0$. Since in the proof of Lemma 3.1, we have shown that $|\sum_{t=1}^{T} u_t d_t| \leq 4(m+1)\sup_{\theta\times r}|\sum_{t=1}^{Tr} u_t f_t(\theta)|$, it follows that:

$$\sum_{t=1}^{T} u_t d_t = O_p(T^{1/2+\nu}),$$

uniformly over all partitions and θ. □

Proof of Lemma 5.2:

The proof of Lemma 5.2 is the same as for Lemma 3.2. There, we let:

$$\sum_{1*} = \sum_{t\in[[T\lambda_j^0]-T\eta+1,\ [T\lambda_j^0]]} \quad \text{and} \quad \sum_{2*} = \sum_{t\in[[T\lambda_j^0]+1,\ [T\lambda_j^0]+T\eta]}, \text{ where } T\eta \text{ is an integer.}$$

If $\hat{\lambda}_j \overset{p}{\nrightarrow} \lambda_j^0$, then with probability $\geq \epsilon$, we had:

$$\begin{aligned}
\sum_{t=1}^{T} d_t^2 &\geq \sum_{1*} d_t^2(\hat{\theta}_k, \theta_j^0) + \sum_{2*} d_t^2(\hat{\theta}_k, \theta_{j+1}^0) \\
&\geq \inf_{\theta}\left[\sum_{1*} d_t^2(\theta, \theta_j^0) + \sum_{2*} d_t^2(\theta, \theta_{j+1}^0)\right] \\
&= \inf_{\theta}\left[(\theta-\theta_j^0)'\sum_{1*} F_t(\bar{\theta}_{t,j})F_t(\bar{\theta}_{t,j})'(\theta-\theta_j^0) + (\theta-\theta_{j+1}^0)'\sum_{2*} F_t(\bar{\theta}_{t,j+1})F_t(\bar{\theta}_{t,j+1})'(\theta-\theta_{j+1}^0)\right],
\end{aligned}$$

where $\bar{\theta}_{t,i}$ is in the hyperplane joining θ and θ_i^0, such that: $d_t(\theta, \theta_i^0) = (\theta-\theta_i^0)'F_t(\bar{\theta}_{t,i})$, $i \in \{j, j+1\}$.

Now $\sum_{1*} F_t(\bar{\theta}_{t,j})F_t(\bar{\theta}_{t,j})'$ and $\sum_{2*} F_t(\bar{\theta}_{t,j+1})F_t(\bar{\theta}_{t,j+1})'$ involve $T\eta$, an infinite number of terms. Both sums are $O_p(T)$ by similar arguments employed before. Hence, there is some $C > 0$, such that, with positive probability greater than $\epsilon > 0$,

$$\sum_{t=1}^{T} d_t^2 \geq \inf_{\theta} \left[\, \|\theta - \theta_j^0\|^2 + \|\theta - \theta_{j+1}^0\|^2 \, \right] O_p(T) \geq \frac{1}{2}\|\theta_j^0 - \theta_{j+1}\|^2 O_p(T) > CTv_T^2,$$

which completes the proof of Lemma 5.2. □

We conclude that the break fractions are consistently estimated. We are now in place to compare A5.3, which says that $v_T \to 0$ and $T^{1/2-\nu}v_T^2 \to \infty$, with less restrictive assumptions found in the literature (e.g. Bai and Perron (1998) impose $T^{1/2-\nu}v_T \to \infty$). First, both us and the previous authors show that $\sum_{t=1}^{T} d_T^2 \geq O_p(Tv_T^2)$. However, in linear models, Bai and Perron (1998) found that $\sum_{t=1}^{T} u_t d_t \leq O_p(T^{1/2+\nu}v_T)$, for some $\nu \in (0, \frac{1}{2})$. The rate of convergence they impose for v_T is smaller than our rate $O_p(T^{1/2+\nu})$. Their smaller rate allows $\sum_{t=1}^{T} d_T^2 + 2\sum_{t=1}^{T} u_t\, d_t \to \infty$ while assuming only $T^{1/2-\nu}v_T \to \infty$. Meanwhile, we need $T^{1/2-\nu}v_T^2 \to \infty$ to ensure that $\sum_{t=1}^{T} d_T^2$ dominates $2\sum_{t=1}^{T} u_t\, d_t$.

Heuristically, consider a linear model. The usual OLS estimator $\hat{\theta}$ can be written as $\theta^0 + (X'X)^{-1}X'u$, for some regressor matrix X, error vector u and parameter vector θ^0. This allows rewriting partial sums of $\sum_{t=1}^{T} u_t d_t$ as $(\theta_i^0 - \theta_{i-1}^0)O_p(T^{1/2+\nu})$, where $i = (2, \ldots, m+1)$. Since $\theta_i^0 - \theta_{i-1}^0 = \delta_{i,T}v_T$, one finds $\sum_{t=1}^{T} u_t d_t \leq O_p(T^{1/2+\nu}v_T)$. In nonlinear models, no exact formula for NLS is available. Resorting to MVT would not provide us with extra terms of shrinking sort $(\theta_i^0 - \theta_{i-1}^0) = O(v_T)$, because we do not know the properties of our parameter estimates in terms of the true values. This restricts the assumptions we can make about v_T to ensure consistency of break-fraction estimates.

5.2.2 Convergence Rate of Break-Fraction Estimates

To derive the rate of convergence of the break-fraction estimates, we need to strengthen A3.1:

Assumption A 5.4 (Dependence and Memory of Processes).

(i) Let $v_t = (x_t, u_t)$. Then $\{v_t\}$ is a strictly stationary α-mixing process of size $-\ell - \delta$, where

$\ell = 4s/(s-2)$, $s > 2$, $\delta > 0$;

(ii) The errors are uncorrelated with the regression function, i.e. $E\left[u_t f_t(\theta)\right] = 0$ *for all* θ, t.

Note that the above assumption implies A3.1, but imposes a faster rate of memory decay. Additionally, we need to show the following extension to Hajek-Renyi inequality:

Lemma 5.3 (**Extended Hayek-Renyi Inequality**). *Let* $\mathfrak{F}_{-\infty}^{t}$ *be an increasing sequence of* σ*-fields generated by* $\{\Psi_i\}_{i=-\infty}^{t}$ *in the space of uniformly bounded and continuous functions* $\mathbf{C} = \{\Psi : \Theta \to \mathbb{R}^p\}$.[2] *Denote by* $\| \Psi_t \|_s$ *the* $\mathcal{L}^s$ *norm of* Ψ_t, *i.e.* $\| \Psi_t \|_s = E^{1/s} \| \Psi_t \|^s$. *If* $\{\Psi_t(\theta), \mathcal{F}_{-\infty}^{t}\}$ *is an* $\mathcal{L}^2$ *mixingale, then there exists an* $L < \infty$, *not depending on* θ, *such that for every* $b > 0$ *and* $m > 0$, *uniformly in* θ,

$$P\left(\sup_{k \geq n} \frac{1}{k} \left\| \sum_{t=1}^{k} \Psi_t(\theta) \right\| \geq b \right) \leq \frac{L}{b^2 n}. \tag{5.32}$$

Note that this inequality is the same as in Bai and Perron (1998), except that ours holds uniformly in θ. Since the extended Hajek-Renyi inequality we would like to prove is defined in terms of mixingales, we will first show:

Lemma 5.4. *Under A1,* $\{\Psi_i(\theta), \mathfrak{F}_i\}$ *is in fact an* $\mathcal{L}^2$ *mixingale of size* $\xi = -2 - \delta(\frac{1}{2} - \frac{1}{s})$, *for* $\delta > 0$ *and* $s > 2$ *defined in Assumption A1.*

A similar lemma can be found in the Supplement to Qu and Perron (2007).

Proof of Lemma 5.4:

The proof verifies the conditions in Definition 2.1 of a mixingale.

Proof of (i). $\Psi_t(\theta)$ is α-mixing of the same size as $\{v_t\}$. We will use McLeish's (1975) inequality[3] to link the concepts of α-mixing and mixingales. Recall that the α-measure of dependence is defined as:

$$\alpha(\mathcal{F}, \mathcal{G}) = \sup_{F \in \mathcal{F}, G \in \mathcal{G}} |P(F \cap G) - P(F)P(G)|,$$

[2] This space is studied in Billingsley (1968), Ch. 3.
[3] Lemma 2.1. in McLeish (1975).

where $\mathcal{F}, \mathcal{G}$ are two σ-algebras.

McLeish's (1975) Inequality.

Suppose Ψ *is a random vector measurable with respect to* $\mathcal{G}$, $1 \leq w_1 \leq w_2 \leq \infty$. *Then:*

$$\| E(\Psi \mid \mathcal{F}) - E\Psi \|_{w_1} \leq 2(2^{1/w_1} + 1)\alpha^{\frac{1}{w_1} - \frac{1}{w_2}}(\mathcal{F}, \mathcal{G}) \| \Psi \|_{w_2} \tag{5.33}$$

The α-mixing coefficients, as defined in McLeish (1975), are:

$$\bar{\alpha}_j = \sup_i \alpha(\mathcal{F}^i_{-\infty}, \mathcal{F}^{i+j}_{i+j}) \leq \alpha_j = \sup_i \alpha(\mathcal{F}^i_{-\infty}, \mathcal{F}^{\infty}_{i+j}),$$

where the latter are our α-coefficients and $i, j > 0$. Hence, let in the inequality above, $\mathcal{F} \equiv \mathcal{F}^i_{-\infty}$, an increasing sequence of σ-algebras generated by $\{\Psi_t(\theta)\}^i_{-\infty}$. In the same inequality, let $\mathcal{G} \equiv \mathcal{F}^i_i$ be the σ-algebra generated by Ψ_i, $w_1 = 2$ and $w_2 = s$. By A1, $\Psi_i(\theta)$ is measurable with respect to $\mathcal{F}^i_{-\infty}$, hence also with respect to $\mathcal{F}^i_i \subset \mathcal{F}^i_{-\infty}$. Then, according to (5.33), we get:

$$\begin{aligned}\| E(\Psi_i(\theta) \mid \mathcal{F}_{i-j}) - E[\Psi_i(\theta)] \|_2 &\leq 2(2^{1/2} + 1)\bar{\alpha}_j^{\frac{1}{2} - \frac{1}{s}} \| \Psi_i(\theta) \|_s \\ &\leq 2(2^{1/2} + 1)\alpha_j^{\frac{1}{2} - \frac{1}{s}} \| \Psi_i(\theta) \|_s\end{aligned}$$

We know $E[\Psi_i(\theta)] = 0$. Also, $\bar{\alpha}_j \leq \alpha_j = O[j^{-\ell-\delta}]$. Hence:

$$\alpha_j^{\frac{1}{2} - \frac{1}{s}} = Dj^{(-\frac{4s}{s-2} - \delta)(\frac{1}{2} - \frac{1}{s})} = Dj^{-2-\delta(\frac{1}{2} - \frac{1}{s})} = D(j^{\xi}),$$

where ξ the size of our mixingale given in Lemma 5.4, and D is a positive constant. McLeish's

(1975) inequality becomes:

$$\begin{aligned}\| E(\Psi_i(\theta) \mid \mathfrak{F}_{i-j}) \|_2 &\leq 2(2^{1/2}+1)Dj^{\xi} \| \Psi_i(\theta) \|_s \\ &\leq 2(2^{1/2}+1)Dj^{\xi} \sup_{\theta} \| \Psi_i(\theta) \|_s \\ &\leq 2MD(2^{1/2}+1)j^{\xi},\end{aligned}$$

where M is a positive constant, and the last inequality follows from the boundedness assumption A3.3(iv). Let ϕ_j in the definition of mixingales be equal to j^{ξ}; then it is positive and of order j^{ξ}, satisfying the requirements stated in that definition. Also let $c_i = M2(2^{1/2}+1)D = K > 0$; then:

$$\| E(\Psi_i(\theta) \mid \mathfrak{F}_{i-j}) \|_2 \leq c_i \phi_j,$$

satisfying requirement (i) in our mixingale definition. Note that neither c_i nor ϕ_j depend on θ. This holds because we have defined our α-coefficients uniformly in θ. Alternatively, the α-coefficients might be functions of θ, case assumed away for simplicity of exposition.

Proof of (ii) Since $\Psi_i(\theta)$ is $\mathcal{F}^{i}_{-\infty}$-measurable, by Andrews (1988), it follows that $E(\Psi_i \mid \mathfrak{F}_{i+j}) = \Psi_i(\theta)$ almost surely (a.s.). It follows:

$$E|\Psi_i(\theta) - E(\Psi_i \mid \mathfrak{F}_{i+j})|^2 = E[0 \text{ a.s }]^2 = 0 \leq c_i \phi_{j+1},$$

which is part (ii) of the mixingale definition.

Proof of (iii) Since $c_i = K > 0$, it follows that $\max_i c_i = K < \infty$.

Proof of (iv) Pick $\kappa = \frac{\delta}{2}(\frac{1}{2} - \frac{1}{s}) > 0$. Then $j^{1+\kappa}\phi_j = j^{1+\kappa-2-2\kappa} = j^{-1-\kappa}$. Hence:

$$\sum_{j=1}^{\infty} j^{1+\kappa}\phi_j = \sum_{j=1}^{\infty} j^{-1-\kappa} < \infty,$$

because this inequality is true for any power of j less than -1. Combining conditions (i)-(iv), we find that indeed $\Psi_t(\theta)$ is an $\mathcal{L}^2$ mixingale of size $\xi = -2 - \delta(\frac{1}{2} - \frac{1}{s})$. □

We are now ready to prove the Hajek-Renyi inequality stated before.

Proof of Lemma 5.3:

By the mixingale property, we can write $\Psi_t(\theta) = \sum_{j=-\infty}^{\infty} \Psi_{jt}(\theta)$, where $\Psi_{jt}(\theta) = E(\Psi_t(\theta)|\mathfrak{F}_{-\infty}^{t-j}) - E(\Psi_t(\theta)|\mathfrak{F}_{-\infty}^{t-j-1})$, and for each j, $\{\Psi_{jt}, \mathfrak{F}_{-\infty}^{t-j}\}$ form a sequence of martingale differences (see Bai and Perron (1998), Proof of Lemma A3.4). Then $\sum_{t=1}^{k} \Psi_t(\theta) = \sum_{j=-\infty}^{\infty} \sum_{t=1}^{k} \Psi_{jt}(\theta)$. Hence, for each $N > 0$,

$$\begin{aligned} P\Big(\sup_{N\geq k\geq n} \frac{1}{k}\Big\|\sum_{t=1}^{k} \Psi_t(\theta)\Big\| > b\Big) &\leq P\Big(\sum_{j=-\infty}^{\infty} \sup_{N\geq k\geq n} \frac{1}{k}\Big\|\sum_{t=1}^{k} \Psi_{jt}(\theta)\Big\| > b\Big) \\ &\leq \sum_{j=-\infty}^{\infty} P\Big(\sup_{N\geq k\geq n} \frac{1}{k}\Big\|\sum_{t=1}^{k} \Psi_{jt}(\theta)\Big\| > b\Big) \\ &\leq \sum_{j=-\infty}^{\infty} P\Big(\sup_{N\geq k\geq n} \frac{1}{k}\Big\|\sum_{t=1}^{k} \Psi_{jt}(\theta)\Big\| > a_j b\Big), \end{aligned} \tag{5.34}$$

where $0 < a_j < 1$ is such that $\sum_{j=-\infty}^{\infty} a_j = 1$. Recall that for each j, $\{\Psi_{jt}, \mathfrak{F}{-\infty}^{t-j}\}$ is a sequence of martingale differences. We will pick a_j later; for now, by Hajek-Renyi inequality for martingale differences (same as Bai and Perron (1998), Proof of Lemma A6):

$$\begin{aligned} &\sum_{j=-\infty}^{\infty} P\Big(\sup_{N\geq k\geq n} \frac{1}{k}\Big\|\sum_{t=1}^{k} \Psi_{jt}(\theta)\Big\| > a_j b\Big) \\ &\qquad \leq \frac{1}{c^2} \sum_{j=-\infty}^{\infty} \frac{1}{a_j^2}\Big(\frac{1}{n^2}\sum_{i=1}^{n} E\|\Psi_{ji}(\theta)\|^2 + \sum_{i=n+1}^{N} \frac{1}{i^2} E\|\Psi_{ji}(\theta)\|^2\Big). \end{aligned}$$

By property (i) of the mixingale $\Psi_t(\theta)$, we have that $E\|\Psi_{tj}(\theta)\|^2 = E\|\Psi_t(\theta) \mid \mathfrak{F}_{t-j}\|^2 \leq c_i^2\phi_j^2$, where none of the latter two terms, c_i or ϕ_j, depend on θ. Then $\dfrac{1}{n^2}\displaystyle\sum_{i=1}^{n} E\|\Psi_i(\theta) \mid \mathfrak{F}_{i-j}\|^2 \leq \dfrac{1}{n^2}\phi_{|j|}^2 \displaystyle\sum_{i=1}^{n} c_i^2$. We assigned earlier $c_i = K$, yielding $\dfrac{1}{n^2}\displaystyle\sum_{t=1}^{n} E\|\Psi_i(\theta) \mid \mathfrak{F}_{i-j}\|^2 \leq \dfrac{K^2\phi_{|j|}^2}{n}$. Similarly, we have $\displaystyle\sum_{i=n+1}^{N} \frac{1}{i^2} E\|\Psi_{ji}(\theta)\|^2 \leq K^2\phi_{|j|}^2 \sum_{i=n+1}^{N} \frac{1}{i^2} \leq K^2\phi_{|j|}^2 \sum_{i=n+1}^{\infty} \frac{1}{i^2} \leq \frac{2K^2\phi_{|j|}^2}{n}$ (last inequality is in Bai and Perron (1998), can be proven by mathematical induction for $n \geq 1$). The inequality above becomes:

$$\sum_{j=-\infty}^{\infty} P\left(\sup_{N\geq k\geq n}\frac{1}{k}\left\|\sum_{t=1}^{k}\Psi_{jt}(\theta)\right\| > a_j b\right) \leq \frac{3K^2}{c^2 n}\sum_{-\infty}^{+\infty}\frac{\phi_{|j|}^2}{a_j^2}.$$

Now let $L = 3K^2\sum_{-\infty}^{+\infty}\frac{\phi_{|j|}^2}{a_j^2}$; if we pick $a_j > 0$ such that $\sum_{-\infty}^{+\infty} a_j = 1$ and L is finite, then, for fixed N:

$$\sum_{j=-\infty}^{\infty} P\left(\sup_{N\geq k\geq n}\frac{1}{k}\left\|\sum_{t=1}^{k}\Psi_{jt}(\theta)\right\| > a_j b\right) \leq \frac{L}{c^2 n}. \tag{5.35}$$

Combining (5.34) with (5.35), we get the desired inequality for each $N > 0$,

$$P\left(\sup_{N\geq k\geq n}\frac{1}{k}\left\|\sum_{t=1}^{k}\Psi_{t}(\theta)\right\| > b\right) \leq \sum_{j=-\infty}^{\infty} P\left(\sup_{N\geq k\geq n}\frac{1}{k}\left\|\sum_{t=1}^{k}\Psi_{jt}(\theta)\right\| > a_j b\right) \leq \frac{L}{c^2 n},$$

completing the proof of Lemma 5.3.

