Long Memory and Structural Breaks in Time Series Models

Štěpána Lazarová

Ph.D. Thesis

Department of Economics
London School of Economics and Political Science

2006

This thesis has been submitted to the University of London in partial fulfilment of the requirements for the Ph.D. degree in Economics at the London School of Economics and Political Sciences (LSE).

No part of this doctoral dissertation had been presented to any university for any degree.

Chapter 3 was undertaken as joint work with Professor Javier Hidalgo.

Štěpána Lazarová

Štěpána Lazarová

Abstract

This thesis examines structural breaks in time series regressions where both regressors and errors may exhibit long range dependence. Statistical properties of methods for detecting and estimating structural breaks are analysed and asymptotic distribution of estimators and test statistics are obtained. Valid bootstrap methods of approximating the limiting distribution of the relevant statistics are also developed to improve on the asymptotic approximation in finite samples or to deal with the problem of unknown asymptotic distribution. The performance of the asymptotic and bootstrap methods are compared through Monte Carlo experiments. A background of the concepts of structural breaks, long memory and bootstrap is offered in Introduction where the main contribution of the thesis is also outlined. Chapter 1 proposes a fluctuation-type test procedure for detecting instability of slope coefficients. A first-order bootstrap approximation of the distribution of the test statistic is proposed. Chapter 2 considers estimation and testing of the time of the structural break. Statistical properties of the estimator are examined under a range of assumptions on the size of the break. Under the assumption of shrinking break, a bootstrap approximation of the asymptotic test procedure is proposed. Chapter 3 addresses a drawback of the assumption of fixed size of break. Under this assumption, the asymptotic distribution of the estimator of the breakpoint depends on the unknown underlying distribution of data and thus it is not available for inference purposes. The proposed solution is a bootstrap procedure based on a specific type of deconvolution.

Contents

List of Tables

Acknowledgements

I would like to thank Javier Hidalgo for his generous supervision during my doctoral studies at LSE, for the key contribution to a part of the thesis, and for inspiring me by having a high expectation of me since the beginning. I have also received advice and encouragement from Liudas Giraitis, Oliver Linton and Peter Robinson, for which I am very grateful.

Thanks also go to my friends and colleagues at the Department of Economics, in particular Jesús Maria Carro, David Jacho-Chávez, Xiaohong Chen, Piotr Eliasz, Javier Hualde, Fabrizio Iacone, Dennis Kristensen, José Sanz-Vidal, Marcia Schafgans and Ayan Sen, for making my years at LSE an enjoyable experience.

During my doctoral research and in the years leading to it, I greatly benefitted from stimulating discussions with Anindya Banerjee, Robert Chirinko, Saul Estrin, Paul Geroski, Stephen Hall, Marie Hušková, Chihwa Kao, Pietro Muliere, Lorenzo Trapani and Fabio Trojani. Special thanks to Giovanni Urga for his continuous encouragement and support.

I gratefully acknowledge financial support from the Dennis Sargan Memorial Fund and from the Economic and Social Research Council through grants R000238212 and R000239936. My studies were also made possible by the financial assistance through the following scholarships: Royal Economics Society Junior Fellowship, LSE Research Studentship, Department of Economics Award (LSE), Overseas Research Studentship, CK Hobson Scholarship (LSE) and British Council Chevening Scholarship.

Finally, an affectionate thanks to my family for their support.

Introduction

This thesis examines structural breaks in time series regressions. The main contribution to the literature in the field is twofold. First, statistical properties of methods for detecting and estimating structural breaks are analysed when both regressors and errors are allowed to exhibit long range dependence. Second, valid bootstrap methods of approximating the limiting distribution of relevant estimators are developed under possible long range dependence.

The principal keywords of the thesis are structural breaks, long memory and bootstrap. The following sections offer some background notes on these concepts and explain in what way this thesis adds to the body of knowledge in respective fields. The last section outlines the notation used throughout the thesis.

Structural breaks

Structural stability is a desirable property of any econometric model. Models that are structurally unstable tend to lead both to erroneous in-sample analysis and out-of-sample forecasts. Tests of parameter instability and structural change have therefore been a subject of a large body of statistical and econometric literature. The maintained hypothesis of parameter stability has been tested against both specific and general forms of alternative hypothesis.

When employed as a model-diagnostic tool, stability tests are frequently constructed against all possible functions describing the evolution of parameters over time. In the linear regression context, such tests are based on the behaviour of regression residuals, as in CUSUM tests of Brown et al. (1975) and Ploberger and Krämer (1990, 1992), or on the behaviour of parameter estimates, as in the fluctuation tests of Sen (1980) or Ploberger et al. (1989).

Alternatively, parameter stability tests can be designed against a specific alternative. Example of specific alternatives are one-time change in parameters as in the papers by Quandt (1960) or Andrews (1993), or parameters following random walk (Nyblom (1989)). Though constructed to detect a specific parameter behaviour, these tests are usually shown to have power against a broader range of departures from the null of parameter constancy.

Beside specification testing, the presence or absence of structural stability may be of interest in itself. If a structural change is detected, an inquiry into the character of the change may reveal factors that caused the structural shift and may lead to a successful revision of the original model. A prime example of this is the article of Perron (1989) who argues that many key economic variables should be modelled as stationary around a deterministic trends with breaks. Such models imply that the majority of shocks in economy are transitory and only few shocks have a permanent effect. The structural break model of Perron is an answer to the stochastic trend model of Nelson and Plosser (1982) which imply that all random shocks have a permanent effect on the economy. The work of Perron brought a change in the common view of the nature of dynamics of economic variables and inspired further investigation of instability in economic systems.

When it is known that the parameters of a model undergo a break, the knowledge of the date of break is often relevant to researchers, for example when judging a delay in reaction of agents to a change in economic policy. There is a steadily growing body of literature on estimating the time of change. Hinkley (1970), Yao (1987) and Bhattacharya (1987) deal with maximum likelihood estimation of time of a shift in mean of otherwise identically distributed independent observations. In the context of dependent observations, Bai (1994, 1997b) allows for a linear process with short memory while Bai (1997a), Bai and Perron (1998) and Fiteni (2002, 2004) analyse estimators of the time of break in parameters of linear regression model with mixing data.

An interesting observation is that in order to obtain a distribution-free asymptotic theory for estimators and test statistics, the magnitude of the structural change is assumed small in a majority of work on both detecting and locating structural change. More specifically, the size of change is assumed to decrease with increasing sample size. Examples of articles that adopt this assumption are Ploberger et al. (1989) and Andrews (1993) for detecting the structural change and Picard (1985) and Bai and Perron (1998) for estimating the date of break.

In the context of testing for parameter instability, this assumption may be regarded as innocuous since it can be argued that if a test procedure is capable of detecting small changes in the structure of the model, it will also be capable of detecting large changes. However, in the context of estimating the date of break, the assumption is no longer incontroversial. The gain in information due to the increase in the sample size is not sufficiently large to offset the loss of information due to the decrease in the magnitude of the break. The dispersion of the breakpoint estimator grows to infinity and tests of hypotheses about the date of change against fixed alternatives lose power in growing samples.

The only solution to this problem is to model the break as having a fixed size. Under the fixed break assumption, however, the limiting distribution of the location estimator depends on the distribution of the data and is therefore generally unknown and unavailable for the purposes of statistical inference. Hinkley (1970) attempts to circumvent the problem of intractability by assuming that the distribution of data is known. His method is difficult or impossible to implement in any but the most simple settings and in any case the assumption that the distribution of data is known is unrealistic. Since then, attempts to reconcile the advantages in assuming fixed breaks with the need for a tractable asymptotic distribution have been largely abandoned, with an exception of Antoch et al. (1995) who devise a bootstrap method for regression with independent identically distributed errors.

The current state of the research on structural changes in linear models with time series has been reviewed by Banerjee and Urga (2005) and Perron (2006). Other overviews of the work on structural breaks include the article by Stock (1994) and the special issue of the Journal of Econometrics on "Recent developments in the econometrics of structural change" edited by Dufour and Ghysels (1996).

Regarding models with strongly dependent data, the effect of long range dependence on estimators of time of break has been examined by Antoch et al. (1995, 1997), Horváth and Kokoszka (1997) and Kuan and Hsu (1998) in the framework of linear processes with a break in mean. Hidalgo and Robinson (1996) propose tests for a change in parameter values at a known time point in linear regression models with long-memory errors while Hidalgo (2003b) designs a test for the presence of breaks in nonparametric regression function with possibly strongly dependent errors.

This thesis

This thesis examines stability of slope coefficients in the linear regression model. An important distinction from the majority of existing literature is that we allow both regressors and errors to be possibly long range dependent. We are interested in two aspects of the problem of parameter stability. First, we examine methods of detecting structural instability and estimating the date of structural change. Second, we analyse in some detail the effect of the assumed size of break on statistical properties of estimators and test statistics, and attempt to resolve the difficulties arising from the imposition of the standard assumption of shrinking breaks.

Chapter 1 proposes a fluctuation-type method of testing for structural stability. The procedure is based on a process of least-squares slope coefficient estimators. The fluctuation of the process is measured by a continuous functional and the presence of instability is indicated by large fluctuation. Though the test is constructed to have power against the alternative of a structural break, it is shown to be powerful against a broader range of alternatives, such as multiple breaks, smooth transition between two steady levels of parameters and a continual change of parameter. The functional defining the test statistic can be chosen to reflect beliefs about the form of alternative and so improve the power of the test procedure.

The limiting distribution of the test statistics considered in the literature is typically a functional of Brownian motion. The main contribution of the Chapter 1 is to confirm that this fact remains true in linear model with stationary long memory series. This may seem a simple and straightforward conclusion but it is actually somewhat surprising since when dealing with long memory time series, fractional Brownian motion processes are usually expected to appear in expressions describing the asymptotic distribution. In our stochastic regressor framework, the effect may be viewed as a mutual stochastic dampening of regressors and errors where the individual series may exhibit long memory but their product displays short memory.

In Chapter 2, we examine a least squares method of estimating the time of the break. Here again the results from the short memory literature carry over to our long memory setting under the assumption of a break of both fixed and shrinking size. The magnitude of break, however, turns out to be crucial in determining the qualitative properties of the asymptotic distribution of the breakpoint estimator. One of the main contribution of Chapter 2 is therefore an analysis of the asymptotic behaviour of the estimator under various assumptions on the size of break, ranging from a fixed size of break through a size shrinking at a certain rate to zero size. While the assumptions of fixed and shrinking breaks have been examined in a variety settings with short memory data, and we extend the analysis to the long-memory time-series regression setting, the assumption of a weak break, that is break of a rapidly decreasing size, has not been analysed yet in the literature.

The conclusion is that when the size of the break is fixed, the asymptotic distribution depends on the entire joint distribution of the regressors and the error term. When the size of the break is shrinking but more slowly than the square root of the sample size, the asymptotic distribution of breakpoint is free of nuisance parameters and is explicitly known. When the size of the break is shrinking faster than the square root of the sample size, or when there is no break in the data generating process, the question of estimating the location of the break becomes vacuous because in this circumstance the break is not detectable. In the borderline case of the size of break decreasing with exactly the square root of the sample size, the break can be detected but there is insufficient information for estimating its location.

We argue that to obtain an efficient breakpoint estimator and a powerful test procedure, only the assumption of fixed break should be used. Since under this assumption the asymptotic distribution of the estimator of the breakpoint depends on the unknown underlying distribution of data and thus it is not available for inference purposes, a method of estimating the distribution is called for. One such procedure, based on the bootstrap, is proposed in Chapter 3.

Long memory

The phenomenon of the slow decay of correlation between observations that are far apart had been observed in various scientific fields since well over one hundred years ago. One of the first important statistical treatments of long memory has been in hydrology by Hurst (1951) who considered the rescaled adjusted range statistic and found its behaviour inconsistent with short range dependence assumption. In economics, one of the first to observe the long-memory properties of economic time series has been Adelman (1965) who observed peaks of estimated spectral functions around zero frequency. The peaked spectral density has been claimed by Granger (1966) to be the typical spectral shape of an economic variable.

For stochastic processes, the property of possessing long memory has been variously defined through the behaviour of the autocorrelation function as having hyperbolically decaying autocorrelations or having autocorrelations that are nor absolutely summable, through the behaviour of the spectral density as having a pole at zero frequency, or through the behaviour of the partial sums of the process as having variance that is increasing faster than the sample size. These definitions are closely related but not equivalent.

The degree of memory of a process may be described by the parameter d which we take to be the order of the singularity of the spectral density at zero. Estimation of d is a well-researched topic and a wide array of estimators is available to practitioners. Recent surveys of developments in long memory estimation and testing include Robinson (2003) and Banerjee and Urga (2005). Various results are collected in surveys by Robinson (1994a) and Baillie (1996), in a book by Beran (1994) and in a recent special issue of the Journal of Econometrics on "Long Memory and Non-Linear Time Series" edited by Davidson and Teräsvirta (2002).

It has long been known that certain classes of processes can mimic long memory behaviour. Among such processes are regime switching processes

(Diebold and Inoue (2001)), processes with certain type of deterministic trend (Bhattacharya et al. (1983)), error duration processes (Parke (1999)) and, importantly from our point of view, processes with structural breaks (Engle and Smith (1999) and Granger and Hyung (2004)). Accordingly, there is a growing body of literature on distinguishing genuine long memory time series from those with other features (Künsch (1986), Shimotsu (2005) and Berkes et al. (2006)).

There is also work that nests both long memory and some of the above features, but focuses on the analysis of one feature only. Iacone (2006) examines the degree of memory of a given series with possible presence of the nuisance deterministic components including broken trends. In the opposite direction, Hidalgo and Robinson (1996) allow for the presence of long memory but regard it as a nuisance phenomenon and concentrate on testing for structural breaks in the framework of linear regression. A similar approach is taken in articles by Antoch et al. (1995, 1997), Horváth and Kokoszka (1997) and Kuan and Hsu (1998) mentioned in the previous section.

This thesis

In this thesis, we are interested in structural breaks and view long memory as a nuisance. Our aim is to develop methods in which the user does not need to know the degree of memory of the data, as long as the data are stationary. As a result, we do not discuss the estimators of the long memory parameter and we only use the existing estimators.

In Chapters 1 and 2, we show that the classical least squares methods for detecting and locating breaks, devised originally under the assumption of no or short memory in regressors and errors, can be used without change under long memory, and the degree of memory does not need to be estimated. We also show that the statistical properties of the estimators remain unchanged. The only place where an allowance for possible presence of long memory needs to be made is when the user wishes to conduct a bootstrap test since it seems convenient to carry out the bootstrap procedure in the frequency domain.

In Chapter 3, the bootstrap procedure involves estimation of the degree of memory of regression residuals as a preliminary step. We suggest that the researcher uses the local Whittle estimator proposed by Robinson (1995b) and we show its consistency when the underlying series is replaced by residuals.

Bootstrap

When the distribution of an estimator or a test statistic is unknown or if it is difficult to calculate, it can be approximated by the bootstrap. Bootstrap methods can also be used to provide more accurate approximation of the finite sample distribution than the approximation obtained from first order asymptotic theory.

The core idea of the bootstrap is to replace the unknown distribution of a random variable by the empirical distribution of a random sample drawn from that distribution. However, when the data are not independent and identically distributed, the basic bootstrap of Efron (1979) is not valid and the bootstrap procedure needs to be modified to reflect the dependence or heterogeneity structure of the data. In the time series context, an early adaptation of the basic bootstrap method rests on the assumption that the data are generated by a finite-order stationary ARMA process with independent identically distributed innovations (Freedman (1984), Efron and Tibshirani (1986)). In a direction towards nonparametric methods, Bühlmann (1997, 1998) approximates the linear infinite-dimensional process by a sieve of finite-dimensional autoregressive processes whose order is growing with the sample size. Diebold et al. (1998) propose a purely nonparametric bootstrap method based on the Cholesky factorization.

A different way of approximately preserving the temporal dependence structure of the data is to resample blocks of data. Carlstein (1986) and Künsch (1989) propose to resample nonoverlapping and overlapping blocks of data, respectively, and to concatenate the blocks to generate a bootstrap sample. Politis and Romano (1992) introduce an idea of subsampling, regarding blocks of data – subseries – as new pseudo-samples.

A problem shared by nonparametric bootstrap methods is that they require an intervention by the researcher in choosing a dimension parameter of the procedure, be it lag length, bandwidth or block length. The performance of time-series bootstrap can be highly sensitive to the choice of the dimension parameter, particularly in samples of moderate size. Although automatic procedures for choosing the dimension have been devised for some methods, they can be computationally expensive.

Nonparametric bootstrap procedures can alternatively be carried out in the frequency domain where either frequency domain data, that is the discrete Fourier coefficients, or their squares, that is the periodograms, can be bootstrapped. This approach is motivated by the observation that converting a stochastic process from the time domain to the frequency domain reduces serial correlation of the process though it induces heteroskedasticity. Bootstrap method of Ramos (1988) for Fourier coefficients or Franke and Härdle (1992) and Dahlhaus and Janas (1996) for periodograms require a consistent estimate of the spectral density and therefore a choice of a bandwidth. Local periodogram bootstrap of Paparoditis and Politis (2000) avoids the need for estimating the spectrum but again demands a bandwidth choice.

Hidalgo (2003a) proposes a method that eliminates the choice of lag length or bandwidth. He suggests to bootstrap OLS residuals in frequency domain. His bootstrap procedure is easy to implement and computationally inexpensive. His approach is the only one among the methods cited so far that has been shown to be valid for strongly dependent data.

This thesis

One of the goals of this thesis is to design bootstrap procedures that are valid for short as well as long memory time series. As with parameter estimation, we aim to avoid the need for the researcher to know or estimate the degree of memory of the data, as long as the data are stationary. We first propose a bootstrap procedure that is useful for approximation of the distribution both of test statistics for detecting the break in Chapter 1 and of the estimator of the date of break under shrinking break in Chapter 2.

The main idea of the proposed bootstrap procedure is to transform a given series into the frequency domain and thereby to translate the problem of dependent bootstrap to a problem of heteroskedastic bootstrap. The quantity to be resampled are the scaled frequency domain regression residuals. The heteroskedasticity is accounted for by re-scaling the resampled values.

The procedure is essentially that of Hidalgo (2003a). However, to prove the validity of the method in the context of time series regression with structural breaks, it is necessary to show that the method is successfully estimating not only the distribution of the normalized sum of a time series, but also the distribution of the entire process of its partial sums. The relevant concept here is the bootstrap weak convergence of Giné and Zinn (1990). The proof of bootstrap weak convergence of the partial-sum process is one of the main contributions of the thesis.

The proposed procedure inherits the advantage of the original method of Hidalgo (2003a) of not requiring a user-chosen parameter such as the block length in the block bootstrap of Carlstein (1986) or the lag length in the sieve bootstrap of Bühlmann (1997, 1998).

While the bootstrap inference procedure in Chapters 1 and 2 is an optional and advantageous alternative to asymptotic inference procedures, in Chapter 3 the limiting distribution of the breakpoint estimator depends on the unknown joint distribution of data and the use of the bootstrap or some other estimating procedure becomes a necessity if inference is to be carried out. The bootstrap procedure proposed in Chapters 1 and 2 asymptotically matches the covariance structure of the underlying process. The ability to estimate the second moment dependence structure is sufficient for approximating distributions that are entirely described by the second-order structure, for example the Gaussian distribution. However, it does not suffice for estimating a general joint distribution of a process.

In Chapter 3, we therefore design a more refined bootstrap method. The idea behind the bootstrap procedure is to fractionally difference the series in question and to approximate the resulting short memory process by an autoregressive process. The values to be resampled here are the estimated innovations of the process, and the bootstrap sample is created by refiltering and fractionally integrating the resampled innovations.

We find it convenient to execute both stages of prewhitening, that is fractional differencing and filtering, in the frequency domain. After obtaining a preliminary estimate of the memory parameter d, an estimate of the linear coefficients of the fractionally differenced process is achieved via the canonical spectral decomposition of a smoothed estimate of the spectral density corresponding to the differenced process.

In contrast to the bootstrap procedure in Chapters 1 and 2, estimation of d cannot be avoided though d remains a nuisance parameter. Moreover, the user needs to choose a bandwidth parameter for smoothing and a number of lags of the truncated linear process. Interestingly, however, it turns out that the two parameters are directly related so that effectively the user chooses only one bandwidth parameter, and the value of this parameter can be determined by a cross-validation procedure. In return for the additional user input, the bootstrap delivers approximation of finite-dimensional joint distributions of the process.

To our knowledge, there is currently no bootstrap procedure available that approximates the joint distribution of data while allowing for strong serial dependence. Construction of such a bootstrap procedure is therefore one of our main contributions to the literature.

Notation

Throughout the thesis, W denotes a p-dimensional vector of independent standard Brownian motion processes on $[0,1]$ or on a set $\Lambda \subset (0,1)$, "$\Longrightarrow$" denotes weak convergence in the space $D(\Lambda)^p$ of p-vectors of right-continuous functions with left-hand limits, endowed with the uniform metric $\rho(x,y) = \sup_{\tau\in\Lambda} \|x(\tau) - y(\tau)\|$ for $x, y \in D(\Lambda)^p$. The statement $y_T \sim x_T$ is equivalent to the statement $\frac{y_T}{x_T} \to 1$ as $T \to \infty$. For σ-algebras $\mathcal{F}$, $\mathcal{G}$, $\mathcal{F} \vee \mathcal{G}$ is their union, that is the smallest σ-algebra containing all elements of $\mathcal{F}$ and $\mathcal{G}$.

For any real numbers a and b, $a \vee b = \max\{a,b\}$ and $a \wedge b = \min\{a,b\}$. For any integers j and k, $|j-k|_+ = \max\{1, |j-k|\}$. For a set S and a constant a, $S \cdot a = \{xa : x \in S\}$. For nonnegative numbers l, m,

$$\sum_t^{l,m} a_t = \begin{cases} \sum_{t=l+1}^{m} a_t & l < m, \\ 0 & l = m, \\ \sum_{t=m+1}^{l} a_t & l > m. \end{cases}$$

For integers j, k and l, we write $j = k \mod l$ if $j - k$ is divisible by l. Notation

$[\cdot]$ signifies integer part and $\mathbb{I}(\cdot)$ is the indicator function of a set.

For a Hermitian matrix A, $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the smallest and the largest eigenvalue of A, respectively. Inequalities $A \geq B$ and $A > B$ among two matrices hold if all the eigenvalues of $A - B$ are nonnegative and positive, respectively. For any matrix A, $\|\cdot\|$ denotes the maximum-eigenvalue norm, that is $\|A\| = \sup_{\|x\|=1} \|Ax\| = \lambda_{\max}^{1/2}(A'A)$. We have $\|A\|^2 \leq \operatorname{tr} A'A$ and due to the equivalence of norms also $\operatorname{tr} A'A \leq C\|A\|^2$ for a constant $C > 0$.

For a generic function φ, we denote $\varphi_j = \varphi(\lambda_j)$, where $\lambda_j = 2\pi j/T$, $j = 1, \ldots, T$ are Fourier frequencies. For sequences $\{a_t\}_{t=1}^T$ and $\{b_t\}_{t=1}^T$ of p-dimensional vectors,

$$w_a(\lambda) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} a_t e^{it\lambda}, \qquad \lambda \in [-\pi, \pi],$$

is the discrete Fourier transform of $\{a_t\}$ and

$$I_{ab}(\lambda) = w_a(\lambda)\, \bar{w}_b'(\lambda), \qquad \lambda \in [-\pi, \pi],$$

is the cross-periodogram matrix of $\{a_t\}$ and $\{b_t\}$. Notation f_{vv} is reserved for the spectral density of a process $\{v_t\}$.

Starred notation in $\hat{k}^*$, $\hat{u}^*$, E^*, O_{p^*} and similar refers to quantities conditional on data, taken with respect to the corresponding bootstrap probability measure. In particular, notation P^* distinguishes the probability conditional on the σ-algebra $\mathcal{F}_T \vee \mathcal{G}_T$. For example,

$$P^*\left(|\hat{k}^* - \hat{k}| \leq x\right) = P\left(|\hat{k}^* - \hat{k}| \leq x \middle| \mathcal{F}_T \vee \mathcal{G}_T\right).$$

Similarly, E^*, var^* and cov^* denote expectation, variance and covariance conditional on $\mathcal{F}_T \vee \mathcal{G}_T$, respectively. For a random variable X and a sequence $\{X_T\}$ of random variables, the statement $X_T \xrightarrow{d^*} X$ is equivalent to the statement

$$P^*(X_T \leq x) \xrightarrow{p} P(X \leq x) \qquad \text{as } T \to \infty$$

for each x which is a continuity point of $F(x) = P(X \leq x)$. When the limiting variable X is a constant, we write $X_T \xrightarrow{p^*} X$. Further, for a stochastic process

Y and a sequence of stochastic processes $\{Y_T\}$, $Y_T \overset{p}{\Longrightarrow} Y$ stands for the weak convergence in probability as defined by Giné and Zinn (1990).

Stochastic orders of magnitude o_{p^*}, O_{p^*} are defined as follows. Let $\{\varphi_T\}$ be a sequence of positive finite numbers. We say that $X_T = O_{p^*}(\varphi_T)$ as $T \to \infty$ if and only if for every $\varepsilon > 0$ and $\eta > 0$ there exist finite M and T_0 such that for all $T \geq T_0$,

$$P\left(P^*\left(|X_T| > M\varphi_T\right) > \eta\right) < \varepsilon.$$

We say that $X_T = o_{p^*}(\varphi_T)$ as $T \to \infty$ if and only if for every $\eta > 0$,

$$P^*\left(|X_T| > \eta\varphi_T\right) = o_p(1).$$

It is easy to verify some useful relations for the orders of magnitude o_{p^*} and O_{p^*}. For example, $o_{p^*}(1) \cdot O_p(1) = o_{p^*}(1)$ or $o_{p^*}(1) + o_p(1) = o_{p^*}(1)$.

Finally, C and D stand for generic constants. Unless specified otherwise, all limits are taken as $T \to \infty$, where T is the sample size.

Chapter 1

Testing for structural change in regression with long memory processes

1.1 Introduction

Parameter instability and structural change have been a subject of a large body of statistical and econometric literature. The maintained hypothesis of parameter stability has been tested against both specific and general forms of alternative hypothesis. When employed as a model-diagnostic tool, stability tests are constructed against all possible functions describing the evolution of parameters over time. Such tests are based on the behaviour of regression residuals, as in CUSUM tests of Brown et al. (1975) and Ploberger and Krämer (1990, 1992), or on the behaviour of parameter estimates, as in the fluctuation tests of Sen (1980) or Ploberger et al. (1989).

Alternatively, parameter stability tests can be designed against a specific alternative. Example of specific alternatives are one-time change in parameters as in the papers by Quandt (1960) or Andrews (1993), or parameters following random walk (Nyblom (1989)). Though constructed to detect specific parameter behaviour, these tests are usually shown to have power against a broader range of departures from the null of parameter constancy.

This chapter considers tests for stability in slope coefficients in linear regression model where both regressors and errors are allowed to be long range dependent. The main contribution of the chapter is twofold. First, the limiting distribution of the test statistics considered in the literature is typically a functional of Brownian motion. It is shown that this remains true for test statistics based on the slope coefficient estimator in linear model with stationary long memory series. Secondly, as an alternative to computing the critical values for the test statistic, a first-order bootstrap approximation of the distribution of the test statistic is proposed and the validity of the bootstrap procedure is shown.

The chapter is organized as follows. Section 1.2 describes the model and the hypotheses of interest and states distributional results for the test statistic. Section 1.3 proposes a bootstrap approximation of the testing procedure and shows its validity. Section 1.4 offers a Monte Carlo study of the small sample performance of the bootstrap testing procedure. Section 1.5 concludes. The proofs of the results stated in the text are gathered in Section 1.A. Section 1.B contains some auxiliary results.

1.2 Model and asymptotic results

We are interested in testing for structural change in regression models with processes that may possess long memory. We consider the model

$$y_t = \alpha + \beta_t' x_t + u_t, \qquad t = 1, \ldots, T, \tag{1.1}$$

where y_t is the observed dependent variable, α is an unknown intercept, β_t is a p-dimensional vector of unknown parameters, x_t is a p-dimensional vector of the explanatory variables and u_t is an unobserved stochastic disturbance. Our hypothesis of interest is whether the parameter vector β_t stays constant,

$$\mathrm{H}_0\colon \quad \beta_t = \beta \quad \text{for some } \beta, \text{ for all } t = 1, \ldots, T.$$

The alternative is that of general parameter instability,

$$\mathrm{H}_1: \quad \beta_t \neq \beta_s \quad \text{for some } 1 < t, s < T.$$

Test procedures for the hypothesis of structural stability of general models are based on test statistics that can be written as

$$Z_T = \phi\left(E_T\right),$$

where E_T is a stochastic process on $[0,1]$ or its subset with values in the space of right-continuous functions with left-hand limits and ϕ is a continuous functional. The process E_T is based on an estimator of parameters of a given model and its form reflects the choice of the testing principle. For example, if $\{e_t, p \leq t \leq T\}$ is the sequence of cumulative recursive residuals from the OLS estimates of the model (1.1) under the null as in the CUSUM test procedure of Brown et al. (1975), the stochastic process E_T can be defined as $E_T = \left\{E_T\left(\tau\right) = e_{[\tau T]}, p/T \leq \tau \leq 1\right\}$. Further examples of processes considered in the literature are Wald-, LM- and LR-like test statistic processes of Andrews (1993), CUSUM of squares process of Brown et al. (1975), OLS CUSUM process of Ploberger and Krämer (1992), OLS parameter estimates process of Ploberger et al. (1989) and Sen (1980) or MOSUM process of Chu et al. (1994).

The functional ϕ measures the excess fluctuation of the process E_T with respect to its hypothesized fluctuation. Depending on the belief about the form of the alternative, the functional ϕ can be chosen to obtain good power of the test. A functional widely used in literature is the supremum functional. The test statistic can also be based on the L_q-distance like Cramér-von Mises test statistic with $q = 2$. The range functional, that is the difference between the maximum and the minimum of a function, can have power advantage over the supremum functional in detecting smaller fluctuations of a process which changes its sign, as argued by Kuan and Hornik (1995). The average exponential functional of Andrews and Ploberger (1994) is shown to enjoy asymptotic optimality with respect to a weighted average power criterion.

In this chapter, we base the test procedure on the OLS estimators of the coefficient δ in the model

$$y_t = \alpha + \beta' x_t + \delta' z_t + u_t, \qquad t = 1, \ldots, T, \tag{1.2}$$

where

$$z_t = z_t(\tau) = \begin{cases} x_t & t \leq [\tau T], \\ 0 & \text{otherwise}, \end{cases} \tag{1.3}$$

where δ is a p-dimensional vector of parameters and where τ lies in a subset Λ of $(0,1)$. In the interest of clarity, the explicit notation of dependence of z_t on τ is sometimes dropped in what follows. The choice $\Lambda = (0,1)$ appears natural but for technical reasons the set Λ needs to be restricted to have closure in $(0,1)$. The grounds for the restriction are discussed after stating Theorem 1.2 and its Corollary 1.1 below. In addition to technical reasons, there may be other motives for restricting the set Λ away from $(0,1)$. It may be suspected that the instability in question occurred in a specific subperiod of a given period. For example, if data for postwar productivity growth are examined, the attention might be focused on testing for an abrupt or gradual change in a period around the 1973 oil price shock.

For any fixed $\tau \in \Lambda$, the OLS estimator of the parameters β and δ in (1.2) is given by

$$\begin{pmatrix} \hat{\beta}(\tau) \\ \hat{\delta}(\tau) \end{pmatrix} = \begin{pmatrix} \sum_{t=1}^{T} (x_t - \bar{x}) x_t' & \sum_{t=1}^{T} (x_t - \overline{x}) z_t' \\ \sum_{t=1}^{T} (z_t - \bar{z}) x_t' & \sum_{t=1}^{T} (z_t - \bar{z}) z_t' \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^{T} (x_t - \bar{x}) y_t \\ \sum_{t=1}^{T} (z_t - \bar{z}) y_t \end{pmatrix}, \tag{1.4}$$

where $\bar{x} = T^{-1} \sum_{t=1}^{T} x_t$ and $\bar{z} = T^{-1} \sum_{t=1}^{[\tau T]} x_t$. Alternatively, model (1.2) can be translated into the frequency domain, becoming

$$w_y(\lambda_j) = \beta' w_x(\lambda_j) + \delta' w_z(\lambda_j) + w_u(\lambda_j), \qquad j = 1, \ldots, T-1. \tag{1.5}$$

Identifying $w_x(\lambda_j)$ and $w_z(\lambda_j)$ as regressors and $w_u(\lambda_j)$ as an error term, the OLS estimate of the parameters β and δ in (1.5) for $\tau \in \Lambda$ is given by

$$\begin{pmatrix} \hat{\beta}(\tau) \\ \hat{\delta}(\tau) \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{T-1} I_{xx}(\lambda_j) & \sum_{j=1}^{T-1} I_{xz}(\lambda_j) \\ \sum_{j=1}^{T-1} I_{zx}(\lambda_j) & \sum_{j=1}^{T-1} I_{zz}(\lambda_j) \end{pmatrix}^{-1} \begin{pmatrix} \sum_{j=1}^{T-1} I_{xy}(\lambda_j) \\ \sum_{j=1}^{T-1} I_{zy}(\lambda_j) \end{pmatrix}. \tag{1.6}$$

Leaving out the zero frequency from the frequency domain regression is equivalent to mean-correcting data before running the regression in the time domain. The estimators defined in (1.4) and (1.6) are therefore identical. Omission of the zero frequency permits inference on the slope parameters when the intercept is unknown. It is worth noting that due to the symmetry of the periodograms, (1.6) is equal to

$$\begin{pmatrix} \hat{\beta}(\tau) \\ \hat{\delta}(\tau) \end{pmatrix} = \left(\operatorname{Re} \begin{pmatrix} \sum_{j=1}^{[T/2]} I_{xx}(\lambda_j) & \sum_{j=1}^{[T/2]} I_{xz}(\lambda_j) \\ \sum_{j=1}^{[T/2]} I_{zx}(\lambda_j) & \sum_{j=1}^{[T/2]} I_{zz}(\lambda_j) \end{pmatrix} \right)^{-1} \times \operatorname{Re} \begin{pmatrix} \sum_{j=1}^{[T/2]} I_{xy}(\lambda_j) \\ \sum_{j=1}^{[T/2]} I_{zy}(\lambda_j) \end{pmatrix} \tag{1.7}$$

for T odd. When T is even, (1.7) differs from (1.6) only by the order of $O_p(1/T)$.

For each τ from a set $\Lambda \subset (0,1)$, an estimator $\hat{\delta}(\tau)$ of δ can be obtained from (1.6) and a process $\hat{\delta}$ can be defined as $\hat{\delta} = \left\{\hat{\delta}(\tau), \tau \in \Lambda\right\}$. For any T and any realization of processes $\{x_t\}$ and $\{u_t\}$ the function $\hat{\delta}$ is bounded and constant on the subintervals $[j/T, (j+1)/T) \cap \Lambda$, $j \in N$, and the process $\hat{\delta}$ is a random element of the space $D(\Lambda)^p$ of $p \times 1$ vectors of right-continuous functions on Λ with left-hand limits endowed with uniform metric.

The test statistic based on the process $\hat{\delta}$ is then $Z_T = \phi\left(\sqrt{T}\hat{\delta}\right)$ for any continuous functional ϕ: $D(\Lambda)^p \mapsto R$. For example, the Kolmogorov-Smirnov (or Bartlett) test statistic is defined as

$$\mathrm{KS}_T = \sup_{\tau \in \Lambda} \sqrt{T} \left|\left|\hat{\delta}(\tau)\right|\right|$$

and the Cramér-von Mises statistic is given by

$$\mathrm{CvM}_T = \int_\Lambda T \left|\left|\hat{\delta}(\tau)\right|\right|^2 d\tau.$$

Under the null hypothesis, the additional regressor z_t has no explanatory power and the process $\hat{\delta}$ is uniformly close to zero, whereas under the alternative, $\hat{\delta}$ can be expected to differ significantly from zero on a set $\Lambda_1 \subset \Lambda$ of Lebesgue measure greater than zero. The norm functionals like KS and CvM constitute one-tailed tests, rejecting H_0 for large values of the test statistic. In principle, two-tailed tests can be constructed for functionals whose range includes both positive and negative values.

It can be expected that the test procedure based on model (1.2) has power mainly against one-time break alternatives of the form

$$\text{H}_1\text{:}\quad \beta_t = \begin{cases} \beta+\delta & t = 1,\dots,[\tau_0 T] \\ \beta & t = [\tau_0 T]+1,\dots,T \end{cases} \tag{1.8}$$

for some $\tau_0 \in \Lambda$ and some constants β and δ with $\delta \neq 0$, but we show that our test procedure has power under a broader range of alternatives.

In our analysis, we assume that $\{x_t\}$ and $\{u_t\}$ are covariance stationary linear processes that satisfy the following conditions.

Condition 1.1

$$x_t = \sum_{j=0}^{\infty} a_j \xi_{t-j}, \quad \sum_{j=0}^{\infty} ||a_j||^2 < \infty, \quad a_0 = I,$$

$$u_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} b_j^2 < \infty, \quad b_0 = 1.$$

Let $\mathcal{F}_t$ and $\mathcal{G}_t$ be the σ-algebras of events generated by ξ_s, $s \leq t$, and ε_s, $s \leq t$, respectively.

Condition 1.2 $\{\xi_t\}$ *is a stochastic process that satisfies*

1. $E\left(\xi_t|\mathcal{F}_{t-1} \vee \mathcal{G}_t\right) = 0$ *a.s.,*
2. $E\left(\xi_t\xi_t'|\mathcal{F}_{t-1} \vee \mathcal{G}_t\right) = E\left(\xi_t\xi_t'\right) = \Xi$ *a.s., and*
3. *the joint fourth cumulants of* $\xi_{t_i j_i}$, $j_i = 1,\dots p$ *and* $i = 1,\dots 4$, *where* ξ_{tj} *denotes the* j*-th component of the vector* ξ_t, *satisfy*

$$cum\left(\xi_{t_1 j_1}, \xi_{t_2 j_2}, \xi_{t_3 j_3}, \xi_{t_4 j_4}|\mathcal{G}_T\right) = \begin{cases} \kappa_{\xi,j_1,j_2,j_3,j_4} \ a.s. & t_1 = t_2 = t_3 = t_4, \\ 0 \ a.s. & otherwise, \end{cases}$$

with $|\kappa_\xi| = \max_{j_i=1,\dots p, i=1,\dots 4} |\kappa_{\xi,j_1,j_2,j_3,j_4}| < \infty$.

Condition 1.3 $\{\varepsilon_t\}$ *is a stochastic process that satisfies*

1. $E\left(\varepsilon_t|\mathcal{F}_t \vee \mathcal{G}_{t-1}\right) = 0$ *a.s.,*

2. $E\left(\varepsilon_t^2|\mathcal{F}_t \vee \mathcal{G}_{t-1}\right) = E\left(\varepsilon_t^2\right) = \sigma_\varepsilon^2$ *a.s., and*

3. *the joint fourth cumulant of* ε_{t_i}, $i = 1, \ldots 4$ *satisfies*

$$cum\left(\varepsilon_{t_1}, \varepsilon_{t_2}, \varepsilon_{t_3}, \varepsilon_{t_4}|\mathcal{F}_T\right) = \begin{cases} \kappa \text{ a.s.} & t_1 = t_2 = t_3 = t_4, \\ 0 \text{ a.s.} & \text{otherwise,} \end{cases}$$

with $|\kappa| < \infty$.

Condition 1.4 *The functions*

$$A\left(e^{i\lambda}\right) = \sum_{j=0}^{\infty} a_j e^{ij\lambda} \text{ and } B\left(e^{i\lambda}\right) = \sum_{j=0}^{\infty} b_j e^{ij\lambda}$$

satisfy the following assumptions:

1. *there exist constants* $0 < C_{x,k}, C_u < \infty$ *and* $d_{x,k}, d \in \left[0, \frac{1}{2}\right)$, $k = 1, 2, \ldots, p$, *such that* $|A_{kk}(\lambda)| \sim C_{x,k}\lambda^{-d_{x,k}}$, $\left|B\left(e^{i\lambda}\right)\right| \sim C_u \lambda^{-d}$ *as* $\lambda \to 0+$,

2. $A\left(e^{i\lambda}\right)$ *and* $B\left(e^{i\lambda}\right)$ *are differentiable on* $(0, \pi]$ *and* $\left|\left|\frac{dA\left(e^{i\lambda}\right)}{d\lambda}\right|\right| = O\left(\frac{||A\left(e^{i\lambda}\right)||}{\lambda}\right)$, $\left|\frac{dB\left(e^{i\lambda}\right)}{d\lambda}\right| = O\left(\frac{\left|B\left(e^{i\lambda}\right)\right|}{\lambda}\right)$ *uniformly over* $(0, \pi]$ *as* $\lambda \to 0+$, *and*

3. $\left|\left|A\left(e^{i\lambda}\right)\right|\right| > 0$ *and* $\left|B\left(e^{i\lambda}\right)\right| > 0$ *for* $\lambda \in (0, \pi]$.