We are left with picking a_j. Let $a_j = a_{-j} = \frac{j^{-1-\kappa}}{(1+2\sum_{l=1}^{\infty} j^{-1-\kappa})}$, for $j \geq 1$, and $a_0 = \frac{1}{(1+2\sum_{l=1}^{\infty} j^{-1-\kappa})}$. Then $a_j > 0$, $\sum_{-\infty}^{+\infty} a_j = 1$. Additionally:

$$\sum_{j=-\infty}^{+\infty}\frac{\phi_{|j|}^2}{a_j^2} = \left(1+2\sum_{j=1}^{+\infty}\phi_j^2 j^{2+2\kappa}\right)\left(1+2\sum_{l=1}^{\infty} j^{-1-\kappa}\right)^2.$$

Now $1+2\sum_{l=1}^{\infty} j^{-1-\kappa} < \infty$ because we chose $\kappa > 0$. Also, $1+2\sum_{j=1}^{+\infty}\phi_j^2 j^{2+2\kappa} \leq \left(\phi_0+2\sum_{j=1}^{\infty}\phi_j j^{1+\kappa}\right)^2 < \infty$ by property (iv) of the mixingale $\Psi_t(\theta)$. It follows that $L < \infty$, confirming that our choice of a_j yields a finite L. Recall that L does not depend on θ as long as ϕ_j, and hence α_j - the mixing coefficients - do not depend on θ. □

Lemma 5.3 is a result that will helps us establish the convergence rates of break-fraction estimates. These rates are different than the ones for fixed shifts, as can be seen from the Theorem below:

Theorem 5.3. *Under Assumptions A3.1-A3.5 and A5.3, for any $\eta > 0$, there is a $C > 0$ such that, for large T, $P(Tv_T^2|\hat{\lambda}_k - \lambda_k^0| > C) < \eta$, for any $k = 1, \ldots, m$.*

Proof of Theorem 5.3.

We will proceed in a similar fashion to the proof of Theorem 3.2 and preserve the notation in its proof.

Step 1.

As in Bai and Perron (1998), without loss of generality, we assume only three breaks. We will focus on proving Theorem 3.2 for $\hat{\lambda}_2$; the analysis for $\hat{\lambda}_1$ and $\hat{\lambda}_3$ is similar. For any $\epsilon > 0$, define $V_\epsilon = \{(T_1, T_2, T_3) :| T_i - T_i^0 |\leq \epsilon T \ (i = 1, 2, 3)\}$. Since $\hat{\lambda}_i \xrightarrow{p} \lambda_i^0$, $\lim P\{| \hat{T}_i - T_i^0 |\leq \epsilon T\} = 1$, hence $\lim P\{(\hat{T}_1, \hat{T}_2, \hat{T}_3) \in V_\epsilon\} = 1$. It follows that we need to examine only the behavior of those breakpoints contained in V_ϵ. Consider, without loss of generality, the case $\hat{T}_2 < T_2^0$. For $C > 0$, define:

$$V_\epsilon(C) = \left\{(T_1, T_2, T_3) :| T_i - T_i^0 |\leq \epsilon T \ (i = 1, 2, 3); T_2^0 - T_2 > \frac{C}{v_T^2}\right\}.$$

Note that $V_\epsilon(C) \subset V_\epsilon$, as in Theorem 3.2. The only difference is that $V_\epsilon(C)$ is defined to reflect the new rate of convergence of the break-fraction estimates under shrinking shifts. We will show, as before, that the probability that the break-points are contained in $V_\epsilon(C)$ is very small.

To that end, denote - as before - by $S_T(T_1, T_2, T_3)$ the minimized sum of squared residuals for a given 3-break-partition of the sample interval. By definition of minimized sum of squared residuals, we have: $S_T(\hat{T}_1, \hat{T}_2, \hat{T}_3) \leq S_T(\hat{T}_1, T_2^0, \hat{T}_3)$. As before, we will show that for any $\eta > 0$, we can pick ϵ and C such that:

$$P\left\{\min_{V_\epsilon(C)} \frac{S_T(T_1, T_2, T_3) - S_T(T_1, T_2^0, T_3)}{T_2 - T_2^0} \leq 0\right\} < \eta, \text{ for } T \geq T(\eta).$$

Preserving the notation in Theorem 3.2, we get:

$$P\left\{\min_{V_\epsilon(C)} \frac{SSR_1 - SSR_2}{\Delta_2} \leq 0\right\} < \eta, \text{ for } T \geq T(\eta), \tag{5.36}$$

where $\Delta_2 = T_2^0 - T_2 > C/v_T^2$. As before, (5.36) suffices for the proof of Theorem 5.3. Recall that:

$$SSR_1 - SSR_2 = (SSR_1 - SSR_3) - (SSR_2 - SSR_3)$$

Step 2:

As before, consider $SSR_1 - SSR_3$:

$$\frac{SSR_1 - SSR_3}{T_2^0 - T_2}$$
$$= \frac{1}{\Delta_2}\sum_{I_2^\Delta}[u_t^2(\theta_3^{**}) - u_t^2(\theta_2^\delta)] + \frac{1}{\Delta_2}\sum_{I_3}[u_t^2(\theta_3^{**}) - u_t^2(\theta_3^*)] = D_1 + D_2 \tag{5.37}$$

Intuitively, D_1 involves a "mismatch" in estimators, because θ_3^{**} is estimating θ_3^0, while θ_2^δ is estimating θ_2^0. This "mismatch" is not present in D_2, because θ_3^{**} and θ_3^* are both estimating θ_3^0. Hence, D_1 should be dominating D_2 for a large enough $\Delta_2 > C/v_T^2$. We will prove this below.

Start with D_1. In interval I_2^Δ, the true parameter value is θ_2^0. We can write:

$$\begin{aligned} D_1 &= \frac{1}{\Delta_2}\sum_{I_2^\Delta}[u_t^2(\theta_3^{**}) - u_t^2(\theta_2^\delta)] \\ &= \frac{1}{\Delta_2}\sum_{I_2^\Delta} d_t^2(\theta_3^{**}, \theta_2^0) - \frac{1}{\Delta_2}\sum_{I_2^\Delta} d_t^2(\theta_2^\delta, \theta_2^0) + \frac{2}{\Delta_2}\sum_{I_2^\Delta} u_t\, d_t(\theta_2^\delta, \theta_3^{**}) \\ &= D_{1,1} - D_{1,2} + D_{1,3}. \end{aligned}$$

We will find the order of each of the terms above. By a MVT of each $f_t(\theta_3^{**})$ around $f_t(\theta_2^0)$, we have:

$$D_{1,1} = [\theta_3^{**} - \theta_2^0]' \frac{1}{\Delta_2} \sum_{I_2^\Delta} F_t(\tilde{\theta}_{t,3,2}^{**,0}) F_t(\tilde{\theta}_{t,3,2}^{**,0})' [\theta_3^{**} - \theta_2^0]. \tag{5.38}$$

Since $F_t(\theta)F_t(\theta)'$ is bounded uniformly in θ, $\frac{1}{\Delta_2} \sum_{I_2^\Delta} F_t(\tilde{\theta}_{t,3,2}^{**,0}) F_t(\tilde{\theta}_{t,3,2}^{**,0})' = O_p^+(1)$, independent of C and ϵ chosen.

Consider the properties of θ_3^{**}, an estimator of θ_3^0. It was obtained by minimizing the sum of squared residuals on the interval $I_3 = [T_2^0, T_3]$. This differs from the true interval $I_3^0 \equiv [T_2^0, T_3^0]$ by ϵT terms. If we estimated θ_3^0 on I_3^0, we would obtain the usual NLS estimator, within $O_p(T^{-1/2})$ of the true value θ_3^0. As long as, for large T, we pick ϵ small enough, the misspecification of the estimation interval will be small, hence so will be the deviation of θ_3^{**} from the properties of the usual NLS estimator. Hence, we can pick ϵ such that

$$\theta_3^{**} - \theta_3^0 \sim O_p(T^{-1/2}).$$

Then, with high probability $\theta_3^{**} - \theta_2^0 = (\theta_3^{**} - \theta_3^0) + (\theta_3^0 - \theta_2^0) \sim O_p(T^{-1/2}) + O_p(v_T)$. According to A5.3, $T^{1/2-\nu} v_T^2 \to \infty$, so $O(v_T) > O(T^{-1/4}) > O(T^{-1/2})$. Hence $O_p(v_T)$ dominates $O_p(T^{-1/2})$, hence $\theta_3^{**} - \theta_2^0 \sim O_p(v_T)$. Equation (5.38) becomes:

$$D_{1,1} \sim O_p(v_T) O_p^+(1) O_p(v_T) \sim O_p^+(v_T^2). \tag{5.39}$$

By a MVT of each $f_t(\theta_2^\delta)$ around $f_t(\theta_2^0)$, we have:

$$D_{1,2} = [\theta_2^\delta - \theta_2^0]' \frac{1}{\Delta_2} \sum_{I_2^\Delta} F_t(\tilde{\theta}_{t,2,2}^{\delta,0}) F_t(\tilde{\theta}_{t,2,2}^{\delta,0})' [\theta_2^\delta - \theta_2^0]. \tag{5.40}$$

By a similar reasoning as for the middle term on the RHS of (5.38), $\frac{1}{\Delta_2} \sum_{I_2^\Delta} F_t(\tilde{\theta}_{t,2,2}^{\delta,0}) F_t(\tilde{\theta}_{t,2,2}^{\delta,0})' \sim O_p^+(1)$.

The properties of θ_2^δ, an estimator of θ_2^0, depend on the length of the sample interval,

$I_2^\Delta = [T_2, T_2^0]$. Since its length is $C/v_T^2 <$ length $I_2^\Delta < \epsilon T$, for any large T, we can choose ϵ such that $I_2^\Delta \subset [T_1^0, T_2^0]$. In this case, θ_2^δ will be a sub-sample NLS estimator of θ_2^0. If we pick C big enough, the sub-sample is large enough to ensure that:

$$\theta_2^\delta - \theta_2^0 \sim O_p(T^{-1/2}).$$

According to the above, for large C and small ϵ,

$$D_{1,2} \sim O_p(T^{-1/2})O_p(1)O_p(T^{-1/2}) \sim O_p(T^{-1}). \tag{5.41}$$

Now

$$D_{1,3} = [\theta_3^{**} - \theta_2^\delta]' \frac{2}{\Delta_2} \sum_{I_2^\Delta} u_t F_t(\tilde{\theta}_{t,2,3}^{\delta,**}) = O_p(v_T) \frac{2}{\Delta_2} \sum_{I_2^\Delta} u_t F_t(\tilde{\theta}_{t,2,3}^{\delta,**}). \tag{5.42}$$

We can determine the order of $D_{1,3}$ by means of the extended Hajek-Renyi inequality stated in Lemma 5.3. In that inequality, consider $t = T_2 + 1$ as the first observation $t = 1$, and T_2^0 as the last $t = k$. Then the length of the summing interval is $k > C/v_T^2$, hence $m = C/v_T^2$. With $b = v_T$, the inequality becomes:

$$P\left[\sup_{T_2^0 - T_2 > C/v_T^2} \frac{2}{\Delta_2} \left\| \sum_{I_2^\Delta} u_t F_t(\tilde{\theta}_{t,2,3}^{\delta,**}) \right\| \geq v_T \right]$$
$$\leq P\left[\sup_{T_2^0 - T_2 > C/v_T^2} \frac{2}{\Delta_2} \left\| \sum_{I_2^\Delta} \sup_\theta u_t F_t(\theta) \right\| \geq v_T \right] \leq \frac{L v_T^2}{T v_T^2 C} = \frac{L}{C} \tag{5.43}$$

In the equation above, by taking supremum over a vector valued function, we mean taking the supremum over each element of it. The last inequality above holds because the bound does not depend on θ. Then a big enough C ensures that $\frac{2}{\Delta_2} \sum_{I_2^\Delta} u_t F_t(\tilde{\theta}_{t,2,3}^{\delta,**}) = O_p(v_T)$. So, with high probability, from (5.42) and (5.43),

$$D_{1,3} = O_p(v_T^2). \tag{5.44}$$

Combining equations (5.39) with (5.41) and (5.44), with large probability for large T, we can pick C large and ϵ small such that:

$$D_1 = D_{1,1} - D_{1,2} + D_{1,3}$$
$$\sim O_p(v_T^2) - O_p(T^{-1}) + O_p(v_T^2) = O_p^+(v_T^2) + O_p(v_T^2) \quad (5.45)$$

Note that the second $O_p(v_T^2)$ term has uncertain sign because the sign of $D_{1,3}$ is unclear. We will now find the order of D_2. As before, we need to consider two cases.

Case 1. $T_3 \leq T_3^0$. In this case, the true parameter value on $I_3 = [T_2^0, T_3] \subset [T_2^0, T_3^0]$ is θ_3^0, hence:

$$\begin{aligned} D_2 &= \frac{1}{\Delta_2} \sum_{I_3} [u_t^2(\theta_3^{**}) - u_t^2(\theta_3^*)] \\ &= \frac{1}{\Delta_2} \sum_{I_3} d_t^2(\theta_3^{**}, \theta_3^0) - \frac{1}{\Delta_2} \sum_{I_3} d_t^2(\theta_3^*, \theta_3^0) + \frac{2}{\Delta_2} \sum_{I_3} u_t \, d_t(\theta_3^*, \theta_3^{**}) \\ &= D_{2,1} - D_{2,2} + D_{2,3} \end{aligned}$$

By a MVT expansion of $f_t(\theta_3^{**})$ around $f_t(\theta_3^0)$,

$$D_{2,1} = [\theta_3^{**} - \theta_3^0]' \frac{1}{\Delta_2} \sum_{I_3} F_t(\tilde{\theta}_{t,3,3}^{**,0}) F_t(\tilde{\theta}_{t,3,3}^{**,0})' [\theta_3^{**} - \theta_3^0]$$
$$= O_p(T^{-1/2}) \frac{1}{\Delta_2} \sum_{I_3} F_t(\tilde{\theta}_{t,3,3}^{**,0}) F_t(\tilde{\theta}_{t,3,3}^{**,0})' O_p(T^{-1/2})$$

Here, $\sum_{I_3} F_t(\tilde{\theta}_{t,3,3}^{**,0}) F_t(\tilde{\theta}_{t,3,3}^{**,0})'$ is close to $O_p(T)$, because the sum involves an infinite number of terms that are close to $O_p(1)$ for large T. Since $\Delta_2 > C/v_T^2$, $\frac{1}{\Delta_2} \sum_{I_3} F_t(\tilde{\theta}_{t,3,3}^{**,0}) F_t(\tilde{\theta}_{t,3,3}^{**,0})' < \frac{v_T^2}{C} O_p(T) = \frac{1}{C} O_p^+(T v_T^2)$. Hence, we have:

$$D_{2,1} \sim O_p(T^{-1/2})' \frac{1}{C} O_p^+(T v_T^2) O_p(T^{-1/2}) = \frac{1}{C} O_p^+(v_T^2) \quad (5.46)$$

Consider now $D_{2,2}$. As long as ϵ is small, the estimator θ_3^* is within $O_p(T^{-1/2})$ of θ_3^0, since it is a sub-sample NLS estimator.

$$D_{2,2} = [\theta_3^* - \theta_3^0]' \frac{1}{\Delta_2} \sum_{I_3} F_t(\tilde{\theta}_{t,3,3}^{*,0}) F_t(\tilde{\theta}_{t,3,3}^{*,0})' [\theta_3^* - \theta_3^0]$$
$$\sim O_p(T^{-1/2}) \frac{1}{\Delta_2} \sum_{I_3} F_t(\tilde{\theta}_{t,3,3}^{*,0}) F_t(\tilde{\theta}_{t,3,3}^{*,0})' O_p(T^{-1/2})$$

We know $\frac{1}{\Delta_2} \sum_{I_3} F_t(\tilde{\theta}_{t,3,3}^{*,0}) F_t(\tilde{\theta}_{t,3,3}^{*,0})' \leq \frac{v_T^2}{C} O_p(T) \sim \frac{1}{C} O_p^+(T v_T^2)$. Then:

$$D_{2,2} \leq O_p(T^{-1/2}) \frac{1}{C} O_p(T v_T^2) O_p(T^{-1/2}) = \frac{1}{C} O_p^+(v_T^2) \tag{5.47}$$

Also, by Proposition 3.1, $\sum_{t=1}^{[Tr]} \Psi_t(\theta) = O_p(T^{1/2})$ uniformly in $\theta \times r$. Then:

$$D_{2,3} = [\theta_3^* - \theta_3^{**}] \frac{2}{\Delta_2} \sum_{I_3} u_t F_t(\tilde{\theta}_{t,3,3}^{*,**})$$
$$\leq \frac{2}{\Delta_2} O_p(T^{-1/2}) \frac{v_T^2}{C} O_p(T^{1/2}) \sim \frac{1}{C} O_p(v_T^2) \tag{5.48}$$

Combining (5.46)-(5.48), we get:

$$D_2 = D_{2,1} - D_{2,2} + D_{2,3} \geq D_{2,1} + D_{2,3} \sim \frac{1}{C} O_p(v_T^2) \tag{5.49}$$