Condition 1.5

$$\int_{-\pi}^{\pi} ||f_{xx}(\lambda) f_{uu}(\lambda)||\, d\lambda < \infty, \qquad E(x_t x_t') > 0,$$

where $f_{xx}(\lambda)$ *and* $f_{uu}(\lambda)$ *are spectral densities of processes* x_t *and* u_t, *respectively.*

The conditions are similar to those used by Robinson (1995a,b, 1998) and Hidalgo (2003a). Conditions 1.1-1.3 imply homoskedasticity of regressors and errors. This assumption could presumably be relaxed to allow for a certain degree of heterogeneity. Conditions 1.1-1.3 also imply that x_t and u_s are uncorrelated for all t and s and that $E\left(x_t u_t x_s' u_s\right) = E\left(x_t x_s'\right) E\left(u_t u_s\right)$ for all t and s and therefore that the spectral density of $x_t u_t$ at frequency zero is $2\pi \int_{-\pi}^{\pi} f_{xx}(\lambda) f_{uu}(\lambda) d\lambda$ if Condition 1.5 holds. One of the reasons for imposing the condition $E(x_t u_t x_s' u_s) = E(x_t x_s') E(u_t u_s)$ is that it allows us to use

$$\frac{4\pi^2}{T} \sum_{j=1}^{T-1} I_{xx}(\lambda_j) I_{uu}(\lambda_j)$$

of Robinson (1998) to consistently estimate $2\pi \int_{-\pi}^{\pi} f_{xx}(\lambda) f_{uu}(\lambda) d\lambda$ without having to select a bandwidth. If the condition $E(x_t u_t x_s' u_s) = E(x_t x_s') E(u_t u_s)$ is not valid, the long run variance of $x_t u_t$ has an additional component which is a function of the fourth cumulants and which is not estimated by the expression displayed above. When x_t and u_t are short memory processes, the results of Taniguchi (1982) and Keenan (1987) can be used to estimate the additional component of variance, but no estimation methods are available for long memory time series. Relaxing condition $E(x_t u_t x_s' u_s) = E(x_t x_s') E(u_t u_s)$ would thus come at a price of a considerable amount of technical work. Therefore, though assumption of no correlation between regressors and errors is admittedly somewhat restrictive and excludes for example some cases of interest studied by cointegration literature, we do not attempt to relax this assumption.

A further remark on Conditions 1.1-1.3 is that while the fourth moments are assumed constant, the third moments are free to vary and so only second order stationarity is required.

Condition 1.4 allows for a possible singularity at the zero frequency but the results of this chapter could be generalized to the case of a singularity at a nonzero frequency or of more than one singularity. The validity of the bound $\left|dB\left(e^{i\lambda}\right)/d\lambda\right| = O\left(\left|B\left(e^{i\lambda}\right)\right|/\lambda\right)$ implies that $|df_{uu}(\lambda)/d\lambda| = O\left(f_{uu}(\lambda)/\lambda\right)$ since $f_{uu} = \left|B\left(e^{i\lambda}\right)\right|^2 \sigma_\varepsilon^2/(2\pi)$. Similar implication holds for the spectral density matrix f_{xx}. Examples of scalar processes that satisfy Condition 1.4 are FARIMA model of Granger and Joyeux (1980) or Hosking (1981), and fractional Gaussian noise of Mandelbrot and van Ness (1968). These models satisfy $f(\lambda) \sim C\lambda^{-2d}$ as $\lambda \to 0+$ for some memory parameter $d \in \left[0, \frac{1}{2}\right)$. An example of a model with singularities at nonzero frequencies is the Gegenbauer model of Gray et al. (1989).

Condition 1.5 has been used by Robinson (1994b) and Robinson and Hidalgo (1997). The condition restricts the collective memory of regressors and errors. For regressors with long memory parameter d_x and errors with long memory parameter d, Condition 1.5 imposes restriction $d_x + d < \frac{1}{2}$. This condition ensures that the standard least squares estimation procedure of the slope coefficients is $\sqrt{T}$-consistent and leads to a Gaussian limit distribution (Robinson (1994b)). As Hidalgo (2003a) remarks, the first part of Condition 1.5 seems to be very mild and appears to be necessary and minimal for the central limit theorem for OLS estimates of slope coefficient to hold. In a related proposition of Giraitis and Surgailis (1990) an analogous condition is required for convergence of quadratic forms in linear processes. The validity of the CLT carries over to the functional CLT in the present chapter. The restriction $d_x + d < \frac{1}{2}$ could be relaxed by employing estimators of a class of weighted least squares estimators proposed by Robinson and Hidalgo (1997) or a class of generalized least squares estimators proposed by Hidalgo and Robinson (2002), but for notational simplicity we keep Condition 1.5 as it stands.

The main result of this section can now be stated.

Theorem 1.1 *Under Conditions 1.1-1.5 and under the null hypothesis,*

$$\sqrt{T}\begin{pmatrix} \hat{\beta}(\tau) - \beta \\ \hat{\delta}(\tau) \end{pmatrix} \Longrightarrow \frac{1}{\tau(1-\tau)}\begin{pmatrix} \Sigma^{-1}\Omega^{\frac{1}{2}}\left(\tau W(1) - \tau W(\tau)\right) \\ \Sigma^{-1}\Omega^{\frac{1}{2}}\left(W(\tau) - \tau W(1)\right) \end{pmatrix}$$

on Λ, *where* $\Omega = 2\pi \int_{-\pi}^{\pi} f_{xx}(\lambda) f_{uu}(\lambda) d\lambda$ *and* $\Sigma = E(x_t x_t')$.

Theorem 1.1 implies in particular that

$$\sqrt{T}\hat{\delta}(\tau) \Longrightarrow \frac{1}{\tau(1-\tau)}\Sigma^{-1}\Omega^{\frac{1}{2}}\left(W(\tau) - \tau W(1)\right)$$

so that for each fixed $\tau \in \Lambda$,

$$\sqrt{T}\hat{\delta}(\tau) \xrightarrow{d} N\left(0, V_{\hat{\delta}}(\tau)\right), \tag{1.9}$$

where

$$V_{\hat{\delta}}(\tau) = \frac{1}{\tau(1-\tau)}\Sigma^{-1}\Omega\Sigma^{-1}.$$

It is interesting to note that when x_t or u_t are long memory processes, the limiting distribution remains to be a function of a Brownian motion rather than of a fractional Brownian motion that often arises in asymptotic results in long memory environment. A result that is crucial for validity of Theorem 1.1 is that $T^{-1/2}\sum_{j=1}^{T-1} I_{zu}(\lambda_j) = T^{-1/2}\sum_{t=1}^{[\tau T]} x_t (u_t - \bar{u}(\tau))$, where $\bar{u}(\tau) = T^{-1}\sum_{t=1}^{[\tau T]} u_t$, converges weakly to a Brownian motion. When a strongly dependent process x_t is considered separately, normalization by $T^{-\frac{1}{2}-d}$ is required to achieve weak convergence of the partial sum $\sum_{t=1}^{[\tau T]} x_t$ and the limiting process is a fractional Brownian motion. However, the case of the partial sum $\sum_{t=1}^{[\tau T]} x_t (u_t - \bar{u}(\tau))$ is different. Intuitively, while the memory of the processes x_t and u_t is of a long range, the product $x_t (u_t - \bar{u}(\tau))$ displays short memory behaviour. This phenomenon may be regarded as analogous to that of Robinson (1998) where the sample autocovariances of processes x_t and u_t are stochastically dampening each other in his estimator of Ω.

To assess the power of the test procedure, we examine limiting behaviour of the process $\left(\hat{\beta}(\tau)', \hat{\delta}(\tau)'\right)'$ under alternatives. We restrict ourselves to the local alternatives

$$\beta_t = \beta + \frac{1}{\sqrt{T}} h\left(\frac{t}{T}\right), \qquad t = 1, \ldots, T, \tag{1.10}$$

for some β, where h is a p-dimensional vector of bounded variation functions on $[0,1]$. This class of alternatives comprises many types of structural change that may be of interest. For instance, a function $h(\tau) = \delta \mathbb{I}(\tau_0 \leq \tau)$ describes the alternative of an abrupt break of size δ at time τ_0. A step function h defines multiple structural breaks. A function h consisting of two constant segments connected by a smooth curve depicts smooth transition between two steady levels of a parameter, while a general smooth function h captures continual change of the parameter.

For the limiting distribution under local alternatives the following result is obtained.

Theorem 1.2 *Under Conditions 1.1-1.5 and under the local alternative hy-*

pothesis (1.10),

$$\sqrt{T}\begin{pmatrix}\hat{\beta}(\tau)-\beta\\ \hat{\delta}(\tau)\end{pmatrix} \Longrightarrow \frac{1}{\tau(1-\tau)}\begin{pmatrix}\Sigma^{-1}\Omega^{\frac{1}{2}}(\tau W(1)-\tau W(\tau))\\ \Sigma^{-1}\Omega^{\frac{1}{2}}(W(\tau)-\tau W(1))\end{pmatrix}$$
$$+\frac{1}{\tau(1-\tau)}\begin{pmatrix}\tau\int_\tau^1 h(u)\,du\\ \left(\int_0^\tau h(u)\,du-\tau\int_0^1 h(u)\,du\right)\end{pmatrix}$$

for $\tau \in \Lambda$.

By the continuous mapping theorem, an immediate consequence of Theorem 1.2 is the following corollary.

Corollary 1.1 *Let* ϕ *be a continuous functional on* $D(\Lambda)^p$. *Let* $Z_T = \phi\left(\sqrt{T}\hat{\delta}(\tau)\right)$ *and*

$$\begin{aligned} Z_h &= \phi\left(\frac{1}{\tau(1-\tau)}\Sigma^{-1}\Omega^{\frac{1}{2}}(W(\tau)-\tau W(1)) + \right.\\ &\left.\frac{1}{\tau(1-\tau)}\left(\int_0^\tau h(u)\,du-\tau\int_0^1 h(u)\,du\right)\right). \end{aligned}$$

Under the conditions of Theorem 1.2,

$$Z_T \xrightarrow{d} Z_h.$$

The corollary shows that the test based on Z_T has nontrivial local power against a broad range of alternatives. The limiting random variable Z_h is indexed by functions h specifying local alternatives. Under the null, when $h \equiv 0$, the test statistic Z_T converges in distribution to Z_0,

$$\phi\left(\sqrt{T}\hat{\delta}(\tau)\right) \xrightarrow{d} \phi\left(\frac{1}{\tau(1-\tau)}\Sigma^{-1}\Omega^{\frac{1}{2}}(W(\tau)-\tau W(1))\right).$$

The asymptotic test at a significance level α is based on a critical region C_α constructed from the asymptotic null distribution, $P(Z_0 \in C_\alpha) = \alpha$. The asymptotic test rejects the null when $Z_T \in C_\alpha$.

The form of the limiting distributions in Theorems 1.1 and 1.2 explains the reason for the necessity of bounding the set Λ away from 0 and 1. The restriction on Λ guarantees that the convergence of the estimator $\hat{\delta}$, which is the basis of the test statistic, is uniform. Moreover, it can be shown that for $\Lambda = (0, 1)$ many functionals, including the sup- and L_q-norms, diverge to infinity in probability.

The trimming restriction on Λ can be avoided by allowing the limiting distribution of the test statistic to be of a different form than a functional of the Brownian bridge. The results of Jaeschke (1979) and Eicker (1979) suggest that the supremum of $\hat{\delta}(\tau)$, taken over subsets of $(0, 1)$ that are increasing towards $(0, 1)$ at an appropriate speed and that are normalized by a suitable centering and rescaling sequences, should converge to an extreme value distribution. However, relaxing the restriction on Λ in such a way comes at a cost. The convergence of the test statistics to the extreme value distribution can be expected to be very slow. The results of Hall (1979) indicate that the rate of convergence could be as slow as $\log T$. The asymptotic critical values are therefore not appropriate for tests in samples of moderate size and an elaborate bootstrap procedure would be required to improve on the performance of the asymptotic test. We do not pursue this possibility in this thesis.

It is interesting to note that

$$\operatorname{var} \frac{W(\tau) - \tau W(1)}{\tau(1-\tau)} = \frac{1}{\tau(1-\tau)}$$

is not constant across Λ which means that under the null, the probability that the process $\left\|\hat{\delta}(\tau)\right\|$ crosses any horizontal line above the real axis is smallest at $\tau = \frac{1}{2}$. This may lead us to inquire whether the power of the test based on supremum and other functionals can be improved by levelling the variance of the estimated process $\hat{\delta}$ across Λ. Given the restriction of Λ away from $(0, 1)$, we may normalize the process $\hat{\delta}$ by multiplying it by $[\tau(1-\tau)]^{1/2}$. By Theorem 1.1, under the null,

$$[\tau(1-\tau)]^{\frac{1}{2}} \sqrt{T}\hat{\delta}(\tau) \Longrightarrow \frac{1}{[\tau(1-\tau)]^{\frac{1}{2}}} \Sigma^{-1}\Omega^{\frac{1}{2}} \left(W(\tau) - \tau W(1)\right),$$

where the variance of the limiting distribution is equal to $\Sigma^{-1}\Omega\Sigma^{-1}$ across Λ. The rejection probabilities of the test based on the levelled process $\hat{\delta}$ in samples of moderate size is examined in a Monte Carlo experiment in Section 1.4.

Our test procedure is based on the behaviour of the OLS estimator of the δ coefficient. At the core of the limit behaviour of the test statistics lies the fact that $T^{-1/2}\sum_{j=1}^{T-1} w_{z(\tau)}(\lambda_j)\,\bar{w}_{\hat{u}}(\lambda_j)$ converges weakly to a Brownian motion process. Using this fact, the asymptotic behaviour of other tests based on the behaviour of OLS slope coefficient estimators can be obtained. For example, if $\hat{\beta}_{t_1}^{t_2}$ is the OLS estimator of β in the regression $y_t = \alpha + \beta' x_t + u_t$ for $t = t_1, \ldots, t_2$, then under the local alternative (1.10)

$$\begin{aligned}\tau\sqrt{T}\left(\hat{\beta}_1^{[\tau T]} - \hat{\beta}_1^{T}\right) &\implies \Sigma^{-1}\Omega^{\frac{1}{2}}\left(W(\tau) - \tau W(1)\right) \\ &\quad + \left(\int_0^\tau h(u)\,du - \tau\int_0^1 h(u)\,du\right)\end{aligned}$$

in correspondence with the results of Ploberger et al. (1989). If $\hat{\Sigma}$ and $\hat{\Omega}$ are consistent estimates of Σ and Ω, then the Wald-statistic process based on partial sample slope estimators,

$$T\left(\hat{\beta}_1^{[\tau T]} - \hat{\beta}_{[\tau T]+1}^{T}\right)'\left(\frac{\Sigma^{-1}\Omega\Sigma^{-1}}{\tau(1-\tau)}\right)^{-1}\left(\hat{\beta}_1^{[\tau T]} - \hat{\beta}_{[\tau T]+1}^{T}\right),$$

has limiting distribution $J(\tau)' J(\tau)$, where

$$\begin{aligned}J(\tau) &= \frac{1}{\left[\tau(1-\tau)\right]^{\frac{1}{2}}}\left(W(\tau) - \tau W(1)\right) \\ &\quad + \frac{1}{\left[\tau(1-\tau)\right]^{\frac{1}{2}}}\Omega^{-\frac{1}{2}}\Sigma\left(\int_0^\tau h(u)\,du - \tau\int_0^1 h(u)\,du\right)\end{aligned}$$

as in Andrews (1993).

On the other hand, the limiting distribution of tests based on behaviour of the OLS residuals depends crucially on the weak convergence of $\sum_{t=1}^{[\tau T]}\hat{u}_t$ to a limiting process. Under long memory, the appropriately normalized partial sum can be expected to converge to a fractional Brownian motion and thus to be different than under short memory.

1.3 Bootstrap procedure

The limiting distribution of the process $\hat{\delta}$ in (1.9) depends on unknown parameters Ω and Σ. The process $\hat{\delta}$ can be normalized by consistent estimates $\hat{\Omega}$, $\hat{\Sigma}$ of these parameters. Such consistent estimates are for example

$$\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T} x_t x_t' \tag{1.11}$$

and

$$\hat{\Omega} = \frac{4\pi^2}{T}\sum_{j=1}^{T-1} I_{xx}\left(\lambda_j\right) I_{\hat{u}\hat{u}}\left(\lambda_j\right). \tag{1.12}$$

Consistency of $\hat{\Sigma}$ follows from ergodicity of x_t in the variance implied by Conditions 1.1 and 1.2. The estimator $\hat{\Omega}$ is based on results of Robinson (1998) and its consistency is asserted in the following theorem.

Theorem 1.3 *Under Conditions 1.1-1.5 and under the local alternative,*

$$\hat{\Omega} \xrightarrow{p} \Omega.$$

The normalized process $\tilde{\delta}\left(\tau\right) = \hat{\Omega}^{-\frac{1}{2}}\hat{\Sigma}\hat{\delta}\left(\tau\right)$ has a limiting distribution which is free of nuisance parameters,

$$\sqrt{T}\tilde{\delta}\left(\tau\right) \Longrightarrow \frac{W\left(\tau\right) - \tau W\left(1\right)}{\tau\left(1-\tau\right)}. \tag{1.13}$$

In special cases, distributions of functionals of Brownian motion are known analytically and quantiles of the distributions can be easily computed. Examples are supremum of a Brownian motion and supremum of a Brownian bridge. In other instances, critical values have been computed by simulation and tabulated, as in case of the supremum of the square of a standardized tied-down Bessel process in Andrews (1993). However, in majority of cases, including non-supremum functionals of the limiting distribution in (1.13), the critical values of the test statistic need to be simulated by the researcher.

An alternative to computing asymptotical critical values by simulation is to employ a bootstrap procedure. The basic bootstrap of Efron (1979) estimates the unknown distribution of a random variable by the empirical distribution of a random sample drawn from that distribution. However, time series data cannot be regarded as a random sample, and the bootstrap procedure needs to be modified to accommodate the time dependence structure of the data.

A number of time-domain bootstrap procedures for time series has been proposed, ranging from parametric procedures such as those of Freedman (1984) or Efron and Tibshirani (1986) to nonparametric methods, such as the block bootstrap of Carlstein (1986) and Künsch (1989), subsampling algorithms introduced by Politis and Romano (1992), or sieve bootstrap proposed by Kreiss (1988) and explored by Bühlmann (1997, 1998). Frequency domain approaches have also been examined. Among others, we can cite the periodogram ordinates bootstrap of Franke and Härdle (1992) or Dahlhaus and Janas (1996). The validity of all of these procedures, however, is subject to the assumption that the dependence between distant observations is sufficiently weak. This assumption excludes processes with long memory. Moreover, all the nonparametric methods cited above require a user intervention in the form of choosing a lag length, a bandwidth or a block length. The performance of time-series bootstrap can be highly sensitive to the choice of the dimension parameter, particularly in samples of moderate size. Although automatic procedures for choosing the dimension have been devised for some methods, they can be computationally expensive.

Hidalgo (2003a) proposes a method that eliminates the dimension choice. He suggests to bootstrap OLS residuals in frequency domain. His bootstrap procedure is easy to implement and computationally inexpensive. Moreover, it is one the first bootstrap procedures shown to be valid for long memory time series in a fairly general context, adding to a still thin body of the literature on long memory time series bootstrap.

In this chapter we propose to approximate the critical values of the testing procedure described in Section 1.2 by a bootstrap procedure based on the ideas of Hidalgo (2003a). The procedure consists of the following steps.

Step 1 Compute OLS estimates $\hat{\beta}(\tau)$ and $\hat{\delta}(\tau)$ from (1.4) or (1.6) for $\tau \in \Lambda$. Compute $\hat{\tau} = \arg\max_{\tau \in \Lambda} \left\| \hat{\delta}(\tau) \right\|$, the OLS estimates $\hat{\beta} = \hat{\beta}(\hat{\tau})$ and

$\hat{\delta} = \hat{\delta}(\hat{\tau})$ and the OLS residuals

$$\hat{u}_t = y_t - \hat{\beta}' x_t - \hat{\delta}' z_t(\hat{\tau}), \qquad t = 1, \ldots, T.$$

Step 2 Compute

$$w_{\hat{u}}(\lambda_j) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} \hat{u}_t e^{it\lambda_j}, \qquad j = 1, \ldots, T-1,$$

and

$$\tilde{w}_{\hat{u}}(\lambda_j) = \frac{w_{\hat{u}}(\lambda_j) - \frac{1}{T-1}\sum_{k=1}^{T-1} w_{\hat{u}}(\lambda_k)}{\left(\frac{1}{T-1}\sum_{j=1}^{T-1}\left|w_{\hat{u}}(\lambda_j) - \frac{1}{T-1}\sum_{k=1}^{T-1} w_{\hat{u}}(\lambda_k)\right|^2\right)^{\frac{1}{2}}}, \qquad j = 1, \ldots, T-1.$$

Step 3 Draw a random sample $\eta_1^*, \ldots, \eta_{[T/2]}^*$ from the distribution $P^*\left(\eta_j^* = \tilde{w}_{\hat{u}}(\lambda_k)\right) = \frac{1}{[T/2]}$ for $k = 1, \ldots, [T/2]$ and generate a bootstrap sample

$$w_y^*(\lambda_j) = \hat{\beta}_0' w_x(\lambda_j) + |w_{\hat{u}}(\lambda_j)|\eta_j^*, \qquad j = 1, \ldots, [T/2],$$

where $\hat{\beta}_0$ is the estimate of β from the null regression of $w_y(\lambda_j)$ on $w_x(\lambda_j)$ alone.

Step 4 Compute $(\hat{\beta}^*(\tau)', \hat{\delta}^*(\tau)')'$ as

$$\begin{pmatrix} \hat{\beta}^*(\tau) \\ \hat{\delta}^*(\tau) \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{T-1} I_{xx,j} & \sum_{j=1}^{T-1} I_{xz,j} \\ \sum_{j=1}^{T-1} I_{zx,j} & \sum_{j=1}^{T-1} I_{zz,j} \end{pmatrix}^{-1} \times 2\,\mathrm{Re}\begin{pmatrix} \sum_{j=1}^{[T/2]} w_{x,j}\,|w_{\hat{u},j}|\,\eta_j^* \\ \sum_{j=1}^{[T/2]} w_{z(\tau),j}\,|w_{\hat{u},j}|\,\eta_j^* \end{pmatrix},$$

where the right-hand side depends on τ through the definition $z_t = x_t \mathbb{I}(t \leq [\tau T])$ in (1.3).

Step 5 Compute the value of the functional used for the original data, $Z_T^* = \phi\left(\sqrt{T}\hat{\delta}^*\right)$.

The distribution of the bootstrap test statistic Z_T^* can be used to approximate the asymptotic null distribution of Z_T, that is to construct a bootstrap test. To show the validity of the bootstrap procedure, we need to prove that the bootstrap process

$$\begin{pmatrix} \hat{\beta}^*(\tau) - \hat{\beta}_0 \\ \hat{\delta}^*(\tau) \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{T-1} I_{xx,j} & \sum_{j=1}^{T-1} I_{xz,j} \\ \sum_{j=1}^{T-1} I_{zx,j} & \sum_{j=1}^{T-1} I_{zz,j} \end{pmatrix}^{-1} \times 2\operatorname{Re} \begin{pmatrix} \sum_{j=1}^{[T/2]} w_{x,j} \left| w_{\hat{u},j} \right| \eta_j^* \\ \sum_{j=1}^{[T/2]} w_{z(\tau),j} \left| w_{\hat{u},j} \right| \eta_j^* \end{pmatrix} \tag{1.14}$$

consistently estimates the null behaviour of the process $\left(\hat{\beta}(\tau)' - \beta', \hat{\delta}(\tau)'\right)'$. It must be shown that under the null and under the local alternative the process $2\operatorname{Re} T^{-1/2} \sum_{j=1}^{[T/2]} w_{z(\tau),j} \left| w_{\hat{u},j} \right| \eta_j^*$, conditionally on data, converges weakly in probability to the same process as $T^{-1/2} \sum_{j=1}^{T-1} I_{zu}(\lambda_j)$, that is,

$$2\operatorname{Re} \frac{1}{\sqrt{T}} \sum_{j=1}^{[T/2]} w_{z(\tau),j} \left| w_{\hat{u},j} \right| \eta_j^* \overset{p}{\Longrightarrow} \frac{1}{2\pi} \Omega^{\frac{1}{2}} W(\tau).$$

The consistency of the bootstrap is asserted in the following theorem.

Theorem 1.4 *Under Conditions 1.1-1.5 and under both the null and the local alternative hypotheses,*

$$\sqrt{T} \begin{pmatrix} \hat{\beta}^*(\tau) - \hat{\beta} \\ \hat{\delta}^*(\tau) \end{pmatrix} \overset{p}{\Longrightarrow} \frac{1}{\tau(1-\tau)} \begin{pmatrix} \Sigma^{-1}\Omega^{\frac{1}{2}} (\tau W(1) - \tau W(\tau)) \\ \Sigma^{-1}\Omega^{\frac{1}{2}} (W(\tau) - \tau W(1)) \end{pmatrix}.$$

A straightforward consequence of Theorem 1.4 and the continuous mapping theorem is the following corollary.

Corollary 1.2 *Let ϕ be a continuous functional on $D(\Lambda)^p$. Let $Z_T^* = \phi\left(\sqrt{T}\hat{\delta}^*(\tau)\right)$ and let Z_0 be Z_h of Corollary 1.1 with $h = 0$, i.e.*

$$Z_0 = \phi\left((\tau(1-\tau))^{-1} \Sigma^{-1}\Omega^{\frac{1}{2}} (W(\tau) - \tau W(1))\right).$$

Under the conditions of Theorem 1.4,

$$Z_T^* \xrightarrow{d^*} Z_0.$$

The bootstrap test is constructed using a critical region C_α^* based on the bootstrap distribution in such a way that $P(Z_T^* \in C_\alpha^*) = \alpha$, where α is a level of significance. The bootstrap test rejects when $Z_T \in C_\alpha^*$. Let $F_T^*(x) = P(Z_T^* \leq x | \mathcal{F}_T \vee \mathcal{G}_T)$ denote the distribution function of Z_T^* conditional on data and $F(x) = P(Z_0 \leq x)$ the null asymptotic distribution function. The bootstrap p-value for a one-tailed test is $p_T = 1 - F_T^*(Z_T)$. The bootstrap test rejects H_0 when Z_T is large, that is when p_T is small. By Corollaries 1.1 and 1.2, $Z_T \xrightarrow{d} Z_h$ and $F_T^* \xRightarrow{p} F$. The continuous mapping theorem implies that $p_T \xrightarrow{d^*} 1 - F(Z_h)$. The p-values based on the bootstrap distribution F_T^* are therefore asymptotically equivalent to the p-values based on the distribution F.

It should be noted that the proposed bootstrap is not the only possibility. The variables η_j^* in Step 3 are drawn from the empirical distribution of normalized discrete Fourier transform of the OLS residuals. Alternatively, external bootstrap can be carried out by drawing η_j^* from any complex-valued distribution with zero mean, unit variance and $E\eta_j^{*2} = 0$. A natural choice is a complex normal distribution. The proof of validity of the external bootstrap procedure remains identical to the current proof. Another valid modification is to multiply η_j^* in Step 3 by the value of $w_{\hat{u}}(\lambda_j)$ instead of its modulus. The proof of validity in this case goes through with only minor alterations as noted at the end of the proof of Lemma 1.12 in Section 1.A below. A simulation study suggests that none of the methods above dominates the others in performance.

Hidalgo (2003a) interchanges the resampling with the Fourier transformation, first resampling the normalized time-domain residuals and then transforming the resampled data into the frequency domain. His simulation results seem to suggest that there is no substantial advantage in exchanging the order of the operations. In the simulation experiments in this chapter we use the procedure given in Steps 1-5.

1.4 Finite sample properties

In order to assess the performance of the bootstrap procedure in finite samples, a small Monte Carlo study is conducted. Data are generated according to a simple linear model

$$y_t = \alpha + \beta_t x_t + u_t, \qquad t = 1, \ldots, T,$$

where scalar series $\{x_t\}$ and $\{u_t\}$ follow a FARIMA$(0, d, 0)$ process and where $\alpha = 0$. The long memory parameters d_x and d for the regressor x_t and errors u_t are either 0 (short memory) or 0.2 (stationary long memory). The series x_t and u_t are generated using the Davies-Harte (1987) algorithm. The set Λ of feasible break dates is taken to be the interval $[\varepsilon T, (1-\varepsilon) T]$ where $\varepsilon = 0.05$, so that approximately 5% of potential break dates are discarded from each side of the $1, \ldots, T$ range. The sample sizes considered are 32, 64, 128, 256. While a sample of length 32 may be too short to yield satisfactory results in the long memory case, the Monte Carlo simulation can still offer useful insights into the performance of the method for the short memory case. Two functionals are chosen on which to base the test procedure: a Kolmogorov-Smirnov- (or Bartlett-) type statistic, whose discrete version is

$$\mathrm{KS} = \sup_{[\varepsilon T] \leq j \leq [(1-\varepsilon)T]} \sqrt{T} \left| \hat{\delta}\left(\frac{j}{T}\right) \right|,$$

and a Cramér-von Mises-type statistic based on L_2-distance, with a discrete version

$$\mathrm{CvM} = \sum_{j=[\varepsilon T]}^{[(1-\varepsilon)T]} \hat{\delta}^2 \left(\frac{j}{T}\right).$$

The bootstrap test is based on the estimated process $\hat{\delta}$ obtained from (1.4) or (1.6). Since the limiting variance of the process $\hat{\delta}(\tau)$ varies with τ, we also consider a normalized version $[\tau(1-\tau)]^{\frac{1}{2}} \hat{\delta}(\tau)$, whose variance is level across Λ.

The asymptotic test is based on the process $\tilde{\delta}(\tau) = \hat{\Omega}^{-\frac{1}{2}} \hat{\Sigma} \hat{\delta}(\tau)$, where $\hat{\Sigma}$ and $\hat{\Omega}$ are computed as in (1.11) and (1.12), respectively. A levelled version $[\tau(1-\tau)]^{\frac{1}{2}} \tilde{\delta}(\tau)$ is also considered. The values of the Kolmogorov-Smirnov and Cramér-von Mises test statistics are compared with quantiles of their asymptotic distribution. These quantiles are estimated by approximating the limiting processes by their discrete versions over a grid of 10 000 points spaced equally across the interval $[0, 1]$ and by simulating the distribution of functionals of these processes by Monte Carlo. The number of Monte Carlo replications is 10^6.

The results in each of the tables are all obtained conditionally on a set of 5000 replications of a 256×2 matrix of independent identically distributed $N(0, 1)$ elements. Within each replication, 1000 bootstrap samples are generated. The rejection probabilities are based on 5% nominal significance level.

Table 1.1 gives the results of the examination of the level of the bootstrap and asymptotic tests. In this table and in Table 1.2, the heading "raw" denotes the size of the test based on the original process $\hat{\delta}(\tau)$ defined in (1.4) or (1.6) whereas the heading "norm" refers to the size of the test based on the levelled process $[\tau(1-\tau)]^{\frac{1}{2}} \hat{\delta}(\tau)$. The bootstrap test is non-conservative, with level approaching the nominal value from above as the sample size increases. Overall, neither KS nor CvM test statistic can be said to generate better test as far as level is concerned. The actual level tends to be closer to the nominal value when the memory of the error is of short range. Levelling the variance of the process $\hat{\delta}$ does not seem to bring substantial changes in the size.

The asymptotic test performs poorly for the range of sample sizes under consideration. Again, neither of the Kolmogorov-Smirnov and Cramér-von Mises tests dominates the other. Levelling the variance of the process $\tilde{\delta}$ actually seems to slightly damage the null rejection probabilities for a range of sample sizes.

In order to explore the power of the test under the alternative, the alternative is set up as a break in the middle of the sample, $\tau_0 = 1/2$, with unit size of the jump, $\delta = 1$. In the experiment the alternative is fixed, that is the size of break does not change with the sample size. The outcome of the simulation of rejection probabilities under the alternative is reported in Table 1.2. In terms of rejection probabilities under the alternative, the CvM test appears to be strictly preferable to the KS test for both the bootstrap and the asymptotic test. This is in agreement with expectation of Ploberger and Krämer (1992) who suspect that L_2-norm CvM test might perform better than sup-norm KS test in case of the one-time structural break. The rejection probabilities of the asymptotic test are larger than those of the bootstrap test in a majority of parameter combinations. However, such a comparison is not informative since the actual critical values have not been corrected to yield 5% level of the tests. An important observation is that levelling the variance of the process $\hat{\delta}$ unambiguously and substantially improves the power of all forms of the test.

Overall, the outcome of the simulation exercise provides evidence that the bootstrap procedure proposed in the chapter performs reasonably well already for samples of moderate size. The results of the exercise further seem to suggest that (a) the bootstrap test is preferable to the asymptotic test for small to moderately sized samples, (b) Cramér-von Mises-type of test statistic is preferable to the Kolmogorov-Smirnov-type, at least for one-time change alternatives, and (c) levelling the variance of the test process $\hat{\delta}$ across Λ may be recommended, at least for some forms of the alternative hypothesis.

1.5 Conclusions

The chapter examines a test for parameter instability in a linear model where memory of both regressors and errors is allowed to be of a long range. The testing procedure is based on a process of OLS slope coefficient estimators. The choice of a continuous functional of this process for constructing the test statistic can reflect beliefs about the form of alternative and can improve the power of the test procedure.

A bootstrap procedure is proposed to approximate the distribution of the test statistic to the first order. The procedure is carried out in frequency domain and does not require choice of any tuning parameter such as block length in block bootstrap methods. A Monte Carlo study suggests that the bootstrap produces good results and is superior over the asymptotic test for moderate size samples.

There are several natural directions in which the current work can be extended. First, the condition that $\Omega < \infty$ could be relaxed to allow for greater degree of collective memory of regressors and errors. In this case,

		Bootstrap test				Asymptotic test			
		KS		CvM		KS		CvM	
d_x	d	**raw**	**norm**	**raw**	**norm**	**raw**	**norm**	**raw**	**norm**
T=32									
0	**0**	9.9	9.9	9.4	9.3	46.7	41.5	52.3	34.6
0	**0.2**	12.3	12.2	11.9	10.5	48.8	43.1	54.6	36.1
0.2	**0**	9.9	10.4	10.2	9.4	49.7	44.6	56.9	41.0
0.2	**0.2**	12.2	12.6	12.3	11.0	50.5	45.5	58.8	42.9
T=64									
0	**0**	9.1	9.2	8.8	7.7	17.9	15.0	15.7	9.4
0	**0.2**	10.2	9.6	8.3	7.5	18.7	15.8	17.1	10.1
0.2	**0**	8.8	8.6	8.7	8.1	20.7	18.2	21.6	13.5
0.2	**0.2**	10.1	9.4	9.3	8.5	19.9	18.2	22.6	15.5
T=128									
0	**0**	6.5	6.3	6.7	6.5	7.6	4.6	6.4	4.7
0	**0.2**	6.9	6.4	6.7	6.5	8.2	5.0	6.8	4.5
0.2	**0**	6.4	6.5	6.9	6.7	9.5	6.2	8.5	6.1
0.2	**0.2**	7.4	7.3	7.2	7.1	8.7	6.4	9.6	7.4
T=256									
0	**0**	5.3	5.7	5.9	5.9	3.7	1.7	4.0	3.3
0	**0.2**	5.8	5.5	5.9	5.9	4.0	1.9	4.2	3.4
0.2	**0**	5.4	5.3	6.1	6.1	4.8	2.4	5.2	4.1
0.2	**0.2**	6.3	6.0	6.1	6.0	4.0	2.2	5.5	4.7

Table 1.1: Size of the bootstrap and asymptotic test at 5% nominal level

		Bootstrap test				Asymptotic test			
		KS		CvM		KS		CvM	
d_x	d	**raw**	**norm**	**raw**	**norm**	**raw**	**norm**	**raw**	**norm**
T=32									
0	**0**	11.0	19.8	24.9	36.9	48.0	52.1	75.8	70.0
0	**0.2**	13.0	20.9	26.5	38.4	49.9	53.0	76.8	71.0
0.2	**0**	11.9	21.0	26.2	37.9	52.3	58.9	81.1	76.2
0.2	**0.2**	14.3	22.5	27.9	38.4	52.6	58.9	80.5	75.8
T=64									
0	**0**	15.5	53.1	68.3	80.9	17.5	54.7	78.9	82.4
0	**0.2**	15.6	51.4	66.6	79.6	11.7	54.4	78.4	81.8
0.2	**0**	16.8	53.4	68.9	81.5	22.4	65.0	84.3	87.5
0.2	**0.2**	17.5	50.8	66.0	77.1	21.5	60.5	81.0	83.8
T=128									
0	**0**	32.3	91.9	97.5	99.1	16.0	91.5	97.1	98.5
0	**0.2**	31.0	90.3	96.5	98.7	16.0	89.9	96.3	98.0
0.2	**0**	34.7	92.5	98.1	99.4	24.5	94.8	98.3	99.2
0.2	**0.2**	32.3	89.2	95.0	97.9	19.9	88.8	95.7	99.2
T=256									
0	**0**	79.4	100.0	100.0	100.0	61.7	100.0	100.0	100.0
0	**0.2**	74.6	99.9	100.0	100.0	100.0	100.0	100.0	100.0
0.2	**0**	81.5	100.0	100.0	100.0	73.7	100.0	100.0	100.0
0.2	**0.2**	71.6	100.0	100.0	100.0	49.3	99.9	99.8	100.0

Table 1.2: Rejection probabilities of the bootstrap and asymptotic test under the alternative at 5% nominal level

the OLS estimation procedure could be replaced by a GLS-type procedure. Second, partial structural change could be considered and gains in efficiency from allowing partial change evaluated. Third, a bootstrap procedure might be shown to approximate the distribution of the test statistics to an order higher than first. These topics can be examined in future research.

Further, under the assumption that the alternative hypothesis holds and is of the one-time structural break form, the date of break could be estimated and, based on the distribution of the break date estimator, inference conducted. Estimation of the date of break is the topic of Chapters 2 and 3.

1.A Proofs

For notational simplicity, the process $\{x_t\}$ in Theorems 1.1-1.4 is taken to be scalar. Asymptotic results for vector processes can be obtained using Cramér-Wold device for stochastic processes as defined for example in Lemma A4 of Andrews (1993). We denote $A_j = A\left(e^{i\lambda_j}\right)$ and $B_j = B\left(e^{i\lambda_j}\right)$.

Validity of Theorems 1.1-1.4 rests on the fact that under Conditions 1.1-1.5,

$$\frac{1}{\sqrt{T}}\sum_{j=1}^{T-1} I_{zu,j} \Longrightarrow \frac{1}{2\pi}\Omega^{\frac{1}{2}}W(\tau), \tag{1.15}$$

$$\frac{1}{T}\sum_{j=1}^{T-1} w_{h_1x,j}\bar{w}'_{h_2x,j} \xrightarrow{p} \Sigma\int_0^1 h_1(t)\,h_2(t)\,dt \tag{1.16}$$

and

$$2\,\mathrm{Re}\frac{1}{\sqrt{T}}\sum_{j=1}^{[T/2]} w_{z(\tau),j}\left|w_{\hat{u},j}\right|\eta_j^* \overset{p}{\Longrightarrow} \frac{1}{2\pi}\Omega^{\frac{1}{2}}W(\tau) \tag{1.17}$$

over $[0,1]$, where for any function h, $\{w_{hx,j}, j=1,\ldots,T\}$ is the discrete Fourier transform of the sequence $\{h(t/T)\,x_t, t=1,\ldots,T\}$ and where the random variables η_j^* are defined in Step 3 of the bootstrap procedure. In all three cases, the convergence is shown in two steps. First, convergence is proved for weighted innovation processes $\{\xi_t\}$ and $\{\varepsilon_t\}$. The result for the processes $\{x_t\}$, $\{u_t\}$ is then established by showing that the difference between the left-hand sides of (1.15)-(1.17) and their weighted-innovation analogues converges to zero in probability uniformly over $[0,1]$. Lemmas 1.1-1.12 in Section 1.B below establish convergence in (1.15)-(1.17). The validity of Theorems 1.1-1.4 is argued employing Lemmas 1.1-1.12.

Proof of Theorems 1.1 and 1.2. Under the local alternative,

$$\sqrt{T}\begin{pmatrix} \hat{\beta}(\tau)-\beta \\ \hat{\delta}(\tau) \end{pmatrix} = \begin{pmatrix} \frac{1}{T}\sum_{j=1}^{T-1} I_{xx,j} & \frac{1}{T}\sum_{j=1}^{T-1} I_{xz,j} \\ \frac{1}{T}\sum_{j=1}^{T-1} I_{zx,j} & \frac{1}{T}\sum_{j=1}^{T-1} I_{zz,j} \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{j=1}^{T-1} I_{xu,j} \\ \frac{1}{\sqrt{T}}\sum_{j=1}^{T-1} I_{zu,j} \end{pmatrix}$$

$$+\begin{pmatrix} \frac{1}{T}\sum_{j=1}^{T-1} I_{xx,j} & \frac{1}{T}\sum_{j=1}^{T-1} I_{xz,j} \\ \frac{1}{T}\sum_{j=1}^{T-1} I_{zx,j} & \frac{1}{T}\sum_{j=1}^{T-1} I_{zz,j} \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{j=1}^{T-1} w_{x,j}\bar{w}_{hx,j} \\ \frac{1}{\sqrt{T}}\sum_{j=1}^{T-1} w_{z,j}\bar{w}_{hx,j} \end{pmatrix}. \tag{1.18}$$

By Lemma 1.9 with $h_1(x)=1$ and $h_2(x)=\mathbb{I}(x\le\tau)$,

$$\frac{1}{T}\sum_{j=1}^{T-1} I_{xz,j} \xrightarrow{p} \frac{\tau}{2\pi}\Sigma.$$

Similarly, $\frac{1}{T}\sum_{j=1}^{T-1} I_{xx,j} \xrightarrow{p} \frac{1}{2\pi}\Sigma$ and $\frac{1}{T}\sum_{j=1}^{T-1} I_{zz,j} \xrightarrow{p} \frac{\tau}{2\pi}\Sigma$, and therefore

$$\begin{pmatrix} \frac{1}{T}\sum_{j=1}^{T-1} I_{xx,j} & \frac{1}{T}\sum_{j=1}^{T-1} I_{xz,j} \\ \frac{1}{T}\sum_{j=1}^{T-1} I_{zx,j} & \frac{1}{T}\sum_{j=1}^{T-1} I_{zz,j} \end{pmatrix} \Longrightarrow \begin{pmatrix} 1 & \tau \\ \tau & \tau \end{pmatrix} \otimes \frac{1}{2\pi}\Sigma$$

over $[0,1]$. Since matrix inverse is a continuous function for $\tau\in\Lambda$,

$$\begin{pmatrix} \frac{1}{T}\sum_{j=1}^{T-1} I_{xx,j} & \frac{1}{T}\sum_{j=1}^{T-1} I_{xz,j} \\ \frac{1}{T}\sum_{j=1}^{T-1} I_{zx,j} & \frac{1}{T}\sum_{j=1}^{T-1} I_{zz,j} \end{pmatrix}^{-1} \Longrightarrow \begin{pmatrix} 1 & \tau \\ \tau & \tau \end{pmatrix}^{-1} \otimes 2\pi\Sigma^{-1}.$$

over Λ. Under the null, that is when $h\equiv 0$, the second term on the right of (1.18) vanishes. By Lemma 1.7,

$$\sqrt{T}\begin{pmatrix} \hat{\beta}(\tau)-\beta \\ \hat{\delta}(\tau) \end{pmatrix} \Longrightarrow \frac{1}{\tau(1-\tau)}\begin{pmatrix} \Sigma^{-1}\Omega^{\frac{1}{2}}(\tau W(1)-\tau W(\tau)) \\ \Sigma^{-1}\Omega^{\frac{1}{2}}(W(\tau)-\tau W(1)) \end{pmatrix}$$

and Theorem 1.1 is proved.

Under the alternative, $h \neq 0$ and by Lemma 1.9

$$\frac{1}{\sqrt{T}}\sum_{j=1}^{T-1} w_{z,j}\bar{w}_{hx,j} \Longrightarrow \frac{1}{2\pi}\Sigma\int_0^{\tau} h(t)\,dt$$

over $[0,1]$. Therefore the second term in (1.18) converges to

$$\frac{1}{\tau(1-\tau)}\begin{pmatrix} \tau\int_{\tau}^{1} h(u)\,du \\ \left(\int_0^{\tau} h(u)\,du - \tau\int_0^1 h(u)\,du\right) \end{pmatrix}$$

and Theorem 1.2 is established. ∎

Proof of Theorem 1.3. By Theorem 1 of Robinson (1998),

$$\frac{4\pi^2}{T}\sum_{j=1}^{T-1} I_{xx,j}I_{uu,j} \xrightarrow{p} \Omega. \tag{1.19}$$

Proceeding as in part (c) of the proof of Lemma 1.12, write

$$w_{u,j} - w_{\hat{u},j} = \left(\hat{\beta}-\beta\right)w_{x,j} + \hat{\delta}w_{z(\hat{\tau}),j} - \frac{1}{\sqrt{T}}w_{hx,j}.$$

Therefore

$$I_{\hat{u}\hat{u},j} - I_{uu,j} = \left|w_{u,j} - w_{\hat{u},j}\right|^2 - 2\,\mathrm{Re}\left(w_{\hat{u},j} - w_{u,j}\right)\bar{w}_{u,j}$$

and

$$\begin{aligned}\left|\frac{1}{T}\sum_{j=1}^{T-1} I_{xx,j}\left(I_{\hat{u}\hat{u},j} - I_{uu,j}\right)\right| \leq\; & \frac{1}{T}\sum_{j=1}^{T-1}\left|w_{x,j}\right|^2\left|w_{u,j} - w_{\hat{u},j}\right|^2 \\ & + \frac{1}{T}\sum_{j=1}^{T-1}\left|w_{x,j}\right|^2\left|w_{u,j} - w_{\hat{u},j}\right|\left|w_{u,j}\right|.\end{aligned}$$

The first term is $o_p(1)$ as shown for (1.42) in Lemma 1.12 part (c). By the Cauchy-Schwarz inequality, the second term is bounded by

$$\left(\frac{1}{T}\sum_{j=1}^{T-1}\left|w_{x,j}\right|^2\left|w_{u,j} - w_{\hat{u},j}\right|^2\right)^{\frac{1}{2}}\left(\frac{1}{T}\sum_{j=1}^{T-1}\left|w_{x,j}\right|^2\left|w_{u,j}\right|^2\right)^{\frac{1}{2}}$$

whose second factor is $O_p(1)$ because of (1.19). Therefore indeed $\hat{\Omega} \xrightarrow{p} \Omega$. ∎

Proof of Theorem 1.4. Write

$$\sqrt{T}\begin{pmatrix} \hat{\beta}^*(\tau)-\hat{\beta} \\ \hat{\delta}^*(\tau) \end{pmatrix} = \begin{pmatrix} \frac{1}{T}\sum_{j=1}^{T-1} I_{xx,j} & \frac{1}{T}\sum_{j=1}^{T-1} I_{xz,j}) \\ \frac{1}{T}\sum_{j=1}^{T-1} I_{zx,j} & \frac{1}{T}\sum_{j=1}^{T-1} I_{zz,j} \end{pmatrix}^{-1} \times$$
$$\times 2\,\mathrm{Re}\begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{j=1}^{[T/2]} w_{x,j}\left|w_{\hat{u},j}\right|\eta_j^* \\ \frac{1}{\sqrt{T}}\sum_{j=1}^{[T/2]} w_{z(\tau),j}\left|w_{\hat{u},j}\right|\eta_j^* \end{pmatrix}.$$

Applying Lemmas 1.9 and 1.12, it can be seen that Theorem 1.4 holds. ∎

1.B Lemmas

This section contains some results employed in Section 1.A.

Lemma 1.1 *Let g be a complex-valued function on $[0,\pi]$ which satisfies (a) $|g|^2$ is integrable on $[0,\pi]$, (b) $g(\lambda) = O\left(\lambda^{-d}\right)$ for $\lambda \to 0+$ for some $d < \frac{1}{2}$ and (c) g is bounded on any subinterval of $(0,\pi]$. Then for any $\alpha \geq 0$, $\beta \geq 0$ such that $2d\alpha + \beta < 1$,*

$$\frac{1}{T}\sum_{j=1}^{[T/2]} \frac{|g_j|^{2\alpha}}{\lambda_j^\beta} \to \frac{1}{2\pi}\int_0^\pi \frac{|g(\lambda)|^{2\alpha}}{\lambda^\beta}d\lambda.$$

Proof. Fix $\varepsilon > 0$. For any small η,

$$\begin{aligned} &\left|\frac{1}{T}\sum_{j=1}^{[T/2]} \frac{|g_j|^{2\alpha}}{\lambda_j^\beta} - \frac{1}{2\pi}\int_0^\pi \frac{|g(\lambda)|^{2\alpha}}{\lambda^\beta}d\lambda\right| \\ \leq\; & \frac{1}{T}\sum_{j=1}^{[\eta T]} \frac{|g_j|^{2\alpha}}{\lambda_j^\beta} + \frac{1}{2\pi}\int_0^\eta \frac{|g(\lambda)|^{2\alpha}}{\lambda^\beta}d\lambda \\ &+\left|\frac{1}{T}\sum_{j=[\eta T]+1}^{[T/2]} \frac{|g_j|^{2\alpha}}{\lambda_j^\beta} - \frac{1}{2\pi}\int_\eta^\pi \frac{|g(\lambda)|^{2\alpha}}{\lambda^\beta}d\lambda\right|. \end{aligned} \tag{1.20}$$

By assumption, for small enough $\eta > 0$ and $0 < \lambda < \eta$,

$$\frac{|g(\lambda)|^{2\alpha}}{\lambda^\beta} \leq C\lambda^{-2d\alpha-\beta} < C\lambda^{-1+\delta}$$

for some $\delta > 0$. Therefore

$$\frac{1}{T}\sum_{j=1}^{[\eta T]}\frac{|g_j|^{2\alpha}}{\lambda_j^{\beta}} \leq \frac{C}{T}\sum_{j=1}^{[\eta T]}\left(\frac{j}{T}\right)^{-1+\delta} \leq C\eta^{\delta}.$$

Similarly

$$\frac{1}{2\pi}\int_0^{\eta}\frac{|g(\lambda)|^{2\alpha}}{\lambda^{\beta}}d\lambda \leq C\eta^{\delta}.$$

The third term in (1.20) converges to zero by integrability of $|g|^2$. For small enough η and large enough T, the left-hand side of (1.20) is smaller than ε. ■

Lemma 1.2 *Let g be a complex-valued function on $[0,\pi]$ which satisfies (a) $g(-\lambda) = \bar{g}(\lambda)$ for all $\lambda \in (0,\pi]$, (b) $|g|^2$ is integrable on $[0,\pi]$, (c) $g(\lambda) = O(\lambda^{-d})$ for $\lambda \to 0+$ for some $d < \frac{1}{2}$ and (d) g is bounded on any subinterval of $(0,\pi]$. Under Conditions 1.1-1.3,*

$$\frac{2\pi}{\sqrt{T}}\sum_{j=1}^{T-1} g_j w_{\zeta(\tau),j}\bar{w}_{\varepsilon,j} \Longrightarrow \left(\sigma_{\xi}^2\sigma_{\varepsilon}^2\frac{1}{2\pi}\int_{-\pi}^{\pi}|g(\lambda)|^2 d\lambda\right)^{\frac{1}{2}} W(\tau) \qquad (1.21)$$

on $[0,1]$, where the sequence $\{\zeta_t(\tau)\}$ is defined as $\{\zeta_t(\tau)\} = \{\xi_t\mathbb{I}(t \leq [\tau T]),$ $t = 1,\ldots,T\}$.

Proof. The left-hand side of (1.21) can be written as

$$G_T(\tau) = \frac{1}{\sqrt{T}}\sum_{t=1}^{[\tau T]}\xi_t\left(\sum_{s=1}^{T}\varepsilon_s c_{t-s}\right),$$

where

$$c_t = \frac{1}{T}\sum_{j=1}^{T-1} g_j e^{it\lambda_j}.$$

Denoting $d_t = T^{-1/2}\sum_{s=1}^{T}\varepsilon_s c_{t-s}$, the process G_T can be written as

$$G_T(\tau) = \sum_{t=1}^{[\tau T]}\xi_t d_t.$$

The realizations of the process G_T belong to the space $D[0,1]$ of real functions which are right continuous with left hand limits. The sequence $\{\xi_t d_t, \mathcal{F}_{t-1} \vee \mathcal{G}_T,$

$1 \le t \le T\}$ is a martingale difference sequence. The first two moments of the process G_T are

$$EG_T(\tau) = 0,$$

$$E|G_T(\tau)|^2 = \sigma_\varepsilon^2 \sigma_\xi^2 \tfrac{[\tau T]}{T} \tfrac{1}{T} \textstyle\sum_{j=1}^{T-1} |g_j|^2 = \tfrac{[\tau T]}{T} E|G_T(1)|^2.$$

The variance of the process G_T therefore increases asymptotically linearly in τ and the weak convergence of the process G_T in (1.21) holds if the following two conditions of Scott (1973) are satisfied:

(a) $\sum_{t=1}^T E\left(|d_t\xi_t|^2 |\mathcal{F}_{t-1} \vee \mathcal{G}_T\right) \xrightarrow{p} \sigma_\xi^2 \sigma_\varepsilon^2 \frac{1}{2\pi} \int_{-\pi}^{\pi} |g(\lambda)|^2 d\lambda$ as $T \to \infty$ and

(b) $\sum_{t=1}^T E\left(|d_t\xi_t|^2 \mathbb{I}(|d_t\xi_t| > \delta) |\mathcal{F}_{t-1} \vee \mathcal{G}_T\right) \xrightarrow{p} 0$ for any positive δ.

These two conditions have been checked by Hidalgo (2003a) under similar assumptions on the weight function g and identical assumptions on the processes $\{\xi_t\}$, $\{\varepsilon_t\}$. After making appropriate adjustments for complex weight functions and replacing Lemma 1 there with our Lemma 1.1, the proof remains valid in our case. ■

Lemma 1.3 *Let h be a bounded variation function on $[0,1]$. Let $H(\lambda) = \sum_{t=1}^T h(t/T) e^{it\lambda}$. Then for a constant $0 < C < \infty$ independent of T,*

(a) $|H(\lambda)| \le \frac{C}{|\lambda|}$ for $\lambda \in (0,\pi]$,

(b) $\int_0^{\lambda_j} |H(\lambda)| d\lambda = O(\log j)$ uniformly over $1 \le j \le [T/2]$.

Proof. (a) Letting $D_t(\lambda) = \sum_{k=1}^t e^{ik\lambda}$, noting that

$$|D_t(\lambda)| = \left| e^{i\lambda(1+t)/2} \frac{\sin \frac{\lambda t}{2}}{\sin \frac{\lambda}{2}} \right| \le \frac{\pi}{|\lambda|}$$

for $0 < \lambda \le \pi$, and using summation by parts, we have

$$\begin{aligned} |H(\lambda)| &\le \sum_{t=1}^{T-1} |D_t(\lambda)| \left| h\left(\frac{t}{T}\right) - h\left(\frac{t+1}{T}\right) \right| + |D_T(\lambda)| |h(1)| \\ &\le \frac{\pi}{|\lambda|} \left(\sum_{t=1}^{T-1} \left| h\left(\frac{t}{T}\right) - h\left(\frac{t+1}{T}\right) \right| + |h(1)| \right) \\ &\le \frac{C}{|\lambda|} \end{aligned}$$

due to the boundedness of the total variation of the function h.

(b)

$$\begin{aligned}\int_0^{\lambda_j} |H(\lambda)|\, d\lambda &= \int_0^{\frac{1}{T}} |H(\lambda)|\, d\lambda + \int_{\frac{1}{T}}^{\lambda_j} |H(\lambda)|\, d\lambda \le T \int_0^{\frac{1}{T}} d\lambda + \int_{\frac{1}{T}}^{\lambda_j} \frac{C}{\lambda} d\lambda \\ &= O(\log j).\end{aligned}$$

■

Lemma 1.4 *Let h be a bounded variation function on $[0,1]$. Let $\{x_t\}$ be a covariance stationary process satisfying Conditions 1.1, 1.2 and 1.4. Let $H_T(\lambda) = \sum_{t=1}^{T} h(t/T)\, e^{it\lambda}$ and $K_{h,T}(\lambda) = \frac{1}{2\pi T} |H_T(\lambda)|^2$. Then*

$$\int_{-\pi}^{\pi} \left| \frac{A(e^{i\lambda})}{A_j} - 1 \right|^2 K_{h,T}(\lambda - \lambda_j)\, d\lambda = O\left(\frac{1}{j}\right) \quad \text{as } T \to \infty$$

uniformly over integers $1 \le j \le [T/2]$.

Proof. The function A satisfies assumptions A1, A2' of Robinson (1995b). Furthermore, the kernel H_T has the property

$$|H_T(\lambda)| \le \frac{\pi}{|\lambda|}, \qquad 0 < \lambda \le \pi,\ T \ge 1,$$

by Lemma 1.3. Therefore the lemma is valid by the arguments of Robinson (1995b) in the proof of his Lemma 3. ■

Lemma 1.5 *Let $\{x_{1t}\}$, $\{x_{2t}\}$ be scalar covariance stationary processes satisfying Conditions 1.1, 1.2 and 1.4. Let h_1, h_2 be bounded variation functions on $[0,1]$. Denote by A_k the transfer functions of the processes $\{x_{kt}\}$, $k = 1,2$. Let $v_k(\lambda_j) = \sqrt{2\pi} w_{k,j} / \left(\sigma_\xi^2 A_k\left(e^{i\lambda_j}\right)\right)$, where $\{w_{k,j}, j = 1, \ldots, T\}$ is the discrete Fourier transform of the sequence $\{h_k(t/T)\, x_{kt}, t = 1, \ldots, T\}$. Then*

(a) $E\{v_k(\lambda_j)\, \bar{v}_l(\lambda_j)\} = \frac{1}{T} \sum_{t=1}^{T} h_k\left(\frac{t}{T}\right) h_l\left(\frac{t}{T}\right) + O\left(\frac{\log j}{j}\right)$ and

(b) $E\{v_k(\lambda_j)\, v_l(\lambda_j)\} = O(1)$

uniformly over integers $1 \le j \le [T/2]$, for k, $l = 1,2$.

Proof. (a) Denote $A_k\left(e^{i\lambda_j}\right) = A_{k,j}$. We have

$$\begin{aligned}
& Ew_{k,j}\bar{w}_{l,j} - \frac{\sigma_\xi^2}{2\pi}\left(\frac{1}{T}\sum_{t=1}^{T} h_k\left(\frac{t}{T}\right) h_l\left(\frac{t}{T}\right)\right) A_{k,j}\bar{A}_{l,j} \\
= & \frac{\sigma_\xi^2}{2\pi}\int_{-\pi}^{\pi}\left(A_k(\lambda)\bar{A}_l(\lambda) - A_{k,j}\bar{A}_{l,j}\right) K_{kl}(\lambda - \lambda_j)\, d\lambda,
\end{aligned}$$

where

$$K_{kl}(\lambda) = \frac{1}{2\pi T}\bar{H}_k(\lambda) H_l(\lambda)$$

and

$$H_k(\lambda) = \sum_{t=1}^{T} h_k\left(\frac{t}{T}\right) e^{it\lambda}, \qquad k = 1, 2.$$

Condition 1.4 implies that we can choose $\eta > 0$ such that for $\lambda \in (-\eta, 0) \cup (0, \eta)$, for some d_k, $d_l \in \left[0, \frac{1}{2}\right)$ and for some $0 < C < \infty$, $\left|A_k(\lambda)\bar{A}_l(\lambda)\right| \leq C|\lambda|^{-(d_k+d_l)}$ and $\left|\frac{d}{d\lambda}A_k(\lambda)\bar{A}_l(\lambda)\right| \leq C|\lambda|^{-(d_k+d_l)-1}$. Furthermore by Lemma 1.3 the kernels K_{kl} and H_k display properties required in the proof of Theorem 2 of Robinson (1995a), namely $K_{kl}(\lambda) = O\left(T^{-1}\lambda^{-2}\right)$ for $0 < |\lambda| \leq \pi$ and $\int_{-D\lambda_j}^{D\lambda_j}|H_k(\lambda)|\, d\lambda = O(\log j)$, $k = 1, 2$. The proof of part (a) therefore follows as in the first part of case (a) of Theorem 2 of Robinson (1995a). We obtain

$$Ew_{k,j}\bar{w}_{l,j} - \frac{\sigma_\xi^2}{2\pi}\left(\frac{1}{T}\sum_{t=1}^{T} h_k\left(\frac{t}{T}\right) h_l\left(\frac{t}{T}\right)\right) A_{k,j}\bar{A}_{l,j} = O\left(\frac{\log j}{j}\lambda_j^{-(d_k+d_l)}\right)$$

from which it can be deduced that

$$Ev_{k,j}\bar{v}_{l,j} = \frac{1}{T}\sum_{t=1}^{T} h_k\left(\frac{t}{T}\right) h_l\left(\frac{t}{T}\right) + O\left(\frac{\log j}{j}\right)$$

as required.

Part (b) follows from part (a) by the Cauchy-Schwarz inequality. ■

Lemma 1.6 *Let g be a function satisfying assumptions of Lemma 1.1. Then for any $\alpha > 0, \beta \geq 0$, $\delta \geq 0$ and $\gamma \geq 1$,*

(a) $\frac{1}{T^\alpha}\sum_{j=1}^{[T/2]}|g_j|^{2\alpha}\frac{\log^\delta j}{j^\beta} = \begin{cases} o(1) & 2d\alpha + \beta \geq 1, \\ O\left(T^{-\alpha-\beta+1}\log^\delta T\right) & 2d\alpha + \beta < 1, \end{cases}$

(b) $\frac{1}{T^{\alpha}}\sum_{j=1}^{[T/2]}\sum_{k=1}^{[T/2]} g_j^{\alpha}\bar{g}\left(\lambda_k\right)^{\alpha}\left(\frac{\log^{\delta} j}{j^{\beta}}\frac{\log^{\delta} k}{k^{\beta}}\right)^{\frac{1}{2}}\frac{1}{|j-k|_+^{\gamma}}$

$$= \begin{cases} o(1) & 2d\alpha+\beta\geq 1, \\ O\left(T^{-\alpha-\beta+1}\log^{\delta+1} T\right) & 2d\alpha+\beta<1. \end{cases}$$

Proof. (a) The assumptions of the lemma imply that $|g(\lambda)| \leq C\lambda^{-d}$ for $0<\lambda\leq\pi$ for some C. Therefore

$$\frac{1}{T^{\alpha}}\sum_{j=1}^{[T/2]}|g_j|^{2\alpha}\frac{\log^{\delta} j}{j^{\beta}} \leq C\frac{\log^{\delta} T}{T^{\alpha}}\sum_{j=1}^{[T/2]}\lambda_j^{-2d\alpha}j^{-\beta} = CT^{-\alpha+2d\alpha}\log^{\delta}\sum_{j=1}^{[T/2]}j^{-2d\alpha-\beta}$$

$$= \begin{cases} O\left(T^{\alpha(2d-1)}\log^{\delta} T\right) & 2d\alpha+\beta>1, \\ O\left(T^{\alpha(2d-1)}\log^{\delta+1} T\right) & 2d\alpha+\beta=1, \\ O\left(T^{1-\alpha-\beta}\log^{\delta} T\right) & 2d\alpha+\beta<1. \end{cases} \qquad (1.22)$$

From here, the part (a) follows easily.

(b) By the Cauchy-Schwarz inequality, the sum in question is bounded by

$$\begin{aligned} \frac{1}{T^{\alpha}}\sum_{j=1}^{[T/2]}|g_j|^{2\alpha}\frac{\log^{\delta} j}{j^{\beta}}\sum_{k=1}^{[T/2]}\frac{1}{|j-k|_+^{\gamma}} &\leq \frac{C}{T^{\alpha}}\sum_{j=1}^{[T/2]}|g_j|^{2\alpha}\frac{\log^{\delta} j}{j^{\beta}}\sum_{k=1}^{[T/2]}\frac{1}{k} \\ &\leq C\log^{\delta+1} T\frac{1}{T^{\alpha}}\sum_{j=1}^{[T/2]}\frac{|g_j|^{2\alpha}}{j^{\beta}} \end{aligned}$$

and part (b) now follows from (1.22). ∎

Lemma 1.7 *Under Conditions 1.1-1.5,*

$$\begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{j=1}^{T-1} I_{xu,j} \\ \frac{1}{\sqrt{T}}\sum_{j=1}^{T-1} I_{zu,j} \end{pmatrix} \Longrightarrow \begin{pmatrix} \frac{1}{2\pi}\Omega^{\frac{1}{2}}W(1) \\ \frac{1}{2\pi}\Omega^{\frac{1}{2}}W(\tau) \end{pmatrix}$$

over $[0,1]$.

Proof. It suffices to show that $T^{-1/2}\sum_{j=1}^{T-1} I_{zu,j} \Longrightarrow (2\pi)^{-1}\Omega^{\frac{1}{2}}W(\tau)$ over $[0,1]$. The function

$$g(\lambda) = \frac{1}{2\pi}A\left(e^{i\lambda}\right)\bar{B}\left(e^{i\lambda}\right)$$

satisfies the conditions of Lemma 1.2. The present lemma is then proved if

$$\frac{1}{\sqrt{T}}\sum_{j=1}^{T-1} A\left(e^{i\lambda_j}\right)\bar{B}\left(e^{i\lambda_j}\right)\left(\frac{w_{z(\tau)}\left(\lambda_j\right)\bar{w}_u\left(\lambda_j\right)}{A\left(e^{i\lambda_j}\right)\bar{B}\left(e^{i\lambda_j}\right)} - w_{\zeta(\tau)}\left(\lambda_j\right)\bar{w}_\varepsilon\left(\lambda_j\right)\right) \Longrightarrow 0. \tag{1.23}$$

The left of (1.23) can be written as

$$Y_1(\tau) + Y_2(\tau) + Y_3(\tau),$$

where

$$Y_1(\tau) = \frac{1}{\sqrt{T}}\sum_{j=1}^{T-1} A_j\bar{B}_j\left(\frac{w_{z(\tau),j}}{A_j} - w_{\zeta(\tau),j}\right)\left(\frac{\bar{w}_{u,j}}{\bar{B}_j} - \bar{w}_{\varepsilon,j}\right),$$

$$Y_2(\tau) = \frac{1}{\sqrt{T}}\sum_{j=1}^{T-1} A_j\bar{B}_j\left(\frac{w_{z(\tau),j}}{A_j} - w_{\zeta(\tau),j}\right)\bar{w}_{\varepsilon,j}$$

and

$$Y_3(\tau) = \frac{1}{\sqrt{T}}\sum_{j=1}^{T-1} A_j\bar{B}_j w_{\zeta(\tau),j}\left(\frac{\bar{w}_{u,j}}{\bar{B}_j} - \bar{w}_{\varepsilon,j}\right). \tag{1.24}$$

Processes Y_1, Y_2 and Y_3 are of the form

$$Y_i(\tau) = \frac{1}{\sqrt{T}}\sum_{j=1}^{T-1} g_j V_j(\tau)\bar{W}_j, \quad i = 1, 2, 3,$$

where $V_j(\tau)$ and $\bar{W}_j$ stand for the third and the fourth factor, respectively, of the summands of the processes Y_i. To prove that $Y_i \Longrightarrow 0$ for $\tau \in [0,1]$ it suffices to show that finite dimensional distributions of the process Y_i converge to zero in probability and that the process Y_i is tight. Take any $n \in N$, any numbers $\tau_1, \ldots, \tau_n$ from the interval $[0,1]$ and any finite complex constants $\alpha_1, \ldots, \alpha_n$. The first moment of $\sum_{l=1}^n \alpha_l Y_i(\tau_l)$ is zero for $i = 1, 2, 3$. The second moment is

$$\frac{1}{T}\sum_{j,k=1}^{T-1} g_j\bar{g}_k Es_{jk}EW_j\bar{W}_k \leq \frac{4}{T}\sum_{j,k=1}^{[T/2]} \left|g_j\bar{g}_k\right|\left|Es_{jk}\right|\left|EW_j\bar{W}_k\right|,$$

where

$$s_{jk} = \sum_{l,m=1}^{n} \alpha_l \bar{\alpha}_m V_j(\tau_l) \bar{V}_k(\tau_m).$$

For $i = 1, 2$ the factor $V_j(\tau)$ is equal to $w_{z(\tau),j}/A_j - w_{\zeta(\tau),j}$. The total variation of functions $h_\tau(x) = \mathbb{I}(0 \le x \le \tau)$, $\tau \in (0,1]$ is equal to one, therefore by Lemma 1.5 part (a)

$$\sup_{\tau\in[0,1]} E\left|\frac{w_{z(\tau),j}}{A_j} - w_{\zeta(\tau),j}\right|^2 \le \frac{D\log j}{j} \tag{1.25}$$

as $T \to \infty$ uniformly over integers $1 \le j \le [T/2]$. Using the Cauchy-Schwarz inequality,

$$|Es_{jk}| \le D\left(\frac{\log j}{j}\frac{\log k}{k}\right)^{\frac{1}{2}} \sum_{l,m=1}^{n} |\alpha_l|\,|\alpha_m| \le D\left(\frac{\log j}{j}\frac{\log k}{k}\right)^{\frac{1}{2}}.$$

When $i = 3$ the factor $V_j(\tau)$ equals $w_{\zeta(\tau),j}$. For any τ, $\sigma \in [0,1]$,

$$Ew_{\zeta(\tau),j}\bar{w}_{\zeta(\sigma),k} = \frac{1}{2\pi T}\sum_{t,s=1}^{[\tau T]} E\xi_t\xi_s e^{it\lambda_j - is\lambda_k} = \frac{\sigma_\xi^2}{2\pi}\frac{1}{T}\sum_{t=1}^{[(\tau\wedge\sigma)T]} e^{it(\lambda_j-\lambda_k)}.$$

For $j = k$, the last expression is equal to $\frac{\sigma_\xi^2}{2\pi}\frac{[(\tau\wedge\sigma)T]}{T}$, while for $j \ne k$,

$$\begin{aligned}\frac{1}{T}\left|\sum_{t=1}^{[(\tau\wedge\sigma)T]} e^{it(\lambda_j-\lambda_k)}\right| &= \frac{1}{T}\left|\frac{\sin\left([(\tau\wedge\sigma)T]\frac{\lambda_j-\lambda_k}{2}\right)}{\sin\left(\frac{\lambda_j-\lambda_k}{2}\right)}\right| \le \frac{1}{T}\frac{1}{\left|\sin\left(\frac{\lambda_j-\lambda_k}{2}\right)\right|} \\ &\le \frac{1}{T}\frac{\pi}{|\lambda_j-\lambda_k|} = \frac{1}{2|j-k|}.\end{aligned}$$

In sum,

$$Ew_{\zeta(\tau),j}\bar{w}_{\zeta(\sigma),k} = O\left(\frac{1}{|j-k|_+}\right) \tag{1.26}$$

uniformly over $(\tau, \sigma) \in [0,1]^2$ and $1 \leq j, k \leq [T/2]$, where $|j-k|_+ = \max\{1, |j-k|\}$. Therefore when $i=3$,

$$|Es_{jk}| \leq D \sum_{l,m=1}^{n} |\alpha_l| |\alpha_m| \frac{1}{|j-k|_+} \leq \frac{D}{|j-k|_+}.$$

Turning to the factor W_j, for $i = 1, 3$ it is equal to $w_{u,j}/B_j - w_{\varepsilon,j}$. By Lemma 1.5 part (a),

$$E\left|\frac{w_{u,j}}{B_j} - w_{\varepsilon,j}\right|^2 \leq \frac{D \log j}{j} \tag{1.27}$$

as $T \to \infty$ and by the Cauchy-Schwarz inequality,

$$\left|EW_j \bar{W}_k\right| \leq D \left(\frac{\log j}{j} \frac{\log k}{k}\right)^{\frac{1}{2}}.$$

In case $i=2$, $W_j = w_{\varepsilon,j}$ and

$$E\bar{w}_{\varepsilon,j} w_{\varepsilon,k} = \frac{1}{2\pi T} \sum_{t,s=1}^{T} E\varepsilon_t \varepsilon_s e^{-it\lambda_j + is\lambda_k} = \frac{\sigma_\varepsilon^2}{2\pi T} \sum_{t=1}^{T} e^{-it(\lambda_j - \lambda_k)} = \frac{\sigma_\varepsilon^2}{2\pi} \mathbb{I}(j=k).$$

Collecting the bounds obtained for moments of the factors $V_j(\tau)$ and W_j and using Lemma 1.6, the following results are obtained:

$$E\left|\sum_{l=1}^{n} \alpha_l Y_1(\tau_l)\right|^2 \leq \frac{D}{T} \sum_{j,k=1}^{[T/2]} |g_j g_k| \frac{\log j}{j} \frac{\log k}{k} = o(1),$$

$$\begin{aligned} E\left|\sum_{l=1}^{n} \alpha_l Y_2(\tau_l)\right|^2 &\leq \frac{D}{T} \sum_{j,k=1}^{[T/2]} |g_j g_k| \left(\frac{\log j}{j} \frac{\log k}{k}\right)^{\frac{1}{2}} \mathbb{I}(j=k) \\ &= \frac{D}{T} \sum_{j=1}^{[T/2]} |g_j|^2 \frac{\log j}{j} = o(1), \end{aligned}$$

$$E\left|\sum_{l=1}^{n} \alpha_l Y_3(\tau_l)\right|^2 \leq \frac{D}{T} \sum_{j,k=1}^{[T/2]} |g_j g_k| \left(\frac{\log j}{j} \frac{\log k}{k}\right)^{\frac{1}{2}} \frac{1}{|j-k|_+} = o(1).$$

An application of the Cramér-Wold device together with the Markov inequality establishes convergence of finite dimensional distributions of processes Y_i, $i = 1, 2, 3$, to zero in probability.

Tightness of the processes Y_i is implied by the moment condition of Billingsley (1999), Theorem 13.5, page 142,

$$E\left|Y_i(\rho) - Y_i(\sigma)\right|^2 \left|Y_i(\tau) - Y_i(\rho)\right|^2 \leq \left(F(\tau) - F(\sigma)\right)^{2\alpha}, \quad i = 1, 2, 3, \tag{1.28}$$

where $\alpha > \frac{1}{2}$, $\sigma \leq \rho \leq \tau$ and F is a nondecreasing, continuous function on $[0, 1]$. The fourth moment of the difference $Y_i(\tau) - Y_i(\sigma)$ is given by

$$\begin{aligned} & E\left|Y_i(\tau) - Y_i(\sigma)\right|^4 \\ \leq\ & \frac{16}{T^2} \sum_{j,k,l,m=1}^{[T/2]} \left|g_j \bar{g}_k g_l \bar{g}_m\right| \left|EV_j \bar{V}_k V_l \bar{V}_m\right| \left|EW_j \bar{W}_k W_l \bar{W}_m\right|, \end{aligned}$$

where $V_j = V_j(\tau) - V_j(\sigma)$. For $i = 1, 2$, $V_j = \left(w_{z(\tau),j} - w_{z(\sigma),j}\right)/A_j - \left(w_{\zeta(\tau),j} - w_{\zeta(\sigma),j}\right)$ and

$$\begin{aligned} & \operatorname{cum}\left(V_j, \bar{V}_k, V_l, \bar{V}_m\right) \\ =\ & \frac{\kappa_\xi}{(2\pi)^5} \frac{1}{T^2} \iiint_{-\pi}^{\pi} \left(\frac{A\left(e^{i\lambda}\right)}{A_j} - 1\right) \left(\frac{A\left(e^{i\mu}\right)}{\bar{A}_k} - 1\right) \\ & \times \left(\frac{A\left(e^{i\zeta}\right)}{A_l} - 1\right) \left(\frac{A\left(e^{i(-\lambda-\mu-\zeta)}\right)}{\bar{A}_m} - 1\right) \\ & \times H(\lambda + \lambda_j) H(\mu - \lambda_k) H(\zeta + \lambda_l) H(-\lambda - \mu - \zeta - \lambda_m)\, d\lambda d\mu d\zeta, \end{aligned}$$

where $\kappa_\xi = \operatorname{cum}(\xi_t, \xi_t, \xi_t, \xi_t)$, $H(\lambda) = \sum_{t=1}^{T} h(t/T) e^{it\lambda}$ and $h(x) = \mathbb{I}(\sigma \leq x \leq \tau)$. Proceeding as in the proof of (4.8) in Robinson (1995b), we get

$$\left|\operatorname{cum}\left(V_j, \bar{V}_k, V_l, \bar{V}_m\right)\right| \leq D P_j^{\frac{1}{2}} P_k^{\frac{1}{2}} P_l^{\frac{1}{2}} P_m^{\frac{1}{2}},$$

where

$$P_j = \int_{-\pi}^{\pi} \left|\frac{A\left(e^{i\lambda}\right)}{A_j} - 1\right|^2 \frac{1}{2\pi T} \left|H(\lambda + \lambda_j)\right|^2 d\lambda.$$

Denoting $K_{h,T}(\lambda) = (2\pi T)^{-1} |H(\lambda)|^2$, it can be seen that

$$\begin{aligned} P_j &= (\tau - \sigma) \int_{-\pi}^{\pi} \left| \frac{A(e^{i\lambda})}{A_j} - 1 \right|^2 K_{1,(\tau-\sigma)T}(\lambda + \lambda_j)\, d\lambda \\ &= (\tau - \sigma) O\left(\frac{1}{j}\right) \end{aligned}$$

uniformly over $(\tau, \sigma) \in [0,1]^2$ and $1 \le j \le [T/2]$ by Lemma 1.4.

Likewise

$$\begin{aligned} E|V_j|^2 &= \frac{\sigma_\xi^2}{2\pi} \int_{-\pi}^{\pi} \left| \frac{A(e^{i\lambda})}{A_j} - 1 \right|^2 \frac{1}{2\pi T} |H(\lambda + \lambda_j)|^2\, d\lambda \\ &= DP_j = (\tau - \sigma) O\left(\frac{1}{j}\right). \end{aligned}$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned} & \left| E V_j \bar{V}_k V_l \bar{V}_m \right| \\ \le\ & \left| \operatorname{cum}\left(V_j, \bar{V}_k, V_l, \bar{V}_m\right) \right| + 3 \left(E|V_j|^2 E|V_k|^2 E|V_l|^2 E|V_m|^2 \right)^{\frac{1}{2}} \\ \le\ & C(\tau - \sigma)^2 j^{-\frac{1}{2}} k^{-\frac{1}{2}} l^{-\frac{1}{2}} m^{-\frac{1}{2}}. \end{aligned} \tag{1.29}$$

For $i = 3$, $V_j = w_{\xi(\tau)-\xi(\sigma),j}$ and

$$\begin{aligned} \operatorname{cum}\left(V_j, \bar{V}_k, V_l, \bar{V}_m\right) &= \frac{\kappa_\xi}{(2\pi)^5} \frac{1}{T^2} \iiint_{-\pi}^{\pi} H(\lambda + \lambda_j) \\ & \times H(\mu - \lambda_k) H(\zeta + \lambda_j) H(-\lambda - \mu - \zeta - \lambda_k)\, d\lambda d\mu d\zeta \end{aligned}$$

which by using periodicity of H and the Cauchy-Schwarz inequality can be shown to be $(\tau - \sigma)^2 O(1)$ uniformly over $(\tau, \sigma) \in [0,1]^2$ and $1 \le j, k, l, m \le [T/2]$. Similarly,

$$E|V_j|^2 = \frac{\sigma_\xi^2}{4\pi^2 T} \int_{-\pi}^{\pi} |H(\lambda + \lambda_j)|^2\, d\lambda \le C(\tau - \sigma),$$

and so for $i = 3$, $\left| E V_j \bar{V}_k V_l \bar{V}_m \right| = (\tau - \sigma)^2 O(1)$.

Regarding the factor W_j, for $i = 1, 3$ we have $W_j = w_{u,j}/B_j - w_{\varepsilon,j}$ and reasoning as in case of V_j $(i = 1, 2)$ we obtain

$$\left|\operatorname{cum}\left(W_j, \bar{W}_k, W_l, \bar{W}_m\right)\right| \leq DP_{B,j}^{\frac{1}{2}} P_{B,k}^{\frac{1}{2}} P_{B,l}^{\frac{1}{2}} P_{B,m}^{\frac{1}{2}}$$

and

$$E\left|W_j\right|^2 = DP_{B,j},$$

where

$$P_{B,j} = \int_{-\pi}^{\pi} \left|\frac{B\left(e^{i\lambda}\right)}{B_j} - 1\right|^2 \frac{1}{2\pi T} \left|H\left(\lambda + \lambda_j\right)\right|^2 d\lambda$$

with $h \equiv 1$ in the definition of $H(\lambda)$. By Lemma 1.4, $P_{B,j} = O\left(j^{-1}\right)$, therefore

$$EW_j\bar{W}_k W_l \bar{W}_m = O\left(j^{-\frac{1}{2}} k^{-\frac{1}{2}} l^{-\frac{1}{2}} m^{-\frac{1}{2}}\right)$$

uniformly over $1 \leq j, k, l, m \leq [T/2]$.

Finally, when $i = 2$, $W_j = w_{\varepsilon,j}$,

$$\operatorname{cum}\left(W_j, \bar{W}_k, W_l, \bar{W}_m\right) = \frac{\kappa_\xi}{4\pi^2} \frac{1}{T^2} \sum_{t=1}^{T} e^{it(\lambda_j - \lambda_k + \lambda_l - \lambda_m)} = O\left(\frac{1}{T}\right)$$

and

$$EW_j\bar{W}_k = \frac{1}{2\pi}\mathbb{I}\left(j = k\right) = O\left(1\right)$$

uniformly over $1 \leq j, k, l, m \leq [T/2]$.