Case 2 $T_3 > T_3^0$. Let $I_3 = [T_2^0, T_3^0] \cup [T_3^0 + 1, T_3] = I_3^0 + I_3^\epsilon$. Then D_2 can be written as:

$$D_2 = \frac{1}{\Delta_2} \sum_{I_3^0} [u_t^2(\theta_3^{**}) - u_t^2(\theta_3^*)] + \frac{1}{\Delta_2} \sum_{I_3^\epsilon} [u_t^2(\theta_3^{**}) - u_t^2(\theta_3^*)] =\sim \frac{1}{C} v_T^2 + D_2^*,$$

by arguments similar as for D_2 in case 1. It remains to analyze $D_{2,2}^*$. On the interval I_3^ϵ, the true value of the parameter is θ_4^0. For each $t \in I_3^\epsilon$,

$$D_2^* = \frac{1}{\Delta_2}\sum_{I_3^\epsilon} d_t^2(\theta_3^{**}, \theta_4^0) - \frac{1}{\Delta_2}\sum_{I_3^\epsilon} d_t^2(\theta_3^{**}, \theta_4^0) + \frac{2}{\Delta_2}\sum_{I_3^\epsilon} u_t\, d_t(\theta_3^*, \theta_3^{**})$$
$$= D_{2,1}^* - D_{2,2}^* + D_{2,3}^*.$$

For small ϵ and large C,

$$D_{2,1}^* = [\theta_3^* - \theta_4^0]' \frac{1}{\Delta_2}\sum_{I_3^\epsilon} F_t(\tilde{\theta}_{t,3,4}^{*,0}) F_t(\tilde{\theta}_{t,3,4}^{*,0})' [\theta_3^* - \theta_4^0]' \tag{5.50}$$
$$\sim O_p(v_T)\frac{1}{\Delta_2} O_p(T\epsilon) O_p(v_T) \leq O_p(v_T^2)\frac{v_T^2}{C} O_p(T\epsilon) O_p(v_T) \sim \frac{\epsilon}{C} O_p(T v_T^4)$$

Similarly,

$$D_{2,2}^* \sim \frac{\epsilon}{C} O_p(T v_T^4) \tag{5.51}$$

$$D_{2,3}^* = [\theta_3^* - \theta_3^{**}]' \frac{1}{\Delta_2}\sum_{I_3^\epsilon} u_t F_t(\theta_{t,I_3^\epsilon})$$
$$= O_p(T^{-1/2})\frac{1}{\Delta_2} O_p([T\epsilon]^{1/2}) \sim \frac{v_T^2}{C} O_p(1) = \frac{\epsilon^{1/2}}{C} O_p(v_T^2) \tag{5.52}$$

Then, in case 2, from (5.50)-(5.52), we get:

$$D_{2,2}^* = D_{2,1}^* - D_{2,2}^* + D_{2,3}^* \sim \frac{\epsilon}{C} O_p(T v_T^4) - \frac{\epsilon}{C} O_p(T v_T^4) + \frac{\epsilon^{1/2}}{C} O_p(v_T^2) = \frac{\epsilon}{C} O_p(T v_T^4) \tag{5.53}$$

From (**??**) and (5.53)

$$D_2 = \sim \frac{1}{C} O_p(v_T^2) + \frac{\epsilon}{C} O_p(T v_T^4) = \frac{1}{C} O_p(v_T^2),$$

if we choose ϵ small enough, such that the magnitude of $\frac{\epsilon}{C} O_p(T v_T^4)$ can be made small compared to $\frac{1}{C}O_p(v_T^2)$. Hence, in case 2,

$$D_2 \sim \frac{1}{C}O_p(v_T^2) \tag{5.54}$$

Hence, in both cases, according to (5.49), respectively (5.54), we can write:

$$D_2 \sim \frac{1}{C}O_p(v_T^2) \tag{5.55}$$

Considering (5.45) and (5.55),

$$\begin{aligned}\frac{SSR_1 - SSR_3}{\Delta_2} &\geq D_1 + D_2 \geq D_{1,1} + D_{1,3} + D_2 \\ &\sim O_p^+(v_T^2) + O_p(v_T^2) + \sim \frac{1}{C}O_p(v_T^2) \sim O_p^+(v_T^2) + O_p(v_T^2)\end{aligned} \tag{5.56}$$

if we pick C large enough and ϵ small enough.

Step 3.

We will analyze $SSR_2 - SSR_3$ in a similar fashion to $SSR_1 - SSR_3$.

$$\begin{aligned}\frac{SSR_2 - SSR_3}{T_2^0 - T_2} &= \frac{1}{\Delta_2}\sum_{I_2}[u_t^2(\theta_2^{**}) - u_t^2(\theta_2^{\delta})] + \frac{1}{\Delta_2}\sum_{I_2^{\Delta}}[u_t^2(\theta_2^{**}) - u_t^2(\theta_2^{*})] \\ &= D_3 + D_4.\end{aligned} \tag{5.57}$$

Intuitively, the terms above should be smaller than $SSR_1 - SSR_2$, as in the proof of Theorem 3.2. We will show this below.

As long as ϵ is small, $I_2 = [T_1 + 1, T_2]$ will not differ much from $I_2^0 = [T_1^0 + 1, T_2^0]$. Then θ_2^* will be within $O_p(T^{-1/2})$ of the true value θ_2^0 with high probability for large T and small ϵ. Since θ_2^{**} is obtained from sample $[T_1, T_2^0]$, similar properties apply. Start with D_4 because it is easier to analyze:

$$\begin{aligned} D_4 &= \frac{1}{\Delta_2} \sum_{I_2^\Delta} [u_t^2(\theta_2^*) - u_t^2(\theta_2^\delta)] \\ &= \frac{1}{\Delta_2} \sum_{I_2^\Delta} d_t^2(\theta_2^*, \theta_2^0) - \frac{1}{\Delta_2} \sum_{I_2^\Delta} d_t^2(\theta_2^\delta, \theta_2^0) + \frac{2}{\Delta_2} \sum_{I_2^\Delta} u_t \, d_t(\theta_2^\delta, \theta_2^*) \\ &= D_{4,1} - D_{4,2} + D_{4,3}. \end{aligned}$$

Then, by MVT,

$$D_{4,1} = [\theta_2^* - \theta_2^0]' \frac{1}{\Delta_2} \sum_{I_2^\Delta} F_t(\tilde{\theta}_{t,2,2}^{*,0}) F_t(\tilde{\theta}_{t,2,2}^{*,0})' [\theta_2^* - \theta_2^0] \sim O_p(T^{-1/2}) O_p(1) O_p(T^{-1/2}) = O_p(T^{-1}).$$

Similarly, $D_{4,2} \sim O_p(T^{-1})$.

$$D_{4,3} = [\theta_2^* - \theta_2^\delta]' \frac{2}{\Delta_2} \sum_{I_2^\Delta} u_t F_t(\tilde{\theta}_{t,2,2}^{*,\delta}) \sim O_p(T^{-1/2}) O_p(v_T) = O_p(T^{-1/2} v_T),$$

where the latter follows by Lemma 5.3, applied in the same way as for $D_{1,3}$. Then

$$D_4 = D_{4,1} - D_{4,2} + D_{4,3} \sim O_p(T^{-1}) + O_p(T^{-1/2} v_T) \sim O_p(T^{-1/2} v_T) \tag{5.58}$$

The order of D_3 can be found by similar arguments as for D_2. If ϵ is small enough for large T,

$$D_3 \sim \frac{1}{C} O_p(v_T^2) \tag{5.59}$$

Then, for large C and small ϵ,

$$\frac{SSR_2 - SSR_3}{T_2^0 - T_2} = D_3 + D_4 \sim \frac{1}{C} O_p(v_T^2) + O_p(T^{-1/2} v_T) = \frac{1}{C} O_p(v_T^2) \tag{5.60}$$

Combine (5.56) and (5.60):

$$\frac{SSR_1 - SSR_3}{T_2^0 - T_2} - \frac{SSR_2 - SSR_3}{T_2^0 - T_2} \geq O_p^+(v_T^2) + O_p(v_T^2) + \frac{1}{C} O_p(v_T^2)$$
$$= O_p^+(v_T^2) + O_p(v_T^2) = av_T^2 + bv_T^2 \quad (5.61)$$

for large C with large probability. Here $a > 0$ comes from $D_{1,1}$, while b pertains to $D_{1,3} = \frac{2}{\Delta_2} \sum_{I_2^\Delta} u_t \, d_t(\theta_2^\delta, \theta_3^{**}) = 2[\theta_2^\delta - \theta_3^{**}]' \frac{1}{\Delta_2} \sum_{I_2^\Delta} u_t F_t(\tilde{\theta}_{t,2,3}^{\delta,**}) = O_p(v_T) \frac{1}{\Delta_2} \sum_{I_2^\Delta} u_t F_t(\tilde{\theta}_{t,2,3}^{\delta,**}) = cv_T \frac{1}{\Delta_2} \sum_{I_2^\Delta} u_t F_t(\tilde{\theta}_{t,2,3}^{\delta,**})$. Here c is a constant, not necessarily positive. We can write:

$$\frac{SSR_1 - SSR_2}{T_2^0 - T_2} \geq av_T^2 + cv_T \sum_{I_2^\Delta} \frac{2}{\Delta_2} u_t F_t(\tilde{\theta}_{t,2,3}^{\delta,**}).$$

Now, for any $\eta > 0$,

$$\begin{aligned}
&P\left[\min_{V_\epsilon(C)} \frac{SSR_1 - SSR_2}{T_2^0 - T_2} \leq 0\right] \\
&\leq P\left[av_T^2 + \min_{V_\epsilon(C)} cv_T \frac{1}{\Delta_2} \sum_{I_2^\Delta} u_t F_t(\tilde{\theta}_{t,2,3}^{\delta,**}) \leq 0\right] \\
&\leq P\left[av_T^2 \leq 0\right] + P\left[\max_{V_\epsilon(C)} cv_T \frac{1}{\Delta_2} \sum_{I_2^\Delta} u_t F_t(\tilde{\theta}_{t,2,3}^{\delta,**}) \geq av_T^2\right] \\
&\leq \frac{\eta}{2} + P\left[\max_{V_\epsilon(C)} \frac{1}{\Delta_2} \left\| \sum_{I_2^\Delta} u_t F_t(\tilde{\theta}_{t,2,3}^{\delta,**}) \right\| \geq \frac{av_T}{|c|}\right] \\
&\leq \frac{\eta}{2} + \frac{c^2 L v_T^2}{C a^2 v_T^2} = \frac{\eta}{2} + \frac{Lc^2/a^2}{C},
\end{aligned}$$

where the latter follows from Lemma 5.3 applied to $\max \left\| \sum_{I_2^\Delta} \frac{1}{\Delta_2} u_t F_t(\theta) \right\|$. We can always pick C large enough such that $\frac{Lb^2/a^2}{C} < \frac{\eta}{2}$. Then, any $\eta > 0$, for large T, we can always pick C large enough and ϵ small enough such that:

$$P\left[\min_{V_\epsilon(C)} \frac{SSR_1 - SSR_2}{T_2^0 - T_2} \leq 0\right] < \eta, \text{ which completes the proof of Theorem 5.3.}$$

□

5.2.3 Asymptotic Distributions of Parameter and Break-Fraction Estimates

According to Theorem 3.2, the differences between the estimated and true sub-samples are of maximum length $C/\,v_T^2$. In this case, the asymptotic distributions of the parameter estimates are the same as for fixed shifts:

Theorem 5.4. *Under A3.1-A3.5 and A5.3,* $T^{1/2}(\hat{\theta} - \theta^0) \xrightarrow{d} \mathcal{N}(0, \Phi)$,
where $\Phi = diag\,[\Phi_1, \ldots, \Phi_{m+1}]$*, and* $\Phi_i = \bar{D}^{-i}\bar{A}_{i,i}\bar{D}_i^{-1}$.

Proof of Theorem 5.4:

The Proof of Theorem 5.4 is similar to that of Theorem 4.2. The key argument in the latter was that terms of the form $\left[\sum_i - \sum_{\hat{i}}\right] T^{-1}F_t(\theta)F_t(\theta)$, $i = 1, \ldots, m+1$. and $\left[\sum_i - \sum_{\hat{i}}\right] T^{-1/2}u_t F_t(\theta)$ $i = 1, \ldots, m+1$ are $o_p(1)$ uniformly in θ. These terms were shown to disappear in the limit because they involved, according to Proposition 3.2, a finite number of terms that were $o_p(1)$ uniformly in θ.

In the case of shrinking shifts, by Theorem 3.2, $\left[\sum_i - \sum_{\hat{i}}\right] T^{-1}F_t(\theta)F_t(\theta)$ involves at most $C/\,v_T^2$ terms that are $O_p(T^{-1})$. Hence, the order of the aforementioned term is $O_p\left(\frac{1}{Tv_T^2}\right) = o_p(1)$, by A5.3. Similarly, we have that $\left[\sum_i - \sum_{\hat{i}}\right] T^{-1/2}u_t F_t(\theta) = O_p\left(\frac{1}{T^{1/2}v_T^2}\right) = o_p(1)$. Since the orders match the ones in Theorem 4.2, the proof of Theorem 5.4 is exactly the same. □

Knowing the distribution of the parameter estimates allows us to derive the distribution of the break-fraction estimates. We will start, as in Section 5.1, with a single break-model. Henceforth preserving the notation of that section, we obtain:

Theorem 5.5. *Under Assumptions A3.1-A3.5, A5.2 and A5.3, for* $m = 1$,

$$v_T^2[\hat{k} - k_0] \xrightarrow{d} \underset{m}{\operatorname{argmax}}\, Z(m) \equiv \underset{m}{\operatorname{argmax}}[\, W(m) - 0.5|m|\,]$$

where $W(m) = W_1(-m), m \leq 0$, $W(m) = W_2(m), m > 0$ *and* $W_1(m), W_2(m)$ *are two independent standard scalar Gaussian processes defined on* $[0, \infty]$.

The distribution described above is similar to the one in Corollary 2 to Theorem 5.1, except for the convergence rate. Unlike in linear models, weak stationarity of the regressors does not suffice. This is because, for proving consistency, we rely on Assumption A3.1, which imposes strict stationarity.

The process $\text{argmax}_m Z(m)$ is symmetric about the origin. The properties of its density function can be found in Picard (1985) and Yao (1987). This allows us to obtain confidence intervals for the break-fractions.

Proof of Theorem 5.5:

Recall that:

$$\hat{k} = \underset{k}{\text{argmin}}\, V_T(k, \hat{\theta}_1(k), \hat{\theta}_2(k)).$$

Step 1. The key to deriving the asymptotic distribution of $V_T(k, \hat{\theta}_1(k), \hat{\theta}_2(k))$ lies in knowing the convergence rates: $\hat{k} = k_0 \pm C/v_T^2$ and $\hat{\theta}_i(k) = \theta_i^0 + O_p(T^{-1/2})$. Then we define the minimization problem over a neighborhood of (k, θ_1, θ_2).

Step 2. Given Step 1, we split the objective function into three parts. Assume $k < k_0$; the proof for $k \geq k_0$ is similar.

$$V_T(k, \hat{\theta}_1(k), \hat{\theta}_2(k)) = \sum_{t=1}^{k} [u_t^2(\hat{\theta}_1) - u_t^2(\theta_1^0)] + \sum_{t=k+1}^{k_0} [u_t^2(\hat{\theta}_2) - u_t^2(\theta_1^0)] \tag{5.62}$$

$$+ \sum_{t=k_0+1}^{T} [u_t^2(\hat{\theta}_2) - u_t^2(\theta_2^0)] = \Sigma_1 + \Sigma_2 + \Sigma_3. \tag{5.63}$$

Step 3. Given that $k \in V_\epsilon \setminus V_\epsilon(C)$, one can show that Σ_1 and Σ_3 have asymptotic distributions that are independent of k. We are left with Σ_2, representing magnitude of the mismatch between the objective function evaluated at $[k, \hat{\theta_1}(k), \hat{\theta}_2(k)]$. This mismatch is essential to the distribution of the objective function in a neighborhood of the true values, hence Σ_2 governs the distribution of $\hat{k}$. The complete proof follows.

Case 1 : $k < k_0$. Let $v = k_0 - k$, $0 < v \leq C/v_T^2$. The latter is a restatement of Theorem 5.3. We can show by similar arguments as for Theorem 3.2 that $\Sigma_1 + \Sigma_3 = \mathfrak{D} + o_p(1)$, a distribution independent of v. The key arguments go through the same way as for the distribution of parameter estimates. Hence,

$$V_T(k, \hat{\theta}_1(k), \hat{\theta}_2(k)) = \mathfrak{D} + o_p(1) + \Sigma_2, \tag{5.64}$$

where the $o_p(1)$ term is uniform in v. Turn to Σ_2, the term that will govern the distribution of the minimand for shrinking shifts.

$$\begin{aligned} \Sigma_2 = & \sum_{t=k+1}^{k_0} d_t^2(\hat{\theta}_2(k), \theta_2^0) + \sum_{t=k+1}^{k_0} d_t^2(\theta_2^0, \theta_1^0) + \\ & + 2 \sum_{t=k+1}^{k_0} d_t(\hat{\theta}_2(k), \theta_2^0)\, d_t(\theta_2^0, \theta_1^0) - 2 \sum_{t=k+1}^{k_0} u_t\, d_t(\hat{\theta}_2(k), \theta_2^0) - 2 \sum_{t=k+1}^{k_0} u_t\, d_t(\theta_2^0, \theta_1^0) \\ & = \Sigma_{2,1} + \Sigma_{2,2} + \Sigma_{2,3} + \Sigma_{2,4} + \Sigma_{2,5}. \end{aligned} \tag{5.65}$$

Consider $\Sigma_{2,1}$.

$$\begin{aligned} \Sigma_{2,1} &= \sum_{t=k+1}^{k_0} d_t^2(\hat{\theta}_2(k), \theta_2^0) = \tilde{\theta}_2(v)' \left[\sum_{t=k+1}^{k_0} T^{-1} F_t(\bar{\theta}_{t,2,2}) F_t(\bar{\theta}_{t,2,2})' \right] \tilde{\theta}_2(v) \\ &= O_p(1) O_p\left(\frac{1}{T v_T^2}\right) O_p(1) = o_p(1), \end{aligned}$$

uniformly in v. Now turn to $\Sigma_{2,2}$:

$$\begin{aligned} \Sigma_{2,2} &= \sum_{t=k+1}^{k_0} d_t^2(\theta_2^0, \theta_1^0) = \sum_{t=k_0-v+1}^{k_0} d_t^2(\theta_2^0, \theta_1^0) \\ &= \sum_{t=n+1}^{0} d_t^2(\theta_2^0, \theta_1^0) = (\theta_2^0 - \theta_1^0)' \sum_{t=n+1}^{0} F_t(\bar{\theta}_{t,1,2}^{0,0}) F_t(\bar{\theta}_{t,1,2}^{0,0})\, (\theta_2^0 - \theta_1^0), \text{ where } n < 0. \end{aligned}$$

Now $\theta_2^0 - \theta_1^0 = \delta_{1,T} = o(1)$. Also, $\sum_{t=n+1}^{0} F_t(\theta)F_t(\theta) - |n|D(\theta) \xrightarrow{p} 0$, for any fixed θ, by A3.4(iii). Hence,

$$\Sigma_{2,2} = |n|\delta_{1,T}' D(\theta_1^0)\delta_{1,T} + o_p(1)$$

Consider $\Sigma_{2,3}$.

$$\begin{aligned}\Sigma_{2,3} &= 2\sum_{t=k+1}^{k_0} d_t(\hat{\theta}_2(k),\theta_2^0)\, d_t(\theta_2^0,\theta_1^0) = 2\tilde{\theta}_2(v)'\left[\sum_{t=k+1}^{k_0} T^{-1/2}F_t(\bar{\theta}_{t,2,2})F_t(\tilde{\theta}_{t,2,1}^{0,0})'\right]\delta_{1,T}\\ &= O_p(1)O_p\left(\frac{1}{T^{1/2}v_T^2}\right)O(v_T) = o_p(1)\end{aligned}$$

We are left with $\Sigma_{2,4}$ and $\Sigma_{2,5}$:

$$\begin{aligned}\Sigma_{2,4} &= -2\sum_{t=k+1}^{k_0} u_t\, d_t(\hat{\theta}_2(k),\theta_2^0) = -2\tilde{\theta}_2(v)\left[T^{-1/2}\sum_{t=k+1}^{k_0} u_t F_t(\bar{\theta}_{t,2,2})\right]\\ &= O_p(1)o_p\left(\frac{1}{T^{1/2}v_T^2}\right) = o_p(1)\end{aligned}$$

Also,

$$\Sigma_{2,5} = -2\sum_{t=k_0-v+1}^{k_0} u_t\, d_t(\theta_2^0,\theta_1^0)$$

We have already shown that

$$\Sigma_{2,5} \xrightarrow{d} (a*)^{1/2}W_1(-n),$$

where $a*$ was defined in Corolarry 2 of Theorem 5.1. By MVT and the fact that $\theta_2^0 = \theta_1^0 + \delta_{1,T}v_T$, we get:

$$a* = \lim E[d_t^2(\theta_2^0,\theta_1^0)] = \lim[\theta_2^0 - \theta_1^0]'E[u_t^2 F_t(\bar{\theta}_{t,1,2}^{0,0})F_t(\bar{\theta}_{t,1,2}^{0,0})'][\theta_2^0 - \theta_1^0] = \delta_{i,T}' A(\theta_1^0)\delta_{i,T}. \quad (5.66)$$

Since $\Sigma_2 = \Sigma_{2,1} + \Sigma_{2,2} + \Sigma_{2,3} + \Sigma_{2,4} + \Sigma_{2,5}$, we have, uniformly in v:

$$\Sigma_2 = |n|\delta'_{1,T}D(\theta_1^0)\delta_{1,T} - 2[\delta'_{1,T}A(\theta_1^0)\delta_{1,T}]^{1/2}W_1(-n) + o_p(1), \text{ for } n \leq 0. \tag{5.67}$$

Equations (5.64) and (5.67) imply that the minimand is, uniformly in $0 \leq k_0 - k = v \leq C/v_T^2$,

$$V(k, \hat{\theta}_1, \hat{\theta}_2) = \mathfrak{D} + |s|\delta'_{1,T}D(\theta_1^0)\delta_{1,T} - 2[\delta'_{1,T}A(\theta_1^0)\delta_{1,T}]^{1/2}W_1(-n) + o_p(1). \tag{5.68}$$

Since $C/v_T^2 \to \infty$,

$$\begin{aligned}\hat{k} - k_0 &= \underset{-C/v_T^2 \leq v < 0}{\operatorname{argmin}} V(k, \hat{\theta}_1, \hat{\theta}_2) \qquad (5.69)\\ &= \underset{n \leq 0}{\operatorname{argmax}} \left[[\delta'_{1,T}A(\theta_1^0)\delta_{1,T}]^{1/2}W_1(-n) - 0.5|n|\delta'_{1,T}D(\theta_1^0)\delta_{1,T} + o_p(1) + \mathfrak{D} \right].\end{aligned}$$

As for fixed shifts, $\mathfrak{D}$ is independent of n, by uniform convergence in $v = -n$. Hence,

$$\hat{k} - k_0 = \underset{n \leq 0}{\operatorname{argmax}} \left[[\delta'_{1,T}A(\theta_1^0)\delta_{1,T}]^{1/2}W_1(-n) - 0.5|n|\delta'_{1,T}D(\theta_1^0)\delta_{1,T} \right] + o_p(1) \tag{5.70}$$

Let $\varpi_2 = \delta'_1 D(\theta_1^0)\delta_1 v_T^2$ and $\varpi_1 = \delta'_{1,T}A(\theta_1^0)\delta_{1,T}v_T^2$. By a similar change in variable as before, we get:

$$\begin{aligned}\hat{k} - k_0 &= \underset{n<0}{\operatorname{argmax}} \left[\varpi_1^{1/2}W_1(-n) - 0.5|n|\varpi_2 \right] + o_p(1)\\ \hat{k} - k_0 &= \underset{n<0}{\operatorname{argmax}} \frac{\varpi_2}{\varpi_1^2}[W_1(n) - 0.5|n| + o_p(1)]\\ \frac{\varpi_1^2}{\varpi_2}v_T^2[\hat{k} - k_0] &\xrightarrow{d} \underset{n<0}{\operatorname{argmax}} Z_1(-n)\end{aligned}$$

Case 2: $k \geq k_0$. Similarly, it can be shown that:

$$\frac{\varpi_1^2}{\varpi_2}v_T^2[\hat{k} - k_0] \xrightarrow{d} \underset{n \geq 0}{\operatorname{argmax}} Z_2(n)$$

Hence, the proof of Theorem 5.5 is complete. □

Consider the more general multiple breaks model. The theorem above is readily generalizable:

Corrolary 1 to Theorem 5.5. *Under Assumptions A3.1-A3.5, A5.2 and A5.3,*

$$\frac{\delta_i' A(\theta_1^0)\delta_i}{[\delta_i' D(\theta_1^0)\delta_i]^2}[\hat{T}_i - T_i^0] \xrightarrow{d} \underset{n}{\operatorname{argmax}}\, Z(n) = [W^i(n) - 0.5|n|]$$

where $W^i(n) = W_1^i(-n), s \leq 0$, $W^i(s) = W_2^i(n), n > 0$ *and* $W_j^i(n)$ *are standard p-dimensional Gaussian processes defined on* $[0, \infty]$ *and independent for all* $j = 1, 2$ *and* $i = 1, \ldots, m$.

Proof of Corollary to Theorem 5.5:

Under Assumption A3.1, the sub-samples from the estimated partitions are asymptotically independent. Hence, the proof is the same as for one break. $D(\theta_1^0)$ appears in the distribution of each break-point estimate because, under Assumption A5.3, $\theta_1^0 - \theta_i^0 = o_p(1)$, for each $i = 2, \ldots, m$. □

Chapter 6

Tests for Stability

The analysis in previous chapters was based on the assumption that the number of change-points, m, is known. In practice however, m may be unknown or there may be no breaks. In this chapter, we reconsider the test statistics described in Bai and Perron (1998). They can be used to find the number of breaks when unknown. This could potentially be done by criteria such as Akaike's Information Criterion (AIC) or Bayesian Information Criterion (BIC).

Given that our estimation is based on LS methods, it is natural to consider LS principles for these tests. We propose similar tests as in Bai and Perron (1998). The distribution of these tests are non-standard and carry through to our nonlinear setting. The critical values are already tabulated in Bai and Perron (1998) and Bai and Perron (2003a).

In Section 1, we will present a test of one against a known number of breaks. Section 2 extends these tests to one versus an unknown number of breaks. Finally, Section 3 provides a test of ℓ versus $\ell + 1$ breaks, also suggesting a sequential procedure for estimating the change-points and parameters, rather than the global procedure presented in Chapter 3.

6.1 A Test of No Breaks versus a Fixed Number of Breaks

Consider the following hypothesis:

$$H_0 : m = 0 \qquad vs. \qquad H_A : m = k. \tag{6.1}$$

where k is a fixed finite positive integer. For this purpose, consider a partition $(T_1, \ldots, T_k)$ of the $[1, T]$ interval such that $T_i = [T\lambda_i]$. We also need to restrict each change point to be asymptotically distinct and bounded away from the end-points of the sample. To this end, define $\Lambda_\epsilon = \{(\lambda_1, \ldots, \lambda_k) : |\lambda_{i+1} - \lambda_i| \geq \epsilon, \lambda_1 \geq \epsilon, \lambda_k \leq 1 - \epsilon\}$, where ϵ is a small number, usually ranging from 0.05 to 0.15 in practice. As in Bai and Perron (1998), consider a generalized version of the sup F-type tests proposed in Andrews (1993):

$$\sup_{(\lambda_1,\ldots,\lambda_k)\in\Lambda_\epsilon} F_T(k;p) = \sup_{(\lambda_1,\ldots,\lambda_k)\in\Lambda_\epsilon} \frac{(SSR_0 - SSR_k)/kp}{SSR_k/[T-(k+1)p]} \tag{6.2}$$

where SSR_0 and SSR_k are the sums of squared residuals under the null, respectively under the alternative hypothesis. Let $B_p(\cdot)$ be a p-vector of independent Brownian motions. The following proposition describes the distribution of the test under H_0:

Theorem 6.1. *Under A3.2-A4.1 and* $H_0 : m = 0$*,*

$$\sup_{(\lambda_1,\ldots,\lambda_k)\in\Lambda_\epsilon} F_T(k;p) \Rightarrow \frac{1}{kp} \sup_{(\lambda_1,\ldots,\lambda_k)\in\Lambda_\epsilon} \sum_{i=1}^{k} \frac{\|\lambda_i B_p(\lambda_{i+1}) - \lambda_{i+1} B_p(\lambda_i)\|^2}{\lambda_i \lambda_{i+1}(\lambda_{i+1} - \lambda_i)}$$

It is essential to note that the distribution of the sup F test under H_0 above does not depend on any nuisance parameters. As Bai and Perron (1998) show, the test described above is consistent for its alternative.

Start by noting that $SSR_k = \sum_{i=1}^{k+1} \sum_{T_{i-1}}^{T_i} [y_t - f_t(\hat{\theta}_i[\bar{T}^k])]^2$ depends on the particular k-partition of the interval $[1, T]$, call it $\bar{T}^k$. Now let $\hat{\theta}_i[\bar{T}^k] = \hat{\theta}_i$. The two main steps in this proof involve analyzing the behavior of SSR_k and $SSR_0 - SSR_k$ under H_0. We will first show :

Lemma 6.1. *Under A3.2-A4.1 and* $H_0 : m = 0$, $\dfrac{SSR_k}{T-(k+1)p} \xrightarrow{p} \sigma^2$.

Proof of Lemma 6.1:

The proof is very similar to the steps used in Theorem 3.1 (a trivial consequence of Lemma 3.1 is that the estimated sum of squared residuals converges in probability to σ^2). To that end, we follow the same steps. Begin by defining $d_t^* = \hat{u}_t - u_t$. Then:

$$\begin{aligned} \frac{SSR_k}{T-(k+1)p} &\equiv O(1) \times T^{-1} SSR_k \\ &= T^{-1} \sum_{t=1^T} u_t^2 + T^{-1} \sum_{t=1^T} (d_t^*)^2 + T^{-1} \sum_{t=1}^{T} u_t d_t^* \\ &\leq T^{-1} \sum_{t=1}^{T} u_t^2 \end{aligned} \tag{6.3}$$

According to Lemma 3.1,

$$\sum_{t=1}^{T} u_t d_t^* = o_p(1), \tag{6.4}$$

independent of the properties of the break-fractions. From (6.3), it follows that $T^{-1} \sum_{t=1}^{T} (d_t^*)^2 \leq 0$ with large probability, which cannot be the case unless $T^{-1} \sum_{t=1}^{T} (d_t^*)^2 \xrightarrow{p} 0$. This in turn implies $\dfrac{SSR_k}{T-(k+1)p} - T^{-1} \sum_{t=1}^{T} u_t^2 \xrightarrow{p} 0$, hence $\dfrac{SSR_k}{T-(k+1)p} \xrightarrow{p} \sigma^2$. □

Now we turn to $SSR_0 - SSR_k$. To this end, let $\hat{\theta}_{1,i}$ and $\hat{\theta}_i$ denote the parameter estimates using observations in intervals $[1, T_i]$, respectively $[T_{i-1}+1, T_i]$, and let the sums squared residuals based on these estimates be $S_{1,i}$, respectively S_i. It is important to realize that, under this notation, $\hat{\theta}_{1,1} = \hat{\theta}_1$, hence $S_{1,1} = S_1$. Also denote $\sum_{j,i} \equiv \sum_{t=T_{j-1}+1}^{T_i}$.

For the purpose of analyzing $SSR_0 - SSR_k$, we need the properties of $\hat{\theta}_{1,i}$ and $\hat{\theta}_i$.

Lemma 6.2. *Under A3.2-A4.1 and* $H_0 : m = 0$,
$T^{1/2}(\hat{\theta}_{1,i} - \theta^0) \Rightarrow \sigma \lambda_i^{-1} D^{-1/2}(\theta^0) B_p(\lambda_i)$, *where* $D(\theta)$ *was defined in A3.4(iii).*

Proof of Lemma 6.2:

First, $\hat{\theta}_{1,i} \xrightarrow{p} \theta^0$ because it is just a sub-sample NLS estimator of θ^0 in stable models.

Using the MVT and letting $\bar{\theta}_{1,i}$ interpolate between $\hat{\theta}_{1,i}$ and θ^0,

$$\left.\frac{\partial S_{1,i}}{\partial \theta}\right|_{\hat{\theta}_{1,i}} = \left.\frac{\partial S_{1,i}}{\partial \theta}\right|_{\theta^0} + \left.\frac{\partial^2 S_{1,i}}{\partial \theta \partial \theta'}\right|_{\bar{\theta}_{1,i}} (\hat{\theta}_{1,i} - \theta^0) = 0$$

Then:

$$\begin{aligned} T^{1/2}(\hat{\theta}_{1,i} - \theta_0) &= - \left[\left.\frac{\partial^2 S_{1,i}}{\partial \theta \partial \theta'}\right|_{\bar{\theta}_{1,i}} \right]^{-1} \left.\frac{\partial S_{1,i}}{\partial \theta}\right|_{\theta^0} \\ &= \left[T^{-1} \sum_{t=1}^{[T\lambda_i]} F_t(\bar{\theta}_{1,i}) F_t(\bar{\theta}_{1,i})' - T^{-1} \sum_{t=1}^{[T\lambda_i]} u_t f_t^{(2)}(\bar{\theta}_{1,i}) \right]^{-1} \times \\ &\times T^{-1/2} \sum_{t=1}^{[T\lambda_i]} u_t F_t(\theta^0), \end{aligned}$$

where $f_t^{(2)}(\cdot)$ is the $p \times p$ second derivative of $f_t(\cdot)$. Now $\{u_t F_t(\theta^0)\}$ satisfies the conditions in FCLT. Moreover, θ is fixed here, and:

$$\begin{aligned} \text{Cov}\ [u_t F_t(\theta^0), u_s F_s(\theta^0)'] &= E[u_t u_s | x_t, x_s]\ E[F_t(\theta^0) F_s(\theta^0)'] = O_{p \times p} \\ \text{Var}\ [u_t F_t(\theta^0)] &= E[u_t^2 | x_t] E[F_t(\theta^0) F_t(\theta^0)'] = \sigma^2 D(\theta^0) \end{aligned}$$

because $E[u_t u_s | x_t, x_s]$ by A4.1(ii) and $\|F_t(\theta^0) F_s(\theta^0)'\| \leq \sup_t \|F_t(\theta^0)\|^s < \infty$ by A3.3(iv). Hence, we can apply FCLT 2 to obtain:

$$T^{-1/2} \sum_{t=1}^{[T\lambda_i]} u_t F_t(\theta^0) \xrightarrow{p} \sigma D^{-1/2}(\theta^0) B_p(\lambda_i) \qquad (6.5)$$

On the other hand, because $F_t(\cdot)$ is a continuous function, and $\bar{\theta}_{1,i} \xrightarrow{p} \theta^0$ by consistency of $\hat{\theta}_{1,i}$, we get:

$$T^{-1}\sum_{t=1}^{[T\lambda_i]} F_t(\bar{\theta}_{1,i})F_t(\bar{\theta}_{1,i})' - T^{-1}\sum_{t=1}^{[T\lambda_i]} F_t(\theta^0)F_t(\theta^0) = o_p(1)$$

But $T^{-1}\sum_{t=1}^{[T\lambda_i]} F_t(\theta^0)F_t(\theta^0) \xrightarrow{p} \lambda_i D(\theta^0)$ by A3.4(iii), so

$$T^{-1}\sum_{t=1}^{[T\lambda_i]} F_t(\bar{\theta}_{1,i})F_t(\bar{\theta}_{1,i})' \xrightarrow{p} \lambda_i D(\theta^0) \tag{6.6}$$

By similar arguments,

$$T^{-1}\sum_{t=1}^{[T\lambda_i]} u_t f_t^{(2)}(\bar{\theta}_{1,i}) - T^{-1}\sum_{t=1}^{[T\lambda_i]} u_t f_t^{(2)}(\theta^0) = o_p(1) \tag{6.7}$$

But $T^{-1}\sum_{t=1}^{[T\lambda_i]} u_t f_t^{(2)}(\theta^0) = o_p(1)$ by the law of large numbers for strictly stationary processes, since $E[u_t f_t^{(2)}(\theta^0)] = 0$.[1] Hence:

$$T^{-1}\sum_{t=1}^{[T\lambda_i]} u_t f_t^{(2)}(\bar{\theta}_{1,i}) = o_p(1) \tag{6.8}$$

Putting together (6.5), (6.6) and (6.8), together with Slutsky's Theorem, we obtain the desired result:

$$T^{1/2}(\hat{\theta}_{1,i} - \theta^0) \Rightarrow \sigma\lambda_i^{-1}D^{-1/2}(\theta^0)B_p(\lambda_i)$$

□

The properties of $\hat{\theta}_i$ are similar to those of $\hat{\theta}_{1,i}$:

Lemma 6.3. *Under A4.1, A3.2-A3.5 and $H_0 : m = 0$,*
$T^{1/2}(\hat{\theta}_i - \theta^0) \Rightarrow \sigma[\lambda_i - \lambda_{i-1}]^{-1}D^{-1/2}(\theta^0)B_p(\lambda_i)$.