Due to the bounds obtained above for moments of V_j and W_j, the following inequalities hold:

$$\begin{aligned}
E\left|Y_1(\tau) - Y_1(\sigma)\right|^4 &\leq \frac{D}{T^2} \sum_{j,k,l,m=1}^{[T/2]} \left|g_j g_k g_l g_m\right| (\tau - \sigma)^2 j^{-1} k^{-1} l^{-1} m^{-1} \\
&= D(\tau - \sigma)^2 \left(\frac{1}{T^{\frac{1}{2}}} \sum_{j=1}^{[T/2]} \frac{\left|g_j\right|}{j}\right)^4 = D(\tau - \sigma)^2 o(1)
\end{aligned}$$

uniformly over $(\tau, \sigma) \in [0,1]^2$ by Lemma 1.1,

$$\begin{aligned} E\left|Y_2(\tau) - Y_2(\sigma)\right|^4 &\leq \frac{D}{T^2} \sum_{j,k,l,m=1}^{[T/2]} \left|g_j g_k g_l g_m\right| (\tau-\sigma)^2 j^{-\frac{1}{2}} k^{-\frac{1}{2}} l^{-\frac{1}{2}} m^{-\frac{1}{2}} \\ &= D(\tau-\sigma)^2 \left(\frac{1}{T^{\frac{1}{2}}} \sum_{j=1}^{[T/2]} \frac{|g_j|}{j^{\frac{1}{2}}} \right)^4 = (\tau-\sigma)^2 o(1) \end{aligned}$$

by the Cauchy-Schwarz inequality and Lemma 1.1. The same bound applies to $E\left|Y_3(\tau) - Y_3(\sigma)\right|^4$. By the Cauchy-Schwarz inequality,

$$E\left|Y_i(\rho) - Y_i(\sigma)\right|^2 \left|Y_i(\tau) - Y_i(\rho)\right|^2 \leq D\left((\rho-\sigma)^2 (\tau-\sigma)^2\right)^{\frac{1}{2}} \leq D(\tau-\sigma)^2$$

for $i = 1, 2, 3$ and the moment condition (1.28) is verified with $\alpha = 2$ and $F(\tau) = D\tau^2$. This proves the uniform convergence in (1.23). The lemma is established. ■

Lemma 1.8 *Let g be a function satisfying the assumptions of Lemma 1.2. Let h_1, h_2 be bounded variation functions on $[0,1]$. Let $\{w_{h\xi,j}, j = 1, \ldots, T\}$ be the discrete Fourier transform of the sequence $\{h(t/T)\xi_t, t = 1, \ldots, T\}$. Under Conditions 1.1-1.2,*

$$\frac{2\pi}{T} \sum_{j=1}^{T-1} |g_j|^2 w_{h_1\xi,j} \bar{w}_{h_2\xi,j} \xrightarrow{p} \frac{\sigma_\xi^2}{2\pi} \int_{-\pi}^{\pi} |g(\lambda)|^2 d\lambda \int_0^1 h_1(t) h_2(t) dt.$$

Proof. Denote $h_{kt} = h_k(t/T)$ and

$$Z = \frac{2\pi}{T} \sum_{j=1}^{T-1} |g_j|^2 w_{h_1\xi,j} \bar{w}_{h_2\xi,j}.$$

We have

$$\begin{aligned} EZ &= \frac{2\pi}{T} \sum_{j=1}^{T-1} |g_j|^2 \frac{1}{2\pi T} \sum_{t,s=1}^{T} E\xi_t \xi_s h_{1t} h_{2s} e^{i(t-s)\lambda_j} \\ &= \frac{\sigma_\xi^2}{T^2} \sum_{j=1}^{T-1} |g_j|^2 \sum_{t=1}^{T} h_{1t} h_{2t} \to \frac{\sigma_\xi^2}{2\pi} \int_{-\pi}^{\pi} |g(\lambda)|^2 d\lambda \int_0^1 h_1(t) h_2(t) dt \end{aligned}$$

by Lemma 1.1. Further,

$$E|Z|^2 = \frac{1}{T^4}\sum_{j,k=1}^{T-1}|g_jg_k|^2 \sum_{t,s,r,v=1}^{T} E\left(\xi_t\xi_s\xi_r\xi_v\right) h_{1t}h_{2s}h_{1r}h_{2v}e^{i(t-s)\lambda_j}e^{-i(r-v)\lambda_k}$$

$$\begin{aligned} &= \frac{\kappa_\xi}{T^4}\sum_{j,k=1}^{T-1}|g_jg_k|^2\sum_{t=1}^{T}h_{1t}^2h_{2t}^2 + \frac{\sigma_\xi^4}{T^4}\sum_{j,k=1}^{T-1}|g_jg_k|^2\left(\sum_{t=1}^{T}h_{1t}h_{2t}\right)^2 \\ &+\frac{\sigma_\xi^4}{T^4}\sum_{j,k=1}^{T-1}|g_jg_k|^2\sum_{t=1}^{T}h_{1t}^2e^{it(\lambda_j-\lambda_k)}\sum_{t=1}^{T}h_{2t}^2e^{-it(\lambda_j-\lambda_k)} \\ &+\frac{\sigma_\xi^4}{T^4}\sum_{j,k=1}^{T-1}|g_jg_k|^2\left|\sum_{t=1}^{T}h_{1t}h_{2t}e^{it(\lambda_j+\lambda_k)}\right|^2. \end{aligned}$$

The first term is $O\left(T^{-1}\right)$ by Lemma 1.1. Proceeding as in the computations leading to (1.26), it can be seen that

$$\sum_{t=1}^{T}h_{lt}^2e^{it(\lambda_j-\lambda_k)} \le \frac{CT}{|j-k|_+}, \qquad l=1,2.$$

Therefore the third term is bounded in absolute value by

$$\frac{D}{T^2}\sum_{j,k=1}^{[T/2]}|g_jg_k|^2\frac{1}{|j-k|_+^2}$$

which is $o(1)$ by Lemma 1.6.

Similarly, the fourth term is $o(1)$. The second term is dominant and converges to

$$\left(\frac{\sigma_\xi^2}{2\pi}\int_{-\pi}^{\pi}|g(\lambda)|^2d\lambda\int_0^1 h_1(t)\,h_2(t)\,dt\right)^2$$

by Lemma 1.1. In sum,

$$EZ \to \frac{\sigma_\xi^2}{2\pi}\int_{-\pi}^{\pi}|g(\lambda)|^2d\lambda\int_0^1 h_1(t)\,h_2(t)\,dt$$

and $E|Z|^2 \to |EZ|^2$. An application of the Markov inequality completes the proof. ∎

Lemma 1.9 *Let h_1, h_2 be bounded variation functions on $[0,1]$. Let $\{w_{hx,j}, j=1,\ldots,T\}$ be the discrete Fourier transform of the sequence $\{h(t/T)x_t,$ $t=1,\ldots,T\}$. Under Conditions 1.1-1.5,*

$$\frac{1}{T}\sum_{j=1}^{T-1} w_{h_1x,j}\bar{w}_{h_2x,j} \xrightarrow{p} \frac{1}{2\pi}\Sigma\int_0^1 h_1(t)\,h_2(t)\,dt.$$

Proof. The function $g(\lambda) = A\left(e^{i\lambda}\right)/\sqrt{2\pi}$ satisfies the conditions of Lemma 1.2. It is sufficient to prove that

$$\frac{1}{T}\sum_{j=1}^{T-1}|A_j|^2\left(\frac{w_{h_1x,j}\bar{w}_{h_2x,j}}{|A_j|^2} - w_{h_1\xi,j}\bar{w}_{h_2\xi,j}\right) \xrightarrow{p} 0. \tag{1.30}$$

The left-hand side of (1.30) is equal to

$$\begin{aligned}
&\frac{1}{T}\sum_{j=1}^{T-1}|A_j|^2\left(\frac{w_{h_1x,j}}{A_j} - w_{h_1\xi,j}\right)\left(\frac{\bar{w}_{h_2x,j}}{\bar{A}_j} - \bar{w}_{h_2\xi,j}\right)\\
&+\frac{1}{T}\sum_{j=1}^{T-1}|A_j|^2\left(\frac{w_{h_1x,j}}{A_j} - w_{h_1\xi,j}\right)\bar{w}_{h_2\xi,j}\\
&+\frac{1}{T}\sum_{j=1}^{T-1}|A_j|^2 w_{h_1\xi,j}\left(\frac{\bar{w}_{h_2x,j}}{\bar{A}_j} - \bar{w}_{h_2\xi,j}\right).
\end{aligned}$$

By the Cauchy-Schwarz inequality, the expectation of the modulus of the first term is bounded by

$$\begin{aligned}
&\frac{2}{T}\sum_{j=1}^{[T/2]}|A_j|^2\left(E\left|\frac{w_{h_1x,j}}{A_j} - w_{h_1\xi,j}\right|^2\right)^{\frac{1}{2}}\left(E\left|\frac{w_{h_2x,j}}{A_j} - w_{h_2\xi,j}\right|^2\right)^{\frac{1}{2}}\\
\leq\ &\frac{D}{T}\sum_{j=1}^{[T/2]}|A_j|^2\frac{\log j}{j} = o(1)
\end{aligned}$$

by Lemma 1.5 part (a) and Lemma 1.6. A bound for the expectation of the absolute value of the second term is

$$
\begin{aligned}
& \frac{2}{T}\sum_{j=1}^{[T/2]} |A_j|^2 \left(E\left|\frac{w_{h_1 x,j}}{A_j} - w_{h_1 \xi,j}\right|^2 \right)^{\frac{1}{2}} \left(E\left|w_{h_2 \xi,j}\right|^2\right)^{\frac{1}{2}} \\
\leq & \frac{D}{T}\sum_{j=1}^{[T/2]} |A_j|^2 \left(\frac{\log j}{j}\right)^{\frac{1}{2}} = o(1)
\end{aligned}
$$

by Lemma 1.5 part (a) and Lemma 1.6. The third term can be bounded in the same way as the second term. Therefore (1.30) holds and by Lemma 1.8,

$$
\begin{aligned}
\frac{1}{T}\sum_{j=1}^{T-1} w_{h_1 x,j}\bar{w}_{h_2 x,j} & \xrightarrow{p} \frac{\sigma_\xi^2}{2\pi}\int_0^\pi \tfrac{1}{2\pi}\left|A\left(e^{i\lambda}\right)\right|^2 d\lambda \int_0^1 h_1(t)\, h_2(t)\, dt \\
& = \frac{1}{2\pi}\Sigma \int_0^1 h_1(t)\, h_2(t)\, dt.
\end{aligned}
$$

■

Lemma 1.10 *Under Conditions 1.1-1.5, with a function g satisfying the conditions of Lemma 1.2,*

$$
\frac{1}{T}\sum_{j=1}^{T-1} |g_j|^2 w_{\zeta(\tau),j}\bar{w}_{\zeta(\sigma),j}\left|w_{\varepsilon,j}\right|^2 \Longrightarrow (\tau\wedge\sigma)\frac{\sigma_\xi^2\sigma_\varepsilon^2}{4\pi^2}\frac{1}{2\pi}\int_{-\pi}^{\pi} |g(\lambda)|^2 d\lambda \qquad (1.31)
$$

uniformly over $(\tau,\sigma) \in [0,1]^2$.

Proof. First moment of the expression on the left of (1.31) is

$$
\begin{aligned}
\frac{1}{T}\sum_{j=1}^{T-1} |g_j|^2 E w_{\zeta(\tau),j}\bar{w}_{\zeta(\sigma),j} E\left|w_{\varepsilon,j}\right|^2 & = \frac{[(\tau\wedge\sigma)T]}{T}\frac{\sigma_\xi^2\sigma_\varepsilon^2}{4\pi^2}\frac{1}{T}\sum_{j=1}^{T-1}|g_j|^2 \\
& \to (\tau\wedge\sigma)\frac{\sigma_\xi^2\sigma_\varepsilon^2}{4\pi^2}\frac{1}{2\pi}\int_{-\pi}^{\pi}|g(\lambda)|^2 d\lambda
\end{aligned}
$$

by Lemma 1.1 because

$$Ew_{\zeta(\tau),j}\bar{w}_{\zeta(\sigma),j} = \frac{1}{2\pi T}\sum_{t,s=1}^{[\tau T]} E\xi_t\xi_s e^{i(t-s)\lambda_j} = \frac{[(\tau\wedge\sigma)T]}{T}\frac{\sigma_\xi^2}{2\pi}. \tag{1.32}$$

Second moment of the expression on the left of (1.31) is

$$\frac{1}{T^2}\sum_{j,k=1}^{T-1}|g_j g_k|^2 Ew_{\zeta(\tau),j}\bar{w}_{\zeta(\sigma),j}\bar{w}_{\zeta(\tau),k}w_{\zeta(\sigma),k}E|w_{\varepsilon,j}|^2|w_{\varepsilon,k}|^2$$

$$\begin{aligned}
= & \frac{1}{T^2}\sum_{j,k=1}^{T-1}|g_j g_k|^2 \operatorname{cum}\left(w_{\zeta(\tau),j},\bar{w}_{\zeta(\sigma),j},\bar{w}_{\zeta(\tau),k},w_{\zeta(\sigma),k}\right)E|w_{\varepsilon,j}|^2|w_{\varepsilon,k}|^2 \\
& +\frac{1}{T^2}\sum_{j,k=1}^{T-1}|g_j g_k|^2 Ew_{\zeta(\tau),j}\bar{w}_{\zeta(\sigma),j}E\bar{w}_{\zeta(\tau),k}w_{\zeta(\sigma),k}E|w_{\varepsilon,j}|^2|w_{\varepsilon,k}|^2 \\
& +\frac{1}{T^2}\sum_{j,k=1}^{T-1}|g_j g_k|^2 Ew_{\zeta(\tau),j}\bar{w}_{\zeta(\tau),k}E\bar{w}_{\zeta(\sigma),j}w_{\zeta(\sigma),k}E|w_{\varepsilon,j}|^2|w_{\varepsilon,k}|^2 \\
& +\frac{1}{T^2}\sum_{j,k=1}^{T-1}|g_j g_k|^2 Ew_{\zeta(\tau),j}w_{\zeta(\sigma),k}E\bar{w}_{\zeta(\tau),k}\bar{w}_{\zeta(\sigma),j}E|w_{\varepsilon,j}|^2|w_{\varepsilon,k}|^2.
\end{aligned} \tag{1.33}$$

Now

$$\begin{aligned}
&\operatorname{cum}\left(w_{\zeta(\tau),j},\bar{w}_{\zeta(\sigma),j},\bar{w}_{\zeta(\tau),k},w_{\zeta(\sigma),k}\right) = \\
&\quad = \frac{1}{4\pi^2T^2}\sum_{t=1}^{[(\tau\wedge\sigma)T]}\operatorname{cum}(\xi_t,\xi_t,\xi_t,\xi_t) = \frac{\kappa_\xi}{4\pi^2}\frac{1}{T}\frac{[(\tau\wedge\sigma)T]}{T} = O\left(\frac{1}{T}\right)
\end{aligned}$$

uniformly over $(\tau,\sigma)\in[0,1]^2$. The fourth moments of ε_t are finite, therefore the first term of (1.33) is bounded by $DT^{-3}\sum_{j,k=1}^{[T/2]}|g_jg_k|^2$ which is $O(T^{-1})$ by Lemma 1.1. Further, from (1.26),

$$\left|Ew_{\zeta(\tau),j}\bar{w}_{\zeta(\sigma),k}\right| \leq \frac{C}{|j-k|_+}$$

uniformly over $(\tau, \sigma) \in [0,1]^2$ and $1 \leq j,k \leq [T/2]$, and the third term of (1.33) is bounded by $DT^{-2}\sum_{j,k=1}^{[T/2]} |g_j g_k|^2 |j-k|_+^{-2}$ which is $o(1)$ by Lemma 1.6. Similarly, the fourth term is $o(1)$. Therefore we are left with the dominant second term,

$$\begin{aligned}
& E\left|\frac{1}{T}\sum_{j=1}^{T-1} |g_j|^2 w_{\zeta(\tau),j}\bar{w}_{\zeta(\sigma),j} |w_{\varepsilon,j}|^2\right|^2 \\
= & \left(\frac{[(\tau\wedge\sigma)T]}{T}\right)^2 \frac{\sigma_\xi^4}{4\pi^2}\frac{1}{T^2}\sum_{j,k=1}^{T-1} |g_j g_k|^2 E|w_{\varepsilon,j}|^2 |w_{\varepsilon,k}|^2 + o(1).
\end{aligned}$$

Since

$$\operatorname{cum}(w_{\varepsilon,j}, \bar{w}_{\varepsilon,j}, w_{\varepsilon,k}, \bar{w}_{\varepsilon,k}) = \frac{1}{4\pi^2 T^2}\sum_{t=1}^{T}\operatorname{cum}(\varepsilon_t, \varepsilon_t, \varepsilon_t, \varepsilon_t) = \frac{\kappa}{4\pi^2 T}$$

and $Ew_{\varepsilon,j}\bar{w}_{\varepsilon,k} = \mathbb{I}(j=k)\,\sigma_\varepsilon^2/(2\pi)$, we have

$$\begin{aligned}
\frac{1}{T^2}\sum_{j,k=1}^{T-1} |g_j g_k|^2 E|w_{\varepsilon,j}|^2 |w_{\varepsilon,k}|^2 &= \frac{\kappa}{4\pi^2 T}\frac{1}{T^2}\sum_{j,k=1}^{T-1} |g_j g_k|^2 + \frac{\sigma_\varepsilon^4}{4\pi^2}\frac{1}{T^2}\sum_{j,k=1}^{T-1} |g_j g_k|^2 \\
&\quad + \frac{\sigma_\varepsilon^4}{4\pi^2}\frac{1}{T^2}|g_{T/2}|^4\,\mathbb{I}(T\text{ even}) + \frac{\sigma_\varepsilon^4}{4\pi^2}\frac{1}{T^2}\sum_{j=1}^{T-1}|g_j|^4 \\
&= \frac{\sigma_\varepsilon^4}{4\pi^2}\frac{1}{T^2}\sum_{j,k=1}^{T-1} |g_j g_k|^2 + o(1)
\end{aligned}$$

by Lemmas 1.1 and 1.6. That means that

$$E\left|\frac{1}{T}\sum_{j=1}^{T-1} |g_j|^2 w_{\zeta(\tau),j}\bar{w}_{\zeta(\sigma),j} |w_{\varepsilon,j}|^2\right|^2 \to \left((\tau\wedge\sigma)\frac{\sigma_\varepsilon^2\sigma_\xi^2}{4\pi^2}\frac{1}{2\pi}\int_{-\pi}^{\pi}|g(\lambda)|^2 d\lambda\right)^2.$$

The second moment of the process $T^{-1}\sum_{j=1}^{T-1}|g_j|^2 w_{\zeta(\tau),j}\bar{w}_{\zeta(\sigma),j}|w_{\varepsilon,j}|^2$ therefore converges to the square of the limit of its first moment. By Markov inequality,

$$\frac{1}{T}\sum_{j=1}^{T-1} |g_j|^2 w_{\zeta(\tau),j}\bar{w}_{\zeta(\sigma),j} |w_{\varepsilon,j}|^2 \xrightarrow{p} (\tau\wedge\sigma)\frac{\sigma_\varepsilon^2\sigma_\xi^2}{4\pi^2}\frac{1}{2\pi}\int_{-\pi}^{\pi}|g(\lambda)|^2 d\lambda$$

for each $(\tau, \sigma) \in [0,1]^2$. Since the limiting function is continuous and increasing in τ and σ, the convergence is uniform. ■

Lemma 1.11 *Let $h_1, \ldots, h_4$ be bounded variation functions on $[0,1]$. Let $\{x_t\}$ be a covariance stationary process satisfying Conditions 1.1, 1.2 and 1.4. Let $\{w_{h_r x,j}, j = 1, \ldots, T\}$ be the discrete Fourier transform of the sequence $\{h_r(t/T)\, x_t, t = 1, \ldots, T\}$, $r = 1, \ldots 4$. Let $I_{h_r x, h_s x, j} = w_{h_r x,j} \bar{w}_{h_s x,j}$. Then*

$$\frac{1}{T} \sum_{j=1}^{T-1} I_{h_1 x, h_2 x, j} I_{h_3 x, h_4 x, j} = o_p(T).$$

Proof. We have

$$\frac{1}{T} \sum_{j=1}^{T-1} I_{h_1 x, h_2 x, j} I_{h_3 x, h_4 x, j} = \frac{1}{T} \sum_{j=1}^{T-1} f_{xx,j}^2 \left(a_j + b_j + c_j + d_j\right), \tag{1.34}$$

where

$$\begin{aligned}
a_j &= \left(\frac{I_{h_1 x, h_2 x, j}}{f_{xx,j}} - 2\pi \frac{I_{h_1 \xi, h_2 \xi, j}}{\sigma_\xi^2}\right) \left(\frac{I_{h_3 x, h_4 x, j}}{f_{xx,j}} - 2\pi \frac{I_{h_3 \xi, h_4 \xi, j}}{\sigma_\xi^2}\right), \\
b_j &= \left(\frac{I_{h_1 x, h_2 x, j}}{f_{xx,j}} - 2\pi \frac{I_{h_1 \xi, h_2 \xi, j}}{\sigma_\xi^2}\right) 2\pi \frac{I_{h_3 \xi, h_4 \xi, j}}{\sigma_\xi^2}, \\
c_j &= 2\pi \frac{I_{h_1 \xi, h_2 \xi, j}}{\sigma_\xi^2} \left(\frac{I_{h_3 x, h_4 x, j}}{f_{xx,j}} - 2\pi \frac{I_{h_3 \xi, h_4 \xi, j}}{\sigma_\xi^2}\right) \text{ and} \\
d_j &= \frac{4\pi^2}{\sigma_\xi^4} I_{h_1 \xi, h_2 \xi, j} I_{h_3 \xi, h_4 \xi, j}.
\end{aligned}$$

The second moment of the first factor of a_j is

$$E \left|u_{1j} \bar{u}_{2j} - v_{1j} \bar{v}_{2j}\right|^2 = a_{1j} + a_{2j},$$

where

$$u_{r,j} = \frac{\sqrt{2\pi}}{\sigma_\xi} \frac{w_{h_r x, j}}{A_j}, \quad v_{r,j} = \frac{\sqrt{2\pi}}{\sigma_\xi} w_{h_r \xi, j},$$

$$
\begin{aligned}
a_{1j} &= \operatorname{cum}\left(u_{1j}, \overline{u}_{2j}, \overline{u}_{1j}, u_{2j}\right) - \operatorname{cum}\left(u_{1j}, \overline{u}_{2j}, \overline{v}_{1j}, v_{2j}\right) \\
&\quad - \operatorname{cum}\left(v_{1j}, \overline{v}_{2j}, \overline{u}_{1j}, u_{2j}\right) + \operatorname{cum}\left(v_{1j}, \overline{v}_{2j}, \overline{v}_{1j}, v_{2j}\right)
\end{aligned}
$$

and, denoting $h_{rs} = \frac{1}{T}\sum_{t=1}^{T} h_r(t/T)\, h_s(t/T)$ for $r, s = 1, 2$,

$$
\begin{aligned}
& a_{2j} = \\
& \quad \left(Eu_{1j}\overline{u}_{2j} - h_{12}\right)\left(E\overline{u}_{1j}u_{2j} - h_{12}\right) + \left(Eu_{1j}\overline{u}_{2j} - h_{12}\right) + \left(E\overline{u}_{1j}u_{2j} - h_{12}\right) \\
& +h_{12}^2 + \left(Eu_{1j}\overline{u}_{1j} - h_{11}\right)\left(E\overline{u}_{2j}u_{2j} - h_{22}\right) + \left(Eu_{1j}\overline{u}_{1j} - h_{11}\right) \\
& + \left(E\overline{u}_{2j}u_{2j} - h_{22}\right) + h_{11}h_{22} + Eu_{1j}u_{2j}E\overline{u}_{2j}\overline{u}_{1j} \\
& \\
& - \left(Eu_{1j}\overline{u}_{2j} - h_{12}\right)\left(E\overline{v}_{1j}v_{2j} - h_{12}\right) - \left(Eu_{1j}\overline{u}_{2j} - h_{12}\right) - \left(E\overline{v}_{1j}v_{2j} - h_{12}\right) \\
& -h_{12}^2 - \left(Eu_{1j}\overline{v}_{1j} - h_{11}\right)\left(E\overline{u}_{2j}v_{2j} - h_{22}\right) - \left(Eu_{1j}\overline{v}_{1j} - h_{11}\right) \\
& - \left(E\overline{u}_{2j}v_{2j} - h_{22}\right) - h_{11}h_{22} - Eu_{1j}v_{2j}E\overline{u}_{2j}\overline{v}_{1j} \\
& \\
& - \left(Ev_{1j}\overline{v}_{2j} - h_{12}\right)\left(E\overline{u}_{1j}u_{2j} - h_{12}\right) - \left(Ev_{1j}\overline{v}_{2j} - h_{12}\right) - \left(E\overline{u}_{1j}u_{2j} - h_{12}\right) \\
& -h_{12}^2 - \left(Ev_{1j}\overline{u}_{1j} - h_{11}\right)\left(E\overline{v}_{2j}u_{2j} - h_{22}\right) - \left(Ev_{1j}\overline{u}_{1j} - h_{11}\right) \\
& - \left(E\overline{v}_{2j}u_{2j} - h_{22}\right) - h_{11}h_{22} - Ev_{1j}u_{2j}E\overline{v}_{2j}\overline{u}_{1j} \\
& \\
& + \left(Ev_{1j}\overline{v}_{2j} - h_{12}\right)\left(E\overline{v}_{1j}v_{2j} - h_{12}\right) + \left(Ev_{1j}\overline{v}_{2j} - h_{12}\right) + \left(E\overline{v}_{1j}v_{2j} - h_{12}\right) \\
& +h_{12}^2 + \left(Ev_{1j}\overline{v}_{1j} - h_{11}\right)\left(E\overline{v}_{2j}v_{2j} - h_{22}\right) + \left(Ev_{1j}\overline{v}_{1j} - h_{11}\right) \\
& + \left(E\overline{v}_{2j}v_{2j} - h_{22}\right) + h_{11}h_{22} + Ev_{1j}v_{2j}E\overline{v}_{2j}\overline{v}_{1j}.
\end{aligned}
$$

The term a_{1j} is equal to

$$
\begin{aligned}
& \frac{\kappa}{\sigma_\xi^4}\frac{1}{(2\pi)^3}\frac{1}{T^2}\iiint_{-\pi}^{\pi}\left(\frac{A\left(e^{i\lambda}\right)A\left(e^{i\mu}\right)}{|A_j|^2} - 1\right)\left(\frac{A\left(e^{i\zeta}\right)A\left(e^{i(-\lambda-\mu-\zeta)}\right)}{|A_j|^2} - 1\right) \\
& \times H_1\left(\lambda + \lambda_j\right) H_2\left(\mu - \lambda_j\right) H_1\left(\zeta + \lambda_j\right) H_2\left(-\lambda - \mu - \zeta - \lambda_j\right) d\lambda d\mu d\zeta,
\end{aligned}
\tag{1.35}
$$

where $H_r(\lambda) = \sum_{t=1}^{T} h_r(t/T)\, e^{it\lambda}$, $r = 1, 2$. Proceeding as in the proof of (4.8) in Robinson (1995b), expression (1.35) can be written as a sum of components of three types. The first component is

$$\frac{\kappa}{\sigma_{\xi}^{4}}\frac{1}{(2\pi)^{3}}\frac{1}{T^{2}}\iiint_{-\pi}^{\pi}\left(\frac{A\left(e^{i\lambda}\right)}{A_j}-1\right)\left(\frac{A\left(e^{i\mu}\right)}{\bar{A}_j}-1\right)$$
$$\times\left(\frac{A\left(e^{i\zeta}\right)}{A_j}-1\right)\left(\frac{A\left(e^{i(-\lambda-\mu-\zeta)}\right)}{\bar{A}_j}-1\right)$$
$$\times H_1\left(\lambda+\lambda_j\right)H_2\left(\mu-\lambda_j\right)H_1\left(\zeta+\lambda_j\right)H_2\left(-\lambda-\mu-\zeta-\lambda_j\right)d\lambda d\mu d\zeta.$$

Using the Cauchy-Schwarz inequality, periodicity of the integrand and the fact that $\int_{-\pi}^{\pi}\left|H_r\left(\lambda\right)\right|^2 d\lambda = O\left(T\right)$, this component can be shown to be bounded in absolute value by

$$CP_{1,j}P_{2,j},$$

where

$$P_{r,j}=\int_{-\pi}^{\pi}\left|\frac{A\left(e^{i\lambda}\right)}{A_j}-1\right|^2 K_r\left(\lambda-\lambda_j\right)d\lambda$$

and $K_r\left(\lambda\right)=\left|H_r\left(\lambda\right)\right|^2/\left(2\pi T\right)$.

A typical representative of the second type of component of (1.35) is

$$\frac{\kappa}{\sigma_{\xi}^{4}}\frac{1}{(2\pi)^{3}}\frac{1}{T^{2}}\iiint_{-\pi}^{\pi}\left(\frac{A\left(e^{i\lambda}\right)}{A_j}-1\right)\left(\frac{A\left(e^{i\mu}\right)}{\bar{A}_j}-1\right)\left(\frac{A\left(e^{i\zeta}\right)}{A_j}-1\right)$$
$$\times H_1\left(\lambda+\lambda_j\right)H_2\left(\mu-\lambda_j\right)H_1\left(\zeta+\lambda_j\right)H_2\left(-\lambda-\mu-\zeta-\lambda_j\right)d\lambda d\mu d\zeta$$

whose absolute value can be similarly shown to be bounded by

$$CP_{1,j}P_{2,j}^{\frac{1}{2}}.$$

The last type of component is exemplified by

$$\frac{\kappa}{\sigma_{\xi}^{4}}\frac{1}{(2\pi)^{3}}\frac{1}{T^{2}}\iiint_{-\pi}^{\pi}\left(\frac{A\left(e^{i\lambda}\right)}{A_j}-1\right)\left(\frac{A\left(e^{i\zeta}\right)}{A_j}-1\right)$$
$$\times H_1\left(\lambda+\lambda_j\right)H_2\left(\mu-\lambda_j\right)H_1\left(\zeta+\lambda_j\right)H_2\left(-\lambda-\mu-\zeta-\lambda_j\right)d\lambda d\mu d\zeta$$

$$
\begin{aligned}
&= \frac{\kappa}{\sigma_\xi^4}\frac{1}{(2\pi)^3}\frac{1}{T^2}\iiint_{-\pi}^{\pi}\left(\frac{A\left(e^{i\lambda}\right)}{A_j}-1\right)\left(\frac{A\left(e^{i\zeta}\right)}{A_j}-1\right) \\
&\quad \times H_1\left(\lambda+\lambda_j\right)H_2\left(-\lambda-\zeta-\theta-\lambda_j\right)H_1\left(\zeta+\lambda_j\right)H_2\left(\theta-\lambda_j\right)d\lambda d\theta d\zeta
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{\kappa}{\sigma_\xi^4}\frac{1}{(2\pi)^2}\frac{1}{T^2}\iint_{-\pi}^{\pi}\left(\frac{A\left(e^{i\lambda}\right)}{A_j}-1\right)\left(\frac{A\left(e^{i\zeta}\right)}{A_j}-1\right) \\
&\quad \times H_1\left(\lambda+\lambda_j\right)H_1\left(\zeta+\lambda_j\right)H_2^{(2)}\left(-\lambda-\zeta-2\lambda_j\right)d\lambda d\zeta \qquad (1.36)
\end{aligned}
$$

since

$$
\int_{-\pi}^{\pi} H_r\left(u+\lambda\right)H_r\left(v-\lambda\right)d\lambda = 2\pi H_r^{(2)}\left(u+v\right),
$$

where $H_r^{(2)}\left(\lambda\right)=\sum_{t=1}^{T}h_r^2\left(t/T\right)e^{i\lambda t}$. Since $\int_{-\pi}^{\pi}\left|H_r^{(2)}\left(\lambda\right)\right|d\lambda=O\left(T\right)$, the modulus of (1.36) is bounded by

$$
CT^{-\frac{1}{2}}P_{1,j}.
$$

By Lemma 1.4 the term a_{1j} is $O\left(j^{-2}+j^{-\frac{3}{2}}+j^{-1}T^{-\frac{1}{2}}\right)$. Applying Lemma 1.5 gives $a_{2j}=O\left(1\right)$. Therefore the first factor of a_j is $O\left(1\right)$. Likewise, the second factor of a_j, and therefore a_j itself, is $O\left(1\right)$.

Denoting $h_{rt}=h_r\left(t/T\right)$, the second moment of $I_{h_1\xi,h_2\xi,j}$ is

$$
\begin{aligned}
E\left|I_{h_1\varepsilon,h_2\varepsilon,j}\right|^2 &= \frac{1}{4\pi^2T^2}\sum_{t,s,r,v=1}^{T}h_{1t}h_{1s}h_{2r}h_{2v}E\varepsilon_t\varepsilon_s\varepsilon_r\varepsilon_v e^{i(t-s+r-v)\lambda_j} \\
&= \frac{1}{4\pi^2T^2}\left(E\varepsilon_t^4\sum_{t=1}^{T}h_{1t}^2h_{2t}^2+\sigma_\varepsilon^4\sum_{t=1}^{T}\sum_{\substack{s=1\\s\neq t}}^{T}h_{1t}^2h_{2s}^2\right. \\
&\quad \left.+\sigma_\varepsilon^4\sum_{t=1}^{T}\sum_{\substack{s=1\\s\neq t}}^{T}h_{1t}h_{1s}h_{2t}h_{2s}e^{i(t-s)2\lambda_j}+\sigma_\varepsilon^4\sum_{t=1}^{T}\sum_{\substack{s=1\\s\neq t}}^{T}h_{1t}h_{1s}h_{2s}h_{2t}\right) \\
&= O\left(1\right)
\end{aligned}
$$

because the fourth moments of ε_t are finite. In the same way, the factor $I_{h_3\xi,h_4\xi,j}$ is $O\left(1\right)$. Using the Cauchy-Schwarz inequality, the sum $a_j+b_j+c_j+d_j$ in (1.34) is $O\left(1\right)$ uniformly over integers $1\leq j\leq[T/2]$. The proof of the lemma is then completed by applying Lemma 1.6 part (a) with $g\left(\lambda\right)=A\left(e^{i\lambda}\right)$.

■

Lemma 1.12 *Under Conditions 1.1-1.5,*

$$2\,\mathrm{Re}\begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{j=1}^{[T/2]} w_{x,j}\left|w_{\hat{u},j}\right|\eta_j^* \\ \frac{1}{\sqrt{T}}\sum_{j=1}^{[T/2]} w_{z(\tau),j}\left|w_{\hat{u},j}\right|\eta_j^* \end{pmatrix} \overset{p}{\Longrightarrow} \begin{pmatrix} \frac{1}{2\pi}\Omega^{\frac{1}{2}}W(1) \\ \frac{1}{2\pi}\Omega^{\frac{1}{2}}W(\tau) \end{pmatrix}$$

over $\tau \in [0,1]$.

Proof. Define $\eta_{T-j}^* = \overline{\eta_j^*}$ for $j = [T/2]+1, \dots, T-1$. Then

$$2\,\mathrm{Re}\frac{1}{\sqrt{T}}\sum_{j=1}^{[T/2]} w_{z(\tau),j}\left|w_{\hat{u},j}\right|\eta_j^* = \frac{1}{\sqrt{T}}\sum_{j=1}^{T-1} w_{z(\tau),j}\left|w_{\hat{u},j}\right|\eta_j^* + r(\tau),$$

where $r(\tau) = T^{-1/2}w_{z(\tau),T/2}\left|w_{\hat{u},T/2}\right|\eta_{T/2}^*\mathbb{I}(T\text{ even}) = O_p\left(T^{-1/2}\right)$ uniformly over $\tau \in [0,1]$. It is therefore sufficient to show that $T^{-1/2}\sum_{j=1}^{T-1} w_{z(\tau),j}\left|w_{\hat{u},j}\right|\eta_j^*$ $\overset{p}{\Longrightarrow} (2\pi)^{-1}\Omega^{\frac{1}{2}}W(\tau)$ over $\tau \in [0,1]$. We need to prove that

(a)

$$\frac{1}{\sqrt{T}}\sum_{j=1}^{T-1} g_j w_{\zeta(\tau),j}\left|w_{\varepsilon,j}\right|\eta_j^* \overset{p}{\Longrightarrow} \frac{1}{2\pi}\Omega^{\frac{1}{2}}W(\tau), \tag{1.37}$$

(b)

$$\frac{1}{\sqrt{T}}\sum_{j=1}^{T-1}\left(g_j w_{\zeta(\tau),j}\left|w_{\varepsilon,j}\right|\eta_j^* - w_{z(\tau),j}\left|w_{u,j}\right|\eta_j^*\right) \overset{p}{\Longrightarrow} 0 \quad \text{and} \tag{1.38}$$

(c)

$$\frac{1}{\sqrt{T}}\sum_{j=1}^{T-1}\left(w_{z(\tau),j}\left|w_{u,j}\right|\eta_j^* - w_{z(\tau),j}\left|w_{\hat{u},j}\right|\eta_j^*\right) \overset{p}{\Longrightarrow} 0 \tag{1.39}$$

over $\tau \in [0,1]$ for any $\varepsilon > 0$, where $g(\lambda) = A\left(e^{i\lambda}\right)\bar{B}\left(e^{i\lambda}\right)$.

To prove the convergence in part (a), we need to show that finite dimensional distributions of the process $Y_T = T^{-1/2}\sum_{j=1}^{T-1} g_j w_{\zeta(\tau),j}\left|w_{\varepsilon,j}\right|\eta_j^*$ converge in probability to the finite dimensional distributions of a centered Gaussian process with covariance function $K(\tau,\sigma) = (\tau\wedge\sigma)\Omega/\left(4\pi^2\right)$ and that the process Y_T is tight. First, $E^*Y_T(\tau) = 0$ and

$$\mathrm{var}^*\,Y_T(\tau) = \frac{1}{T}\sum_{j=1}^{T-1}\left|g_j\right|^2\left|w_{\zeta(\tau),j}\right|^2\left|w_{\varepsilon,j}\right|^2.$$

By Lemma 1.10, the last expression converges in probability to

$$\tau \frac{1}{2\pi} \frac{\sigma_\xi^2 \sigma_\varepsilon^2}{4\pi^2} \int_{-\pi}^{\pi} |g(\lambda)|^2 \, d\lambda = \frac{\tau}{4\pi^2} \Omega.$$

Second, we need to show that the Lindeberg condition is satisfied,

$$\sum_{j=1}^{T-1} E^* \left| T^{-\frac{1}{2}} A_j B_j w_{\zeta(\tau),j} \left| w_{\varepsilon,j} \right| \eta_j^* \right|^2 \mathbb{I} \left(\left| T^{-\frac{1}{2}} A_j B_j w_{\zeta(\tau),j} \left| w_{\varepsilon,j} \right| \eta_j^* \right|^2 > \varepsilon \right) \xrightarrow{p} 0 \tag{1.40}$$

for each $\varepsilon > 0$.

We examine $\sup_\tau T^{-1} |A_j B_j|^2 |I_{\zeta\zeta,j} I_{\varepsilon\varepsilon,j}|$. From An et al. (1983), we have

$$\sup_{j=1,\ldots,[T/2]} \left(\frac{2\pi}{\sigma_\varepsilon^2} \frac{1}{\log T} |w_{\varepsilon,j}|^2 \right) \leq 1 \quad \text{a.s.}$$

and

$$\sup_{j=1,\ldots,[T/2]} \left(\frac{2\pi}{\sigma_\xi^2} \frac{1}{\log T} |w_{\xi,j}|^2 \right) \leq 1 \quad \text{a.s.}$$

Therefore

$$\begin{aligned} \sup_{j=1,\ldots,[T/2]} \frac{1}{T} |A_j B_j|^2 |I_{\zeta\zeta,j} I_{\varepsilon\varepsilon,j}| &\leq D \sup_j \frac{1}{T} |A_j B_j|^2 \log^2 T \quad \text{a.s.} \\ &\leq D T^{2(d_x + d) - 1} \log^2 T \quad \text{a.s.} \end{aligned}$$

As η_j^*, given the data, are independent identically distributed variables, the sum in (1.40) is bounded by

$$E^* |\eta_j^*|^2 \mathbb{I} \left(|\eta_j^*|^2 > \varepsilon T^{1-2d} \log^{-2} T \right) \frac{2}{T} \sum_{j=1}^{[T/2]} |A_j B_j|^2 I_{\zeta\zeta,j} I_{\varepsilon\varepsilon,j}.$$

The first factor converges to zero since η_j^* has finite moments and $1 - 2d > 0$. The second factor is $O_p(1)$ by Lemma 1.10 with $g_j = A_j \bar{B}_j$. Therefore the left-hand side of (1.40) is $o_p(1)$ and by the Lindeberg-Feller central limit theorem the pointwise convergence

$$\frac{1}{\sqrt{T}}\sum_{j=1}^{T-1} g_j w_{\zeta(\tau),j} \left|w_{\varepsilon,j}\right| \eta_j^* \xrightarrow{d^*} N\left(0, \frac{\tau}{4\pi^2}\Omega\right)$$

is proved.

Further,

$$\operatorname{cov}^*\left(Y_T\left(\tau\right), Y_T\left(\sigma\right)\right) = \frac{1}{T}\sum_{j=1}^{T-1} \left|g_j\right|^2 w_{\zeta(\tau),j}\bar{w}_{\zeta(\sigma),j} \left|w_{\varepsilon,j}\right|^2$$

which converges in probability to $(\tau \wedge \sigma)\,\Omega/\left(4\pi^2\right)$ by Lemma 1.10. The proof of convergence of the finite dimensional distributions in part (a) is completed by using the Cramér-Wold device.

We now prove tightness of the process $Y_T\left(\tau\right)$. By Theorem 13.5 of Billingsley (1999) it is sufficient to check the moment condition

$$E^*\left|Y_T\left(\rho\right) - Y_T\left(\sigma\right)\right|^2 \left|Y_T\left(\tau\right) - Y_T\left(\rho\right)\right|^2 \leq \left(1 + o_p\left(1\right)\right)\left(F\left(\tau\right) - F\left(\sigma\right)\right)^{\alpha}, \tag{1.41}$$

where $\alpha > \frac{1}{2}$, $\sigma \leq \rho \leq \tau$, F is a nondecreasing, continuous function on Λ and $o_p\left(1\right)$ is uniform over $(\tau, \sigma) \in \Lambda^2$. Denoting $w_j = w_{\zeta(\tau),j} - w_{\zeta(\sigma),j}$, we have

$$\begin{aligned}
& E^*\left|Y_T\left(\tau\right) - Y_T\left(\sigma\right)\right|^4 \\
= \; & \frac{1}{T^2}\sum_{j=1}^{T-1} \left|g_j\right|^4 \left|w_j\right|^4 \left|w_{\varepsilon,j}\right|^4 E^*\left|\eta_j^*\right|^4 \\
& + \frac{2}{T^2}\sum_{j=1}^{T-1}\sum_{\substack{k=1 \\ k\neq j}}^{T-1} \left|g_j g_k\right|^2 \left|w_j\right|^2 \left|w_k\right|^2 \left|w_{\varepsilon,j}\right|^2 \left|w_{\varepsilon,k}\right|^2 E^*\left|\eta_j^*\right|^2 \left|\eta_k^*\right|^2 \\
& + \frac{1}{T^2}\sum_{j=1}^{T-1}\sum_{\substack{k=1 \\ k\neq j}}^{T-1} g_j^2 \bar{g}_k^2 w_j^2 \bar{w}_k^2 \left|w_{\varepsilon,j}\right|^2 \left|w_{\varepsilon,k}\right|^2 E^*\eta_j^{*2}\bar{\eta}_k^{*2}
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{C}{T^2}\sum_{j,k=1}^{T-1}\left|g_j g_k\right|^2\left|w_j\right|^2\left|w_k\right|^2\left|w_{\varepsilon,j}\right|^2\left|w_{\varepsilon,k}\right|^2 \\
&= C\left(\frac{1}{T}\sum_{j=1}^{T-1}\left|g_j\right|^2\left|w_j\right|^2\left|w_{\varepsilon,j}\right|^2\right)^2.
\end{aligned}
$$

By Lemma 1.10,

$$
C\left(\frac{1}{T}\sum_{j=1}^{T-1}\left|g_j\right|^2\left|w_j\right|^2\left|w_{\varepsilon,j}\right|^2\right)^2 \xrightarrow{p} C\left(\tau-\sigma\right)^2\frac{1}{4\pi^2}\Omega
$$

uniformly over $(\tau,\sigma)\in[0,1]^2$. It follows that by the Cauchy-Schwarz inequality the left-hand side of (1.41) is bounded by $D^2\left(\tau-\sigma\right)^2\left(1+o_p\left(1\right)\right)$ since $\left(\tau-\rho\right)\left(\rho-\sigma\right)\leq\left(\tau-\sigma\right)^2$. The moment condition (1.41) is thus verified with $F\left(\tau\right)=D\tau$ and $\alpha=2$. This establishes tightness in probability of the process Y_T and completes the proof of the uniform convergence in part (a).

For the convergence in part (b), we have

$$
\begin{aligned}
& E^*\left|\frac{1}{\sqrt{T}}\sum_{j=1}^{T-1}\left(A_j\bar{B}_j w_{\zeta(\tau),j}\left|w_{\varepsilon,j}\right|-w_{z(\tau),j}\left|w_{u,j}\right|\right)\eta_j^*\right| \\
\leq\ & \frac{D}{\sqrt{T}}\sum_{j=1}^{T-1}\left|A_jB_j\right|\left|\frac{w_{z(\tau),j}}{A_j}-w_{\zeta(\tau),j}\right|\left|\frac{w_{u,j}}{B_j}-w_{\varepsilon,j}\right| \\
& +\frac{D}{\sqrt{T}}\sum_{j=1}^{T-1}\left|A_jB_j\right|\left|\frac{w_{z(\tau),j}}{A_j}-w_{\zeta(\tau),j}\right|\left|w_{\varepsilon,j}\right| \\
& +\frac{D}{\sqrt{T}}\sum_{j=1}^{T-1}\left|A_jB_j\right|\left|w_{\zeta(\tau),j}\right|\left|\frac{w_{u,j}}{B_j}-w_{\varepsilon,j}\right|
\end{aligned}
$$

and proceeding as in the proof of Lemma 1.7 it can be shown that the last expression is $o_p\left(1\right)$ uniformly over $\tau\in[0,1]$.

To verify the convergence in part (c), we write the difference between errors and residuals under the local alternative as

$$u_t - \hat{u}_t = (\hat{\alpha} - \alpha) + \left(\hat{\beta} - \beta\right) x_t + \hat{\delta}\hat{z}_t - \frac{1}{\sqrt{T}} x_t h_t,$$

where $\hat{z}_t = z_t(\hat{\tau}) = x_t \mathbb{I}(t \leq [\hat{\tau}T])$. Therefore

$$w_{u,j} - w_{\hat{u},j} = \left(\hat{\beta} - \beta\right) w_{x,j} + \hat{\delta} w_{z(\hat{\tau}),j} - \frac{1}{\sqrt{T}} w_{hx,j},$$

$j = 1, \ldots, T-1$, where $w_{hx,j}$ is the discrete Fourier transform of the sequence $\{h_t x_t, 1 \leq t \leq T\}$. Since $||w_{u,j}| - |w_{\hat{u},j}||^2 \leq |w_{u,j} - w_{\hat{u},j}|^2$,

$$\begin{aligned}
& E^* \left| \frac{1}{\sqrt{T}} \sum_{j=1}^{T-1} w_{z(\tau),j} \left(|w_{u,j}| - |w_{\hat{u},j}|\right) \eta_j^* \right|^2 \\
= & \frac{1}{T} \sum_{j=1}^{T-1} \left|w_{z(\tau),j}\right|^2 ||w_{u,j}| - |w_{\hat{u},j}||^2 E^* \left|\eta_j^*\right|^2 \\
\leq & \frac{1}{T} \sum_{j=1}^{T-1} \left|w_{z(\tau),j}\right|^2 \left|w_{u,j} - w_{\hat{u},j}\right|^2 \qquad (1.42)
\end{aligned}$$

$$\begin{aligned}
\leq & \ 3\left(\hat{\beta} - \beta\right)^2 \frac{1}{T} \sum_{j=1}^{T-1} \left|w_{z(\tau),j}\right|^2 \left|w_{x,j}\right|^2 + 3\hat{\delta}^2 \frac{1}{T} \sum_{j=1}^{T-1} \left|w_{z(\tau),j}\right|^2 \left|w_{z(\hat{\tau}),j}\right|^2 \\
& + \frac{3}{T^2} \sum_{j=1}^{T-1} \left|w_{z(\tau),j}\right|^2 \left|w_{hx,j}\right|^2.
\end{aligned}$$

By Theorem 1.2, $\hat{\beta} - \beta = O_p\left(T^{-1/2}\right)$ and $\hat{\delta} = O_p\left(T^{-1/2}\right)$. Also, by Lemma 1.11 with functions $h_1(x) = h_2(x) = \mathbb{I}(0 \leq x \leq \tau)$ and $h_3(x) = h_4(x) = \mathbb{I}(0 \leq x \leq \hat{\tau})$,

$$\frac{1}{T} \sum_{j=1}^{T-1} \left|w_{z(\tau),j}\right|^2 \left|w_{z(\hat{\tau}),j}\right|^2 = o_p(T)$$

uniformly over $\tau \in [0,1]$, and similarly for the other sums. Therefore the right-hand side of the last displayed inequality is $o_p(1)$ uniformly over $[0,1]$. The uniform convergence in (1.39) is established by using the Markov inequality.

■

Replacing $|w_{\varepsilon,j}|$, $|w_{u,j}|$ and $|w_{\hat{u},j}|$ in (1.37)-(1.39) by $w_{\varepsilon,j}$, $w_{u,j}$ and $w_{\hat{u},j}$, and drawing η_j^* from any complex distribution with mean zero, unit variance, finite fourth moment and with $E\eta_j^{*2} = 0$, it can be seen that the proof remains valid with only small modifications. In particular, expressions for $\text{var}^* Y_T(\tau)$ and $\text{cov}^*(Y_T(\tau), Y_T(\sigma))$ do not change, inequalities in part (a) for suprema in the Lindeberg condition and for $E^* |Y_T(\tau) - Y_T(\sigma)|^4$, in part (b) for the conditional first moment and in part (c) for the conditional second moment continue to hold with minor changes in intermediate steps where required. This observation shows that there are several valid modifications of the basic bootstrap procedure described in Section 1.3.

Chapter 2

Locating structural change in regression with long memory processes

2.1 Introduction

When a presence of a structural change is detected in an econometric model, the time of change is frequently of interest. There is a steadily growing body of literature on estimating the time of change. Hinkley (1970), Yao (1987) and Bhattacharya (1987) deal with maximum likelihood estimation of time of a shift in mean of otherwise identically distributed independent observations. In the context of dependent observations, Bai (1994, 1997b) allows for a linear process with short memory while Bai (1997a), Bai and Perron (1998) and Fiteni (2002, 2004) analyse estimators of the time of break in parameters of linear regression model with mixing data. The current state of the research on structural changes in linear models with time series is reviewed by Banerjee and Urga (2005) and Perron (2006).

In the last decades, however, it has been recognized that many economic and financial data possess a dependence structure stronger than that displayed by mixing data. The effect of long range dependence on estimators of time of break has been examined by Antoch et al. (1995, 1997) and Horváth and Kokoszka (1997) in the framework of linear processes with a break in mean.

The first purpose of this chapter is to develop a procedure for estimation and testing of the time of change in slope coefficients in a linear regression model where both regressors and disturbances are allowed to possess long memory. It is shown that estimators employed for weakly dependent data continue to be valid for strongly dependent data, and the researcher does not need to distinguish between the short and long memory type of dependence at any point of the estimation procedure.

It is known that asymptotic properties of parameter estimators in structural change models depend qualitatively on the magnitude of change. The second purpose of this chapter is therefore to examine the asymptotic behaviour of estimators under various assumptions on the size of break, ranging from a fixed size of break through a size shrinking at a certain rate to zero size.

Asymptotic theory for the breakpoint estimator is derived, including consistency, rate of convergence and limiting distribution. Under the assumption of fixed size of the break, the date of break is estimated with highest relative asymptotic efficiency, but the asymptotic distribution of the breakpoint estimator depends on the joint distribution of the regressors and the error term and is not amenable to hypothesis testing. The problem of unknown limiting distribution of the breakpoint estimator under fixed break is the topic of Chapter 3. Breaks of a fixed magnitude can be regarded as large.

To obtain a distribution-free asymptotic theory of the breakpoint estimator, the size of break can be assumed shrinking as the sample size increases but at the slower speed than the square root of the sample size. This has been a mainstream assumption in the literature for the last two decades. Under a slowly shrinking break, this chapter shows that the asymptotic distribution of our breakpoint estimator is invariant to the distribution of data also when the data are strongly dependent.

The case of breaks shrinking with the square root of the sample size or faster is also considered. Breaks shrinking at such a rate can be denominated as weak. The plausible situation where a researcher estimates the date of a presumed break when the parameters of the processes do not break can be analysed as a special case of a weak break. It is shown that if the break is weak, its location is not estimable. Since the breaks can be detected only when their magnitude shrinks at the rate of the square root of the sample size at the fastest, this rate constitutes a borderline case when the break can be detected but cannot be consistently located.

Beside the asymptotic theory of the breakpoint estimator, we also consider asymptotic properties of the slope coefficient estimators. When the break is large, the slope estimators are asymptotically normal and their distribution is the same as if the time of change were known. Asymptotic normality breaks down for a weak break, under which a nonstandard distribution is obtained.

In the case of a shrinking break, the form of the limiting distribution of the breakpoint estimator allows construction of hypothesis tests. Since the limiting distribution function is known, asymptotic tests can be carried out easily. However, it is known that asymptotic tests may not perform well in small samples. For this reason, we propose a bootstrap procedure to approximate the limiting distribution of the break point for the purpose of hypothesis testing. A small Monte Carlo study compares the performance of the bootstrap and asymptotic tests and confirms that in small samples the bootstrap test seems preferable to the asymptotic test.

The chapter is organized in the following way. Section 2.2 introduces a linear regression model with break in the slope parameter and presents a least squares procedure for estimating the time of break and the slope coefficients. Asymptotic properties of estimators are studied in Section 2.3. Section 2.4 comments on the difference in testing hypotheses about the time of break under fixed and shrinking break. Section 2.5 discusses the cases of weak break and no break. In Section 2.6, a bootstrap approximation of the asymptotic test procedure is proposed. Section 2.7 reports the results of a small Monte Carlo simulation conducted to investigate small sample properties of the proposed bootstrap procedure. Section 2.8 concludes. The proofs are collected in Section 2.A which refers to Section 2.B for intermediate results.

2.2 Linear regression with break

We examine a special case of model (1.1) - a linear regression model with a break in the slope parameter. Let

$$y_t = \alpha + \beta' x_t + \delta_T' z_t + u_t, \qquad t = 1, \dots, T, \tag{2.1}$$

where

$$z_t = z_t(k_0) = \begin{cases} x_t & t = 1, \dots, k_0 \\ 0 & t = k_0 + 1, \dots, T \end{cases}$$

and where k_0 is an unknown date of break and β and δ_T are p-dimensional vectors of unknown parameters with $\delta_T \neq 0$. It is assumed that $k_0 = [\tau_0 T]$ for some $\tau_0 \in \Lambda \in (0,1)$, where the set Λ has closure in $(0,1)$. The size of the break δ_T can be assumed either dependent on the sample size T or fixed at $\delta_T = \delta$.

We are interested in estimating the time of the break and the slope coefficients β and δ_T. In addition to the point estimation, we are also interested in testing hypothesis of the form

$$\mathrm{H}_0\colon\ k_0 = k_H$$

for some constant k_H against the alternative

$$\mathrm{H}_1\colon\ k_0 \neq k_H.$$

In this chapter, we focus on breaks in regression coefficients of stochastic regressors. Break in the regression intercept has been analysed by Kuan and Hsu (1998) in a similar setting.

Model (2.1) can be written in the matrix form as

$$y = \alpha\iota + X\beta + Z_0\delta_T + u, \tag{2.2}$$

where $y = (y_1, \dots, y_T)'$, $\iota = (1, \dots, 1)'$, $X = (x_1, \dots, x_T)'$ and where $Z_k = (x_1, \dots, x_k, 0, \dots, 0)'$ is a $T \times p$ matrix comprising first k rows of the matrix X and completed with zeros, Z_{k_0} is denoted as Z_0, and $u = (u_1, \dots, u_T)'$.

We estimate the parameters of the model by the least squares method. Denote $W_k = (X, Z_k)$ and $M_{A,B} = I - P_{A,B}$, where $P_{A,B} = P_C = C\,(C'C)^{-1}\,C'$ is the matrix of orthogonal projection on the column space of a matrix $C = (A, B)$. Let $\hat{u}\,(k)$ be the vector of residuals from the least squares regression of y on X and Z_k,

$$\hat{u}\,(k) = M_{\iota,X,Z_k} y, \qquad k = 1, \ldots, T-1,$$

and let $\hat{\beta}_k$ and $\hat{\delta}_k$ be the least squares estimators of the slope parameters,

$$\begin{pmatrix} \hat{\beta}_k \\ \hat{\delta}_k \end{pmatrix} = \left(W_k' M_\iota W_k\right)^{-1} W_k' M_\iota y, \qquad k = 1, \ldots, T-1.$$

The least squares estimator $\hat{k}$ of the breakpoint k_0 is obtained by minimizing the objective function

$$S_T\,(k) = \left|\left|\hat{u}\,(k)\right|\right|^2, \qquad k = 1, \ldots, T-1, \tag{2.3}$$

that is

$$\hat{k} = \arg \min_{k \in \Lambda \cdot T} S_T\,(k)\,,$$

where $\Lambda \cdot T = \{k\colon\ k/T \in \Lambda\}$. If the point of minimum is not unique, we define $\hat{k} = \min\left\{k : S_T\,(k) = \min_{l \in \Lambda \cdot T} S_T\,(l)\right\}$. While the expressions for $\hat{\beta}_k$ and $\hat{\delta}_k$ are explicit, the breakpoint k_0 is estimated implicitly.

Denote $\hat{u} = \hat{u}(\hat{k})$, $\hat{\beta} = \hat{\beta}_{\hat{k}}$ and $\hat{\delta} = \hat{\delta}_{\hat{k}}$. The quantities $\hat{u}$, $\hat{\beta}$ and $\hat{\delta}$ can be regarded as least squares estimators of errors and slope coefficients of model (2.2) when the location of break is unknown. Beside the estimator of the date of the break, an estimator $\hat{\tau}$ of the relative time of break τ_0 can be defined as

$$\hat{\tau} = \frac{\hat{k}}{T}.$$

Since some of the properties of our estimators are more easily established in the frequency domain, it is useful to transform data from the time domain to the frequency domain. In the frequency domain, model (2.1) is given by

$$w_y(\lambda_j) = \beta' w_x(\lambda_j) + \delta_T' w_{z(k_0)}(\lambda_j) + w_u(\lambda_j), \qquad j = 1, \ldots, T-1. \quad (2.4)$$

Omission of the frequency zero in (2.4) permits the researcher to avoid estimating the unknown intercept α. As the discrete Fourier transform is invariant to location shift of the sequence for $1 \leq j \leq T-1$, the regression (2.4) is equivalent to a time-domain regression in deviations from the mean. Defining F as the $(T-1) \times T$ matrix of the discrete Fourier transform at the frequencies λ_j,

$$F_{jk} = \frac{1}{\sqrt{2\pi T}} e^{ij\lambda_k}, \qquad j = 1, \ldots, T-1,\ k = 1, \ldots, T,$$

model (2.4) can be written in the matrix form as

$$Fy = FX\beta + FZ_0\delta_T + Fu. \quad (2.5)$$

In the least squares regression of Fy on FX and FZ_k, let

$$\widehat{Fu}(k) = M_{FX,FZ_k}Fy = M_{FW_k}Fy \quad (2.6)$$

be the vector of residuals and

$$\begin{pmatrix} \tilde{\beta}_k \\ \tilde{\delta}_k \end{pmatrix} = \left(W_k' \overline{F}' F W_k\right)^{-1} W_k' \overline{F}' F y, \qquad k = 1, \ldots, T-1,$$

be the estimators of the slope coefficients, where now P_A in the definition of M_A is $P_A = A\left(\bar{A}'A\right)^{-1}\bar{A}'$ where $\bar{A}'$ is the complex conjugate of a complex matrix A. The least squares estimator of the date of break is now a point of minimum of the objective function $\tilde{S}_T(k) = ||\widehat{Fu}(k)||^2$. From the definition of F and M_ι it follows that $\overline{F}'F = M_\iota/2\pi$ and so $\left(\tilde{\beta}_k', \tilde{\delta}_k'\right)' = \left(\hat{\beta}_k', \hat{\delta}_k'\right)'$. Moreover, $FM_\iota = F$ and

$$M_{FW_k}F = FM_{M_\iota W_k} = FM_\iota M_{M_\iota W_k} = FM_{\iota, W_k},$$

which implies that

$$\widehat{Fu}(k) = F\hat{u}(k), \qquad k = 1, \ldots, T-1,$$

$$||\widehat{Fu}(k)||^2 = \frac{1}{2\pi}||\hat{u}(k)||^2 \qquad k = 1, \ldots, T-1.$$

Therefore for the purpose of estimating the time of break and the slope coefficients in a linear regression model with unknown break, estimation in the time and frequency domain is equivalent.

In the following analysis, it is assumed that $\{x_t\}$ and $\{u_t\}$ are covariance stationary linear processes that satisfy Conditions 1.1-1.5 and the following additional condition.

Condition 2.1

(a) $\sup_{l\geq 1}\left\|\frac{1}{l}\sum_{t=1}^{l} x_t x_t'\right\| = O_p(1)$, $\sup_{l\geq 1}\left\|\frac{1}{l}\sum_{t=k_0+1}^{k_0+l} x_t x_t'\right\| = O_p(1)$, $\sup_{l\geq 1}\left\|\frac{1}{l}\sum_{t=k_0-l+1}^{k_0} x_t x_t'\right\| = O_p(1)$,

(b) there exists $\lambda > 0$ *such that for every* $\varepsilon > 0$, *there exists* l_0 *such that* $P(\lambda_l^+ < \lambda) < \varepsilon$ *and* $P(\lambda_l^- < \lambda) < \varepsilon$ *for all* $l \geq l_0$, *where* λ_j^+ *and* λ_j^- *are the minimum eigenvalues of the matrices* $\frac{1}{l}\sum_{t=k_0+1}^{k_0+l} x_t x_t'$ *and* $\frac{1}{l}\sum_{t=k_0-l+1}^{k_0} x_t x_t'$, *respectively.*

Condition 2.1 constrains matrices $\sup_{l\geq 1}\frac{1}{l}\sum_{t=1}^{l} x_t x_t'$, $\sup_{l\geq 1}\frac{1}{l}\sum_{t=k_0+1}^{k_0+l} x_t x_t'$ and $\sup_{l\geq 1}\frac{1}{l}\sum_{t=k_0-l+1}^{k_0} x_t x_t'$ to be uniformly stochastically bounded as T increases. Moreover, it constraints the latter two matrices to have minimum eigenvalues bounded away from zero with large probability for large l. This would be implied for example by the strong law of large numbers for the sequence $\{x_t x_t'\}$.

2.3 Asymptotic properties of the breakpoint and slope estimator

In the discussion of the asymptotic properties of the breakpoint and slope coefficient estimators, we first examine the rate of convergence of the breakpoint estimator. Deriving the rate of convergence not only allows us to characterize consistency properties of the estimators, but is also necessary in order to establish the limiting distribution. In this section we consider breaks whose size is fixed or is shrinking but at a speed smaller than the square root of the sample size.

Proposition 2.1 *Assume Conditions 1.1-1.5 and 2.1 are satisfied. If the size of break is fixed, $\delta_T = \delta \neq 0$, or if it is shrinking with $\delta_T \to 0$ and $T\,||\delta_T||^2 \to \infty$, then*

$$\hat{k} - k_0 = O_p\left(||\delta_T||^{-2}\right). \tag{2.7}$$

Proposition 2.1 implies that if the magnitude of break is fixed, $\hat{k} - k_0 = O_p(1)$ and the quantiles of distribution of $\hat{k} - k_0$ remain of the same order as T grows. On the other hand, if the size of break is shrinking, the dispersion of $\hat{k} - k_0$ grows at the rate of $||\delta_T||^{-2}$. Strictly speaking, $\hat{k}$ is not a consistent estimator of k_0. However, if $\hat{\tau} = \hat{k}/T$ defines an estimator of the relative time of break τ_0, the rate in (2.7) implies consistency of $\hat{\tau}$ with the convergence rate of $T^{-1}\,||\delta_T||^{-2} = o(1)$. Whether the size of break δ is fixed or shrinking, Proposition 2.1 guarantees the consistency of the estimator $\hat{\tau}$ of the relative time of break τ_0 as long as the shrinking is not too fast.

The rate of convergence in (2.7) is typical for changepoint problems in general and holds over a range of models and estimators. In the context of estimation of the time of shift in mean in a stochastic processes, this rate of convergence has been obtained earlier by Bhattacharya (1987) and Yao (1987) for maximum likelihood estimators with independent identically distributed data, Antoch et al. (1995) for an estimator with independent identically distributed processes and Bai (1994, 1997b) for a least squares estimator of shift in the mean of linear process under mixing conditions. Regarding estimators of the time of break in linear regression models, Bai (1997a) and Bai and Perron (1998) discuss least squares procedures in regression with mixingale errors and possibly trending regressors and Fiteni (2002, 2004) considers robust estimators in regression with strongly mixing data.

The following proposition characterizes the asymptotic distribution of the slope estimators for the case of a known and unknown date of break, respectively, for the cases of a break whose size is fixed or whose size decreases at a moderate speed as the sample size decreases.

Proposition 2.2 *Assume Conditions 1.1-1.5 and 2.1 hold. If* $T\,||\delta_T||^2 \longrightarrow \infty$, *then*

(a)

$$\sqrt{T}\begin{pmatrix} \hat{\beta}_{k_0}-\beta \\ \hat{\delta}_{k_0}-\delta_T \end{pmatrix} \xrightarrow{d} N\left(0,V\right) \quad \textit{and}$$

(b)

$$\sqrt{T}\begin{pmatrix} \hat{\beta}-\beta \\ \hat{\delta}-\delta_T \end{pmatrix} \xrightarrow{d} N\left(0,V\right),$$

where

$$V=\frac{1}{\tau_0\left(1-\tau_0\right)}\begin{pmatrix} \tau_0 & -\tau_0 \\ -\tau_0 & 1 \end{pmatrix}\otimes\Sigma^{-1}\Omega\Sigma^{-1}$$

and where Σ *and* Ω *are defined in Theorem 1.1.*

The limiting distribution of the slope estimators $\hat{\beta}$ and $\hat{\delta}$ in the case of unknown date of break is the same as if the date of break were known. It is worth noting that neither the rate of convergence nor the form of the asymptotic distribution depends on whether the size of break is assumed fixed or shrinking, as long as the magnitude of break does not decrease too fast.

Similar results have been obtained by Bai (1997a) for breaks in linear regression model with mixingale errors and possibly trending regressors. Fiteni (2002) has also reported asymptotic normality of a robust estimator of regression coefficients. The asymptotic normality found elsewhere in the structural change literature therefore carries over to linear regression where regressors and errors possibly exhibit strong dependence.

In contrast to the rate of convergence of the breakpoint estimator and the asymptotic distribution of the slope estimator, the asymptotic distribution of the estimator of the location of break requires a separate discussion for the cases of fixed and shrinking break. First, we consider the case of a fixed magnitude of break, $\delta_T = \delta \neq 0$. Define the process W^0 on the set of all integers as

$$W^0(s) = \begin{cases} \delta' \sum_{t=1}^{s} x_t x_t' \delta - 2\delta' \sum_{t=1}^{s} x_t u_t & s \geq 1, \\ 0 & s = 0, \\ \delta' \sum_{t=s+1}^{0} x_t x_t' \delta + 2\delta' \sum_{t=s+1}^{0} x_t u_t & s \leq -1. \end{cases} \tag{2.8}$$

The following proposition gives the asymptotic distribution of the breakpoint estimator for the case of fixed break. To avoid dependence of the asymptotic distribution of the estimator on the unknown date of break k_0, we need to ensure shift invariance of the distribution of data by strengthening the assumption of second order stationarity implicit in Conditions 1.1-1.5 to strict stationarity.

Proposition 2.3 *Assume that Conditions 1.1-1.5 and 2.1 hold and that in addition the process $\{x_t, u_t\}$ is strictly stationary. Assume further that $(\delta' x_t)^2 \pm 2\delta' x_t u_t$ has a continuous distribution. If the magnitude of break is fixed, $\delta_T = \delta \neq 0$, then*

$$\hat{k} - k_0 \xrightarrow{d} \arg\min_s W^0(s).$$

The asymptotic distribution of the breakpoint estimator with a break of fixed size therefore depends not only on the nuisance parameter δ but also on the joint distribution of x_t and u_t. While the size of jump δ can be consistently estimated by Proposition 2.2, the distribution of the data is generally unknown. Unless the joint distribution of data is estimated, inference about the time of break cannot be based on the distribution of the limiting random variable. The estimation of the distribution of data is discussed in the next chapter.

It is worth noting that the distribution of the location estimator $\hat{k}$ is discrete. Therefore even when the distribution of $\arg\min W^0$ is known, tests of hypotheses about the time of break cannot be performed at any arbitrary level of significance. In this situation, hypothesis testing can be approached in two ways. One possibility is to carry out tests at the significance levels given by the quantiles of the limiting discrete variable. Alternatively, given a nominal level of confidence, conservative tests can be constructed by taking the next higher quantile of the limiting distribution. The latter approach has been adopted for example by Bai (1997a) and Antoch and Hušková (1999).

The problem of dependence of the asymptotic distribution of the breakpoint estimator on the joint distribution of data can be overcome if we are willing to modify the assumption on the size of the break. Consider the case of a diminishing magnitude of break. Define the process $\bar{W}$ as

$$\bar{W}(\rho) = \begin{cases} W_1(\rho) + \frac{|\rho|}{2} & \rho \geq 0, \\ W_2(-\rho) + \frac{|\rho|}{2} & \rho < 0, \end{cases}$$

where W_1, W_2 are independent standard Brownian motion processes defined on $[0, \infty)$. The following proposition describes the asymptotic distribution of the breakpoint estimator for the case of shrinking break.

Proposition 2.4 *Assume that Conditions 1.1-1.5 and 2.1 hold and that* $\delta_T \to 0$ *and* $T\|\delta_T\|^2 \to \infty$. *Let* $\hat{\Sigma}$ *and* $\hat{\Omega}$ *be consistent estimators of* Σ *and* Ω. *Then*

$$\frac{\left(\hat{\delta}'\hat{\Sigma}\hat{\delta}\right)^2}{\hat{\delta}'\hat{\Omega}\hat{\delta}}\left(\hat{k} - k_0\right) \xrightarrow{d} \arg\min_{\rho\in\mathbb{R}} \bar{W}(\rho).$$

Results similar to those of Propositions 2.3 and 2.4 have been also obtained by Bai (1994, 1997 a,b), Antoch et al. (1995, 1997), Bai and Perron (1998) and Fiteni (2002, 2004).

An example of consistent estimator of Σ is

$$\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T} x_t x_t', \tag{2.9}$$

whose consistency follows from Conditions 1.1 and 1.2 because x_t is ergodic. A consistent estimator of Ω is

$$\hat{\Omega} = \frac{4\pi^2}{T}\sum_{j=1}^{T-1} I_{xx}(\lambda_j)\, I_{\hat{u}\hat{u}}(\lambda_j).$$

Consistency of $\hat{\Omega}$ is asserted by the following proposition.

Proposition 2.5 *Assume that Conditions 1.1-1.5 and 2.1 hold and that* $T^{1/2}\delta_T \to$

$\delta \neq 0$ *or* $T^{1/2} \|\delta_T\| \to \infty$. *Then*

$$\hat{\Omega} \xrightarrow{p} \Omega.$$

The distribution of $\arg\min_\rho \bar{W}(\rho)$ is not only free of nuisance parameters, but also the explicit form of its distribution function G is known,

$$\begin{aligned} G(x) &= 1 + \sqrt{\frac{x}{2\pi}} e^{-\frac{x}{8}} - \frac{1}{2}(x+5)\,\Phi\left(-\frac{\sqrt{x}}{2}\right) + \frac{3}{2} e^{x} \Phi\left(-\frac{3\sqrt{x}}{2}\right) \quad x > 0, \\ G(x) &= 1 - G(-x), \end{aligned}$$

see for example Yao (1987). The two-sided critical values at the 0.1, 0.05 and 0.01 significance level are 7.687, 11.033 and 19.767, respectively.

While assuming shrinking size of break leads to a distribution-free asymptotics for the breakpoint estimator, it has also some disadvantages. One such disadvantage, a loss of information reflected in a loss of power in testing hypotheses on the time of break, is discussed in the following section.

2.4 Hypothesis testing

The results of the previous sections allow us to make inference about the date of break. The null hypothesis of interest is

$$H_0\colon\ k_0 = k_H,$$

where k_0 is the true value of the break date and k_H denotes the hypothesized time of change.

When the size of break is assumed fixed, Proposition 2.3 gives the limiting distribution of $\hat{k} - k_0$ under the null hypothesis. The test of the null hypotheses can be based on the test statistic $Z_T = \hat{k} - k_H$. Under the alternative hypothesis

$$H_1 : k_0 = k_H + \Delta \tag{2.10}$$

with a constant $\Delta \neq 0$, we have

$$Z_T \xrightarrow{d} k_0 - k_H + \arg\min_s W^0(s).$$

Since $k_0 - k_H \neq 0$ and $\arg\min_s W^0(s) = O_p(1)$, the test based on Z_T has asymptotical local power against the alternative hypothesis (2.10). However, since the asymptotic distribution of Z_T under both null and alternative hypotheses depends on the underlying joint distribution of x_t, u_t, the critical values for the test are in general not available.

Under the assumption that the size of break is shrinking with $\delta_T \to 0$ and $T\|\delta_T\|^2 \to \infty$, Proposition 2.4 indicates that the limiting distribution of $\hat{k} - k_0$ normalized by $\widehat{v}_T^2 = \left(\widehat{\delta}'\widehat{\Sigma}\widehat{\delta}\right)^2 / \widehat{\delta}'\widehat{\Omega}\widehat{\delta}$ is invariant to the underlying distribution of data. This suggests to use $Z_T = \widehat{v}_T^2\left(\hat{k} - k_H\right)$ as a test statistic. Proposition 2.4 then gives the asymptotic distribution of Z_T under the null. Under the alternative hypothesis (2.10),

$$Z_T = \widehat{v}_T^2\left(\hat{k} - k_0\right) + \Delta\widehat{v}_T^2 \xrightarrow{d} \arg\min_\rho \bar{W}(\rho).$$

The distribution of Z_T is therefore identical under both the null and alternative hypothesis, and the test based on Z_T has asymptotically no power against the alternative that $k_0 = k_H + \Delta$.

If we consider a sequence of local hypotheses in the form of

$$H_1\colon\ k_0 = k_H + \Delta_T, \tag{2.11}$$

where Δ_T depends on the sample size, the test based on Z_T has asymptotic local power against such alternatives if $\Delta_T^{-1} = O\left(\|\delta_T\|^2\right)$. When $\Delta_T^{-1} = o\left(\|\delta_T\|^2\right)$, the test is consistent, or has global power, against the alternative hypothesis (2.11). For example, if we consider an alternative hypothesis to be H_1: $k_0 = k_H + T\cdot\Delta$, which corresponds to the alternative fixed in terms of the relative time of the break, H_1: $\tau_0 = \tau_H + \Delta$, the test has global power since $\Delta_T^{-1} = T^{-1}\Delta^{-1} = o\left(\|\delta_T\|^2\right)$.

2.5 Weak breaks

In the preceding analysis we have assumed that there is a break of a nonzero magnitude. It is interesting to examine the asymptotic behaviour of the breakpoint estimator when the researcher erroneously estimates the time of break when there is in fact no break in the data generating process, that is when $\delta_T = 0$. More generally, it is of interest to study the statistical properties of the breakpoint estimator when only a weak break is present, that is when the size of the break is nonzero but decreasing fast with T in such a way that $T \|\delta_T\|^2 = O(1)$.

For the purposes of analysing asymptotic properties of estimators under the assumption of weak break, we define

$$G(\tau) = \frac{1}{(\tau(1-\tau))^{\frac{1}{2}}} \Sigma^{-\frac{1}{2}} \Omega^{\frac{1}{2}} (W(\tau) - \tau W(1)) + \frac{m(\tau)}{(\tau(1-\tau))^{\frac{1}{2}}} \Sigma^{\frac{1}{2}} \delta$$

for $\tau \in (0,1)$, where for $\delta \neq 0$ the function m is defined as

$$m(\tau) = \begin{cases} \tau(1-\tau_0) & \tau \leq \tau_0, \\ \tau_0(1-\tau) & \tau \geq \tau_0, \end{cases}$$

and for $\delta = 0$ the function m can be left undefined. Further, we define

$$L = \arg\max_{\tau \in \Lambda} G(\tau)' G(\tau).$$

By the definition of Λ, the random variable L takes values in a subset of $(0,1)$. The following proposition describes the asymptotic distribution of the breakpoint estimator under a weak break.

Proposition 2.6 *Assume that Conditions 1.1-1.5 and 2.1 hold and that* $T^{1/2}\delta_T \to \delta$ *with* $0 \leq \|\delta\| < \infty$. *Then*

$$\hat{\tau} \xrightarrow{d} L.$$

Proposition 2.6 implies that the estimator $\hat{\tau}$ of the relative time of break τ is not consistent when the break is weak. Moreover, since for the cases of both $\delta_T = 0$ and $T \|\delta_T\|^2 \to 0$ the limiting value δ is equal to zero, Proposition

2.6 indicates that the presence of a break shrinking to zero faster than $T^{1/2}$ is asymptotically equivalent to the absence of the break.

The asymptotic properties of the slope coefficient estimators $\hat{\beta}$ and $\hat{\delta}$ under weak breaks are given in the following proposition.

Proposition 2.7 *Assume that Conditions 1.1-1.5 and 2.1 hold and that* $T^{1/2}\delta_T \to \delta$ *with* $0 \leq \|\delta\| < \infty$. *Then*

$$\sqrt{T}\begin{pmatrix} \hat{\beta}-\beta \\ \hat{\delta}-\delta_T \end{pmatrix} \xrightarrow{d} \frac{1}{L(1-L)}\begin{pmatrix} \Sigma^{-1}\Omega^{\frac{1}{2}}\left(LW(1)-LW(L)\right) \\ \Sigma^{-1}\Omega^{\frac{1}{2}}\left(W(1)-LW(1)\right) \end{pmatrix} + \frac{1}{L(1-L)}\begin{pmatrix} L(\tau_0-L)\,\mathbb{I}(L<\tau_0) \\ (\tau_0-L)\left(\mathbb{I}(\tau_0\leq L)-L\right) \end{pmatrix}\otimes\delta.$$

The random variable $W(L)$ has a mixed normal distribution where the mixing variable is L, in the sense that for any real p-vector b, the cumulative distribution function $P(W(L)\leq b)$ is given by $\int_{l\in\Lambda}\Phi\left(b/\sqrt{l}\right)dF_L(l)$, where Φ is the distribution function of a p-dimensional standard normal variable, F_L is the distribution function of L and where the inequality $W(L)\leq b$ is to be taken componentwise.

Proposition 2.7 together with Proposition 2.2 imply that $\hat{\beta}$ remains a $\sqrt{T}$-consistent estimator of β for the whole range of assumptions on the size of the breaks, from breaks of size zero to breaks of a fixed size. Similarly, $\hat{\delta} = \delta_T + O_p\left(T^{-1/2}\right)$ under a break of any size.

While the rate of convergence of the slope coefficient estimators under a fixed or shrinking break continues to hold under a weak break, the asymptotic normality does not. The form of the asymptotic distribution of the slope estimators reflects the fact that the estimation is attempted in the situation where the point of break is not well identified.

The results up to this point imply that the location of the break can be estimated only when the magnitude of break diminishes slower than $T^{-1/2}$. A related question is when the break is detectable, that is, what is the range of alternatives against which tests of the null of no break have nontrivial power. The tests may be based on continuous functionals of the sum of squares $S_T(k)$. As it transpires in the proof of Proposition 2.6 in Section 2.A, if $T^{1/2}\delta_T \to \delta$, then

$$y' M_{\iota,X} y - S_T([\tau T]) \Longrightarrow G(\tau)' G(\tau).$$

Therefore it can be seen that while tests based on continuous functionals of $S_T([\tau T])$ have nontrivial power against the alternatives where $T^{1/2}\delta_T \to \delta \neq 0$, the test have no power against the alternatives where $\delta_T = o_p\left(T^{-1/2}\right)$ since the limiting distribution is identical for cases $\delta_T = 0$ under the null and $T^{1/2}\delta_T \to 0$ under the alternative.

In sum, the analysis of the asymptotic properties of the breakpoint estimator under a range of assumption on the size of break shows that while a break with $T\|\delta_T\|^2 \to \infty$ can be detected and its location can be estimated, a break with $T\|\delta_T\|^2 \to \delta \neq 0$ is detectable but its location is not estimable, and a break with $T\|\delta_T\|^2 \to 0$ cannot be detected.

2.6 Bootstrap under shrinking break

The results of the preceding sections suggest that if the magnitude of change is too small, the changepoint cannot be identified. On the other hand, if the size of break is large, $\delta_T = \delta$, the relative time of break can be estimated T-consistently but its asymptotic distribution is intractable for the purposes of hypothesis testing. The only circumstance when a consistent breakpoint estimator with distribution-free asymptotic properties is available is the case of a break whose magnitude is diminishing but more slowly than the square root of the sample size. In this instance, tests of hypotheses about the time of break can be based on the asymptotic distribution of the breakpoint estimator.