The proof of the lemma above follows exactly the same steps as the proof of Lemma 6.2 and is omitted for simplicity. We are now ready to prove Theorem 6.1.

[1] Such a law of large numbers can be deduced from FCLT 2 since we have shown that $\{u_t f_t^{(2)}(\theta^0)\}$ satisfies its conditions in the Proof of Theorem 4.1.

Proof of Theorem 6.1:

Denote by $\hat{\theta}$ the full sample NLS estimator, under H_0.

$$SSR_0 - SSR_k = \sum_{t=1}^{T}[y_t - f_t(\hat{\theta})]^2 - \sum_{i=1}^{k+1}\sum_{t=T_{i-1}}^{T_i}[y_t - f_t(\hat{\theta}_{i+1})]^2$$
$$= \sum_{i=1}^{k}\left[\sum_{1,i+1} u_t^2(\hat{\theta}_{1,i+1}) - \sum_{1,i} u_t^2(\hat{\theta}_{1,i}) - \sum_{i+1,i+1} u_t^2(\hat{\theta}_{i+1})\right] + \left[\sum_{1} u_t^2(\hat{\theta}_{1,1})^2 - u_t^2(\hat{\theta}_1)\right]$$

Let $F^*_{T,i} = \sum_{1,i+1} u_t^2(\hat{\theta}_{1,i+1})^2 - \sum_{1,i} u_t^2(\hat{\theta}_{1,i})^2 - \sum_{i+1,i+1} u_t^2(\hat{\theta}_{i+1})$, and $\sum_{j,i} \equiv \sum_{t=T_{j-1}+1}^{T_i}$. Hence, we can write:

$$SSR_0 - SSR_k = \sum_{i=1}^{k} F^*_{T,i}$$

$F^*_{T,i}$ can be written as:

$$F^*_{T,i} = D^R(1, i+1) - D^R(1,i) - D^U(i+1, i+1), \tag{6.9}$$

where $D^R(1,i) = \sum_{1,i}[u_t^2(\hat{\theta}_{1,i}) - u_t^2]$ and $D^U(i,i) = \sum_{i,i}[u_t^2(\hat{\theta}_i) - u_t^2]$.

$$D^R(1,i) = \sum_{1,i}[y_t - f_t(\hat{\theta}_{1,i})]^2 - [y_t - f_t(\theta^0)]^2$$
$$= \sum_{1,i}[f_t(\theta^0) - f_t(\hat{\theta}_{1,i})]^2 - 2\sum_{1,i} u_t\ d_t(\hat{\theta}_{1,i}, \theta^0) = I + II.$$

Using the MVT, we have:

$$I = T^{1/2}(\hat{\theta}_{1,i} - \theta^0)' \left[T^{-1}\sum_{1,i} F_t(\bar{\theta}_{1,i,t})F_t(\bar{\theta}_{1,i,t})'\right] T^{1/2}(\hat{\theta}_{1,i} - \theta^0) \tag{6.10}$$
$$II = 2\left[T^{-1/2}\sum_{1,i} u_t F_t(\bar{\theta}_{1,i,t})\right] T^{1/2}(\hat{\theta}_{1,i} - \theta^0),$$

where $\bar{\theta}_{1,i,t}$ lies in the segment line $\hat{\theta}_{1,i}$ and θ^0.

Since $\bar{\theta}_{1,i,t} \xrightarrow{p} \theta^0$ for each t and $E[F_t(\theta)F_t(\theta)']$ has uniform bounds by A3.3(iv), an argument similar to Lemma 6.2 yields:

$$T^{-1}\sum_{1,i} F_t(\bar{\theta}_{1,i,t})F_t(\bar{\theta}_{1,i,t})' \xrightarrow{p} \lambda_i D(\theta^0). \tag{6.11}$$

From (6.11), (6.10) and Lemma 6.2,

$$\begin{aligned} I &\xrightarrow{d} \sigma\lambda_i^{-1}D^{-1/2}(\theta^0)B_p(\lambda_i)' \times \lambda_i D(\theta^0) \times \sigma\lambda_i^{-1}D^{-1/2}(\theta^0)B_p(\lambda_i) \\ &= \sigma^2\|B_p(\lambda_i)\|^2/\lambda_i \end{aligned} \tag{6.12}$$

On the other hand,

$$T^{-1/2}\sum_{1,i} u_t F_t(\bar{\theta}_{1,i,t}) \xrightarrow{d} \sigma D^{1/2}(\theta^0)B(\lambda_i) \tag{6.13}$$

From equations (6.13), (6.10) and Lemma 6.2,

$$\begin{aligned} II &\xrightarrow{d} -2\sigma\lambda_i^{-1}D^{-1/2}(\theta^0)B_p(\lambda_i)' \times \sigma D^{1/2}(\theta^0)B(\lambda_i) \\ &= -2\sigma^2\|B_p(\lambda_i)\|^2/\lambda_i \end{aligned} \tag{6.14}$$

From (6.11) and (6.13),

$$D^R(1,i) = I + II \xrightarrow{d} -\sigma^2\|B_p(\lambda_i)\|^2/\lambda_i$$

Similarly,

$$D^R(1,i+1) \xrightarrow{d} -\sigma^2\|B_p(\lambda_{i+1})\|^2/\lambda_{i+1}$$

Using Lemma 6.3 and similar arguments, we have:

$$D^R(i+1,i+1) \xrightarrow{d} -\sigma^2\|B_p(\lambda_{i+1}) - B_p(\lambda_i)\|^2/[\lambda_{i+1} - \lambda_i]$$

Using the last three equations into (6.9),

$$\begin{aligned}
F^*_{T,i} &= D^R(1, i+1) - D^R(1, i) - D^R(i+1, i+1) \\
&\xrightarrow{d} -\frac{\sigma^2 \|B_p(\lambda_{i+1})\|^2}{\lambda_{i+1}} + \frac{\sigma^2 \|B_p(\lambda_i)\|^2}{\lambda_i} + \frac{\sigma^2 \|B_p(\lambda_{i+1}) - B_p(\lambda_i)\|^2}{\lambda_{i+1} - \lambda_i} \\
&= \sigma^2 \frac{\|\lambda_i B_p(\lambda_{i+1}) - \lambda_{i+1} B_p(\lambda_i)\|^2}{\lambda_i \lambda_{i+1} [\lambda_{i+1} - \lambda_i]}
\end{aligned}$$

Hence,

$$\begin{aligned}
\sup_{(\lambda_1, \ldots, \lambda_k) \in \Lambda_\epsilon} F_T(k; p) &= \sum_{i=1}^{k} F^*_{T,i} \\
&= \sigma^2 \sum_{i=1}^{k} \sup_{(\lambda_1, \ldots, \lambda_k) \in \Lambda_\epsilon} \frac{\|\lambda_i B_p(\lambda_{i+1}) - \lambda_{i+1} B_p(\lambda_i)\|^2}{\lambda_i \lambda_{i+1} [\lambda_{i+1} - \lambda_i]}
\end{aligned}$$

which completes the proof of Theorem 6.1. □

6.2 A Double Maximum Test

Sometimes, the researcher may know little about the number of break-points. In that case, one can consider testing against an unknown number of breaks $m < M$, M being an upper bound on the number of change-points. To that end, consider the hypothesis:

$$H_0^* : m = 0 \qquad vs. \qquad H_A^* : \text{ m unknown}, m < M, M \text{ fixed.} \tag{6.15}$$

As Bai and Perron (1998) point out, it suffices to take the maximum over weighted versions of the test statistics described in the previous section, where the weights are $(a_1, \ldots, a_M)$

$$D \max F_T(M, a_1, \ldots, a_M) = \max_{1 \le m \le M} a_m \sup_{(\lambda_1, \ldots, \lambda_k) \in \Lambda_\epsilon} F_T(m; p) \tag{6.16}$$

The distribution of the test statistic above is:

Corrolary 1 to Theorem 6.1. *Under A3.2-A4.1 and* H_0^*,

$$D\max F_T(M, a_1, \ldots, a_M)$$
$$\xrightarrow{d} \max_{1\leq m\leq M} \frac{a_m}{kp} \sup_{(\lambda_1,\ldots,\lambda_k)\in\Lambda_\epsilon} \sum_{i=1}^{k} \frac{\|\lambda_i B_p(\lambda_{i+1}) - \lambda_{i+1} B_p(\lambda_i)\|^2}{\lambda_i\lambda_{i+1}(\lambda_{i+1}-\lambda_i)} \quad (6.17)$$

This distribution is readily computed from Theorem 6.1, and it is the double-maximum test described in Bai and Perron (1998). As they mention, the choice of weights remains an open question. It may reflect the imposition of some priors on the likelihood of various number of breaks. One possibility is to set all weights equal to unity. We denote this test as $UD\max F_T(M,p) = \max_{1\leq m\leq M} \sup_{(\lambda_1,\ldots,\lambda_k)\in\Lambda_\epsilon} F_T(m;p)$. Recall that, for a fixed m, $F_T(m;p)$ is the sum of m dependent χ^2 variables, each with p degrees of freedom, each divided by m. This scaling by m can be viewed as a prior that, as m increases, a fixed sample becomes less informative about the hypotheses being confronted. Since for any fixed p, the critical values of $\sup_{(\lambda_1,\ldots,\lambda_k)\in\Lambda_\epsilon} F_T(m;p)$ decrease as m increases, this implies that if we have a large number of breaks, we may get a test with low power, because the marginal p-values decrease with m. One way to keep marginal p-values of the tests equal across m is to use weights that depend on p and the significance level of the test, say $\tilde{\alpha}$. More precisely, let $c(p,\tilde{\alpha},m)$ be the asymptotic critical value of the test $\sup_{(\lambda_1,\ldots,\lambda_k)\in\Lambda_\epsilon} F_T(m;p)$. Define, as in Bai and Perron (1998), $a_1 = 1$ and $a_m = c(p,\tilde{\alpha},1)/\, c(p,\tilde{\alpha},m)$ for $1 < m \leq M$. The test obtained this way is:

$$WD\max F_T(M,p) = \max_{1\leq m\leq M} \frac{c(p,\tilde{\alpha},1)}{c(p,\tilde{\alpha},m)} \times \sup_{(\lambda_1,\ldots,\lambda_k)\in\Lambda_\epsilon} F_T(m;p).$$

For consistency of Dmax tests and critical values of both its version, UDmax and WDmax, see Bai and Perron (1998).

6.3 A Test of ℓ Versus $\ell + 1$ Breaks

Consider the following hypothesis of interest:

$$H_0^{**} : m = \ell \qquad vs. \qquad H_A^{**} : m = \ell + 1. \tag{6.18}$$

As Bai and Perron (1998) mention, one would ideally construct such a test based on the difference between the sum of squared residuals obtained with ℓ breaks and that obtained with $(\ell + 1)$ breaks. Considering the different mismatches in end-points of partial sums obtained this way, it would be hard to describe the limiting behavior of such tests. An easier strategy involves imposing ℓ breaks and testing each segment for an additional change-point. The test statistic is:

$$F_T(\ell + 1|\ell) = \{S_T(\hat{T}_1, \ldots, \hat{T}_\ell) - \min_{1 \le i \le \ell+1} \inf_{\tau \in \Delta_{i,\ell}} S_T(\hat{T}_1, \ldots, \hat{T}_{i-1}, \tau, \hat{T}_i, \ldots, \hat{T}_\ell)\}/\hat{\sigma}^2.$$

where

$$\Delta_{i,\ell} = \{\tau : \hat{T}_{i-1} + \left(\hat{T}_i - \hat{T}_{i-1}\right)\eta \le \tau \le \hat{T}_i - \left(\hat{T}_i - \hat{T}_{i-1}\right)\eta\}, \text{ and } \hat{\sigma}^2 \xrightarrow{p} \sigma^2 \text{ under } H_0^* * .$$

Its distribution is the same as in Bai and Perron (1998):

Theorem 6.2. *Under A3.2-A4.1 and H_0^{**},* $\lim P(F_T(l + 1|l) \le x) = G_{p,\eta}^{l+1}$*, where $G_{p,\eta}$ is the distribution function of*

$$\sup_{\eta \le \mu \le 1-\eta} \frac{\|B_p(\mu) - \mu B_p(1)\|^2}{\mu(1-\mu)}$$

As Bai and Perron (1998) mention, this test must be consistent, too. If there are more than ℓ breaks, but we estimated a model with just ℓ breaks, then there must be at least one additional break not estimated. Hence, at least one of the $(\ell + 1)$ segments obtained contains a nontrivial breakpoint, in the sense that both boundaries of this segment are separated from the true break-point by a positive fraction of the total number of observations. For this segment, the $\sup_T F(1,p)$ test statistic diverges to infinity as the sample size increases, since this test in consistent (see Section 6.1). Then so does $F_T(\ell+1|\ell)$ under H_A^{**}, so this test is consistent for its alternative.

Proof of Theorem 6.2:

Under $H_0^{**}: m=\ell$, compute the estimated break-points, and let $SSR(i,j)$ be the minimized sum of squared residuals for the segment containing observations in the interval $[T_i+1, T_j]$. With a small abuse of notation, we can write:

$$F_T(\ell+1|\ell) = \sup_{1\le i\le \ell} \sup_{\tau\in\Delta_{i,\eta}} F_{T,i}^*(\ell+1|\ell)/\hat{\sigma}^2, \tag{6.19}$$

$$\text{where } F_{T,i}^*(\ell+1|\ell) = SSR(\hat{T}_{i-1},\hat{T}_i) - SSR(\hat{T}_{i-1},\tau) - SSR(\tau,\hat{T}_i).$$

For simplicity, use the following notation: $\sum_{\hat{i}} = \sum_{t=\hat{T}_{i-1}+1}^{\hat{T}_i}$, $\sum_{\hat{i},\,\tau} = \sum_{t=\hat{T}_{i-1}+1}^{\tau}$ and $\sum_{\tau,\,\hat{i}+1} = \sum_{t=\tau}^{\hat{T}_i}$. As before, let $\hat{\theta}_i$ be the parameter estimates, using estimated break-fractions, under $H_0^{**}: m=\ell$, and $\hat{\theta}_{i,\tau}$, respectively $\hat{\theta}_{\tau,i+1}$, be the estimator using observations from $[\hat{T}_{i-1}+1,\tau]$, respectively $[\tau,\hat{T}_i]$.

Adding and subtracting $\sum_{\hat{i}} u_t^2$ yields, by a similar argument as in the Proof of Theorem 6.1,

$$F_{T,i}^*(\ell+1|\ell) = E^R(\hat{i},\hat{i}) - E^U(\hat{i},\tau) - E^U(\tau,\hat{i}+1) \tag{6.20}$$

with:

$$E^R(\hat{i},\hat{i}) = \sum_{\hat{i}} [u_t^2(\hat{\theta}_i) - u_t^2];$$
$$E^U(\hat{i},\tau) = \sum_{\hat{i},\,\tau} [u_t^2(\hat{\theta}_{i,\tau}) - u_t^2];$$
$$E^U(\tau,\hat{i}+1) = \sum_{\tau,\,\hat{i}+1} [u_t^2(\hat{\theta}_{\tau,i+1}) - u_t^2].$$

By Theorem 4.1, $\hat{\theta}_i \xrightarrow{p} \theta_i^0$. Similarly, $\hat{\theta}_{i,\tau} \xrightarrow{p} \theta_i^0$ and $\hat{\theta}_{\tau,i+1} \xrightarrow{p} \theta_i^0$ because imposing an additional break when there is none is as if we were imposing one break where there is none in segment $[\hat{T}_{i-1}+1, \hat{T}_i]$, hence the same arguments can be employed as for Lemma 6.3 to show consistency. Then the terms in each of the sums $E^R(\hat{i}, \hat{i})$, $E^U(\hat{i}, \tau)$ and $E^U(\tau, \hat{i}+1)$ are $o_p(1)$. So, by Theorem (3.2), we can change the endpoints of the sums above with the true break-points because the difference is a bounded number of terms that are $o_p(1)$. If we do so, we have:

$$E^R(\hat{i}, \hat{i}) = \sum_{i,i} [u_t^2(\hat{\theta}_i) - u_t^2] + o_p(1) \xrightarrow{d} \sigma^2 \|B_p(1)\|^2 \tag{6.21}$$

by similar arguments as in Proof of Theorem 6.1 for $D^R(1, i)$. Recall that

$$\Delta_{\hat{i},\ell} = \{\tau : \hat{T}_{i-1} + \left(\hat{T}_i - \hat{T}_{i-1}\right)\eta \leq \tau \leq \hat{T}_i - \left(\hat{T}_i - \hat{T}_{i-1}\right)\eta\}, \tag{6.22}$$

and set $\tau = [T\mu]$.

Then by similar arguments as for $D^U(i, i)$ in the Proof of Theorem 6.1,

$$E^U(\hat{i}, \tau) \Rightarrow -\sigma^2 \|B_p(\mu)\|^2/\mu \text{ and } E^U(\tau, \hat{i}+1) \xrightarrow{d} -\sigma^2 \|B_p(1-\mu)\|^2/(1-\mu). \tag{6.23}$$

Putting (6.21) and (6.23) together, we get:

$$F_{T,i}^*(\ell+1|\ell) \Longrightarrow \left[sup_{\eta \leq \mu \leq 1-\eta} \frac{\|B_p(\mu) - \mu B_p(1)\|^2}{\mu(1-\mu)} \right]. \tag{6.24}$$

Now, since the regimes considered in $SSR(\cdot, \cdot)$ are non-overlapping, $F_{T,i}^*(\ell+1|\ell)$ are asymptotically independent for different i. Hence, the result in Theorem 6.2. □

Chapter 7

Simulations

In this chapter, we test the usefulness of our method in small samples. We discuss computational aspects and provide simulations results for one, two and three breaks, with both small and large parameter shifts. In both cases, even for small samples, the method is able to accurately estimate the break locations.