However, it is known that the finite sample distribution of a statistic may not be well approximated by its asymptotic distribution when the sample size is small. The purpose of this section is to obtain a bootstrap procedure that approximates the asymptotic distribution of the breakpoint estimator and that may improve on the performance of the asymptotic distribution in small samples.

To approximate the distribution of the breakpoint estimator, we propose to use a method similar to that employed by Hidalgo (2003a) and ourselves in Chapter 1. The procedure consists of the following steps.

Step 1 Compute the least squares estimate $\hat{k} = \arg\min_{k \in \Lambda \cdot T} S_T(k)$ in equation (2.3). Compute the least squares estimates $\hat{\beta} = \hat{\beta}_{\hat{k}}$ and $\hat{\delta} = \hat{\delta}_{\hat{k}}$ and the least squares residuals

$$\hat{u}_t = y_t - \hat{\beta}' x_t - \hat{\delta}' \hat{z}_t, \qquad t = 1, \ldots, T,$$

where $\hat{z}_t = x_t \mathbb{I}\left(t \leq \hat{k}\right)$.

Step 2 Compute

$$w_{\hat{u}}(\lambda_j) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} \hat{u}_t e^{it\lambda_j}, \qquad j = 1, \ldots, [T/2],$$

and

$$\tilde{w}_{\hat{u}}(\lambda_j) = \frac{w_{\hat{u}}(\lambda_j) - \frac{1}{[T/2]} \sum_{k=1}^{[T/2]} w_{\hat{u}}(\lambda_k)}{\left(\frac{1}{[T/2]} \sum_{j=1}^{[T/2]} \left| w_{\hat{u}}(\lambda_j) - \frac{1}{[T/2]} \sum_{k=1}^{[T/2]} w_{\hat{u}}(\lambda_k) \right|^2\right)^{\frac{1}{2}}}.$$

Step 3 Draw a random sample $\eta_1^*, \ldots, \eta_{[T/2]}^*$ from the distribution $P^*\left(\eta_j^* = \tilde{w}_{\hat{u}}(\lambda_k)\right) = \frac{1}{[T/2]}$ for $k = 1, \ldots, [T/2]$, define $\eta_j^* = \overline{\eta}_{T-j}^*$ for $1 \leq j < T/2$ and generate a bootstrap sample

$$w_y^*(\lambda_j) = \hat{\beta} w_x(\lambda_j) + \hat{\delta} w_{\hat{z}}(\lambda_j) + w_{\hat{u}}(\lambda_j) \eta_j^*, \qquad j = 1, \ldots, T-1.$$

In matrix notation,

$$Fy^* = FX\hat{\beta} + FZ_{\hat{k}}\hat{\delta} + HF\hat{u},$$

where $H = diag\left(\eta_1^*, \ldots, \eta_{T-1}^*\right)$.

Step 4 Let $\hat{\beta}_k^*$ and $\hat{\delta}_k^*$ be the least squares estimators of the slope coefficients and let $\hat{u}^*(k)$ be the vector of residuals from the least squares regression of Fy^* on FX and FZ_k. Let

$$S_T^*(k) = \|\hat{u}^*(k)\|^2.$$

Compute the bootstrap estimator $\hat{k}^*$ of the breakpoint as

$$\hat{k}^* = \arg\min_{k \in \Lambda \cdot T} S_T^*(k)$$

and obtain $\hat{\beta}^* = \hat{\beta}^*_{\hat{k}^*}$, $\hat{\delta}^* = \hat{\delta}^*_{\hat{k}^*}$ and $\hat{u}^* = \hat{u}^*(\hat{k}^*)$.

Step 5 Compute the bootstrap test statistic

$$Z_T^* = \frac{\left(\hat{\delta}^{*\prime}\hat{\Sigma}\hat{\delta}^*\right)^2}{\hat{\delta}^{*\prime}\widehat{\Omega}^*\hat{\delta}^*}\left(\hat{k}^* - \hat{k}\right),$$

where $\hat{\Sigma}$ is defined in (2.9) and where $\widehat{\Omega}^* = \frac{4\pi^2}{T}\sum_{j=1}^{T-1} I_{xx}(\lambda_j)\, I_{\hat{u}^*\hat{u}^*}(\lambda_j)$. Alternatively, compute a nonpivotal statistic

$$\widetilde{Z}_T^* = \hat{k}^* - \hat{k}.$$

The following proposition demonstrates that the proposed bootstrap procedure consistently estimates the distribution of slope estimators $\hat{\beta}$ and $\hat{\delta}$.

Proposition 2.8 *Assume that Conditions 1.1-1.5 and 2.1 hold and that $\delta_T \to 0$ and $T\|\delta_T\|^2 \to \infty$. Then*

$$\sqrt{T}\begin{pmatrix} \hat{\beta}^* - \hat{\beta} \\ \hat{\delta}^* - \hat{\delta} \end{pmatrix} \xrightarrow{d^*} N(0, V),$$

where V is defined in Proposition 2.2.

The distribution of the bootstrap test statistic Z_T^* defined in Step 5 of the bootstrap procedure can be used to construct a bootstrap test as an approximation of the asymptotic test based on the asymptotic null distribution of the test statistic Z_T. The approximation is valid if bootstrap distribution estimator consistently estimates the null distribution of Z_T. Denoting the null distribution of Z_T as $P(Z_T \leq x|\mathrm{H}_0)$ and taking the Kolmogorov-Smirnov distance, consistency requires that

$$\sup_{x\in R}\left|P^*(Z_T^* \leq x) - P(Z_T \leq x|\mathrm{H}_0)\right| \xrightarrow{p} 0.$$

Under the null, Z_T converges in distribution to a continuous distribution function F, $P(Z_T \leq x|\mathrm{H}_0) \to F(x)$, therefore it is sufficient to show that

$$P^*(Z_T^* \leq x) \xrightarrow{p} F(x)$$

pointwise, see for example van der Vaart (1998, p. 329). This observation is exploited in the following proposition asserting the consistency of bootstrap.

Proposition 2.9 *Assume that Conditions 1.1-1.5 and 2.1 hold and that* $\delta_T \to 0$ *and* $T\|\delta_T\|^2 \to \infty$. *Then*

$$Z_T^* = \frac{\left(\hat{\delta}^{*\prime}\hat{\Sigma}\hat{\delta}^*\right)^2}{\hat{\delta}^{*\prime}\hat{\Omega}^*\hat{\delta}^*}\left(\hat{k}^* - \hat{k}\right) \xrightarrow{d^*} \arg\min_{\rho} \bar{W}(\rho).$$

Given the consistency of the bootstrap procedure, a bootstrap test can be constructed to approximate the asymptotic test. The asymptotic α-level critical region C_α based on the asymptotic null distribution, $P(Z_T \in C_\alpha) = \alpha$, is replaced by a critical region C_α^* based on the bootstrap distribution, where C_α^* satisfies $P^*(Z_T^* \in C_{\alpha*}) = \alpha$. Proposition 2.9 guarantees that the bootstrap test has asymptotically correct size.

2.7 Finite sample properties

In this section we assess the performance of the proposed tests in samples of small and moderate size via a small Monte Carlo experiment. Beside the overall assessment of the tests, we are particularly interested in the comparison between bootstrap and asymptotic tests.

The data for the regressor x_t and error term u_t in model (2.1) are generated as scalar ARFIMA$(0, d, 0)$ processes where d is the memory parameter and where the innovations are normally distributed with zero mean and unit variance. Values of 0, 0.2 and 0.4 for d_x and d are considered in admissible combinations such that $0 \leq d_x + d < 1/2$. Samples of size $T = 64$, 128, 256 and 512 are generated by the algorithm of Davies and Harte (1987). Each sample is normalized to have the standard deviation equal to one. Number of

Monte Carlo replications in each experiment is 5000. For bootstrap tests, the number of bootstrap replication is 800.

The break in the model is located in the middle of the sample, $\tau_0 = 1/2$. The size of the break is set to $\delta_T = \delta_0 T^{-1/4}$, where the shrinking rate $T^{-\alpha}$ is chosen such that $\alpha = 1/4$ is the midpoint of the interval $(0, 1/2)$. The parameter δ_0 is equal to 2 so that for $T = 64$, 128, 256 and 512 the size of the break is 0.84, 0.71, 0.60, 0.50, respectively. The value of the slope coefficient is $\beta = 0$.

We examine performance of three tests: two bootstrap tests, of which one is based on the distribution of a nonpivotal bootstrap test statistic $\hat{k}^* - \hat{k}$ and the other on the distribution of a pivotal bootstrap test statistic $\left(\hat{\delta}^{*\prime}\hat{\Sigma}\hat{\delta}^*\right)^2 \times \left(\hat{\delta}^{*\prime}\hat{\Omega}^*\hat{\delta}^*\right)^{-1}\left(\hat{k}^* - \hat{k}\right)$, and an asymptotic test based on the limiting distribution of the test statistic $\left(\hat{\delta}'\hat{\Sigma}\hat{\delta}\right)^2\left(\hat{\delta}'\hat{\Omega}\hat{\delta}\right)^{-1}\left(\hat{k} - k_0\right)$ under the null hypothesis that $k_H = k_0$. Nominal significance levels of 10%, 5% and 1% are considered. The two-sided critical values for the asymptotic test are 7.687 at the 10% level, 11.033 at the 5% level and 19.767 at the 1% level of significance.

Table 2.1 reports the rejection probabilities of the three tests under the null hypothesis $k_H = k_0$. The size of all three tests converges very slowly to the nominal values of 10%, 5% and 1%. Both bootstrap tests approximate the asymptotic test well. The pivotal bootstrap test improves on the performance on the asymptotic test at all sample sizes, and the improvement seems to be more pronounced for higher sample sizes. This indicates that even in relatively large samples it may be beneficial to carry out the bootstrap rather than the asymptotic version of the testing procedure.

The nonpivotal bootstrap test does not fare as well as the pivotal test. This is to be expected, but even the nonpivotal test slightly outperforms the asymptotic test when the sample size is 512.

To examine the power of the tests, we select $\tau_0 = 5/8$. This is an alternative which is fixed in terms of the percentage location of the break, therefore the tests under shrinking breaks have global power, that is the rejection rates of all tests under this alternative should converge to one as the sample size increases. The rejection probabilities of the tests under the alternative are given in Table 2.2.

The rejection rates under the alternative mirror the behaviour of the rejection rates under the null in that the convergence to 100% rate is very slow. Moreover, the convergence is non-monotonic. The rejection rates of the asymptotic test are slightly higher than those of the bootstrap tests. It is however worth noting that the critical values of the tests are not size-adjusted, therefore we cannot conclude that the asymptotic test is more powerful against the chosen alternative.

The results of the simulation exercise suggest that the bootstrap tests offer a good approximation of the asymptotic test. The bootstrap tests, and in particular the pivotal test, can improve on the asymptotic test. Whether the improvement achieved by carrying out the bootstrap test justifies the cost of running the bootstrap will depend on the particular circumstances in which the test is carried out.

2.8 Conclusions

In this chapter, statistical properties of estimators of location of a structural change are examined in the context of a linear regression model under mild conditions on regressors and error term. These conditions avoid the need for specifying the type of mixing conditions that are frequently used in the literature, and include data which display long memory behaviour.

Results of our analysis show that the range of assumptions on the size of the break can be divided into five cases: Break of fixed size, of size shrinking at a rate smaller, equal or bigger than the square root of the sample size, and of zero size.

Under the fixed break, the asymptotic distribution of the breakpoint estimator has the smallest relative order of variance but the distribution is not amenable to hypothesis testing. A tractable asymptotic distribution is obtained only if the magnitude of change is assumed to be shrinking but more slowly than the square root of the sample size. In that case, the asymptotic distribution function is free of nuisance parameters and is explicitly known. When the size of the break is shrinking faster than the square root of the sample size, or when there is no break in the data generating process, the

		Nonpivotal bootstrap test			Pivotal bootstrap test			Asymptotic test		
d_x	d	**10%**	**5%**	**1%**	**10%**	**5%**	**1%**	**10%**	**5%**	**1%**
T=64										
0	**0**	35.3	30.3	26.1	26.8	19.4	12.1	29.5	21.2	12.3
0.2	**0**	35.8	31.0	26.7	27.4	19.9	12.7	30.3	22.0	12.7
0	**0.2**	35.1	30.0	25.7	26.6	19.5	12.1	28.8	20.8	11.8
0.2	**0.2**	37.1	32.1	28.1	28.5	21.2	13.3	28.9	21.7	12.5
0.4	**0**	37.1	32.0	28.1	29.1	21.6	14.3	32.4	24.3	14.2
0	**0.4**	37.6	32.5	28.0	29.3	22.3	14.6	30.1	22.5	13.7
T=128										
0	**0**	25.5	20.6	16.4	19.2	13.3	7.4	23.8	16.7	8.9
0.2	**0**	25.6	20.6	16.7	19.3	13.4	7.1	24.6	17.0	8.9
0	**0.2**	25.4	20.3	16.3	19.4	12.8	7.4	23.6	16.5	8.8
0.2	**0.2**	28.6	23.3	19.5	21.0	14.2	8.1	24.7	17.5	9.2
0.4	**0**	28.2	22.9	18.7	22.2	15.2	8.7	28.2	20.1	10.4
0	**0.4**	27.0	21.3	17.3	21.1	14.2	8.9	24.5	17.5	9.9
T=256										
0	**0**	18.6	12.8	8.5	14.1	8.5	3.8	20.2	13.0	5.9
0.2	**0**	18.7	13.1	9.2	14.7	8.7	4.2	20.4	13.7	6.2
0	**0.2**	18.8	13.4	9.2	15.0	9.0	4.1	20.3	13.5	6.2
0.2	**0.2**	20.9	16.3	12.1	16.9	10.7	5.5	21.9	14.8	7.0
0.4	**0**	20.7	15.1	10.8	17.2	11.0	5.2	23.9	16.5	7.7
0	**0.4**	20.4	14.8	10.3	16.7	10.2	4.9	21.2	14.5	7.0
T=512										
0	**0**	15.3	9.7	4.9	12.6	6.9	2.2	17.8	10.9	4.1
0.2	**0**	15.5	9.8	5.2	13.2	7.2	2.5	18.1	11.4	4.6
0	**0.2**	14.8	9.4	5.1	12.6	6.8	2.5	17.8	11.0	4.6
0.2	**0.2**	16.9	11.7	7.6	13.6	7.9	3.5	18.8	12.0	5.6
0.4	**0**	17.2	11.2	6.1	15.3	9.1	3.4	20.7	13.6	5.8
0	**0.4**	16.3	10.3	5.3	13.6	7.8	2.7	18.1	11.4	4.9

Table 2.1: Size of the bootstrap and asymptotic tests

		Nonpivotal bootstrap test			Pivotal bootstrap test			Asymptotic test		
d_x	d	**10%**	**5%**	**1%**	**10%**	**5%**	**1%**	**10%**	**5%**	**1%**
T=64										
0	**0**	48.3	39.2	31.4	36.7	25.6	14.7	43.8	30.3	14.4
0.2	**0**	48.6	39.5	31.8	37.0	26.0	14.7	45.3	31.6	14.9
0	**0.2**	48.8	39.8	32.2	36.6	25.8	14.8	44.2	30.5	15.4
0.2	**0.2**	49.5	41.0	33.9	37.6	27.0	16.2	42.8	29.8	15.4
0.4	**0**	49.9	42.0	34.2	39.6	28.6	16.6	48.2	34.6	17.8
0	**0.4**	51.5	42.5	34.5	39.8	28.9	17.6	47.0	32.8	17.5
T=128										
0	**0**	47.9	35.2	24.3	37.1	22.5	10.3	52.0	34.7	13.9
0.2	**0**	47.7	35.0	24.3	36.6	22.9	10.4	52.5	35.1	14.5
0	**0.2**	48.0	34.9	24.2	36.7	22.3	10.4	51.6	34.4	14.0
0.2	**0.2**	46.9	35.3	26.5	35.6	23.2	12.1	47.7	31.6	13.8
0.4	**0**	47.3	35.5	25.9	37.7	24.9	12.4	54.3	37.9	17.1
0	**0.4**	49.5	36.4	25.7	37.9	24.4	12.0	52.3	35.7	15.4
T=256										
0	**0**	52.6	34.9	18.3	41.7	24.4	8.3	63.9	43.0	14.6
0.2	**0**	52.9	36.2	18.6	42.6	25.6	8.9	64.5	44.1	15.5
0	**0.2**	52.7	36.2	17.7	42.6	25.0	8.2	63.1	43.0	14.2
0.2	**0.2**	50.1	34.3	20.3	38.6	23.2	9.3	56.2	37.1	13.2
0.4	**0**	52.3	36.3	20.0	44.0	27.6	10.8	65.3	47.5	18.6
0	**0.4**	55.1	38.2	19.0	44.8	26.5	9.1	64.0	44.7	15.6
T=512										
0	**0**	67.6	47.2	18.5	57.7	36.0	9.5	78.7	60.9	21.3
0.2	**0**	67.2	46.7	18.6	58.0	36.0	10.3	78.8	61.4	22.0
0	**0.2**	68.5	47.5	19.4	59.7	36.4	10.1	79.5	61.8	22.2
0.2	**0.2**	59.7	40.5	18.3	48.5	29.0	8.6	68.6	49.2	16.7
0.4	**0**	65.3	46.0	20.1	57.7	38.4	13.3	79.5	62.8	25.7
0	**0.4**	70.2	49.9	19.1	61.2	38.4	10.6	79.7	62.7	24.0

Table 2.2: Rejection probabilities of the bootstrap and asymptotic tests under the alternative

question of estimating the location of the break becomes vacuous because in this circumstance the break is not detectable. In the borderline case of the size of break decreasing with exactly the square root of the sample size, the break can be detected but there is insufficient information for estimating its location.

The asymptotic properties of estimators of the slope coefficients also depend on the assumption on the size of break. Slope estimators are asymptotically normal with identical covariance matrix under fixed as well as slowly shrinking breaks but the distribution is nonstandard for weak breaks.

In addition to the thorough examination of the asymptotic properties of estimators, the chapter proposes a bootstrap approximation of the asymptotic test procedure under the standard assumption of shrinking breaks. A Monte Carlo experiment indicates that the bootstrap procedure improves on the performance of asymptotic test when the sample is of small or moderate size.

There are several natural directions in which the findings of this chapter might be generalized. First, it is desirable to devise a method of estimating location of more than one break for both known and unknown number of breaks. Some methods of locating multiple breaks have been suggested by Bai (1997b), Bai and Perron (1998) or Altissimo and Corradi (2003). Second, to broaden the applicability of our method, the restriction on the collective memory of regressors and errors needs to be relaxed to allow for greater collective range of memory. A natural direction here is to employ the weighted least square estimator of Robinson and Hidalgo (1997) or generalized least squares estimators of Hidalgo and Robinson (2002). These topics are left for possible future research.

Finally, for the case of a fixed magnitude of break, it is of interest to find a method of estimating the distribution of the breakpoint estimator when the underlying distribution of data is unknown in order that confidence intervals could be given for the time of break. Such a method is proposed in Chapter 3.

2.A Proofs

This section contains the proofs of the results in the main body of the chapter. We define $N(K) = \{k\colon\ |k-k_0| \le K\,||\delta_T||^{-2}\}$ and $N^c(K) = \Lambda \cdot T - N(K)$. For integers l, m, we define

$$Z_{\Delta(l,m)} = (Z_m - Z_l)\,\mathrm{sgn}\,(m-l)$$

and denote $Z_\Delta = Z_{\Delta(k_0,k)}$. We let $\iota_k = (1,\ldots 1, 0, \ldots, 0)'$ be a T-vector with the first k elements equal to 1 and the remaining elements equal to 0, so that $\iota = \iota_T$. We denote $\iota_0 = \iota_{k_0}$ and $\iota_\Delta = (\iota_k - \iota_0)\,\mathrm{sgn}\,(k-k_0)$. Further, we define

$$Q_T(k) = \delta_T' Z_0' M_{\iota,X,Z_k} Z_0 \delta_T$$

and

$$R_T(k) = 2\delta_T' Z_0' M_{\iota,X,Z_k} u + u'(M_{\iota,X,Z_k} - M_{\iota,X,Z_0})u$$

so that

$$S_T(k) - S_T(k_0) = Q_T(k) + R_T(k)\,. \tag{2.12}$$

Proof of Proposition 2.1. Fix $\varepsilon > 0$. For any $\lambda > 0$,

$$P\left(\left|\hat{k} - k_0\right| > K\,||\delta_T||^{-2}\right) \le P\left(\inf_{N^c(K)} S_T(k) \le S_T(k_0)\right)$$

$$\le P\left(\inf_{N^c(K)} \frac{Q_T(k)}{|k-k_0|} \le \lambda\,||\delta_T||^2\right) + P\left(\sup_{N^c(K)} \left|\frac{R_T(k)}{k-k_0}\right| \ge \lambda\,||\delta_T||^2\right). \tag{2.13}$$

Lemma 2.4 implies that λ can be chosen such that the first term on the right of (2.13) is smaller than $\varepsilon/2$ for large K. We now show that the second term on the right of (2.13) is smaller than $\varepsilon/2$ for large K. To that end, write

$$\begin{aligned} R_T(k) &= -2\delta_T' Z_\Delta' M_\iota u\,\mathrm{sgn}\,(k-k_0) \\ &\quad +2\delta_T' Z_\Delta' M_\iota W_k \left(W_k' M_\iota W_k\right)^{-1} W_k' M_\iota u\,\mathrm{sgn}\,(k-k_0) \\ &\quad +u'\left(M_{\iota,X,Z_k} - M_{\iota,X,Z_0}\right)u. \end{aligned} \tag{2.14}$$

The contribution of the first term of (2.14) to (2.13) is

$$P\left(\sup_{N^c(K)}\left|\frac{2\delta_T' Z_\Delta' M_\iota u}{k-k_0}\right| \geq \frac{\lambda}{3}\|\delta_T\|^2\right) \leq P\left(\sup_{N^c(K)}\left\|\frac{Z_\Delta' M_\iota u}{k-k_0}\right\| \geq \frac{\lambda}{6}\|\delta_T\|\right),$$

which is bounded by

$$\frac{C}{\lambda^2\|\delta_T\|^2}\frac{1}{K\|\delta_T\|^{-2}} = \frac{C}{\lambda^2 K} < \frac{\varepsilon}{6}$$

for large K by Lemma 2.5. Regarding the second term of (2.14), $\sup_{N^c(K)}\|Z_\Delta' M_\iota W_k/(k-k_0)\| = O_p(1)$, $(W_k' M_\iota W_k)^{-1} = O_p(T^{-1})$ and $W_k' M_\iota u = O_p\left(T^{\frac{1}{2}}\right)$ uniformly on $\Lambda \cdot T$ by Lemma 2.2, and $T^{-1/2}\|\delta_T\| = o\left(\|\delta_T\|^2\right)$. Therefore the contribution of this term to (2.13) is bounded by $\varepsilon/6$ for large K and T. The contribution of the third term of (2.14) to (2.13) is bounded by

$$\begin{aligned} & P\left(\sup_{N^c(K)}\left\|\frac{u'(M_{\iota,X,Z_k} - M_{\iota,X,Z_0})u}{k-k_0}\right\| \geq \frac{\lambda}{3}\|\delta_T\|^2\right) \\ \leq\; & P\left(\sup_{k\in\Lambda\cdot T}\|u'(M_{\iota,X,Z_k} - M_{\iota,X,Z_0})u\| \geq \frac{\lambda K}{3}\right) \leq \frac{\varepsilon}{6} \end{aligned}$$

for large K by Lemma 2.3. This concludes the proof that for large K the second term on the right of (2.13) is bounded by $\varepsilon/2$. The proposition is proved. ■

Proof of Proposition 2.2. (a) Denote $W_0 = (X, Z_0)$. We have

$$\sqrt{T}\begin{pmatrix}\hat{\beta}_{k_0} - \beta \\ \hat{\delta}_{k_0} - \delta_T\end{pmatrix} = \left(\frac{1}{T}W_0' M_\iota W_0\right)^{-1}\frac{1}{\sqrt{T}}W_0' M_\iota u.$$

By Lemma 2.2,

$$\frac{1}{T}W_0' M_\iota W_0 \xrightarrow{p} \begin{pmatrix} 1 & \tau_0 \\ \tau_0 & \tau_0\end{pmatrix} \otimes \Sigma$$

and

$$\frac{1}{\sqrt{T}}W_0' M_\iota u \xrightarrow{d} \begin{pmatrix}\Omega^{\frac{1}{2}}W(1) \\ \Omega^{\frac{1}{2}}W(\tau_0)\end{pmatrix}.$$

Therefore the asymptotic distribution of $\sqrt{T}\left(\left(\hat{\beta}_{k_0}-\beta\right)', \left(\hat{\delta}_{k_0}-\delta_T\right)'\right)'$ is normal with zero mean and variance V.

(b) We have

$$\sqrt{T}\begin{pmatrix}\hat{\beta}-\beta\\ \hat{\delta}-\delta_T\end{pmatrix} = \left(\frac{1}{T}W_{\hat{k}}'M_\iota W_{\hat{k}}\right)^{-1}\left(\frac{1}{\sqrt{T}}W_{\hat{k}}'M_\iota u + \frac{1}{\sqrt{T}}W_{\hat{k}}'M_\iota\left(Z_0-Z_{\hat{k}}\right)\delta_T\right).$$

Write

$$\begin{aligned} Z_{\hat{k}}'M_\iota Z_{\hat{k}} &= Z_0'M_\iota Z_0 + \left(Z_{\hat{k}}-Z_0\right)'M_\iota\left(Z_{\hat{k}}-Z_0\right) \\ &\quad + \left(Z_{\hat{k}}-Z_0\right)'M_\iota Z_0 + Z_0'M_\iota\left(Z_{\hat{k}}-Z_0\right) \end{aligned}$$

and

$$Z_{\hat{k}}'M_\iota X = \left(Z_{\hat{k}}-Z_0\right)'M_\iota X + Z_0'M_\iota X.$$

For any $M > 0$,

$$\begin{aligned} & P\left(\left\|\left(Z_{\hat{k}}-Z_0\right)'M_\iota\left(Z_{\hat{k}}-Z_0\right)\right\| > M\left\|\delta_T\right\|^{-2}\right) \\ \leq\ & P\left(2\sup_{N(K)}\sup_{1\leq l\leq T}\left\|Z_\Delta'M_\iota Z_l\right\| > M\left\|\delta_T\right\|^{-2}\right) + P\left(\left|\hat{k}-k_0\right| > K\left\|\delta_T\right\|^{-2}\right). \end{aligned}$$

By Proposition 2.1, the second term on the right of the last displayed inequality is bounded by $\varepsilon/2$ for large K. The first term on the right of the this inequality is bounded by $\varepsilon/2$ for large M by Lemma 2.2. It follows that $\left(Z_{\hat{k}}-Z_0\right)'M_\iota\left(Z_{\hat{k}}-Z_0\right) = O_p\left(\left\|\delta_T\right\|^{-2}\right)$. In a similar way, $\left(Z_{\hat{k}}-Z_0\right)'M_\iota Z_0$ and $\left(Z_{\hat{k}}-Z_0\right)'M_\iota X$ are $O_p\left(\left\|\delta_T\right\|^{-2}\right)$. Therefore

$$\frac{1}{T}W_{\hat{k}}'M_\iota W_{\hat{k}} = \frac{1}{T}W_0'M_\iota W_0 + O_p\left(T^{-1}\left\|\delta_T\right\|^{-2}\right) = \frac{1}{T}W_0'M_\iota W_0 + o_p(1)$$

and by Lemma 2.2,

$$\left(\frac{1}{T}W_{\hat{k}}'M_\iota W_{\hat{k}}\right)^{-1} \xrightarrow{p} \begin{pmatrix}1 & \tau_0\\ \tau_0 & \tau_0\end{pmatrix}^{-1}\otimes\Sigma^{-1}. \tag{2.15}$$

By the same arguments,

$$\frac{1}{\sqrt{T}} W_{\hat{k}}' M_\iota \left(Z_0 - Z_{\hat{k}} \right) \delta_T = O_p \left(T^{-1/2} \left|\left| \delta_T \right|\right|^{-2} \left|\left| \delta_T \right|\right| \right) = o_p \left(1 \right). \tag{2.16}$$

The term $T^{-1/2} W_{\hat{k}}' M_\iota u$ can be written as

$$\frac{1}{\sqrt{T}} W_{\hat{k}}' M_\iota u = \frac{1}{\sqrt{T}} W_0' M_\iota u + \frac{1}{\sqrt{T}} \left(W_{\hat{k}} - W_0 \right)' M_\iota u, \tag{2.17}$$

where $\left(W_{\hat{k}} - W_0 \right)' M_\iota u = (0', ((Z_{\hat{k}} - Z_0)' M_\iota u)')'$. For any any $M > 0$,

$$\begin{aligned} & P \left(\left\| \left(Z_{\hat{k}} - Z_0 \right)' M_\iota u \right\| > M \left|\left| \delta_T \right|\right|^{-1} \right) \\ \leq \ & P \left(\sup_{N(K)} \left\| Z_\Delta' M_\iota u \right\| > M \left|\left| \delta_T \right|\right|^{-1} \right) + P \left(\left| \hat{k} - k_0 \right| > K \left|\left| \delta_T \right|\right|^{-2} \right) \end{aligned}$$

The terms on the right of the last display are bounded by $\varepsilon/2$ for large M and K by Lemma 2.2 and Proposition 2.1, respectively. Therefore $T^{-1/2} \left(Z_{\hat{k}} - Z_0 \right)' M_\iota u$ is $O_p \left(T^{-1/2} \left|\left| \delta_T \right|\right|^{-1} \right) = o_p \left(1 \right)$ and by (2.17) and Lemma 2.2,

$$\frac{1}{\sqrt{T}} W_{\hat{k}}' M_\iota u \xrightarrow{d} \begin{pmatrix} \Omega^{\frac{1}{2}} W \left(1 \right) \\ \Omega^{\frac{1}{2}} W \left(\tau_0 \right) \end{pmatrix}.$$

The last display together with (2.15) and (2.16) imply that

$$\begin{aligned} \sqrt{T} \begin{pmatrix} \hat{\beta} - \beta \\ \hat{\delta} - \delta_T \end{pmatrix} & \xrightarrow{d} \left(\begin{pmatrix} 1 & \tau_0 \\ \tau_0 & \tau_0 \end{pmatrix}^{-1} \otimes \Sigma^{-1} \right) \begin{pmatrix} \Omega^{\frac{1}{2}} W \left(1 \right) \\ \Omega^{\frac{1}{2}} W \left(\tau_0 \right) \end{pmatrix} \\ \sim \ & N \left(0, \frac{1}{\tau_0 \left(1 - \tau_0 \right)} \begin{pmatrix} \tau_0 & -\tau_0 \\ -\tau_0 & 1 \end{pmatrix} \otimes \Sigma^{-1} \Omega \Sigma^{-1} \right) \end{aligned}$$

as maintained by the proposition. ■

Proof of Proposition 2.3. Let $\tilde{k} = \arg\min_{N(K)} S_T \left(k \right)$, $\hat{m} = \arg\min_m W^0 \left(m \right)$ and $\tilde{m} = \arg\min_{|m| \leq K} W^0 \left(m \right)$. For any $K > 0$ and for any integer j,

$$P \left(\hat{k} - k_0 = j \right) = P \left(\hat{k} - k_0 = j, \left| \hat{k} - k_0 \right| \leq K \right) + P \left(\hat{k} - k_0 = j, \left| \hat{k} - k_0 \right| > K \right). \tag{2.18}$$

Since the event $\left\{\left|\hat{k}-k_0\right| \leq K\right\}$ is equivalent to the event $\left\{\hat{k}=\tilde{k}\right\}$, the first term on the right of (2.18) is equal to

$$P\left(\tilde{k}-k_0=j\right)-P\left(\tilde{k}-k_0=j,\left|\hat{k}-k_0\right|>K\right).$$

By similar arguments,

$$P\left(\hat{m}=j\right)=P\left(\tilde{m}=j\right)-P\left(\tilde{m}=j,\left|\hat{m}\right|>K\right)+P\left(\hat{m}=j,\left|\hat{m}\right|>K\right),$$

and therefore

$$\begin{aligned}
& \left|P\left(\hat{k}-k_0=j\right)-P\left(\hat{m}=j\right)\right| \\
\leq\ & \left|P\left(\tilde{k}-k_0=j\right)-P\left(\tilde{m}=j\right)\right| \\
& +2P\left(\left|\hat{k}-k_0\right|>K\right)+2P\left(\left|\hat{m}\right|>K\right). \qquad (2.19)
\end{aligned}$$

Since $Z_\Delta' M_\iota Z_\Delta = Z_\Delta' Z_\Delta + o_p(1)$ and $Z_\Delta' M_\iota u = Z_\Delta' u + o_p(1)$ uniformly on $N(K)$, Lemma 2.6 implies that

$$S_T(k)-S_T(k_0)=\delta' Z_\Delta' Z_\Delta \delta - 2\delta' Z_\Delta' u \operatorname{sgn}(k-k_0)+o_p(1)$$

uniformly on $N(K)$. It follows from the continuous mapping theorem that for any $K>0$,

$$\tilde{k}=\arg\min_{N(K)}\left(S_T(k)-S_T(k_0)\right) \xrightarrow{d} \arg\min_{N(K)}\left(\delta' Z_\Delta' Z_\Delta \delta - 2\delta' Z_\Delta' u \operatorname{sgn}(k-k_0)\right)$$

which has the same distribution as $\arg\min_{|m|\leq K} W^0(m)$ under strict stationarity. The first term of (2.19) is therefore equal to 0 when $|j|>K$ by definition and smaller than $\varepsilon/3$ for large T when $|j|\leq K$. The second term of (2.19) is smaller than $\varepsilon/3$ for large K by Proposition 2.1. Since by Conditions 1.1 and 1.2, $W^0(m) \xrightarrow{p} \infty$ for $|m|\to\infty$, we have that $\hat{m}=O_p(1)$ and that the third term of (2.19) is smaller than $\varepsilon/3$ for large K. It follows that for each j, $P\left(\hat{k}-k_0=j\right)-P\left(\hat{m}=j\right)\to 0$. ■

Proof of the Proposition 2.4. Let $v_T^2=\left(\delta_T'\Sigma\delta_T\right)^2/\delta_T'\Omega\delta_T$. For any $K>0$

and for any real x,

$$\begin{aligned} P\left(v_T^2\left(\hat{k}-k_0\right)\leq x\right) &= P\left(\hat{k}-k_0\leq xv_T^{-2}, \left|\hat{k}-k_0\right|\leq Kv_T^{-2}\right) \\ &\quad +P\left(\hat{k}-k_0\leq xv_T^{-2}, \left|\hat{k}-k_0\right|> Kv_T^{-2}\right). \end{aligned}$$

Let $\tilde{k} = \arg\min_{|k-k_0|\leq Kv_T^{-2}} S_T(k)$, $\hat{\rho} = \arg\min_{\rho\in\mathbb{R}} \bar{W}(\rho)$ and
$\tilde{\rho} = \arg\min_{\rho\in[-K,K]} \bar{W}(\rho)$. Reasoning as in the proof of Proposition 2.3, we obtain

$$\begin{aligned} \left|P\left(v_T^2\left(\hat{k}-k_0\right)\leq x\right)-P(\hat{\rho}\leq x)\right| &\leq \left|P\left(v_T^2\left(\tilde{k}-k_0\right)\leq x\right)-P(\tilde{\rho}\leq x)\right| \\ &\quad +2P\left(\left|\hat{k}-k_0\right|> Kv_T^{-2}\right)+2P(|\hat{\rho}|>K). \end{aligned} \tag{2.20}$$

Since $v_T^2\left(\tilde{k}-k_0\right) = \arg\min_{|\rho|\leq[Kv_T^{-2}]/v_T^{-2}} S_T\left(k_0+\rho v_T^{-2}\right)$, Lemma 2.7 implies that $v_T^2\left(\tilde{k}-k_0\right) \xrightarrow{d} \arg\min_{|\rho|\leq K} \bar{W}(\rho)$. Therefore the first term of (2.20) is bounded by $\varepsilon/3$ for large T. Since $\hat{k}-k_0 = O_p\left(||\delta_T||^{-2}\right)$, $v_T^{-2} = O\left(||\delta_T||^{-2}\right)$ and since the properties of the Brownian motion with drift imply that $\hat{\rho} = O_p(1)$, the last two terms of (2.20) are smaller than $\varepsilon/3$ for large K and T. Inequality (2.20) then implies that

$$\frac{\left(\delta_T'\Sigma\delta_T\right)^2}{\delta_T'\Omega\delta_T}\left(\hat{k}-k_0\right) \xrightarrow{d} \arg\min_{\rho} \bar{W}(\rho). \tag{2.21}$$

Since $\hat{\Sigma}$, $\hat{\Omega}$ are consistent estimators of Σ, Ω and since $\hat{\delta} = \delta_T + O_p\left(T^{-\frac{1}{2}}\right)$, convergence in (2.21) remains valid with the quantities δ_T, Σ and Ω replaced by their estimators $\hat{\delta}$, $\hat{\Sigma}$ and $\hat{\Omega}$. ■

Proof of Proposition 2.5. Assume for simplicity that $\{x_t\}$ is a scalar process. By Theorem 1 of Robinson (1998),

$$\frac{4\pi^2}{T}\sum_{j=1}^{T-1} I_{xx,j}I_{uu,j} \xrightarrow{p} \Omega. \tag{2.22}$$

It is therefore sufficient to prove that

$$\frac{1}{T}\sum_{j=1}^{T-1} I_{xx,j}\left(I_{\hat{u}\hat{u},j} - I_{uu,j}\right) \xrightarrow{p} 0. \tag{2.23}$$

Using the fact that $I_{\hat{u}\hat{u},j} - I_{uu,j} = \left|w_{\hat{u},j} - w_{u,j}\right|^2 - 2\,\mathrm{Re}\left(w_{\hat{u},j} - w_{u,j}\right)\bar{w}_{u,j}$, we obtain that the left-hand side of (2.23) is bounded on absolute value by

$$\frac{1}{T}\sum_{j=1}^{T-1}\left|w_{x,j}\right|^2\left|w_{\hat{u},j} - w_{u,j}\right|^2 + \frac{2}{T}\sum_{j=1}^{T-1}\left|w_{x,j}\right|^2\left|w_{\hat{u},j} - w_{u,j}\right|\left|w_{u,j}\right|. \tag{2.24}$$

From (2.5) and (2.6) and the definition of $\hat{u}$,

$$w_{\hat{u},j} - w_{u,j} = w_{x,j}\left(\beta - \hat{\beta}\right) + \left(w_{z_0,j} - w_{z_{\hat{k}},j}\right)\delta_T + w_{z_{\hat{k}},j}\left(\delta_T - \hat{\delta}\right).$$

Using the c_r-inequality, the first term of (2.24) can be bounded by

$$\begin{aligned}&\frac{3}{T}\left(\beta - \hat{\beta}\right)^2\sum_{j=1}^{T-1}\left|w_{x,j}\right|^4 + \frac{3}{T}\delta_T^2\sum_{j=1}^{T-1}\left|w_{x,j}\right|^2\left|w_{z_0,j} - w_{z_{\hat{k}},j}\right|^2\\ &+\frac{3}{T}\left(\delta_T - \hat{\delta}\right)^2\sum_{j=1}^{T-1}\left|w_{x,j}\right|^2\left|w_{z_{\hat{k}},j}\right|^2.\end{aligned} \tag{2.25}$$

By Propositions 2.2 and 2.7, $\beta - \hat{\beta} = O_p\left(T^{-1/2}\right)$, and by Lemma 1.11, $T^{-1}\sum_{j=1}^{T-1}\left|w_{x,j}\right|^4 = o_p(T)$, therefore the first term of (2.25) is $o_p(1)$. The second term of (2.25) is bounded by

$$\frac{3}{T}\delta_T^2\left(\sum_{j=1}^{T-1}\left|w_{x,j}\right|^4\right)^{\frac{1}{2}}\left(\sum_{j=1}^{T-1}\left|w_{z_0,j} - w_{z_{\hat{k}},j}\right|^4\right)^{\frac{1}{2}}$$

by the Cauchy-Schwarz inequality. The expression in the last bracket is bounded by

$$\frac{1}{(2\pi T)^2}\sum_{t,s,r,v}^{k_0,\hat{k}}|x_t x_s x_r x_v|\left|\sum_{j=1}^{T-1}e^{i(t-s+r-v)\lambda_j}\right|$$

$$\leq \frac{1}{(2\pi T)^2}\sum_{t,s,r,v}^{k_0,\hat{k}}|x_t x_s x_r x_v|\left(1+T\mathbb{I}\left(t-s+r-v=0 \mod T\right)\right)$$

$$\leq \frac{1}{(2\pi T)^2}\left(\sum_t^{k_0,\hat{k}}|x_t|\right)^4+\frac{1}{(2\pi)^2 T}\sum_t^{k_0,\hat{k}}|x_t|\sum_s^{k_0,\hat{k}}|x_s|\left(\sum_r^{k_0,\hat{k}}|x_r|^2\sum_v^{k_0,\hat{k}}|x_{t-s+v}|^2\right)^{\frac{1}{2}},$$

where the last inequality follows from the Cauchy-Schwarz inequality. For any $K>0$ and $M>0$,

$$P\left(\sum_t^{k_0,\hat{k}}|x_t|\geq M\|\delta_T\|^{-2}\right) \leq P\left(\sum_{t=k_0-K\|\delta_T\|^{-2}}^{k_0+K\|\delta_T\|^{-2}}|x_t|\geq M\|\delta_T\|^{-2}\right) + P\left(\left|\hat{k}-k_0\right|>K\|\delta_T\|^{-2}\right). \tag{2.26}$$

By Proposition 2.1, the second term on the right of (2.26) is smaller than $\varepsilon/2$ for large K. By the Markov inequality and by Conditions 1.1 and 1.2, the first term on the right of (2.26) is bounded by CK/M which is smaller than $\varepsilon/2$ for large M. Therefore $\sum_t^{k_0,\hat{k}}|x_t| = O_p\left(\|\delta_T\|^{-2}\right)$. By similar arguments, $\sum_t^{k_0,\hat{k}}|x_t|^2$ is $O_p\left(\|\delta_T\|^{-2}\right)$. Since

$$\sup_{t,s\in N(K)}\sum_{v=k_0-K\|\delta_T\|^{-2}}^{k_0+K\|\delta_T\|^{-2}}|x_{t-s+v}|^2\leq\sum_{v=k_0-3K\|\delta_T\|^{-2}}^{k_0+3K\|\delta_T\|^{-2}}|x_v|^2,$$

it follows that the second term of (2.25) is

$$O_p\left(T^{-1}\|\delta_T\|^2\right)o_p(T)\,O_p\left(T^{-1}\|\delta_T\|^{-4}+T^{-1/2}\|\delta_T\|^{-3}\right)=o_p(1).$$

The sum in the third term of (2.25) is bounded by

$$\sum_{j=1}^{T-1}|w_{x,j}|^2|w_{z0,j}|^2+\sum_{j=1}^{T-1}|w_{x,j}|^2\left|w_{z0,j}-w_{z_{\hat{k}},j}\right|^2, \tag{2.27}$$

where the first term is $o_p(T^2)$ by Lemma 1.11 and where the second term is $O_p\left(T^{-1/2}\|\delta_T\|^{-3}\right)$ by the previous discussion. Noting that $\delta_T - \hat{\delta} = O_p\left(T^{-1/2}\right)$ by Propositions 2.2 and 2.7, we conclude that the third term of (2.25), and hence the first term of (2.24), is $o_p(1)$.