At first sight, searching over all possible partitions is burdensome, because for m breaks, there are $O(T^m)$ partitions to be searched over. Fortunately, from Bai and Perron (2003b) it becomes evident that with a sample size of T, the number of possible segments of a sample to be considered is only $T(T+1)/2$. For example, if we have $m = 2$ breaks, the possible partitions can include the first and second segments changing while the third one is kept fixed. Hence we can store the sum of squared residuals in that segment and do not have to compute it each time. Eventually, we will have $T(T+1)/2$ stored sum of squared residuals corresponding to $T(T+1)/2$ repeating segments, and for all other partitions, the total sum of squared residuals is necessarily a linear combination of those.[1] We also want each segment to be of size $[hT]$, that is, large enough to provide reasonable estimates.[2] This implies a reduction in the number of segments to be considered by $(h-1)T - (h-2)(h-1)/2$. The largest segments need to be short enough to allow m other segments, before and/or after. For example, when the segment starts at a

[1] For more details, see Bai and Perron (2003b).
[2] In practice, h is usually set to 0.05, 0.10 or 0.15.

date between 1 an h, the maximal length of this segment is $T - hm$ when we allow for m breaks. This allows another reduction in the number of segments considered by $h^2 m(m+1)/2$. Finally, a segment cannot start at dates between 2 and h, because no other segment of a minimum length is allowed before it. Hence, a further reduction in the number of segments considered occurs. In conclusion, there are some clear computational advantages of this method.

In linear models, there exists a recursion formula that links the sums of squared residuals with l and $l+1$, greatly reducing computational time. This formula is not available for nonlinear models, only an approximation to it is. The approximation is based on:

$$T^{1/2}[\hat{\theta}_i - \theta_i^0] = (T^{-1}\bar{F}_i'\,\bar{F}_i)^{-1}\,T^{-1/2}\bar{F}_i'\,\bar{U}_i + o_p(1)$$

Since this approximation may not work very well in finite samples, we do not perform it, therefore trading off computational time for accuracy.

Our simulations are based on the following data generation process.

$$f(x_t, \theta) = \theta_i^1 + \theta_i^2\, e^{x_t \theta_i^3}, \text{ with t } \in [T_i^0, T_{i+1}^0], \qquad \text{for } i = 1, \ldots, m+1;\; m = 1, 2, 3.$$

The true data was generated such that $x_t \sim \mathcal{N}(0,1), u_t \sim \mathcal{N}(0, 0.25)$ and $X \perp U$. The regression function we picked is one of the most simple nonlinear regression models, and simulations for NLS based on this function can be found in Gallant (1989). We set $h = 0.15$ and all simulation results tabulated in the Appendix are Monte Carlo average estimates from 100 simulations.

From Tables 1-8, we can see that the break-fraction estimates are very accurate for $m = 1, 2, 3$ breaks. The break-fraction estimation works equally well for one and multiple breaks, if the breaks are large or small. From 100 simulations, we obtained each time a maximum of one wrong estimate. Note that this accuracy is obtained even with very small sample sizes of $T = 50$.

We also see that when the true parameter shifts are far apart, as in Tables 2, 5 and 7, independent of the number of breaks, the estimates very close to their true values and their estimated standard deviation (based on the formula suggested at the end of Chapter 4) is very small. As expected, the standard deviation decreases with a sample size increase, rendering increasingly accurate parameter estimates.

The MC averaged parameter estimates become less accurate as the parameter shifts become small. This problem is an artifact of one wrong estimate of the break-point each time. On the other hand, the NLS method we proposed is designed for a limited number of large breaks, and for smaller shifts different modeling techniques may be required.

Overall, we find that our estimation method works pretty well for finite samples with single or multiple breaks.

Chapter 8

An Application to the US Interest Rate Reaction Function

This chapter illustrates our methodology in the context of interest rate reaction functions (also called monetary policy rules). It is structured as follows: first, we discuss connections between a popular interest rate reaction model, the smooth transition autoregressive (STAR) model, and ours. Section 2 reviews empirical evidence regarding nonlinearity and breaks in the interest rate reaction function. Finally, Section 3 discusses the empirical results, with their advantages and caveats.

8.1 STAR versus Unstable NLS

In application to economic time series, models which allow for state-dependent variables or regime switching behavior are popular tools. One of these regime-switching models is the smooth transition model, and among its many variants, it includes the STAR model. Thorough reviews of the literature on STAR modeling can be found in Granger and Teräsvirta (1993), Teräsvirta (1998), Potter (1999) and van Dijk, Teräsvirta, and Franses (2002).

STAR models are an attractive tool for assessing monetary policy because they tie non-linearity in the reaction function to one or more state variables: as soon as one variable hits a threshold, a linear regime slowly changes into another linear regime. The threshold can be estimated together with the smoothness parameter and the other parameters of the model. To illustrate this, consider a simple one-transition STAR model:

$$y_t = x_t'\beta_1^* \left[1 - G(s_t;\gamma,c)\right] + x_t'\beta_2^*\, G^*(s_t;\gamma,c) + u_t \quad (t = 1,\ldots T) \tag{8.1}$$

with a first-order logistic transition function:

$$G^*(s_t;\gamma,c) = \frac{1}{1+\exp[-\gamma(s_t - c)]} \tag{8.2}$$

Here, x_t are $s \times 1$ regressors that can include lagged dependent variables, s_t is the state variable that can be an exogenous or endogenous variable, and the parameter vector is $\theta = (\beta_1^{*'}, \beta_2^{*'}, \gamma, c)'$. It is usually assumed that u_t is a martingale difference sequence with respect to the history of the time series up to time $t-1$, call it $\Omega_{t-1} = \{x_{t-1}, x_{t-2}, \ldots, x_{1-\pi}\}$, with π indicating the (finite) number of lags in the model. Hence, $E\left[u_t|\Omega_{t-1}\right] = 0$. For simplicity, practitioners often assume that $E[u_t^2|\Omega_{t-1}] = \sigma^2$. These conditions, together with the properties of the smooth transition function $G^*(s_t;\cdot)$ and a mixing assumption on the regressors, can be strengthened to fulfill Assumptions 2.8 - 2.11 of the stable NLS model described in Chapter 2. Hence, this model fits well into our NLS framework, and we can combine breaks with STAR models in each regime.

Below, we further discuss the properties of STAR models. First, note that in (8.2), the parameter c can be interpreted as the threshold between two regimes, because the transition function changes monotonically from 0 to 1 as s_t increases, and $G^*(c; s_t, c) = 0.5$. Second, $\gamma \geq 0$ measures the smoothness of the transition. A small γ indicates a small change, while for a very large γ, $G^*(s_t;\gamma,c)$ approaches the indicator function $\mathbb{1}[s_t > c]$, collapsing into a threshold autoregressive model (TAR). If the state variables is time, i.e. $s_t = t$, then for large γ the model approaches a linear model with one break at $T_1 = c$. To clarify this statement, rewrite the model in (8.1) in the following way:

$$y_t = x_t'\beta_1 + x_t'\beta_2 \; G(s_t;\gamma,c) + u_t \quad (t = 1,\ldots T) \tag{8.3}$$

with $G(s_t;\cdot)$ a logistic transition function. If $s_t = t$, then as $\gamma \to \infty$, $G(s_t;\gamma,c) \to 1$, collapsing into:

$$y_t \;=\; \begin{cases} x_t'\,\beta_1 + u_t & t = 1,\ldots,c \\ x_t'\,\beta_2 + u_t & t = c+1,\ldots,T \end{cases} \tag{8.4}$$

If we obtain a large γ, Eitrheim and Teräsvirta (1996) point out that the testing procedures for STAR can be pretty unstable and suggest remedies. Moreover, it is unclear whether a large γ value would also spill into inaccurate estimates for the other parameters. To avoid this potential outcome, since methods for testing breaks in linear models are available - see Bai and Perron (1998), it seems reasonable to first test for breaks. If s_t is not time, a TAR modeling strategy should be explored prior to STAR modeling.

Similar to (8.3), consider a two-transition model, that is, a model with two state variables:

$$y_t = x_t'\beta_1 + x_t'\beta_2 \; G_1(s_{t,1};\gamma_1,c_1) + x_t'\beta_3 \; G_2(s_{t,2};\gamma_2,c_2) + u_t \quad (t = 1,\ldots T) \tag{8.5}$$

Assume, as in Kesriyeli, Osborn, and Sensier (2004), that one state variable is time, say $s_{t,2} = t$. For a very large γ_2, the model reduces to:

$$y_t \;=\; \begin{cases} x_t'\beta_1 + x_t'\beta_2 \; G_1(s_{t,1};\gamma_1,c_1) + u_t & t = 1,\ldots,c_2 \\ x_t'\,(\beta_1+\beta_3) + x_t'\beta_2 G_1(s_{t,1};\gamma_1,c_1) + u_t & t = c_2+1,\ldots,T \end{cases} \tag{8.6}$$

If we let in $\theta_1 = (\beta_1', \beta_2, \gamma_1, c_1)'$ and $\theta_2 = (\beta_3', \beta_2, \gamma_1, c_1)'$,

$$f_t(\theta_1) = x_t'\beta_1 + x_t'\beta_2 \; G_1(s_{t,1};\gamma_1,c_1) \; ; \; f_t(\theta_2) = x_t'(\beta_1+\beta_3) + x_t'\beta_2 \; G_1(s_{t,1};\gamma_1,c_1)$$

and $T_1 = c_2$, then we obtain our NLS model with one break:

$$y_t = \begin{cases} f_t(\theta_1) + u_t & t = 1, \ldots, T_1 \\ f_t(\theta_2) + u_t & t = T_1 + 1, \ldots, T \end{cases} \tag{8.7}$$

The only difference is that some of the parameters in this model are constant across the two regimes (that is, we have partial structural change). Since we did not allow for it in our theoretical results, we can abstract from this and estimate the model by allowing all parameters to change at an unknown break.[1] In fact, this can be done in the STAR framework, too, by adding to the regression function a term like $G_1^*(s_{t,1}; \gamma_1^*, c_1^*)$. Clearly, this analysis regarding large smoothness parameters can be extended to multiple regime STAR models that may collapse to NLS models with multiple breaks.

The discussion above illustrates the fact that a STAR model with time as one state variable and an associated large smoothness parameter is not that far from a model that allows for breaks, but nests STAR models[2] in each sub-sample, before and after a certain break. Hence, it may be useful to use our break tests as an initial specification tests of STAR, even though without further analysis it is unclear why we would reject a STAR model: maybe the transition is too rough, but maybe the state variable and/or the model is misspecified. It may also be the case that when γ_2 is large, c_2 is not accurately estimated, hence it is useful to compare break-point estimates with the time threshold estimates from the STAR models.

[1]Admittedly, it is desirable to have an asymptotic theory and tests for partial structural change. This is a subject of future research.

[2]Of course, as soon as breaks are accounted for, time should not be anymore included a a state variable.

8.2 Evidence for Nonlinearity and Breaks in the Interest Rate Reaction Function

Until recently, interest rate reaction functions were modeled by means of so called linear Taylor rules, which capture the response of monetary authority to inflation and output gap. These rules are derived from a linear Phillips curve and a quadratic loss function for the central banker - see *interalia* Clarida and Gertler (1997), Clarida, Galí, and Gertler (1998), Clarida, Galí, and Gertler (2000). In these models, parameter linearity is imposed in conjunction with parameter stability.

Several recent theoretical and empirical studies question the assumptions of linearity and/or parameter stability underlying monetary policy rules. For example, Schaling (2004) examines the theoretical implications of a nonlinear Phillips curve. He shows that the resulting optimal interest rate rule is a nonlinear function of the deviation of inflation from its target and of output from its potential level. Dolado, María-Dolores, and Naveira (2005) find empirical evidence for nonlinearity related to the interaction between expected inflation and the output gap for four European countries but not for US. However, when the assumption of a quadratic loss function is dropped, Dolado, María-Dolores, and Ruge-Murcia (2004) find evidence of nonlinearity in the post-1982 (Volcker-Greenspan era) US monetary policy reaction function. Other studies that support nonlinearity for different time periods are *inter alia* Kim, Osborn, and Sensier (2005) and Florio (2006) for US, Martin and Milas (2004) for UK and Bec, Salem, and Collard (2002) for US, France and Germany.

In most of the studies, nonlinearity is modeled as a smooth transition between regimes associated with different inflation gaps (deviations of inflation from target), output gaps (deviations of output from their potential) or both. However, there is widely acknowledged evidence for changes in the interest rate reaction function associated with the tenure of each Fed Chairman - see Judd and Rudebusch (1997) and Dolado, María-Dolores, and Ruge-Murcia (2004). Kesriyeli, Osborn, and Sensier (2004) follow up on this evidence by allowing for nonlinearity associated with time. They estimate a STAR model with two transitions (hence two type of nonlinearities), one associated with three month interest rate changes and another with time. As every STAR model, estimation is done by a NLS method with constant parameters, hence can be nested into our unstable NLS framework. While their modeling strategy proves to be a good fit to the data, their estimate of smoothness indicates that time-transition from one regime to another is done almost instantaneously (more specifically, their data is monthly and the transition occurs in two months).

As discussed in Section 1, our framework allows for both nonlinearity and breaks, hence we use it to compare the performance of the two-transition model above with an unstable NLS model in which we allow for both breaks and one-transition models associated with other variables within each regime. Our empirical results are presented in the next section.

8.3 Empirical Results

Following evidence of nonlinearity and time change, Kesriyeli, Osborn, and Sensier (2004) use monthly data from $1984:1-2002:12$ and employ the following two-transition model:

$$r_t = x_t'\beta_1 + x_t'\beta_2 G_1(\Delta_3 r_{t-1}; \gamma_1, c_1) + x_t'\beta_3 G_2(t; \gamma_2, c_2) + u_t$$

where r_t is the Federal Funds Rate,

$$x_t' = (1, r_{t-1}, r_{t-2}, \pi g_{t-1}, \pi g_{t-2}, \pi g_{t-3}, og_{t-1}, og_{t-2}, \Delta wcp_{t-3}) \tag{8.8}$$

with $\pi g_t, og_t$ denoting inflation gap, respectively output gap, while Δwcp_t stands for the change in the world commodity prices at time t.[3] Here $G_1(\Delta_3 r_{t-1}; \gamma_1, c_1)$ is a logistic transition function associated with a three month change in the interest rate, i.e. $s_{t,1} = \Delta_3 r_{t-1} = r_{t-1} - r_{t-4}$, and

[3]For details on how these series are constructed at a monthly frequency, see Kesriyeli, Osborn, and Sensier (2004).

$G_2(t; \gamma_2, c_2)$ another logistic transition function associated with time, i.e. $s_{t,2} = t$:

$$G_i(s_{t,i}; \gamma_i, c_i) = \frac{1}{1 + \exp[-\gamma(s_{t,i} - c_i) \ / \ \hat{\sigma}(s_{t,i})]}, \ i = 1, 2 \tag{8.9}$$

Note the scaling parameter $\hat{\sigma}(s_{t,i})$ in the transition function. Their parameter estimates, reported in Table 9, indicate that $\hat{\gamma}_2 = 1082$, which is a large number resulting in a two-month only transition period. They estimate the time threshold parameter at $\hat{c}_2 = 14.14$, indicating some abrupt change in the coefficients at the beginning of 1985, soon after the beginning of their sample.

We use the same data as theirs[4], but for computational purposes we restrict ourselves to $1982 : 1 - 1990 : 12$, comprising $T = 104$ observations. Admittedly, the period before 1984:1 includes a part of the Volcker era that was quite unstable, but given that we are exploring instability, it makes sense to include this period. Since a one transition model with multiple breaks:

$$r_t = x_t'\beta_1^{(i)} + x_t'\beta_2^{(i)} \ G_1(\Delta_3 r_{t-1}; \gamma_1^{(i)}, c_1^{(i)}) + u_t, \qquad t \in [T_{i-1} + 1, T_i] \qquad i = 1, \ldots, 3 \tag{8.10}$$

would have - according to specification (8.8) - 22 parameters in each regime, it would be difficult to speed up such a program to find even one break. We would like to have a more parsimonious model. At the same time, for comparison purposes, we would like to keep $\Delta_3 r_{t-1}$ as a state variable in our model, even though with one-transition models, Kesriyeli, Osborn, and Sensier (2004) find a different state variable. After some backward selection in a linear model, we pick

$$x_t' = (1, r_{t-1}, r_{t-2}, \pi g_{t-2}, og_{t-2}) \tag{8.11}$$

This specification makes more or less sense because Kesriyeli, Osborn, and Sensier's (2004) estimates show that Δwcp_{t-3} does not seem to be a relevant factor for the interest rate after

[4]We would like to thank Marianne Sensier for providing us with the data.

the beginning of 1985. We also eliminate the first lags of inflation and output gap as for our sample size they do not seem to be very significant once we account for the second lags in a one-transition STAR model.

Our model in (8.10) and (8.11) is in line with Assumptions 3.1-3.4 because STAR models in each regime are models than can be estimated via NLS. Moreover, Kesriyeli, Osborn, and Sensier (2004) assume i.i.d. errors in their analysis, while we employ a similar assumption to obtain our tests.

With specification (8.11), we still have to estimate $\theta^{(i)} = (\beta_1^{(i)\prime}, \beta_2^{(i)\prime}, \gamma_1^{(i)}, c_1^{(i)})'$, implying $p = 12$ parameters in each regime i. Even though the critical values for the $supF(1;p)$ and $supF(\ell+1|\ell)$ are only tabulated by Bai and Perron (1998) for up to $p = 10$ parameters, they increase in the number of parameters (see Table 10), and that suffices for our purposes.

We start with a linear model, then estimate a one transition model with no breaks, and after carefully searching over different starting values, including the values estimated by Kesriyeli, Osborn, and Sensier (2004), we obtain the models in Table 9. This table also includes the estimates for the one-transition model with one break, as well as Kesriyeli, Osborn, and Sensier's (2004) estimates, even though the latter uses a different regressor set. The OLS fit provides us with a reference point. For example, we see that the coefficient on the output gap is positive, indicating that if output was two month before below potential, ceteris paribus, the interest rate may decrease to reduce the gap. As seen from the table, a one-transition model does not fit the data very well. This may be due to additional nonlinearity or to breaks.

Next, we test for multiple breaks in our data. Our tests in Table 10 indicate that both $supF(1;p)$[5] and its scaled version $supF(1|0)$[6] reject the null of no breaks at a 1% level. Given the critical values from the same table, it is more or less clear that $supF(2|1)$ also rejects the null hypothesis, offering evidence for two breaks. Finally, the low value of $supF(3|2)$ does not reject anymore. Hence, we may conclude that for our sample, we have one or two breaks. As a

[5] This is the test of zero against a fixed number of breaks m, in this case $m = 1$.

[6] This is the sequential test of ℓ versus $\ell + 1$ breaks; here, $\ell = 0$.

robustness check, we calculate the AIC and BIC for linear, one-transition, two-transition STAR with no breaks, and STAR with breaks[7] as follows:

$$\text{AIC}_s = \log \frac{SSR}{T} + \frac{2p_s}{T}$$
$$\text{BIC}_s = \log \frac{SSR}{T} + \frac{p_s \log(T)}{T}$$
$$\text{AIC}_{us} = \log \frac{SSR}{T} + \frac{2[(m+1)p_{us} + m]}{T}$$
$$\text{BIC}_{us} = \log \frac{SSR}{T} + \frac{[(m+1)p_{us} + m]\log(T)}{T}$$

where subscripts s, us refer to models with or without breaks, at $m = 1, 2, 3$. The results in Table 11 indicate that the worst performing models in terms of both AIC and BIC is the linear model, followed by the one-transition three breaks model as indicated by BIC. The model with one transition and two breaks is superior to all others in terms of both BIC and AIC, and the second best is the one transition model with one break. We also find that the latter two models outperform the two-transition model with one state variable time. However, note that the two transition model is not the best fit of the data, but rather a model where we impose the transitions to be similar to the ones employed by Kesriyeli, Osborn, and Sensier (2004), except for a time threshold at 1984 : 9, as indicated from our estimation. This serves our purposes of comparing a one-transition STAR with breaks to a two-transition STAR with no breaks.

In terms of break-point estimates, we obtain, as mentioned above, the first break at $\hat{T}_1 = 29$, that is 1984:9. This is similar (but not exactly the same) to Kesriyeli, Osborn, and Sensier's (2004) who find a break at the beginning of 1985. To check the robustness of our results, we ran a global search for one break for the first $60, 70, 80, 90, 100$ observations in our sample. Each time, the break-point estimate is exactly 1984:9, hence a robust finding. Next, we impose this break and search for an additional break in the second part of the sample. We find the second break at 1986:12, indicating evidence of a new regime at the beginning of 1997. The estimates

[7]The latter BIC is suggested by Bai and Perron (1998) in the companion GAUSS code to their paper.

for the model with two breaks are not reported for brevity purposes, but they are quite similar to the regime from 1984:10 - 1986:12 with the exception of the inflation gap coefficient.

Next, we provide a brief interpretation of the one-transition one-break model. Figure 1 and 2 plot the transition functions for the two sub-regimes (there, $F(s) = G_1(\Delta_3 r_{t-1}; \hat{\gamma}_1^{(i)}, \hat{c}_1^{(i)})$ for each sub-sample $i = 1, 2$). Figure 2 indicates a similar transition function as obtained Kesriyeli, Osborn, and Sensier (2004), and indeed the threshold value of the interest rate is similar. Figure 1 indicates that the transition is only 5 periods, and a small negative $\hat{\gamma}_1^{(1)}$ indicates that a linear fit may work better for the first part of the sample. In any case, as reported in Table 9, the parameters are quite different in the two sub-samples. Prior to 1984:9, we find that output gap has the right sign, and so is the case after 1984:9 except for a small period.

Finally, in Figure 3 we plot the true values of the interest rate against the fitted values from a two-transition model with time and from our model with one break. Overall, we can see a good fit of our model to the true interest rates. Unlike the two-transition model with large smoothness parameter, our model also fits the data between 1982:5-1984:9.

To summarize, we provided a short illustration of how to use STAR in the presence of breaks. We find that, even in the presence of nonlinearity, there is evidence of 2 breaks in the interest rate reaction function up to 1990:12, and these breaks are in 1984:9 and 1986:12.

Chapter 9

Conclusions

We considered a univariate NLS models with multiple parameter changes that occur at unknown dates. If the number of breaks is known, we showed that, similar to the linear model of Bai and Perron (1998), a minimization over all possible breaks and partitions can be used for estimating the change-points and parameters. Our contribution lies in combining the unstable linear framework with stable nonlinear asymptotic theory to yield an unstable nonlinear asymptotic framework for estimation, suitable for our problem. To our knowledge, this hasn't been done before for multiple breaks in any class of nonlinear models. We made fairly general assumptions about the data generation process, and maintained that both errors and regressors are correlated.

If the number of breaks is not known, we provided tests to detect it. Once it is inferred, we showed that the researcher can use the knowledge about the number of breaks and a sequential procedure to estimate their location and the parameters in different regimes. By means of nonlinear asymptotic theory, we showed that both the break-fraction and the parameter estimates are consistent and we derived their asymptotic distributions. We found that, while the break-fraction distributions are non-standard, the parameter estimates remain asymptotically normal and their distributions are the same as if the change-points were known. Knowledge of distributions helps the researcher in determining confidence intervals for the estimates, and we briefly describe methods for constructing these intervals. We also provided simulations that validate the properties of our estimates for shifts of moderate size.

Finally, our method can be a powerful tool for economic modeling. First, we suggest using our stability tests as a misspecification test for STAR models, since the latter are extensively used in macroeconomics. For example, there is widespread evidence of both nonlinearity and instability in the interest rate reaction functions. In our application, we tested a STAR specification of the monetary policy rule for breaks, and indeed found evidence of two breaks in 1984:9 and 1986:12.

From an empirical point of view, it is clearly desirable to further check the robustness of our estimates. There are many other theoretical issues on the agenda: extending this procedure to more general nonlinear models[1], multivariate models, nonstationary models. We would also like to provide tests that are more robust to serially correlated errors than the sup F-tests. These would include, for example, sup-Wald tests.

[1] Note that NLS is also equivalent to ML under certain conditions.

Bibliography

Andreou, E., and Ghysels, E. (2002). ‘Detecting Multiple Breaks in Financial Market Volatility Dynamics’, *Journal of Applied Econometrics*, 17: 579–600.

Andrews, D. W. K. (1988). ‘Laws of Large Numbers for Dependent Nonidentically Distributed Random Variables’, *Econometric Theory*, 4: 458–467.

——— (1993). ‘Tests for Parameter Instability and Structural Change with Unknown Change Point’, *Econometrica*, 61: 821–856.

——— (1994). ‘The Large Sample Correspondence between Classical Hypothesis Tests and Bayesian Posterior Odds Tests’, *Econometrica*, 62: 1207–1232.

——— (2003). ‘End-of-Sample Instability Tests’, *Econometrica*, 71: 1661–1694.

Andrews, D. W. K., and Fair, R. C. (1988). ‘Inference in Nonlinear Econometric Models with Structural Change’, *The Review of Economic Studies*, 55: 615–639.

Andrews, D. W. K., and Ploberger, W. (1994). ‘Optimal Tests When a Nuisance Parameter is Present Only Under the Alternative’, *Econometrica*, 62: 1383–1414.

Aue, A., Horváth, L., and Horváth, Z. (2006). ‘Discriminating between Level Shifts and Random Walks: A Delay Time Approach’, in H. M. and M. Janzura (eds.), *Conference Proceedings: Prague Stochastics*, pp. 73–80.

Bahadur, R. R. (1960). 'Stochastic Comparison of Tests', *Annals of Mathematical Statistics*, 31: 276–295.

Bai, J. (1997). 'Estimation of a Change Point in Multiple Regression Models', *Review of Economics and Statistics*, 79: 551–563.

Bai, J., and Perron, P. (1998). 'Estimating and Testing Linear Models with Multiple Structural Changes', *Econometrica*, 66: 47–78.

——— (2003a). 'Critical Values for Multiple Structural Change Tests', *The Econometrics Journal*, 6: 72–78.

——— (2003b). 'Computation and Analysis of Multiple Structural Change Models', *Journal of Applied Econometrics*, 18: 1–23.

Banerjee, A., Lumsdaine, R. L., and Stock, J. H. (1992). 'Recursive and Sequential Tests of the Unit-Root and Trend-Break Hypotheses: Theory and International Evidence', *Journal of Business and Economic Statistics*, 10: 271–287.

Banerjee, A., and Urga, G. (2005). 'Modeling Structural Breaks, Long Memory and Stock Market Volatility: An Overview', *Journal of Econometrics*, 129: 1–34.

Barry, D., and Hartigan, J. A. (1993). 'A Bayesian Analysis for Change Point Problems', *Journal of the American Statistical Association*, 88: 309–319.

Bec, F., Salem, M., and Collard, F. (2002). 'Asymmetries in Monetary Policy Reaction Function: Evidence for the US, French and German Central Banks', *Studies in Nonlinear Dynamics and Econometrics*, 6: Art. 3.

Bhattacharya, P. K. (1994). 'Some Aspects of Change-Point Analysis', in E. Carlstein, H.-G. Müller, and D. Siegmund (eds.), *Change-Point Problems*, vol. 23 of *IMS Lecture Notes Monograph Series*, pp. 28–56. Institute of Mathematical Statististics.

Bierens, H. J. (1982). 'A Uniform Weak Law of Large Numbers under φ-Mixing with Application to Nonlinear Least Squares Estimation', *Statistica Neerlandica*, 36: 81–86.

Billingsley, P. (1968). *Convergence of Probability Measures.* John Wiley & Sons, New York, 1st edn.

Brodsky, B. E., and Darkovsky, B. S. (1993). *Nonparametric Methods in Change-Point Problems*, Mathematics and Its Application. Kluwer Academic Publishers, Netherlands.

Brown, R. L., Durbin, J., and Evans, J. M. (1975). 'Techniques for Testing the Constancy of Regression Relationships over Time', *Journal of the Royal Statistical Society. Series B. Methodological*, 37: 149–192.

Caner, M. (2007). 'Boundedly Pivotal Structural Change Tests in Continuous Updating GMM with Strong, Weak identification and Completely Unidentified Cases', *Journal of Econometrics*, 137: 28–67.

Carrasco, M., and Chen, X. (2000). 'Beta-Mixing and Moment Properties of RCA Models with Application to GARCH(p,q)', *Comptes Rendus de l'Académie des Sciences. Série I. Mathématique*, 331: 85–90.

Chib, S. (1998). 'Estimation and Comparison of Multiple Change-Point Models', *Journal of Econometrics*, 86: 221–241.

Chow, G. C. (1960). 'Tests of Equality Between Sets of Coefficients in Two Linear Regressions', *Econometrica*, 28: 591–605.

Chu, C. S., and White, H. (1992). 'A Direct Test for a Changing Trend', *Journal of Business and Economic Statistics*, 10: 289–299.

Clarida, R., Galí, J., and Gertler, M. (1998). 'Monetary Policy Rules in Practice: Some International Evidence', *European Economic Review*, 42: 1033–1067.

——— (2000). ‘Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory’, *Quarterly Journal of Economics*, 115: 145–180.

Clarida, R., and Gertler, M. (1997). ‘How the Bundesbank Conducts Monetary Policy’, in C. Romer and D. Romer (eds.), *Reducing Inflation: Motivation and Strategy*, pp. 363–412.

Csörgö, M., and Horváth, L. (1997). *Limit Theorems for Change Point Analysis*. Chichester-Wiley.

Diebold, F. X., and Inoue, A. (2001). ‘Long Memory and Regime Switching’, *Journal of Econometrics*, 105: 131–159.

Dolado, J., María-Dolores, R., and Naveira, M. (2005). ‘Are Monetary-Policy Reaction Functions Asymmetric?: The Role of Nonlinearity in the Phillips Curve’, *European Economic Review*, 49: 485–503.

Dolado, J., María-Dolores, R., and Ruge-Murcia, M. (2004). ‘Nonlinear Monetary Policy Rules: Some New Evidence from US’, *Studies in Nonlinear Dynamics and Econometrics*, 8: Art. 2.

Dufour, J.-M., and Ghysels, E. (1996). ‘Editor‘s Introduction. Recent Developments in the Econometrics of Structural Change’, *Journal of Econometrics*, 70: 1–8.

Dufour, J.-M., Ghysels, E., and Hall, A. R. (1994). ‘Generalized Predictive Tests and Structural Change Analysis in Econometrics’, *International Economic Review*, 35: 199–229.

Eitrheim, Ø., and Teräsvirta, T. (1996). ‘Testing the Adequacy of the Smooth Transition Autoregressive Models’, *Journal of Econometrics*, 74: 59–75.

Fan, J., and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer Series in Statistics, New York.

Florio, A. (2006). ‘Asymmetric Interest Rate Smoothing: The Fed Approach’, *Economic Letters*, 93: 190–195.

Gallant, A. R. (1989). *Nonlinear Statistical Models*, Wiley Series in Probability and Mathematical Statistics. John Wiley and Sons, New York.

Gallant, A. R., and White, H. (1988). *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Basil Blackwell, Oxford, UK.

Ghysels, E., Guay, A., and Hall, A. R. (1998). 'Predictive Tests for Structural Change with Unknown Breakpoint', *Journal of Econometrics*, 82: 209–233.

Ghysels, E., and Hall, A. R. (1990). 'A Test for Structural Stability of Euler Conditions Parameters Estimated via the Generalized Method of Moments Estimator', *International Economic Review*, 31: 355–364.

Goldfeld, S. M., and Quandt, R. (1973). 'A Markov Model for Switching Regressions', *Journal of Econometrics*, 1: 3–16.

Granger, C. W., and Teräsvirta, T. (1993). *Modelling Nonlinear Economic Relationships*. McGraw-Hill, New York.

Granger, C. W. . J., and Hyung, N. (2004). 'Occasional Structural Breaks and Long Memory with an Application to the S&P 500 Absolute Stock Returns', *Journal of Empirical Finance*, 11: 399–421.

Hall, A. R., Han, S., and Boldea, O. (2007). 'Inference Regarding Multiple Structural Changes in Linear Models Estimated via 2SLS', *Discussion Paper, Department of Economics, North Carolina State University*.

Hall, A. R., and Sen, A. (1999). 'Structural Stability Testing in Models Estimated by Generalized Method of Moments', *Journal of Business & Economic Statistics*, 17: 335–348.

Hannan, E. J. (1971). 'Non-linear Time Series Regression', *Journal of Applied Probability*, 8: 767–780.

Hansen, B. E. (2000). 'Testing for Structural Change in Conditional Models', *Journal of Econometrics*, 97: 93–115.

Hansen, L. P. (1982). 'Large Sample Properties of Generalized Method of Moments Estimators', *Econometrica*, 50: 1029–1054.

Harrison, M. J., and McCabe, B. P. M. (1979). 'A Test for Heteroscedasticity Based on Ordinary Least Squares Residuals', *Journal of the American Statistical Association*, 74: 494–499.

Hillebrand, E. (2005). 'Neglecting Parameter Changes in GARCH Models', *Journal of Econometrics*, 129: 121–138.

Jennrich, R. I. (1969). 'Asymptotic Properties of Non-Linear Least Squares Estimators', *Annals of Mathematical Statistics*, 40: 633–643.

Judd, J., and Rudebusch, G. (1997). 'Taylor's Rule and the Fed: 1970-1977', in *Federal Reserve Bank of San Francisco Economic Review*, vol. 3.

Kesriyeli, M., Osborn, D., and Sensier, M. (2004). 'Nonlinearity and Structural Change in Interest Rate Reaction Functions for the US, UK and Germany', in *Discussion Paper*. The University of Manchester.

Kim, D., Osborn, D., and Sensier, M. (2005). 'Nonlinearity in the Fed's Monetary Policy Rule', *Journal of Applied Econometrics*, 20: 621–639.

Kim, D., and Perron, P. (2007a). 'Assessing the Relative Power of Structural Break Tetsts Using a Framework Based on the Approximate Bahadur Slope', in *Working Paper*. Boston University.

——— (2007b). 'Unit Root Tests Allowing for a Break in the Trend Function at an Unknown Time Under Both the Null and Alternative Hypotheses', in *Working Paper*. Boston University.

Kim, J., and Pollard, D. (1990). 'Cube Root Asymptotics', *The Annals of Statistics*, 18: 191–219.

Krämer, W., Ploberger, W., and Alt, R. (1988). 'Testing for Structural Change in Dynamic Models', *Econometrica*, 56: 1355–1369.

Krämer, W., and Sibbertsen, P. (2002). 'Testing for Structural Change in the Presence of Long Memory', *International Journal of Business and Economics*, 1: 235–242.

Krishnaiah, P. R., and Miao, B. Q. (1988). 'Review about Estimation of Change Points', in P. R. Krishnaiah and C. P. Rao (eds.), *Handbook of Statistics*, vol. 7, pp. 375–402. New York: Elsevier.

Lavielle, M., and Moulines, E. (2000). 'Least-Squares Estimation of an Unknown Number of Shifts in a Time Series', *Journal of Time Series Analysis*, 21: 33–59.

Lazarová, S. (2005). 'Testing for Structural Change in Regression with Long Memory Processes', *Journal of Econometrics*, 129: 329–372.

Lo, A. W., and Newey, W. K. (1985). 'A Large-Sample Chow Test for the Linear Simultaneous Equation', *Economics Letters*, 18: 351–353.

Lucas, R. (1976). 'Econometric Policy Evaluation: A Critique', in K. Brunner and A. Melzer (eds.), *The Phillips Curve and Labor Markets*, vol. 1 of *Carnegie-Rochester Conference Series on Public Policy*, pp. 19–46.

MacKinnon, J. G. (1989). 'Heteroskedasticity-Robust Tests for Structural Change', in *Econometrics of Structural Change*, Studies in Empirical Economics, pp. 13–28. Physica, Heidelberg.

Martin, C., and Milas, C. (2004). 'Modeling Monetary Policy: Inflation Targeting in Practice', *Economica*, 71: 209–221.