The second term of (2.24) is bounded by

$$C\left(\frac{1}{T}\sum_{j=1}^{T-1}|w_{x,j}|^2\left|w_{\hat{u},j}-w_{u,j}\right|^2\right)^{\frac{1}{2}}\left(\frac{1}{T}\sum_{j=1}^{T-1}|w_{x,j}|^2|w_{u,j}|^2\right)^{\frac{1}{2}} = o_p(1)$$

since the first bracket is $o_p(1)$ as has been just shown, and where the second bracket is $O_p(1)$ by (2.22). ■

Proof of Proposition 2.6. Write

$$\begin{aligned}\frac{\hat{k}}{T} &= \frac{1}{T}\arg\min_{k\in\Lambda\cdot T} S_T(k) = \arg\min_{\tau\in\Lambda} S_T([\tau T]) \\ &= \arg\max_{\tau\in\Lambda}\left(y'M_{\iota,X}y - S_T([\tau T])\right).\end{aligned}$$

The maximand is equal to

$$y'M_{\iota,X}y - y'M_{\iota,X,Z_{[\tau T]}}y = y'M_{\iota,X}Z_{[\tau T]}\left(Z'_{[\tau T]}M_{\iota,X}Z_{[\tau T]}\right)^{-1}Z'_{[\tau T]}M_{\iota,X}y$$

$$\begin{aligned}&= \left(\sqrt{T}\delta'_T\frac{1}{T}Z'_0M_{\iota,X}Z_{[\tau T]} + \frac{1}{\sqrt{T}}u'M_{\iota,X}Z_{[\tau T]}\right)\left(\frac{1}{T}Z'_{[\tau T]}M_{\iota,X}Z_{[\tau T]}\right)^{-1} \\ &\quad\times\left(\frac{1}{T}Z'_{[\tau T]}M_{\iota,X}Z_0\sqrt{T}\delta_T + \frac{1}{\sqrt{T}}Z'_{[\tau T]}M_{\iota,X}u\right) \qquad (2.28)\end{aligned}$$

where δ_T may be equal to zero. By Lemma 2.2,

$$\begin{aligned}\frac{1}{T}Z'_0M_{\iota,X}Z_{[\tau T]} &= \frac{1}{T}Z'_0M_\iota Z_{[\tau T]} - \frac{1}{T}Z'_0M_\iota X\left(\frac{1}{T}X'M_\iota X\right)^{-1}\frac{1}{T}X'M_\iota Z_{[\tau T]} \\ &\Longrightarrow (\tau_0\wedge\tau)\Sigma - \tau_0\Sigma\Sigma^{-1}\tau\Sigma = m(\tau)\Sigma\end{aligned}$$

on $[0,1]$ and similarly $\frac{1}{T}Z'_{[\tau T]}M_{\iota,X}Z_{[\tau T]} \Longrightarrow \tau(1-\tau)\Sigma$ on $[0,1]$. Further, by

Lemma 2.2,

$$\begin{aligned}\frac{1}{\sqrt{T}}Z_{[\tau T]}' M_{\iota,X}u &= \frac{1}{\sqrt{T}}Z_{[\tau T]}' M_\iota u - \frac{1}{T}Z_{[\tau T]}' M_\iota X\left(\frac{1}{T}X' M_\iota X\right)^{-1}\frac{1}{\sqrt{T}}X' M_\iota u \\ &\Longrightarrow \Omega^{\frac{1}{2}}W(\tau) - \tau\Sigma\Sigma^{-1}\Omega^{\frac{1}{2}}W(1) = \Omega^{\frac{1}{2}}\left(W(\tau) - \tau W(1)\right)\end{aligned}$$

on $[0,1]$. Therefore

$$S_T([\tau T]) - y' M_{\iota,X}y \Longrightarrow G(\tau)' G(\tau)$$

on Λ and the proposition follows from the continuous mapping theorem (see for example Theorem 3.2.2 of van der Vaart and Wellner (1996)). ■

Proof of Proposition 2.7. Define

$$Y_T(\tau) = \sqrt{T}\begin{pmatrix}\hat{\beta}_{[\tau T]} - \beta \\ \hat{\delta}_{[\tau T]} - \delta_T\end{pmatrix}$$

and

$$\begin{aligned}Y(\tau)\frac{1}{\tau(1-\tau)} & \left(\begin{pmatrix}\Sigma^{-1}\Omega^{\frac{1}{2}}(\tau W(1) - \tau W(\tau)) \\ \Sigma^{-1}\Omega^{\frac{1}{2}}(W(1) - \tau W(1))\end{pmatrix}\begin{matrix}(\tau)) \\ 1))\end{matrix}\right. \\ & \left. +\frac{1}{\tau(1-\tau)}\begin{pmatrix}\tau(\tau_0 - \tau)\mathbb{I}(\tau < \tau_0) \\ (\tau_0 - \tau)(\mathbb{I}(\tau_0 \leq \tau) - \tau)\end{pmatrix}_{-\tau)}\right)\otimes\delta.\end{aligned}$$

We need to show that $Y_T(\hat{\tau}) \xrightarrow{d} Y(L)$. To that end, write

$$\begin{aligned}Y_T(\tau) &= \left(\frac{1}{T}W_{[\tau T]}' M_\iota W_{[\tau T]}\right)^{-1} \\ &\quad\times\left(\frac{1}{\sqrt{T}}W_{[\tau T]}' M_\iota u + \frac{1}{T}W_{[\tau T]}' M_\iota\left(Z_0 - Z_{[\tau T]}\right)\sqrt{T}\delta_T\right). \qquad (2.29)\end{aligned}$$

Expressions $y' M_{\iota,X}y - y' M_{\iota,X,Z_{[\tau T]}}y$ in (2.28) and $Y_T(\tau)$ in (2.29) are continuous functions of matrices $T^{-1/2}Z_{[\tau T]}' M_\iota u$, $T^{-1}Z_{[\tau T]}' M_\iota Z_{[\tau T]}$, $T^{-1}X' M_\iota Z_{[\tau T]}$ and $T^{-1}Z_0' M_\iota Z_{[\tau T]}$ on $\tau \in \Lambda$, therefore we need to study the joint convergence of

$$\left(\frac{1}{\sqrt{T}}Z_{[\tau T]}' M_\iota u, \frac{1}{T}Z_{[\tau T]}' M_\iota Z_{[\tau T]}, \frac{1}{T}X' M_\iota Z_{[\tau T]}, \frac{1}{T}Z_0' M_\iota Z_{[\tau T]}\right). \qquad (2.30)$$

But by Lemma 2.2, $T^{-1}Z'_{[\tau T]}M_\iota Z_{[\tau T]} \Longrightarrow \tau\Sigma$ and $T^{-1}X'M_\iota Z_{[\tau T]} \Longrightarrow \tau\Sigma$ and $T^{-1}Z'_0 M_\iota Z_{[\tau T]} \Longrightarrow (\tau_0 \wedge \tau)\,\Sigma$ on $[0,1]$. Also by Lemma 2.2, $T^{-1/2}Z'_{[\tau T]}M_\iota u \Longrightarrow \Omega^{\frac{1}{2}}W(\tau)$ on $[0,1]$. Hence (2.30) converges weakly to

$$\left(\Omega^{\frac{1}{2}}W(\tau), \tau\Sigma, \tau\Sigma, (\tau_0 \wedge \tau)\,\Sigma\right)$$

because all but the first component converge weakly to constant functions in the space $C[0,1]$ of p-vectors of continuous functions on $[0,1]$. The continuous mapping theorem, Proposition 2.6 and the assumption $T^{1/2}\delta_T \to \delta$ imply that

$$(Y_T(\tau), \hat{\tau}) \Longrightarrow (Y(\tau), L) \tag{2.31}$$

on $D(\Lambda) \times \Lambda$.

For an arbitrarily small $\eta > 0$, choose points $\tau_0, \tau_1, \ldots, \tau_v$ such that $0 = \tau_0 < \tau_1 < \ldots < \tau_v = 1$ and $\sup_{1 \le i \le v} |\tau_i - \tau_{i-1}| < \eta$. For $i = 1, \ldots, v$, denote $D_i = [\tau_{i-1}, \tau_i) \cap \Lambda$. Then for arbitrary $\rho > 0$ and any x,

$$\begin{aligned}
& P(Y_T(\hat{\tau}) \le x) \\
\le\ & \sum_{i=1}^{v} P\left(Y_T(\tau_{i-1}) \le x + \rho, \hat{\tau} \in D_i, \sup_{t,s \in D_i} |Y_T(t) - Y_T(s)| < \rho\right) \\
& + \sum_{i=1}^{v} P\left(\hat{\tau} \in D_i, \sup_{t,s \in D_i} |Y_T(t) - Y_T(s)| \ge \rho\right).
\end{aligned}$$

By (2.31) and by the portmanteau lemma (see for example Theorem 2.1 of Billingsley (1999, p. 16)), the right-hand side of the last displayed inequality converges to

$$\begin{aligned}
& \sum_{i=1}^{v} P\left(Y(\tau_{i-1}) \le x + \rho, L \in D_i, \sup_{t,s \in D_i} |Y(t) - Y(s)| < \rho\right) \\
& + \sum_{i=1}^{v} P\left(L \in D_i, \sup_{t,s \in D_i} |Y(t) - Y(s)| \ge \rho\right) \\
\le\ & P(Y(L) \le x + 2\rho) + P\left(\sup_{|t-s| \le \eta} |Y(t) - Y(s)| \ge \rho\right)
\end{aligned}$$

as $T \to \infty$. For any $\varepsilon > 0$, the second term on the right of the last displayed inequality is bounded by ε for sufficiently small η because the random function Y takes values in $C[0,1]$. Therefore

$$\limsup_{T\to\infty} P\left(Y_T\left(\hat{\tau}\right)\leq x\right)\leq P\left(Y\left(L\right)\leq x+2\rho\right)+\varepsilon$$

for small η. Proceeding similarly, we obtain that

$$P\left(Y\left(L\right)\leq x-2\rho\right)-\varepsilon\leq \liminf_{T\to\infty} P\left(Y_T\left(\hat{\tau}\right)\leq x\right).$$

Since the distribution of the random variable $Y\left(L\right)$ is continuous and since ρ and ε are arbitrarily small, we conclude that

$$P\left(Y_T\left(\hat{\tau}\right)\leq x\right)\to P\left(Y\left(L\right)\leq x\right)$$

as desired. ∎

Proof of Proposition 2.8. We have

$$\begin{aligned} Fy^* &= FX\hat{\beta}+FZ_{\hat{k}}\hat{\delta}+HF\hat{u} \\ &= FX\hat{\beta}+FZ_{\hat{k}^*}\hat{\delta}+\tilde{u}, \end{aligned}$$

where $\tilde{u}=HF\hat{u}+F\left(Z_{\hat{k}}-Z_{\hat{k}^*}\right)\hat{\delta}$. Therefore

$$\begin{aligned} \sqrt{T}\begin{pmatrix} \hat{\beta}^*-\hat{\beta} \\ \hat{\delta}^*-\hat{\delta} \end{pmatrix} &= \left(\frac{1}{T}W_{\hat{k}^*}'\overline{F}'FW_{\hat{k}^*}\right)^{-1}\left(\frac{1}{\sqrt{T}}W_{\hat{k}^*}'\overline{F}'HF\hat{u}\right. \\ &\quad \left.+\frac{1}{\sqrt{T}}W_{\hat{k}^*}'\overline{F}'F\left(Z_{\hat{k}}-Z_{\hat{k}^*}\right)\hat{\delta}\right). \end{aligned}$$

Write

$$\begin{aligned} Z_{\hat{k}^*}'\overline{F}'FZ_{\hat{k}^*} &= Z_0'\overline{F}'FZ_0+\left(Z_{\hat{k}^*}-Z_0\right)'\overline{F}'F\left(Z_{\hat{k}^*}-Z_0\right) \\ &\quad +\left(Z_{\hat{k}^*}-Z_0\right)'\overline{F}'FZ_0+Z_0'\overline{F}'F\left(Z_{\hat{k}^*}-Z_0\right) \end{aligned}$$

and

$$X'\overline{F}'FZ_{\hat{k}^*}=X'\overline{F}'FZ_0+X'\overline{F}'F\left(Z_{\hat{k}^*}-Z_0\right).$$

For any $K>0$, expression $P^*\left(\left\|\left(Z_{\hat{k}^*}-Z_0\right)'\overline{F}'F\left(Z_{\hat{k}^*}-Z_0\right)\right\|>M\left\|\delta_T\right\|^{-2}\right)$ is bounded by

$$P^*\left(\sup_{N(K)}\left\|Z_\Delta'\overline{F}'FZ_\Delta\right\| > M\left\|\delta_T\right\|^{-2}\right) + P^*\left(\left|\hat{k}^* - k_0\right| > K\left\|\delta_T\right\|^{-2}\right). \tag{2.32}$$

Fix $\varepsilon, \eta > 0$. Expectation of the first term of (2.32) is smaller than $\varepsilon/3$ for large M and T by Lemma 2.2. The second term of (2.32) is bounded by

$$P^*\left(\left|\hat{k}^* - \hat{k}\right| > \frac{K}{2}\left\|\delta_T\right\|^{-2}\right) + P^*\left(\left|\hat{k} - k_0\right| > \frac{K}{2}\left\|\delta_T\right\|^{-2}\right),$$

where the first term exceeds η with probability smaller than $\varepsilon/3$ for large K by Lemma 2.12 and where expectation of the second term is smaller than $\varepsilon/3$ for large K by Proposition 2.1. Therefore

$$\frac{1}{T}\left(Z_{\hat{k}^*} - Z_0\right)'\overline{F}'F\left(Z_{\hat{k}^*} - Z_0\right) = O_{p^*}\left(T^{-1}\left\|\delta_T\right\|^{-2}\right) = o_{p^*}(1).$$

By Lemma 2.2, $T^{-1}X'\overline{F}'FX$ and $T^{-1}Z_0'\overline{F}'FZ_0$ are $O_p(1)$. The Cauchy-Schwarz inequality implies that the terms $T^{-1}\left(Z_{\hat{k}^*} - Z_0\right)'\overline{F}'FZ_0$ and $T^{-1}\left(Z_{\hat{k}^*} - Z_0\right)'\overline{F}'FX$ are $o_{p*}(1)$. It follows that

$$\frac{1}{T}W_{\hat{k}^*}'\overline{F}'FW_{\hat{k}^*} = \frac{1}{T}W_0'\overline{F}'FW_0 + o_{p^*}(1)$$

and by Lemma 2.2,

$$\left(\frac{1}{T}W_{\hat{k}^*}'\overline{F}'FW_{\hat{k}^*}\right)^{-1} \xrightarrow{p^*} 2\pi\begin{pmatrix} 1 & \tau_0 \\ \tau_0 & \tau_0 \end{pmatrix}^{-1} \otimes \Sigma^{-1}.$$

Second, write

$$\frac{1}{\sqrt{T}}W_{\hat{k}^*}'\overline{F}'HF\hat{u} = \frac{1}{\sqrt{T}}W_0'\overline{F}'HF\hat{u} + \left(\frac{1}{\sqrt{T}}\left(W_{\hat{k}^*} - W_0\right)\overline{F}'HF\hat{u}\right). \tag{2.33}$$

Expression $P^*\left(\left\|\left(W_{\hat{k}^*} - W_0\right)'\overline{F}'HF\hat{u}\right\| > M\left\|\delta_T\right\|^{-1}\right)$ is bounded by

$$P^*\left(\sup_{N(K)}\left\|Z_\Delta'\overline{F}'HF\hat{u}\right\| > M\left\|\delta_T\right\|^{-1}\right) + P^*\left(\left|\hat{k}^* - k_0\right| > K\left\|\delta_T\right\|^{-2}\right). \tag{2.34}$$

By Lemma 2.8, the first term of (2.34) is smaller than η with probability at least $1-\varepsilon/2$ for large M. Arguing as in (2.32), the second term of (2.34) is smaller than η with probability at least $1-\varepsilon/2$ for large K. Therefore

$$\frac{1}{\sqrt{T}}\left(W_{\hat{k}^*}-W_0\right)'\overline{F}'HF\hat{u}=O_{p^*}\left(T^{-\frac{1}{2}}\left\|\delta_T\right\|^{-1}\right)=o_{p^*}(1)$$

and by (2.33) and Lemma 2.8,

$$\frac{1}{\sqrt{T}}W_{\hat{k}^*}'\overline{F}'HF\hat{u}\xrightarrow{d^*}\frac{1}{2\pi}\Omega^{\frac{1}{2}}\begin{pmatrix}W(1)\\W(\tau_0)\end{pmatrix}.$$

Finally, for any $M>0$,

$$\begin{aligned}
&P^*\left(\left\|\left(Z_{\hat{k}^*}-Z_{\hat{k}}\right)'\overline{F}'FW_{\hat{k}^*}\right\|>M\left\|\delta_T\right\|^{-2}\right)\\
\leq\ &P^*\left(4\sup_{N(K/2)}\sup_{1\leq l\leq T}\left\|Z_{\Delta}'\overline{F}'FW_l\right\|>M\left\|\delta_T\right\|^{-2}\right)\\
&+P^*\left(\left|\hat{k}^*-k_0\right|>K\left\|\delta_T\right\|^{-2}\right)\\
&+P^*\left(\left|\hat{k}-k_0\right|>K\left\|\delta_T\right\|^{-2}\right).
\end{aligned}\tag{2.35}$$

By Lemma 2.2, expectation of the first term of (2.35) is smaller than $\varepsilon/3$ for large M. Proceeding as before, we obtain

$$\frac{1}{\sqrt{T}}\left(Z_{\hat{k}^*}-Z_{\hat{k}}\right)'\overline{F}'FW_{\hat{k}^*}\hat{\delta}=T^{-\frac{1}{2}}O_{p^*}\left(\left\|\delta_T\right\|^{-2}\right)O_p\left(\left\|\delta_T\right\|\right)=o_{p^*}(1),$$

where the bound for $\hat{\delta}$ is due to Proposition 2.2. By the continuous mapping theorem,

$$\sqrt{T}\begin{pmatrix}\hat{\beta}^*-\hat{\beta}\\\hat{\delta}^*-\hat{\delta}\end{pmatrix}\xrightarrow{d^*}2\pi\left(\begin{pmatrix}1&\tau_0\\\tau_0&\tau_0\end{pmatrix}^{-1}\otimes\Sigma^{-1}\right)\frac{1}{2\pi}\begin{pmatrix}\Omega^{\frac{1}{2}}W(1)\\\Omega^{\frac{1}{2}}W(\tau_0)\end{pmatrix}.$$

This implies that the proposition holds true. ■

Proof of Proposition 2.9. For any $0<K<\infty$, let $\hat{\rho}$, $\tilde{\rho}$ and v_T^2 be defined as in the proof of Proposition 2.4 and let $\tilde{k}^*=\arg\min_{\left|k-\hat{k}\right|\leq Kv_T^{-2}}S_T^*(k)$.

Proceeding as in the proof of Proposition 2.4, write

$$\left|P^*\left(v_T^2\left(\hat{k}^*-\hat{k}\right)\leq x\right)-P\left(\hat{\rho}\leq x\right)\right|\leq\left|P^*\left(v_T^2\left(\tilde{k}^*-\hat{k}\right)\leq x\right)-P\left(\tilde{\rho}\leq x\right)\right|$$
$$+2P^*\left(\left|\hat{k}^*-\hat{k}\right|>Kv_T^{-2}\right)+2P\left(|\hat{\rho}|>K\right). \tag{2.36}$$

Fix $\varepsilon,\eta>0$. Since $v_T^2\left(\tilde{k}^*-\hat{k}\right)=\arg\min_{|\rho|\leq\left[Kv_T^{-2}\right]/v_T^{-2}}S_T^*\left(\hat{k}+\left[\rho v_T^{-2}\right]\right)$, Lemma 2.14 implies that

$$v_T^2\left(\tilde{k}^*-\hat{k}\right)\xrightarrow{d^*}\arg\min_{|\rho|\leq K}\bar{W}\left(\rho\right)=\tilde{\rho}$$

and so the first term on the right of (2.36) is smaller than $\eta/3$ with probability at least $1-\varepsilon/3$ for large T. By Lemma 2.12, the second term on the right of (2.36) is smaller than $\eta/3$ with probability no smaller than $1-\varepsilon/3$ for large K and T. The third term on the right of (2.36) is bounded by $\eta/3$ for large K because $\hat{\rho}=O_p\left(1\right)$. Therefore the right-hand side of (2.36) is $o_p\left(1\right)$ and

$$\frac{\left(\delta_T'\Sigma\delta_T\right)^2}{\delta_T'\Omega\delta_T}\left(\hat{k}^*-\hat{k}\right)\xrightarrow{d^*}\arg\min_{\rho}\bar{W}\left(\rho\right).$$

Since $\hat{\delta}^*=\hat{\delta}+O_{p^*}\left(T^{-\frac{1}{2}}\right)=\delta_T+O_{p^*}\left(T^{-\frac{1}{2}}\right)$ by Propositions 2.7 and 2.8, $\hat{\Sigma}\xrightarrow{p}\Sigma$ by Conditions 1.1 and 1.2, and $\hat{\Omega}^*\xrightarrow{p^*}\Omega$ by Lemma 2.15, it follows that

$$\frac{\left(\hat{\delta}^{*\prime}\hat{\Sigma}\hat{\delta}^*\right)^2}{\hat{\delta}^{*\prime}\hat{\Omega}^*\hat{\delta}^*}\left(\hat{k}^*-\hat{k}\right)\xrightarrow{d^*}\arg\min_{\rho}\bar{W}\left(\rho\right).$$

■

2.B Lemmas

This section contains the some auxiliary results and their proofs. Throughout this section, it is assumed that Conditions 1.1-1.5 and 2.1 hold.

Lemma 2.1 *For any matrix*

$$\begin{pmatrix} A_1 & B_1 \\ A_2 & B_2 \end{pmatrix}$$

and $A = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}$ *and* $B = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}$,

$$A' M_B A \geq A_1' M_{B_1} A_1.$$

Proof. The inequality is related to the fact that in the context of a projection of vectors on the space spanned by the columns of matrix B, the sum of squared residuals is nondecreasing as the number of observations increases. For a proof, see for example Lemma A.1 of Bai and Perron (1998) or Lemma 2 of Brown et al. (1975). ■

Lemma 2.2 *For any* $0 < K < \infty$,

(a) $\frac{1}{T} Z_{[\tau T]} M_\iota Z_{[\sigma T]} \Longrightarrow (\tau \wedge \sigma) \Sigma$ *on* $(\tau, \sigma) \in [0,1]^2$,

(b) $\frac{1}{\sqrt{T}} Z_{[\tau T]}' M_\iota u \Longrightarrow \Omega^{\frac{1}{2}} W(\tau)$ *on* $\tau \in [0,1]$,

(c) $\sup_{k \in \Lambda \cdot T} \left\| (W_k' M_\iota W_k)^{-1} \right\| = O_p(T^{-1})$,

(d) $\sup_{1 \leq k \leq T} \| W_k' M_\iota u \| = O_p\left(T^{\frac{1}{2}}\right)$,

(e) $\sup_{1 \leq k, l \leq T} \| Z_k' M_\iota Z_l \| = O_p(T)$,

(f) $\sup_{k \in N(K)} \sup_{1 \leq l \leq T} \| Z_\Delta' M_\iota Z_l \| = O_p\left(\|\delta_T\|^{-2}\right)$,

(g) if $T \|\delta_T\|^2 \to \infty$ *then* $\sup_{N(K)} \| Z_\Delta' M_\iota u \| = O_p\left(\|\delta_T\|^{-1}\right)$,

(h) $\sup_{k \in N^c(K)} \sup_{1 \leq l \leq T} \left\| \frac{Z_\Delta' M_\iota Z_l}{k - k_0} \right\| = O_p(1)$.

Proof. Parts (a) and (b) follow from Lemmas 1.9 and 1.7, respectively, after noting that

$$\frac{1}{T} Z_{[\tau T]}' M_\iota Z_{[\sigma T]} = \frac{2\pi}{T} Z_{[\tau T]}' \overline{F}' F Z_{[\sigma T]} = \frac{2\pi}{T} \sum_{j=1}^{T-1} w_{z([\tau T]),j} \bar{w}_{z([\sigma T]),j}$$

and

$$\frac{1}{\sqrt{T}} Z_{[\tau T]}' M_\iota u = \frac{2\pi}{\sqrt{T}} Z_{[\tau T]}' \overline{F}' F u = \frac{2\pi}{\sqrt{T}} \sum_{j=1}^{T-1} w_{z([\tau T]),j} \bar{w}_{u,j}.$$

Parts (c), (d) and (e) are implied by parts (a) and (b). Part (f) follows from the bound

$$\|Z'_\Delta M_\iota Z_l\| = \left\| Z'_\Delta Z_l - \frac{1}{T} Z'_\Delta \iota_\Delta \iota'_l Z_l \right\| \leq \|Z_\Delta\|^2 + |k-k_0|^{\frac{1}{2}} \|Z_\Delta\| \frac{l^{\frac{1}{2}}}{T} \|Z_l\| \tag{2.37}$$

and from Conditions 1.1, 1.2 and 2.1. Turning to part (g), for any given K and $\rho \in [-K, K]$, write $k = k_0 + \left[\rho \|\delta_T\|^{-2}\right]$. Part (b) implies that

$$\frac{1}{\sqrt{T}} Z'_{\Delta([\tau_0 T],[(\tau_0+\rho)T])} M_\iota u \Longrightarrow \begin{cases} \Omega^{\frac{1}{2}} \left(W(\tau_0+\rho) - W(\tau_0)\right) & \rho \geq 0, \\ \Omega^{\frac{1}{2}} \left(W(\tau_0) - W(\tau_0+\rho)\right) & \rho \leq 0, \end{cases}$$

from which it follows that

$$\|\delta_T\| Z'_\Delta M_\iota u \Longrightarrow \begin{cases} \Omega^{\frac{1}{2}} W_1(\rho) & \rho \geq 0, \\ \Omega^{\frac{1}{2}} W_2(-\rho) & \rho < 0, \end{cases}$$

where W_1, W_2 are independent p-vectors of independent standard Brownian motion processes on $[0, \infty)$. Therefore $\sup_{k \in N(K)} Z'_\Delta M_\iota u = O_p\left(\|\delta_T\|^{-1}\right)$. Finally, part (h) follows from (2.37) after noting that Condition 2.1 implies that $\sup_{k \in N^c(K)} |k-k_0|^{-1} \|Z_\Delta\|^2 = O_p(1)$. ■

Lemma 2.3 *As* $T \to \infty$,

(a) $u'\left(M_{\iota,X,Z_k} - M_{\iota,X,Z_0}\right) u = O_p(1)$ *uniformly on* $k \in \Lambda \cdot T$,

(b) if $T \|\delta_T\|^2 \to \infty$ *then* $u'\left(M_{\iota,X,Z_k} - M_{\iota,X,Z_0}\right) u = o_p(1)$ *uniformly on* $N(K)$.

Proof. Denote $W_\Delta = (W_k - W_0) \operatorname{sgn}(k-k_0) = (0, Z_\Delta)$. For any nonsingular matrices A and $A+B$,

$$P_{A+B} - P_A = B\left(\bar{A}'A\right)^{-1} \bar{A}' + A\left(\bar{A}'A\right)^{-1} \bar{B}' + B\left(\bar{A}'A\right)^{-1} \bar{B}'$$

$$- (A+B)\left(\left(\bar{A}+\bar{B}\right)'(A+B)\right)^{-1} \left(\bar{A}'B + \bar{B}'A + \bar{B}'B\right)\left(\bar{A}'A\right)^{-1} \left(\bar{A}+\bar{B}\right)'. \tag{2.38}$$

Let $A = M_\iota W_0$ and $B = M_\iota W_\Delta \operatorname{sgn}(k-k_0)$. Then

$$u'\left(M_{\iota,X,Z_0} - M_{\iota,X,Z_k}\right) u = (M_\iota u)' \left(P_{M_\iota W_k} - P_{M_\iota W_0}\right) M_\iota u$$

$$
\begin{aligned}
&= 2u'M_\iota W_\Delta \left(W_0'M_\iota W_0\right)^{-1} W_0'M_\iota u \operatorname{sgn}(k-k_0) \\
&\quad + u'M_\iota W_\Delta \left(W_0'M_\iota W_0\right)^{-1} W_\Delta' M_\iota u \\
&\quad - u'M_\iota W_k \left(W_k'M_\iota W_k\right)^{-1} \{(W_0'M_\iota W_\Delta + W_\Delta' M_\iota W_0) \operatorname{sgn}(k-k_0) \\
&\quad + W_\Delta' M_\iota W_\Delta\} \left(W_0'M_\iota W_0\right)^{-1} W_k' M_\iota u.
\end{aligned}
$$

The bounds in part (a) and (b) are implied by Lemma 2.2. ■

Lemma 2.4 *There exists $\rho > 0$ such that for every $\varepsilon > 0$, there exists $K < \infty$ such that*

$$P\left(\inf_{N^c(K)} \frac{Q_T(k)}{|k-k_0|} \geq \rho \left\|\delta_T\right\|^2\right) \geq 1 - \varepsilon.$$

Proof. By the definition of $Q_T(k)$, the left-hand side of the last displayed inequality is bounded from below by

$$P\left(\inf_{N^c(K)} \lambda_{\min}\left(\frac{Z_\Delta' M_{\iota,X,Z_k} Z_\Delta}{|k-k_0|}\right) \geq \rho\right). \tag{2.39}$$

Consider first the case $k > k_0$. Since the columns of the matrix (ι, X, Z_k) lie in the column space of matrix $(\iota_0, \iota_\Delta, \iota - \iota_k, Z_k, X - Z_k)$, we have

$$
\begin{aligned}
Z_\Delta' M_{\iota,X,Z_k} Z_\Delta &\geq Z_\Delta' M_{\iota_0,\iota_\Delta,\iota-\iota_k,Z_k,X-Z_k} Z_\Delta \geq Z_\Delta' M_{\iota_0,\iota_\Delta,Z_k} Z_\Delta \\
&= Z_\Delta' M_{\iota_\Delta} Z_\Delta \left(Z_k' M_{\iota_0,\iota_\Delta} Z_k\right)^{-1} Z_0' M_{\iota_0} Z_0,
\end{aligned}
$$

where the last inequality is due to Lemma 2.1 and where the equality follows from a simple algebra. Since for any symmetric matrices A and B, inequality $A \geq B$ implies $\lambda_{\min}(A) \geq \lambda_{\min}(B)$ (see for example Magnus and Neudecker (1988), page 204), we have

$$\lambda_{\min}\left(\frac{Z_\Delta' M_{\iota,X,Z_k} Z_\Delta}{|k-k_0|}\right) \geq \lambda_{\min}\left\{\frac{Z_\Delta' M_{\iota_\Delta} Z_\Delta}{|k-k_0|}\left(\frac{1}{T} Z_k' M_{\iota_0,\iota_\Delta} Z_k\right)^{-1} \frac{1}{T} Z_0' M_{\iota_0} Z_0\right\}. \tag{2.40}$$

Similar inequality is obtained in the case $k < k_0$,

$$\begin{aligned}
&\lambda_{\min}\left(\frac{Z'_\Delta M_{\iota,X,Z_k} Z_\Delta}{|k-k_0|}\right) \\
\geq\; &\lambda_{\min}\left\{\frac{Z'_\Delta M_{\iota_\Delta} Z_\Delta}{|k-k_0|}\left(\frac{1}{T}(X-Z_k)' M_{\iota_\Delta,\iota-\iota_0}(X-Z_k)\right)^{-1}\right. \\
&\left.\times \frac{1}{T}(X-Z_0)' M_{\iota-\iota_0}(X-Z_0)\right\}. \qquad (2.41)
\end{aligned}$$

By Conditions 1.1 and 1.2, as $l \to \infty$ and $l \to -\infty$, $l^{-1}\sum_t^{k_0,k_0+l} x_t \xrightarrow{p} 0$ and $l^{-1}\sum_t^{k_0,k_0+l} x_t x_t' \xrightarrow{p} \Sigma$ with $\Sigma > 0$, therefore the eigenvalues of the matrix

$$\frac{Z'_\Delta M_{\iota_\Delta} Z_\Delta}{|k-k_0|} = \frac{1}{|k-k_0|}\sum_t^{k_0,k} x_t x_t' - \left(\frac{1}{|k-k_0|}\sum_t^{k_0,k} x_t\right)\left(\frac{1}{|k-k_0|}\sum_t^{k_0,k} x_t'\right)$$

are bounded from below by $\rho > 0$ with probability at least $1-\varepsilon/3$ for large K. Similarly, since $T^{-1}\sum_{t=1}^T x_t \xrightarrow{p} 0$ and $T^{-1}\sum_{t=1}^T x_t x_t' \xrightarrow{p} \Sigma$ the eigenvalues of matrix

$$\frac{1}{T} Z_k' M_{\iota_k} Z_k = \frac{1}{T}\sum_{t=1}^k x_t x_t' - \left(\frac{1}{T}\sum_{t=1}^k x_t\right)\left(\frac{1}{k}\sum_{t=1}^k x_t'\right)$$

are bounded and bounded from below by a positive number with a large probability for large T uniformly on $k \in \Lambda \cdot T$. Since $M_{\iota_0,\iota_\Delta} \leq M_{\iota_k}$, the same is true of the eigenvalues of matrix $T^{-1}Z_k' M_{\iota_0,\iota_\Delta} Z_k$, and similarly for the remaining factors of (2.40) and (2.41). Since for a symmetric matrix A, $\lambda_{\min}(A) = \inf_x \|Ax\| / \|x\|$, it is easy to see that for any positive semidefinite matrices A and B, $\lambda_{\min}(AB) \geq \lambda_{\min}(A) \cdot \lambda_{\min}(B)$. It follows that there exists $\rho > 0$ such that (2.39) is greater than $1-\varepsilon$, and the lemma is established. ∎

The following lemma extends the Hájek-Rényi inequality to the cross-product of two mean-adjusted series possibly exhibiting long-memory dependence.

Lemma 2.5 *Let $\overline{u} = T^{-1}\sum_{t=1}^T u_t$. Then for any $\alpha > 0$ and for any integers*

m and T such that $m < T$,

$$P\left(\max_{m\le k\le T}\frac{1}{k}\left\|\sum_{t=1}^{k}x_t\left(u_t-\overline{u}\right)\right\|>\alpha\right)<\frac{D}{\alpha^2 m} \tag{2.42}$$

for some positive $D < \infty$.

Proof. Assume without loss of generality that $\{x_t\}$ is a scalar process. Let $S_k = \sum_{t=1}^{k} x_t\left(u_t - \overline{u}\right)$ and let an event A_k be defined as

$$A_k = \left\{\frac{1}{k}\left|S_k\right| > \alpha, \frac{1}{j}\left|S_j\right| \le \alpha \text{ for } m \le j < k\right\}.$$

Proceeding as in the proof of Theorem 1 of Kounias and Weng (1969) and in the proof of a version of the maximal inequality of Kuan and Hsu (1998), we obtain

$$\begin{aligned}
&P\left(\max_{m\le k\le T}\frac{1}{k}\left|\sum_{t=1}^{k}x_t\left(u_t-\overline{u}\right)\right|>\alpha\right)\\
\le\ &\frac{1}{\alpha^2}\left(\frac{1}{m^2}ES_m^2+\sum_{k=m+1}^{T}\frac{1}{k^2}E\left(S_k^2-S_{k-1}^2\right)\left(1-\mathbb{I}\left(A_{k-1}\right)-\ldots-\mathbb{I}\left(A_m\right)\right)\right)\\
\le\ &\frac{1}{\alpha^2}\left(\frac{1}{m^2}ES_m^2+\sum_{k=m+1}^{T}\frac{1}{k^2}\left(Ex_k^2\left(u_k-\overline{u}\right)^2+2\left|Ex_k\left(u_k-\overline{u}\right)S_{k-1}\right|\right)\right).
\end{aligned} \tag{2.43}$$

We have

$$ES_k^2 = \sum_{t,s=1}^{k} Ex_t x_s u_t u_s - \frac{2}{T}\sum_{t,s=1}^{k}\sum_{r=1}^{T} Ex_t x_s u_t u_r + \frac{1}{T^2}\sum_{t,s=1}^{k}\sum_{r,v=1}^{T} Ex_t x_s u_r u_v.$$

Since

$$\left|Ex_t x_s u_r u_v\right| \le C\varphi_{|t-s|}\psi_{|r-v|}$$

where $\varphi_k = \sum_{j=0}^{\infty} |a_j| \, |a_{j+k}|$ and $\psi_k = \sum_{j=0}^{\infty} |b_j| \, |b_{j+k}|$, we obtain that the first term of ES_k^2 is bounded in absolute value by

$$C \sum_{t=1}^{k} \sum_{s=1}^{k} \varphi_{|t-s|} \psi_{|t-s|} \leq C \sum_{t=1}^{k} \max_{|r|<\infty} \sum_{s=1}^{\infty} \varphi_s \psi_{s-r} \leq Ck$$

by Lemma 2 of Robinson (1998). Similarly, the second term of ES_k^2 is bounded by $CT^{-1} \sum_{t,s=1}^{k} \sum_{r=1}^{T} \varphi_{|t-s|} \psi_{|t-r|}$ and the third term by $CT^{-2} \sum_{t,s=1}^{k} \sum_{r,v=1}^{T} \varphi_{|t-s|} \psi_{|r-v|}$, both of which, proceeding as with the last displayed inequality, can be seen to be bounded by Ck by Lemma 2 of Robinson (1998). Therefore

$$\frac{ES_k^2}{k^2} \leq \frac{C}{k}$$

for all $1 \leq k \leq T$. Further,

$$Ex_k^2 \left(u_k - \overline{u}\right)^2 \leq Ex_k^2 u_k^2 + \frac{C}{T} \sum_{s=1}^{T} \psi_{|k-s|} + \frac{C}{T^2} \sum_{t,s=1}^{T} \psi_{|t-s|} \leq C$$

by the second order stationarity and Lemma 1 of Robinson (1998). Next,

$$\begin{aligned} \left|Ex_k \left(u_k - \overline{u}\right) S_{k-1}\right| \leq & \sum_{t=1}^{k-1} \varphi_{k-t} \psi_{k-t} + \frac{C}{T} \sum_{t=1}^{k-1} \sum_{s=1}^{T} \varphi_{k-t} \left(\psi_{|k-s|} + \psi_{|t-s|}\right) \\ & + \frac{C}{T^2} \sum_{t=1}^{k-1} \sum_{s,r=1}^{T} \varphi_{k-t} \psi_{|s-r|} \leq C \end{aligned}$$

uniformly in k by Lemma 2 of Robinson (1998). Therefore the second term in the bracket on the right of (2.43) is bounded by

$$C \sum_{k=m+1}^{T} \frac{1}{k^2} \leq \frac{C}{m}$$

and thus we conclude that (2.42) holds true. ■

Lemma 2.6 *If $T\|\delta_T\|^2 \to \infty$, then for any $K < \infty$,*

$$S_T(k) - S_T(k_0) = \delta_T' Z_\Delta' M_\iota Z_\Delta \delta_T - 2\delta_T' Z_\Delta' M_\iota u \operatorname{sgn}(k - k_0) + o_p(1),$$

where the $o_p(1)$ term is uniform on $N(K)$.