McLeish, D. L. (1975). 'A Maximal Inequality and Dependent Strong Laws', *Annals of Probability*, 3: 829–839.

Nelson, C. R., and Plosser, C. I. (1982). 'Trends and Random Walks in Macroeconomic Time Series', *Journal of Monetary Economics*, 10: 139–162.

Nyblom, J. (1989). 'Testing for the Constancy of Parameters over Time', *Journal of the American Statistical Association*, 84: 223–230.

Nyblom, J., and Mäkeläinen, T. (1983). 'Comparisons of Tests for the Presence of Random Walk Coefficients in a Simple Linear Model', *Journal of the American Statistical Association*, 78: 856–864.

Perron, P. (1989). 'The Great Crash, the Oil Price Shock and the Unit Root Hypothesis', *Econometrica*, 57: 1361–1401.

Perron, P., and Qu, Z. (2006). 'Estimating Restricted Structural Change Models', *Journal of Econometrics*, 134: 373–399.

——— (2007). 'Estimating and Testing Multiple Structural Changes in Multivariate Regressions', *Econometrica*, 75.

Phillips, G. D. A., and McCabe, B. P. (1983). 'The Independence of Tests for Structural Change in Regression Models', *Economic Letters*, 12: 283–287.

Picard, D. (1985). 'Testing and Estimating Change-Points in Time Series', *Advances in Applied Probability*, 17: 841–867.

Ploberger, W., and Krämer, W. (1992). 'The CUSUM Test with OLS Residuals', *Econometrica*, 60: 271–285.

Ploberger, W., Krämer, W., and Kontrus, K. (1989). 'A New Test For Structural Stability in the Linear Regression Model', *Journal of Econometrics*, 40: 307–318.

Potter, S. (1999). 'Nonlinear Time Series Modelling: An Introduction', *Journal of Economic Surveys*, 13: 505–528.

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.

——— (1960). 'Tests of the Hypothesis that a Linear Regression System Obeys Two Separate Regimes', *Journal of the American Statistical Association*, 55: 324–330.

Rissanen, J. (1989). *Stochastic Complexity in Statistical Inquiry*, vol. 15 of *World Scientific Series in Computer Science*. World Scientific Publishing, Teaneck, NJ.

Schaling, E. (2004). 'The Nonlinear Phillips Curve and Inflation Forecast Targeting: Symmetric versus Asymmetric Monetary Policy Rules', *Journal of Money, Credit and Banking*, 36: 361–386.

Sen, P. K. (1980). 'Asymptotic Theory of Some Tests for a Possible Change in the Regression Slope Occurring at an Unknown Time Point', *Probability Theory and Related Fields*, 52: 203–218.

Sowell, F. (1996). 'Optimal Tests for Parameter Instability in the Generalized Method of Moments Framework', *Econometrica*, 64: 1085–1107.

Teräsvirta, T. (1998). 'Modelling Economic Relationships with Smooth Transition Regression', in M. Dekker (ed.), *Handbook of Applied Economic Statistics*, pp. 507–552, New York.

Toyoda, T. (1974). 'Use of the Chow test under Heteroscedasticity', *Econometrica*, 42: 601–608.

van Dijk, D., Teräsvirta, T., and Franses, P. (2002). 'Smooth Transition Autoregressive Models: A Survey of Recent Developments', *Econometric Reviews*, 21: 1–47.

White, H., and Domowitz, I. (1984). 'Nonlinear Regression with Dependent Observations', *Econometrica*, 52: 143–161.

Wooldridge, J. (2006). *Introductory Econometrics - A Modern Approach*. Thomson, New York, 3rd edn.

Wright, J. H. (1998). 'Testing for a Structural Break at Unknown Date with Long-Memory Disturbances', *Journal of Time Series Analysis*, 19: 369–376.

Yao, Y.-C. (1987). 'Approximating the Distribution of the Maximum Likelihood Estimate of the Change-Point in a Sequence of Independent Random Variables', *The Annals of Statistics*, 15: 1321–1328.

Yao, Y.-C., and Au, S. T. (1989). 'Least-Squares Estimation of a Step Function', *Sankhyā (Statistics). The Indian Journal of Statistics. Series A*, 51: 370–381.

Zacks, S. (1983). 'Survey of Classical and Bayesian Approaches to the Change-Point problem: Fixed Sample and Sequential Procedures of Testing and Estimation', in *Recent Advances in Statistics*, pp. 245–269. Academic Press, New York.

Zivot, E., and Andrews, D. W. K. (1992). 'Further Evidence on the Great Crash, the Oil Price Shock, and the Unit Root Hypothesis', *Journal of Business and Economic Statistics*, 10: 251–270.

Appendix

Table 1: Break Points for $m = 1$, $\lambda_1^0 = 0.4$ and large shifts

Break Points		
T	T_1^0	$\hat{T}_1$
50	20	20
100	40	40

Table 2: Parameters for $m = 1$, $\lambda_1^0 = 0.4$ and large shifts

Parameters		
T	θ_1^0	
	$+1$	-1
	$\hat{I}_1$	$\hat{I}_2$
	$\hat{\theta}_1$	
50	-0.97 (0.29)*	1.02 (0.23)
100	-1.00 (0.19)	1.02 (0.15)
T	θ_2^0	
	-10	10
	$\hat{I}_1$	$\hat{I}_2$
	$\hat{\theta}_2$	
50	-10.02 (0.23)	9.98 (0.29)
100	-9.99 (0.17)	9.97 (0.14)
T	θ_3^0	
	$+1$	-1
	$\hat{I}_1$	$\hat{I}_2$
	$\hat{\theta}_3$	
50	1.00 (0.01)	-1.00 (0.01)
100	1.00 (0.008)	-1.00 (0.006)

*estimated standard deviations

Table 3: Break Points for $m = 1$, $\lambda_1^0 = 0.4$ and small shifts

Parameters		
$\theta_1^0 = (0.1, 0.2)$	$\theta_2^0 = (-3, -1)$	$\theta_3^0 = (0.1, 0.2)$
Break Points		
T	T_1^0	$\hat{T}_1$
50	20	19.98
100	40	39.99
300	120	120

Table 4: Break Points for $m = 2$, $(\lambda_1^0, \lambda_2^0) = (0.4, 0.7)$ and large shifts

Break Points		
T	T_1^0, T_2^0	$\hat{T}_1, \hat{T}_2$
50	20, 35	20, 35
100	40, 70	40, 70

Table 5: Parameters for $m = 2$, $(\lambda_1^0, \lambda_2^0) = (0.4, 0.7)$ and large shifts

Parameters			
T	θ_1^0		
	-1	1	2
	$\hat{I}_1$	$\hat{I}_2$	$\hat{I}_3$
	$\hat{\theta}_1$		
50	-0.98 (0.13)	0.99 (0.15)	2.04 (0.15)
100	-1.02 (0.09)	1.01 (0.10)	2.01 (0.10)
T	θ_2^0		
	-10	10	-5
	$\hat{I}_1$	$\hat{I}_2$	$\hat{I}_3$
	$\hat{\theta}_2$		
50	-10.02 (0.15)	10.01 (0.07)	-5.05 (0.08)
100	-9.98 (0.10)	9.99 (0.04)	-5.01 (0.05)
T	θ_3^0		
	1	-1	2
	$\hat{I}_1$	$\hat{I}_2$	$\hat{I}_3$
	$\hat{\theta}_3$		
50	1.00 (0.006)	-1.00 (0.007)	1.99 (0.007)
100	1.00 (0.003)	-1.00 (0.004)	2.00 (0.004)

Table 6: Break Points for $m = 3$, $(\lambda_1^0, \lambda_2^0, \lambda_3^0) = (0.3, 0.5, 0.7)$ and large shifts

Break Points		
T	T_1^0, T_2^0, T_3^0	$\hat{T}_1, \hat{T}_2, \hat{T}_3$
100	30, 50, 70	30, 50, 70

Table 7: Parameters for $m = 3$, $(\lambda_1^0, \lambda_2^0, \lambda_3^0) = (0.3, 0.5, 0.7)$ and large shifts

Parameters				
T	θ_1^0			
	-1	1	2	-2
	$\hat{I}_1$	$\hat{I}_2$	$\hat{I}_3$	$\hat{I}_4$
	$\hat{\theta}_1$			
100	-0.96 (0.10)	0.93 (0.13)	1.99 (0.13)	-1.99 (0.10)
T	θ_2^0			
	-1	1	-5	5
	$\hat{I}_1$	$\hat{I}_2$	$\hat{I}_3$	$\hat{I}_4$
	$\hat{\theta}_2$			
100	-10.04 (0.04)	10.07 (0.05)	-4.97 (0.05)	4.99 (0.04)
T	θ_3^0			
	1	-1	2	-2
	$\hat{I}_1$	$\hat{I}_2$	$\hat{I}_3$	$\hat{I}_4$
	$\hat{\theta}_3$			
100	1.00 (0.004)	-1.00 (0.005)	2.01 (0.005)	-2.00 (0.004)

Table 8: Break Points for $m = 3$, $(\lambda_1^0, \lambda_2^0, \lambda_3^0) = (0.2, 0.5, 0.7)$ and small shifts

Parameters		
$\theta_1^0 = (0.3, 0.2, 0.4, 0.1)$	$\theta_2^0 = (3, 1, 2, 4)$	$\theta_3^0 = (0.4, 0.1, 0.3, 0.2)$
Break Points		
T	T_1^0, T_2^0, T_3^0	$\hat{T}_1, \hat{T}_2, \hat{T}_3$
100	20, 50, 70	19.91, 49.98, 69.95

Table 9: Contrasting STAR Estimates With and Without Breaks

	Kesriyeli	OLS	STAR One*	STAR Two*	STAR One
			$m = 0$	$m = 0$	$m = 1$
Estimates for $m = 0$ or $m = 1$, first regime starting in 1982:5					
int	2.437	0.702	0.143	0.986	10.724
r_{t-1}	1.098	1.240	1.483	1.130	-0.188
r_{t-2}	-0.375	-0.327	-0.507	-0.296	0.085
πg_{t-2}		-0.018	-0.014	0.092	0.295
og_{t-2}		0.025	-0.037	0.068	0.278
πg_{t-3}	-0.085				
Δwcp_{t-3}	17.63				
G_1	-2.549		22.474	0.037	-8.761
$r_{t-1}G_1$	0.289		-4.038	0.001	0.922
$r_{t-2}G_1$			1.387	-	-0.013
$\pi g_{t-2}G_1$			0.443	-	-0.146
$og_{t-2}G_1$	-0.181		-0.868	-	-0.234
$\pi g_{t-1}G_1$	0.055				
$\Delta wcp_{t-3}G_1$	-17.63				
$s_{t,1} = \Delta_3 r_{t-1}$					
γ_1	2.255		2.223	2.255	0.010
c_1	0.757		1.038	0.757	-2.15
G_2	1.034	-	-	-	-
$r_{t-1}G_2$	-0.731	-	-	0.085	-
$r_{t-2}G_2$	0.605	-	-	-0.064	-
$og_{t-1}G_2$	0.595	-	-	-0.071	-
$og_{t-2}G_2$	0.923	-	-	-0.005	-
$\pi g_{t-1}G_2$	-0.437	-	-	-	-
$og_{t-1}G_2$	-0.688	-	-	-	-
$s_{t,2} = t$					
γ_1	1082			1000	
c_1	14			29	
Estimates for $m = 1$, second regime starting in 1984:10					
int					1.303
r_{t-1}					1.227
r_{t-2}					-0.401
πg_{t-2}					0.049
og_{t-2}					-0.010
G_1					-2.149
$r_{t-1}G_1$					0.193
$r_{t-2}G_1$					0.057
$\pi g_{t-2}G_1$					-0.465
$og_{t-2}G_1$					0.690
$s_{t,1} = \Delta_3 r_{t-1}$					
γ_1					7.070
c_1					0.777

*Here, m indicates the number of breaks. The notation "STAR One" indicates a model with one transition , a "STAR Two with two transitions", *int* indicates the intercept, and "-" is used to indicate very small parameter values.

Table 10: Stability Tests and Critical Values

Critical Values*					
p	α	$p \times Sup\ F\ 0:1$	$Sup\ F\ 1\|0$	$Sup\ F2\ \|1$	$Sup\ F\ 3\|2$
p=8	0.10	22.92	22.92	25.15	26.38
	0.05	25.22	25.22	25.15	26.38
	0.025	27.21	27.21	29.01	30.09
	0.01	29.60	29.60	31.80	32.84
p=9	0.10	24.75	24.75	26.99	28.11
	0.05	27.08	27.08	29.10	30.24
	0.025	29.13	29.13	31.04	32.48
	0.01	31.66	31.66	33.47	34.60
p=9	0.10	26.13	26.13	28.40	29.68
	0.05	28.49	28.49	30.65	31.90
	0.025	30.67	30.67	32.87	34.27
	0.01	33.62	33.62	35.86	36.68
Our test statistics					
p=12		67.92	88.27	42.89	2.06
Potential Conclusions					
p=12	0.01	reject	reject	reject	don't reject

*Critical values are taken from Bai and Perron (1998), and α are confidence levels.

Table 11: AIC and BIC of the Estimated Models

Model	SSR	AIC	BIC
OLS	13.660	-1.934	-1.807
STAR One, $m = 3$	3.590	-2.385	-1.088
STAR Two*, $m = 0$	9.000	-2.159	-1.777
STAR One, $m = 0$	9.560	-2.156	-1.851
STAR One, $m = 1$	5.171	-2.521	-1.885
STAR One, $m = 2$	3.661	-2.616	-2.186

*See definitions in Table 6.

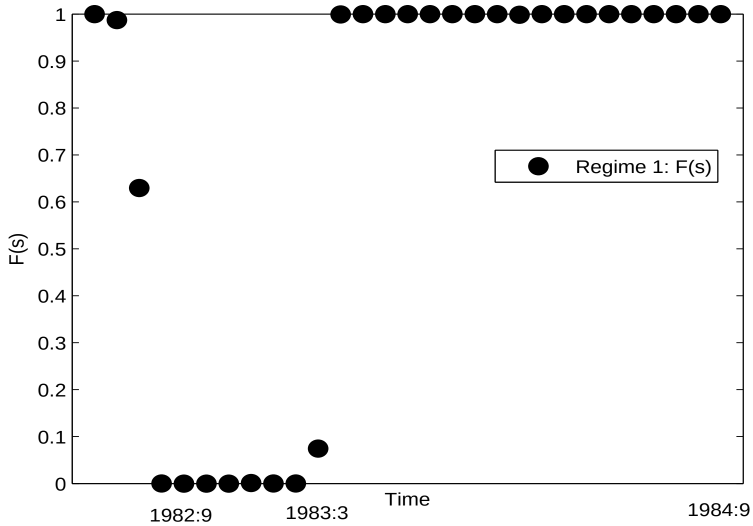

Figure 1: Transition Function 1982:5 - 1984:9

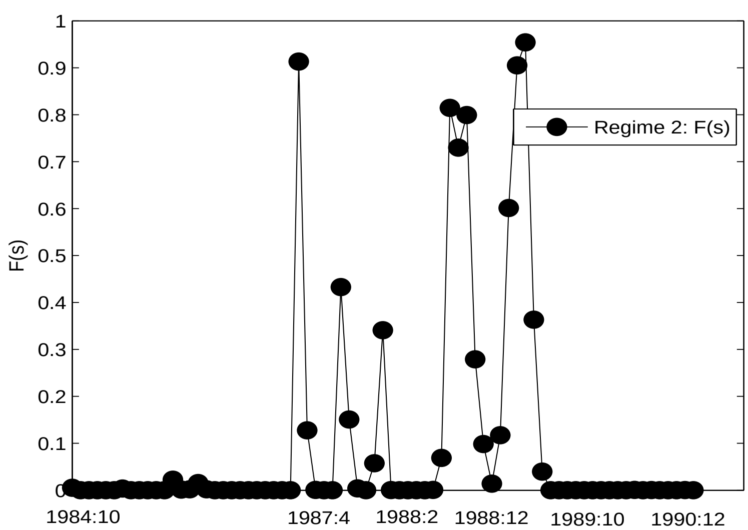

Figure 2: Transition Function 1984:10 - 1990:12

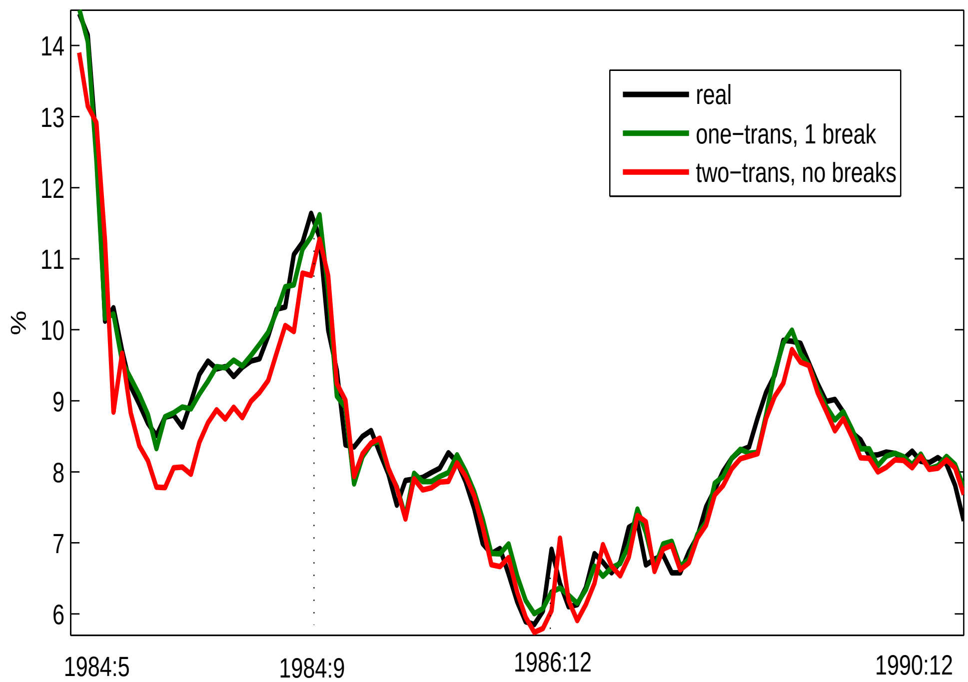

Figure 3: US Interest Rate Reaction Function