Proof. Write

$$Q_T(k) = \delta_T' Z_\Delta' M_\iota Z_\Delta \delta_T - \delta_T' Z_\Delta' M_\iota W_k (W_k' M_\iota W_k)^{-1} W_k M_\iota Z_\Delta \delta_T.$$

Since by Lemma 2.2, the terms and $Z_\Delta' M_\iota W_k$ and $(W_k' M_\iota W_k)^{-1}$ are $O_p\left(\|\delta_T\|^{-2}\right)$ and $O_p(T^{-1})$ uniformly on $N(K)$, respectively, the second term of $Q_T(k)$ is $O_p\left(\|\delta_T\|^2 \|\delta_T\|^{-4} T^{-1}\right) = o_p(1)$. Further, since $W_k' M_\iota u$ is $O_p\left(T^{1/2}\right)$ uniformly on $\Lambda \cdot K$ by Lemma 2.6 and $u'\left(M_{\iota,X,Z_k} - M_{\iota,X,Z_0}\right)u$ is uniformly on $N(K)$ by Lemma 2.3, the decomposition of $R_T(k)$ in (2.14) implies that

$$\begin{aligned} R_T(k) &= -2\delta_T' Z_\Delta' M_\iota u \operatorname{sgn}(k - k_0) + O_p\left(\|\delta_T\| \|\delta_T\|^{-2} T^{-1} T^{1/2}\right) + o_p(1) \\ &= -2\delta_T' Z_\Delta' M_\iota u \operatorname{sgn}(k - k_0) + o_p(1). \end{aligned}$$

The lemma now follows from (2.12). ■

Lemma 2.7 *For $\delta_T \neq 0$, let $v_T^2 = (\delta_T' \Sigma \delta_T)^2 / \delta_T' \Omega \delta_T$. If the conditions of Proposition 2.4 are satisfied then for any $K > 0$,*

$$\arg\min_{|\rho| \le K} S_T\left(k_0 + \left[\rho v_T^{-2}\right]\right) \xrightarrow{d} \arg\min_{|\rho| \le K} \bar{W}(\rho).$$

Proof. For any given K and $\rho \in [-K, K]$, write $k = k_0 + \left[\rho v_T^{-2}\right]$. By Lemma 2.6,

$$S_T\left(k_0 + \left[\rho v_T^{-2}\right]\right) - S_T(k_0) = \delta_T' Z_\Delta' M_\iota Z_\Delta \delta_T - 2\delta_T' Z_\Delta' M_\iota u \operatorname{sgn}\left[\rho v_T^{-2}\right] + o_p(1), \tag{2.44}$$

where the $o_p(1)$ term is uniform on $N(K)$. Lemma 2.2 implies that

$$\frac{1}{T} Z_{\Delta([\sigma T],[(\sigma+\rho)T])}' M_\iota Z_{\Delta([\sigma T],[(\sigma+\rho)T])} \xrightarrow{p} |\rho| \Sigma$$

uniformly on $\{(\sigma, \rho) : 0 \leq \sigma, \sigma + \rho \leq 1\}$, from which it follows that

$$v_T^2 Z_\Delta' M_\iota Z_\Delta \xrightarrow{p} |\rho| \, \Sigma$$

uniformly on $\rho \in [-K, K]$. Proceeding as in the proof of part (g) of Lemma 2.2, it can be seen that

$$v_T Z_\Delta' M_\iota u \overset{p}{\Longrightarrow} \begin{cases} \Omega^{\frac{1}{2}} W_1(\rho) & \rho \geq 0, \\ \Omega^{\frac{1}{2}} W_2(-\rho) & \rho < 0, \end{cases}$$

where W_1, W_2 are independent p-vectors of independent standard Brownian motion processes. Since $-2\delta_T' v_T^{-1} \Omega^{\frac{1}{2}} W_1(\rho)$ and $2\delta_T' v_T^{-1} \Omega^{\frac{1}{2}} W_2(\rho)$ are equal in distribution to $2(\delta_T' \Omega \delta_T)^{\frac{1}{2}} v_T^{-1} W_1(\rho)$ and $2(\delta_T' \Omega \delta_T)^{\frac{1}{2}} v_T^{-1} W_2(\rho)$, respectively, the left-hand side of (2.44) is equal in distribution to

$$\begin{aligned} & |\rho| \frac{\delta_T' \Sigma \delta_T}{v_T^2} + 2\frac{(\delta_T' \Omega \delta_T)^{\frac{1}{2}}}{v_T} \left(W_1(\rho) \mathbb{I}(\rho \geq 0) + W_2(-\rho) \mathbb{I}(\rho < 0)\right) + o_p(1) \\ = \; & 2\frac{\delta_T' \Omega \delta_T}{\delta_T' \Sigma \delta_T} \bar{W}(\rho) + o_p(1) \end{aligned}$$

uniformly on $\rho \in [-K, K]$. Now observing that

$$\arg\min_{|\rho| \leq K} 2\frac{\delta_T' \Omega \delta_T}{\delta_T' \Sigma \delta_T} \bar{W}(\rho) = \arg\min_{|\rho| \leq K} \bar{W}(\rho),$$

the proof of the lemma is completed by an application of the continuous mapping theorem. ■

For the proofs of the statements about bootstrap quantities, define

$$Q_T^*(k) = \hat{\delta}' Z_{\hat{k}}' \overline{F}' M_{FW_k} F Z_{\hat{k}} \hat{\delta}$$

and

$$R_T^*(k) = 2\hat{\delta}' Z_{\hat{k}}' \overline{F}' M_{FW_k} H F \hat{u} + \hat{u}' \overline{F}' \overline{H}' \left(M_{FW_k} - M_{FW_{\hat{k}}}\right) H F \hat{u},$$

so that

$$Q_T^*(k) + R_T^*(k) = \|\hat{u}_k^*\|^2 - \left\|\hat{u}_{\hat{k}}^*\right\|^2 = S_T^*(k) - S_T^*\left(\hat{k}\right). \tag{2.45}$$

Let

$$\hat{N}(K) = \left\{k\colon \left|k - \hat{k}\right| \le K\|\delta\|^{-2}\right\}, \tag{2.46}$$

$$\hat{N}^C(K) = \Lambda \cdot T - \hat{N}(K) \tag{2.47}$$

and denote $\hat{Z}_\Delta = Z_{\Delta(\hat{k},k)}$.

Lemma 2.8 *As $T \to \infty$,*

(a) $\frac{2\pi}{\sqrt{T}} Z'_{[\tau T]} \overline{F}' H F \hat{u} \overset{p}{\Longrightarrow} \Omega^{\frac{1}{2}} W(\tau)$,

(b) for any $K > 0$, $\sup_{k \in N(K)} \sup_{1 \le l \le T} \left\|Z'_\Delta \overline{F}' H F Z_l\right\| = O_{p^*}\left(\|\delta_T\|^{-2}\right)$,

(c) for any $K > 0$, $\sup_{N(K)} \left\|Z'_\Delta \overline{F}' H F \hat{u}\right\| = O_{p^*}\left(\|\delta_T\|^{-1}\right)$,

(d) for any $K > 0$, $\sup_{k \in \hat{N}(K)} \sup_{1 \le l \le T} \left\|\hat{Z}'_\Delta \overline{F}' F Z_l\right\| = O_p\left(\|\delta_T\|^{-1}\right)$,

(e) $W'_k \overline{F}' H F \hat{u} = O_{p^*}\left(T^{1/2}\right)$ *uniformly over $1 \le k \le T$,*

(f) $\left(\frac{1}{T} W'_{\hat{k}} \overline{F}' F W_{\hat{k}}\right)^{-1} \overset{p}{\to} 2\pi \begin{pmatrix} 1 & \tau_0 \\ \tau_0 & \tau_0 \end{pmatrix}^{-1} \otimes \Sigma^{-1}$,

(g) for every $\varepsilon > 0$ there exist $K, M > 0$ such that

$$P\left(\sup_{k \in \hat{N}^C(K)} \sup_{1 \le l \le T} \frac{1}{\left|k - \hat{k}\right|} \left\|\hat{Z}'_\Delta \overline{F}' F Z_l\right\| > M\right) < \varepsilon,$$

(h) for every $\varepsilon, \eta > 0$ there exist $K, M > 0$ such that

$$P\left(P^*\left(\sup_{k \in \hat{N}^C(K)} \sup_{1 \le l \le T} \frac{1}{\left|k - \hat{k}\right|} \left\|\hat{Z}'_\Delta \overline{F}' H F Z_l\right\| > M\right) > \eta\right) < \varepsilon.$$

Proof. Part (a) follows from Lemma 1.12 and from the remark at the end of its proof because

$$Z'_{[\tau T]} \overline{F}' H F \hat{u} = \sum_{j=1}^{T-1} \bar{w}_{z(\tau),j} w_{\hat{u},j} \eta_j^*.$$

To show the validity of part (b), define matrix G as $G = \overline{F}' H F$. By the definition of matrices F and H, matrix G is a real circulant matrix with elements $g_{t,s} = g_{s-t}$, $1 \le t, s \le T$, where $g_t = (2\pi T)^{-1} \sum_{j=1}^{T-1} \eta_j^* e^{i\lambda_j t}$. Let

$g^t = (g_{-t+1}, \ldots, g_{-1}, g_0, g_1, \ldots, g_{T-t-1})$ be the t-th row of matrix G. By the Cauchy-Schwarz inequality,

$$\left\| Z'_\Delta \overline{F}' HFZ_l \right\| = \left\| \sum_t^{k_0,k} x_t g^t Z_l \right\| \le \left(\sum_t^{k_0,k} \|x_t\|^2 \right)^{\frac{1}{2}} \left(\sum_t^{k_0,k} \|g^t Z_l\|^2 \right)^{\frac{1}{2}}.$$

Therefore

$$\sup_{k \in N(K)} \sup_{1 \le l \le T} \left\| Z'_\Delta \overline{F}' HFZ_l \right\|$$
$$\le \left(\sum_{t=k_0-K\|\delta_T\|^{-2}}^{k_0+K\|\delta_T\|^{-2}} \|x_t\|^2 \right)^{\frac{1}{2}} \left(\sum_{t=k_0-K\|\delta_T\|^{-2}}^{k_0+K\|\delta_T\|^{-2}} \sup_{1 \le l \le T} \|g^t Z_l\|^2 \right)^{\frac{1}{2}}. \qquad (2.48)$$

The expression in the first bracket on the right of (2.48) is $O_p\left(\|\delta_T\|^{-2}\right)$ by Conditions 1.1 and 1.2. Further,

$$\|g^t Z_l\|^2 = g^t Z_l Z'_l \left(g^t\right)' \le g^t XX' \left(g^t\right)' = \operatorname{tr} \left(g^t\right)' g^t XX'$$

and so $E^* \sup_{1 \le l \le T} \|g^t Z_l\|^2 \le \operatorname{tr} E^* \left(g^t\right)' g^t XX'$. For any t and s,

$$E^* g_t g_s = \frac{1}{4\pi^2} \left(-\frac{1}{T^2} + \frac{1}{T} I\left(t = s\right) \right),$$

therefore $E^* \left(g^t\right)' g^t = \left(4\pi^2 T\right)^{-1} M_\iota$ and

$$E^* \sup_{1 \le l \le T} \|g^t Z_l\|^2 \le \operatorname{tr} \frac{1}{4\pi^2 T} M_\iota XX' = \frac{1}{4\pi^2} \operatorname{tr} \frac{1}{T} X' M_\iota X = O_p(1) \qquad (2.49)$$

by Lemma 2.2. By the Markov inequality, the expression in the second bracket of (2.48) is $O_{p^*}\left(\|\delta_T\|^{-2}\right)$ and part (b) is established. Part (c) follows from part (a) in the same way as part (g) of Lemma 2.2 follows from part (b) of Lemma 2.2. To prove part (d), write

$$2\pi \left\| \hat{Z}'_\Delta \overline{F}' FZ_l \right\| \le \left\| \hat{Z}_\Delta \right\|^2 + \left| k - \hat{k} \right| \left\| \hat{Z}_\Delta \right\| \frac{l^{\frac{1}{2}}}{T} \|Z_l\|.$$

For any $M > 0$,

$$P\left(\sup_{k\in\hat{N}(K)}\left\|\hat{Z}_\Delta\right\|^2 > M\left\|\delta_T\right\|^{-2}\right) \leq P\left(\sum_{t=\hat{k}-K\|\delta_T\|^{-2}}^{\hat{k}+K\|\delta_T\|^{-2}}\left\|x_t\right\|^2 > M\left\|\delta_T\right\|^{-2}\right)$$

which is bounded by CK/M by the Markov inequality and Conditions 1.1 and 1.2 and which is bounded by ε for large K. From here we can conclude in the same way as in the proof of part (e) of Lemma 2.2. Part (e) follows from part (a) and part (f) follows from (2.15). In part (g), we have

$$\sup_{\substack{k\in\hat{N}^C(K)\\ 1\leq l\leq T}} 2\pi\left\|\frac{\hat{Z}_\Delta'\overline{F}'FZ_l}{k-\hat{k}}\right\| \leq \sup_{k\in\hat{N}^C(K)}\frac{\left\|\hat{Z}_\Delta\right\|^2}{\left|k-\hat{k}\right|} + \sup_{k\in\hat{N}^C(K)}\frac{\left\|\hat{Z}_\Delta\right\|}{\left|k-\hat{k}\right|^{\frac{1}{2}}}\sup_{1\leq l\leq T}\frac{l^{\frac{1}{2}}}{T}\left\|Z_l\right\|.$$

Now

$$\begin{aligned}\sup_{k\in\hat{N}^C(K)}\frac{1}{\left|k-\hat{k}\right|}\left\|\hat{Z}_\Delta\right\|^2 \;&\leq\; 2\sup_{k\in\hat{N}^C(K)}\frac{|k-k_0|}{\left|k-\hat{k}\right|}\sup_{k\in\hat{N}^C(K)}\frac{1}{|k-k_0|}\left\|Z_\Delta\right\|^2 \\ &\quad +2\sup_{k\in\hat{N}^C(K)}\frac{1}{\left|k-\hat{k}\right|}\left\|Z_{\hat{k}}-Z_0\right\|^2.\end{aligned}$$

The factor $\sup_{k\in\hat{N}^C(K)}\left|k-\hat{k}\right|^{-1}|k-k_0|$ is bounded by $\max\left\{1, K^{-1}\left\|\delta_T\right\|^2\right\} \leq C$. For any $M > 0$, the probability that the factor $\sup_{k\in\hat{N}^C(K)}|k-k_0|^{-1}\left\|Z_\Delta\right\|^2$ is greater than M is bounded by

$$P\left(\sup_{|k-k_0|\geq K/2}\frac{1}{|k-k_0|}\left\|Z_\Delta\right\|^2 > M\right) + P\left(\left|\hat{k}-k_0\right| > \frac{K}{2}\left\|\delta_T\right\|^{-2}\right). \quad (2.50)$$

The second term of (2.50) is bounded by $\varepsilon/2$ for large K by Proposition 2.1 and the first term of (2.50) is bounded by $\varepsilon/2$ for large M by Condition 2.1. Further, for any $N > 0$,

$$P\left(\sup_{k\in\hat{N}^C(K)}\frac{1}{\left|k-\hat{k}\right|}\left\|Z_{\hat{k}}-Z_0\right\|^2 > M\right) \leq P\left(\left\|Z_{\hat{k}}-Z_0\right\|^2 > MK\left\|\delta_T\right\|^{-2}\right)$$

$$\leq P\left(\sum_{t=k_0-N\|\delta_T\|^{-2}}^{k_0+N\|\delta_T\|^{-2}} \|x_t\|^2 > MK\|\delta_T\|^{-2}\right) + P\left(\left|\hat{k}-k_0\right| > N\|\delta_T\|^{-2}\right). \tag{2.51}$$

By the Markov inequality the first term on the right of (2.51) is bounded by CN/MK. Proposition 2.1 then implies that both terms on the right of (2.51) can be bounded by $\varepsilon/2$ for large M and N for any $K > 0$. Gathering the results and recalling that the factor $\sup_{1\leq l\leq T} l^{1/2}T^{-1}\|Z_l\|$ is $O_p(1)$ by Condition 2.1, we can conclude that part (g) holds.

Finally, to prove part (h), we follow the steps in part (b) and write

$$\begin{aligned}
&\sup_{\substack{k\in\hat{N}^C(K)\\ 1\leq l\leq T}} \frac{1}{\left|k-\hat{k}\right|}\left\|\hat{Z}_\Delta'\overline{F}'HFZ_l\right\| \\
\leq &\left(\sup_{\hat{N}^C(K)} \frac{1}{\left|k-\hat{k}\right|}\sum_t^{\hat{k},k}\|x_t\|^2\right)^{\frac{1}{2}} \\
&\times\left(\sup_{\hat{N}^C(K)} \frac{1}{\left|k-\hat{k}\right|}\sum_t^{\hat{k},k}\sup_{1\leq l\leq T}\left\|g^tZ_l\right\|^2\right)^{\frac{1}{2}}.
\end{aligned} \tag{2.52}$$

Proceeding as in the proof of part (g), it can be seen that the expression

$$EP^*\left(\sup_{\hat{N}^C(K)} \frac{1}{\left|k-\hat{k}\right|}\sum_t^{\hat{k},k}\|x_t\|^2 > M\right) = P\left(\sup_{\hat{N}^C(K)} \frac{1}{\left|k-\hat{k}\right|}\sum_t^{\hat{k},k}\|x_t\|^2 > M\right)$$

is smaller than $\varepsilon/2$ for large M and K. The expression in the second bracket on the right of (2.52) is equal to $\sup_{1\leq l\leq T}\left\|g^tZ_l\right\|^2$. Part (g) is then implied by (2.49) and by the Markov inequality. ■

Lemma 2.9 *(a)* $\sup_{k\in\Lambda\cdot T}\hat{u}'\overline{F}'\overline{H}'\left(M_{FW_k}-M_{FW_{\hat{k}}}\right)HF\hat{u} = O_{p^*}(1)$,
(b) if $T\|\delta_T\|^2\to\infty$, $\sup_{\hat{N}(K)}\hat{u}'\overline{F}'\overline{H}'\left(M_{FW_k}-M_{FW_{\hat{k}}}\right)HF\hat{u} = o_{p^*}(1)$.

Proof. Denote $\hat{W}_\Delta = (W_k - W_{\hat{k}})\operatorname{sgn}\left(k-\hat{k}\right) = \left(0,\hat{Z}_\Delta\right)$. Proceeding as in Lemma 2.3 and taking $A = FW_{\hat{k}}$ and $B = F\hat{W}_\Delta\operatorname{sgn}\left(k-\hat{k}\right)$ in (2.38), we obtain that

$$\hat{u}'\overline{F}'\overline{H}'\left(M_{FW_{\hat{k}}}-M_{FW_k}\right)HF\hat{u}=\hat{u}'\overline{F}'\overline{H}'\left(P_{FW_k}-P_{FW_{\hat{k}}}\right)HF\hat{u}$$

$$\begin{aligned}
= \; & 2\operatorname{Re}\hat{u}'\overline{F}'\overline{H}'F\hat{W}_\Delta\left(W_{\hat{k}}'\overline{F}'FW_{\hat{k}}\right)^{-1}W_{\hat{k}}'\overline{F}'HF\hat{u}\operatorname{sgn}\left(k-\hat{k}\right)\\
& +\hat{u}'\overline{F}'\overline{H}'F\hat{W}_\Delta\left(W_{\hat{k}}'\overline{F}'FW_{\hat{k}}\right)^{-1}\hat{W}_\Delta'\overline{F}'HF\hat{u}\\
& -\hat{u}'\overline{F}'\overline{H}'FW_k\left(W_k'\overline{F}'FW_k\right)^{-1}\left\{\left(W_{\hat{k}}'\overline{F}'F\hat{W}_\Delta+\hat{W}_\Delta'\overline{F}'FW_{\hat{k}}\right)\operatorname{sgn}\left(k-\hat{k}\right)\right.\\
& \left.+\hat{W}_\Delta'\overline{F}'F\hat{W}_\Delta\right\}\left(W_{\hat{k}}'\overline{F}'FW_{\hat{k}}\right)^{-1}W_{\hat{k}}'\overline{F}'HF\hat{u}.
\end{aligned}$$

Therefore the lemma holds by Lemmas 2.2 and 2.8. ∎

Lemma 2.10 *There exists $\lambda>0$ such that for every $\varepsilon>0$, there exists $K<\infty$ and $T_0<\infty$ such that for all $T\geq T_0$,*

$$EP^*\left(\inf_{\hat{N}^C(K)}\frac{Q_T^*(k)}{\left|k-\hat{k}\right|}\geq\lambda\left\|\hat{\delta}\right\|^2\right)\geq 1-\varepsilon.$$

Proof. If an event A does not depend on η_j^* for $j=1,\ldots,T-1$, then $P^*(A)=\mathbb{I}(A)$ and therefore $EP^*(A)=P(A)$. Since $Q_T^*(k)$, $\hat{k}$ and $\hat{\delta}$ do not involve η_j^*, the left-hand side of the hypothesized inequality is bounded from below by

$$P\left(\inf_{\hat{N}^C(K)}\lambda_{\min}\left(\frac{\hat{Z}_\Delta'M_{\iota,X,Z_k}\hat{Z}_\Delta}{\left|k-\hat{k}\right|}\right)\geq 2\pi\lambda\right).$$

Denote $\hat{\iota}_\Delta=(\iota_k-\iota_{\hat{k}})\operatorname{sgn}\left(k-\hat{k}\right)$. Proceeding as in the proof of Lemma 2.4, we obtain inequality

$$\lambda_{\min}\left(\frac{\hat{Z}_\Delta'M_{\iota,X,Z_k}\hat{Z}_\Delta}{\left|k-\hat{k}\right|}\right)\geq\lambda_{\min}\left\{\frac{\hat{Z}_\Delta'M_{\hat{\iota}_\Delta}\hat{Z}_\Delta}{\left|k-\hat{k}\right|}\left(\frac{1}{T}Z_k'M_{\hat{\iota}_\Delta,\iota_{\hat{k}}}Z_k\right)^{-1}\frac{1}{T}Z_{\hat{k}}'M_{\iota_{\hat{k}}}Z_{\hat{k}}\right\} \tag{2.53}$$

for $k > \hat{k}$ and

$$\begin{aligned}
&\lambda_{\min}\left(\frac{\hat{Z}_{\Delta}' M_{\iota,X,Z_k} \hat{Z}_{\Delta}}{\left|k-\hat{k}\right|}\right) \\
\geq\; &\lambda_{\min}\left\{\frac{\hat{Z}_{\Delta}' M_{\hat{\iota}_{\Delta}} \hat{Z}_{\Delta}}{\left|k-\hat{k}\right|}\left(\frac{1}{T}(X-Z_k)' M_{\hat{\iota}_{\Delta},\iota-\iota_{\hat{k}}}(X-Z_k)\right)^{-1}\right. \\
&\left.\times \frac{1}{T}(X-Z_{\hat{k}})' M_{\iota-\iota_{\hat{k}}}(X-Z_{\hat{k}})\right\}. \qquad (2.54)
\end{aligned}$$

for $k < \hat{k}$. Write the first factor in the curly bracket in (2.53) as

$$\frac{\hat{Z}_{\Delta}' M_{\hat{\iota}_{\Delta}} \hat{Z}_{\Delta}}{\left|k-\hat{k}\right|} = \frac{\hat{Z}_{\Delta}' \hat{Z}_{\Delta}}{\left|k-\hat{k}\right|} - \left(\frac{1}{\left|k-\hat{k}\right|}\sum_{t}^{\hat{k},k} x_t\right)\left(\frac{1}{\left|k-\hat{k}\right|}\sum_{t}^{\hat{k},k} x_t'\right).$$

We have

$$\frac{\hat{Z}_{\Delta}' \hat{Z}_{\Delta}}{\left|k-\hat{k}\right|} = \frac{|k-k_0|}{\left|k-\hat{k}\right|}\frac{Z_{\Delta}' Z_{\Delta}}{|k-k_0|} + \frac{1}{\left|k-\hat{k}\right|}\left\|Z_{\hat{k}} - Z_0\right\|^2.$$

For any $\delta > 0$,

$$\begin{aligned}
P\left(\sup_{\hat{N}^C(K)}\left|\frac{k-k_0}{k-\hat{k}} - 1\right| > \delta\right) &= P\left(\sup_{\hat{N}^C(K)}\left|\frac{\hat{k}-k_0}{k-\hat{k}}\right| > \delta\right) \\
&\leq P\left(\left|\hat{k}-k_0\right| > \delta K \left\|\delta_T\right\|^{-2}\right)
\end{aligned}$$

which is smaller than $\varepsilon/2$ for large K by Proposition 2.1. Also, for any $\delta > 0$ and any $M > 0$,

$$P\left(\sup_{\hat{N}^C(K)}\frac{\left\|Z_{\hat{k}} - Z_0\right\|^2}{\left|k-\hat{k}\right|} > \delta\right) \leq P\left(\sum_{t}^{k_0,\hat{k}}\left\|x_t\right\|^2 > \delta K \left\|\delta_T\right\|^{-2}\right)$$

$$\leq P\left(\sum_{t=k_0-M\|\delta_T\|^{-2}}^{k_0+M\|\delta_T\|^{-2}} \|x_t\|^2 > \delta K \|\delta_T\|^{-2}\right)$$
$$+P\left(\left|\hat{k}-k_0\right| > M\|\delta_T\|^{-2}\right). \tag{2.55}$$

The second term of (2.55) is smaller than $\varepsilon/2$ for large M by Proposition 2.1. By the Markov inequality, the first term of (2.55) is bounded by $CM/\delta K$ which is smaller than $\varepsilon/2$ for large K. Therefore as $K \to \infty$,

$$\frac{\hat{Z}'_\Delta \hat{Z}_\Delta}{\left|k-\hat{k}\right|} = \frac{Z'_\Delta Z_\Delta}{|k-k_0|} + o_p(1).$$

Since by Conditions 1.1 and 1.2, $\lim_{l\to\pm\infty} |l|^{-1} \sum_t^{k_0,k_0+l} x_t x'_t \xrightarrow{p} \Sigma$, the eigenvalues of matrix $\left|k-\hat{k}\right|^{-1} \hat{Z}'_\Delta \hat{Z}_\Delta$ are bounded away from zero with probability at least $1-\varepsilon/2$ for large K. Similarly, it can be shown that as $K \to \infty$,

$$\frac{1}{\left|k-\hat{k}\right|}\sum_t^{\hat{k},k} x_t = \frac{1}{|k-k_0|}\sum_t^{k_0,\hat{k}} x_t + o_p(1) = o_p(1).$$

It follows that the eigenvalues of matrix $\left|k-\hat{k}\right|^{-1} \hat{Z}'_\Delta M_{\hat{\iota}_\Delta} \hat{Z}_\Delta$ are bounded from below by $\lambda > 0$ with probability at least $1-\varepsilon/3$ for large K.

The third factor in the curly bracket in (2.53) can be written as

$$\frac{1}{T} Z'_{\hat{k}} M_{\iota_{\hat{k}}} Z_{\hat{k}} = \frac{1}{T}\sum_{t=1}^{\hat{k}} x_t x'_t - \left(\frac{1}{T}\sum_{t=1}^{\hat{k}} x_t\right)\left(\frac{1}{\hat{k}}\sum_{t=1}^{\hat{k}} x'_t\right).$$

Since $T^{-1}\sum_{t=1}^{k_0} x_t x'_t \xrightarrow{p} \tau_0 \Sigma$, $T^{-1}\sum_{t=1}^{k_0} x_t \xrightarrow{p} 0$ and $\left|\hat{k}-k_0\right| = O_p\left(\|\delta_T\|^{-2}\right)$, the eigenvalues of matrix $T^{-1} Z'_{\hat{k}} M_{\iota_{\hat{k}}} Z_{\hat{k}}$ are bounded from below by a positive number with large probability for large T by the arguments of the proof of Lemma 2.4. Similarly, the eigenvalues of the remaining factors of (2.53) and (2.54) are bounded from below by a positive number. Concluding as in the proof of Lemma 2.4, the current lemma is established. ∎

Lemma 2.11 *Let H be the matrix defined in Step 3 of the bootstrap procedure. Then for any $\alpha > 0$ and any integers m and T such that $m < T$,*

$$EP^* \left(\max_{m \le k \le T} \frac{1}{k} \left\| Z_k' \overline{F}' HFu \right\| > \alpha \right) \le \frac{C}{\alpha^2 m}$$

for some positive $C < \infty$.

Proof. Without loss of generality, assume that $\{x_t\}$ is scalar. Let $S_k^* = Z_k' \overline{F}' HFu = \sum_{t=1}^k x_t d_t$, where $d_t = \sum_{r=1}^T g_{r-t} u_r$ and $g_t = (2\pi T)^{-1} \sum_{j=1}^{T-1} \eta_j^* e^{i\lambda_j t}$. Arguing as in the proof of Lemma 2.5,

$$\begin{aligned} & P^* \left(\max_{m \le k \le T} \frac{1}{k} \left| Z_k' \overline{F}' HFu \right| > \alpha \right) \\ \le \ & \frac{1}{\alpha^2} \left(\frac{1}{m^2} E^* S_m^{*2} + \sum_{k=m+1}^{T} \frac{1}{k^2} \left(E^* x_k^2 d_k^2 + 2 \left| E^* x_k d_k S_{k-1}^* \right| \right) \right). \qquad (2.56) \end{aligned}$$

Because $E^* g_t g_s = (4\pi^2)^{-1} \left(-T^{-2} + T^{-1} \mathbb{I}\,(t = s) \right)$, we have

$$E^* S_k^{*2} = \frac{1}{4\pi^2 T} \sum_{t,s=1}^{k} \sum_{r,v=1}^{T} x_t x_s u_r u_v \mathbb{I}\,(t - s = r - v) - \frac{1}{4\pi^2 T^2} \sum_{t,s=1}^{k} \sum_{r,v=1}^{T} x_t x_s u_r u_v.$$

In a similar way, $E^* x_k^2 d_k^2 = (4\pi^2 T)^{-1} x_k^2 \sum_{r=1}^T u_r^2 - (4\pi^2 T^2)^{-1} x_k^2 \sum_{r,v=1}^T u_r u_v$ and $E^* x_k d_k S_{k-1}^* = (4\pi^2 T)^{-1} x_k \sum_{t=1}^{k-1} x_t \sum_{r,v=1}^T u_r u_v \mathbb{I}\,(k - t = r - v)$ $- (4\pi^2 T^2)^{-1} x_k \sum_{t=1}^{k-1} x_t \sum_{r,v=1}^T u_r u_v$. Proceeding as in the proof of Lemma 2.5, it can be shown that expectation of $E^* S_k^{*2}$ is bounded by Ck and that expectation of $E^* x_k^2 d_k^2$ and $E^* x_k d_k S_{k-1}^*$ is bounded by C for all $1 \le k \le T$. Therefore the expectation of the right-hand side of (2.56) is bounded by

$$\frac{1}{\alpha^2} \left(\frac{C}{m} + \sum_{t=m+1}^{T} \frac{C}{m^2} \right) \le \frac{C}{\alpha^2 m}.$$

■

Lemma 2.12 *If $T \left\| \delta_T \right\|^2 \to \infty$, then*

$$\hat{k}^* - \hat{k} = O_{p^*} \left(\left\| \delta_T \right\|^{-2} \right).$$

Proof. Fix $\varepsilon, \eta > 0$. For any $K < \infty$,

$$\begin{aligned} & P^*\left(\left|\hat{k}^* - \hat{k}\right| > K\left\|\delta_T\right\|^{-2}\right) \\ \leq \; & P^*\left(\inf_{\hat{N}^C(K)} \frac{Q_T^*(k)}{\left|k-\hat{k}\right|} \leq \lambda\left\|\hat{\delta}\right\|^2\right) + P^*\left(\sup_{\hat{N}^C(K)} \frac{\left|R_T^*(k)\right|}{\left|k-\hat{k}\right|} \geq \lambda\left\|\hat{\delta}\right\|^2\right). \end{aligned}$$

By Lemma 2.10, we can choose $\lambda > 0$ and $K < \infty$ such that

$$EP^*\left(\inf_{\hat{N}^C(K)} \frac{Q_T^*(k)}{\left|k-\hat{k}\right|} \leq \lambda\left\|\hat{\delta}\right\|^2\right) \leq \varepsilon \tag{2.57}$$

for large T. Write

$$\begin{aligned} & R_T^*(k) \\ = \; & 2\hat{\delta}'\hat{Z}_\Delta'\overline{F}'M_{FW_k}HF\hat{u} + \hat{u}'\overline{F}'\overline{H}'\left(M_{FW_k} - M_{FW_{\hat{k}}}\right)HF\hat{u} \\ = \; & 2\hat{\delta}'\hat{Z}_\Delta'\overline{F}'HFu - 2\hat{\delta}'\hat{Z}_\Delta'\overline{F}'FW_k\left(W_k'\overline{F}'FW_k\right)^{-1}W_k'\overline{F}'HF\hat{u} \qquad (2.58) \\ & +2\hat{\delta}'\hat{Z}_\Delta'\overline{F}'HF(\hat{u}-u) + \hat{u}'\overline{F}'\overline{H}'\left(M_{FW_k} - M_{FW_{\hat{k}}}\right)HF\hat{u}. \qquad (2.59) \end{aligned}$$

Examining the first term of (2.58), we have

$$\begin{aligned} & P^*\left(\sup_{\hat{N}^C(K)} \frac{\left|2\hat{\delta}'\hat{Z}_\Delta'\overline{F}'HFu\right|}{\left|k-\hat{k}\right|} \geq \lambda\left\|\hat{\delta}\right\|^2\right) \\ \leq \; & P^*\left(\sup_{\hat{N}^C(K)} \left\|\frac{\hat{Z}_\Delta'\overline{F}'HFu}{k-\hat{k}}\right\| \geq \frac{\lambda}{4}\left\|\delta_T\right\|\right) \\ & +P^*\left(\left\|\hat{\delta}\right\| \leq \frac{1}{2}\left\|\delta_T\right\|\right). \end{aligned} \tag{2.60}$$

By Lemma 2.11 and by the second order stationarity, expectation of the first term on the right of (2.60) is bounded by

$$\frac{C}{\lambda^2\left\|\delta_T\right\|^2 K\left\|\delta_T\right\|^{-2}} = \frac{C}{\lambda^2 K} \leq \frac{\varepsilon}{2}$$

for K large enough. Further, since $P^*\left(\left\|\hat{\delta}\right\| \leq \frac{1}{2}\left\|\delta_T\right\|\right) = \mathbb{I}\left(\left\|\hat{\delta}\right\| \leq \frac{1}{2}\left\|\delta_T\right\|\right)$, expectation of the second term on the right of (2.60) is equal to $P\left(\left\|\hat{\delta}\right\| \leq \frac{1}{2}\left\|\delta_T\right\|\right)$ which is smaller than $\varepsilon/2$ for large T because by Proposition 2.2, $\hat{\delta} \xrightarrow{p} \delta$.

Regarding the second term of (2.58), the factor $\left(W_k'\overline{F}'FW_k\right)^{-1}$ is $O_p\left(T^{-1}\right)$ uniformly over $k \in \Lambda \cdot T$ by Lemma 2.2 whereas the factor $W_k'\overline{F}'HF\hat{u}$ is $O_{p^*}\left(T^{1/2}\right)$ uniformly over $1 \leq k \leq T$ by Lemma 2.8. Moreover, for any $M > 0$,

$$\begin{aligned}
& P^*\left(\sup_{\hat{N}^C(K)}\left\|\frac{2\hat{\delta}'}{\sqrt{T}}\frac{\hat{Z}_\Delta'\overline{F}'FW_k}{k-\hat{k}}\right\| \geq \lambda\left\|\hat{\delta}\right\|^2\right) \\
\leq\ & P^*\left(\sup_{\hat{N}^C(K)}\left\|\frac{\hat{Z}_\Delta'\overline{F}'FW_k}{k-\hat{k}}\right\| \geq M\right) + P^*\left(\left\|\hat{\delta}\right\| \leq \frac{2M}{\lambda}\frac{1}{\sqrt{T}}\right) \qquad (2.61)
\end{aligned}$$

Expectation of the first term on the right of (2.61) is bounded by $\varepsilon/2$ for large M by Lemma 2.8 and expectation of the second term on the right of (2.61) is bounded by $\varepsilon/2$ for large T since $\hat{\delta} = \delta_T + O_p\left(T^{-1/2}\right)$ by Proposition 2.2 and since $T^{-\frac{1}{2}}\left\|\delta_T\right\|^{-1} = o(1)$.

Turning to the first term of (2.59), we have from (2.5) and (2.6) and from the definition of $\hat{u}$ that

$$F(\hat{u} - u) = FX\left(\beta - \hat{\beta}\right) + F\left(Z_0 - Z_{\hat{k}}\right)\delta_T + FZ_{\hat{k}}\left(\delta_T - \hat{\delta}\right).$$

Therefore

$$\begin{aligned}
& P^*\left(\sup_{\hat{N}^C(K)}\left|\frac{2\hat{\delta}'\hat{Z}_\Delta'\overline{F}'HF(\hat{u}-u)}{k-\hat{k}}\right| \geq \lambda\left\|\hat{\delta}\right\|^2\right) \\
\leq\ & P^*\left(\sup_{\hat{N}^C(K)}\left\|\frac{\hat{Z}_\Delta'\overline{F}'HFX\left(\beta-\hat{\beta}\right)}{k-\hat{k}}\right\| \geq \frac{\lambda}{6}\left\|\hat{\delta}\right\|\right) \\
& + P^*\left(\sup_{\hat{N}^C(K)}\left\|\frac{\hat{Z}_\Delta'\overline{F}'HF\left(Z_0 - Z_{\hat{k}}\right)\delta_T}{k-\hat{k}}\right\| \geq \frac{\lambda}{6}\left\|\hat{\delta}\right\|\right) \\
& + P^*\left(\sup_{\hat{N}^C(K)}\left\|\frac{\hat{Z}_\Delta'\overline{F}'HFZ_{\hat{k}}\left(\delta_T - \hat{\delta}\right)}{k-\hat{k}}\right\| \geq \frac{\lambda}{6}\left\|\hat{\delta}\right\|\right). \qquad (2.62)
\end{aligned}$$

For any $M > 0$, the first term on the right of (2.62) is bounded by

$$P^*\left(\sup_{\hat{N}^C(K)}\left\|\frac{\hat{Z}_\Delta'\overline{F}'HFX}{k-\hat{k}}\right\| \geq M\right) + P^*\left(\left\|\beta-\hat{\beta}\right\| \geq \frac{\lambda}{6M}\left\|\hat{\delta}\right\|\right) \tag{2.63}$$

By Lemma 2.8, the first term of (2.63) is smaller than $\eta/6$ with probability larger than $1-\varepsilon/6$ for large K and M. Expectation of the second term of (2.63) is smaller than $\varepsilon/6$ for large T because by Proposition 2.2, $\beta-\hat{\beta}$ and $\hat{\delta}-\delta_T$ are $O_p\left(T^{-1/2}\right)$, and because $T^{-1/2} = o\left(\|\delta_T\|\right)$. In a similar way, the third term on the right of (2.62) can be shown to be smaller than $\eta/3$ with probability at least $1-\varepsilon/3$ for large T.

For any $K > 0$, the second term on the right of (2.62) is bounded by

$$\begin{aligned}
& P^*\left(\frac{1}{K}\|\delta_T\|^3 \sup_{1\leq l\leq T}\left\|Z_l'\overline{F}'HF\left(Z_0-Z_{\hat{k}}\right)\right\| \geq \frac{\lambda}{12}\left\|\hat{\delta}\right\|\right) \\
\leq\ & P^*\left(\sup_{N(K)}\sup_{1\leq l\leq T}\left\|Z_\Delta'\overline{F}'HFZ_l\right\| \geq \frac{\lambda K}{24}\|\delta_T\|^{-2}\right) + P^*\left(\left\|\hat{\delta}\right\| < \frac{1}{2}\|\delta_T\|\right) \\
& +P^*\left(\left|\hat{k}-k_0\right| > K\|\delta_T\|^{-2}\right).
\end{aligned} \tag{2.64}$$

By Lemma 2.8, the first term on the right of (2.64) is smaller than $\eta/3$ with probability no smaller than $1-\varepsilon/3$ for large K and T. Expectation of the second and third term on the right of (2.64) is bounded by $\varepsilon/3$ for large T and K, respectively.

Finally, for the second term of (2.59),

$$\begin{aligned}
& P^*\left(\sup_{\hat{N}^C(K)}\left|\frac{\hat{u}'\overline{F}'\overline{H}'\left(M_{FW_k}-M_{FW_{\hat{k}}}\right)HF\hat{u}}{k-\hat{k}}\right| \geq \lambda\left\|\hat{\delta}\right\|^2\right) \\
\leq\ & P^*\left(\sup_{\hat{N}^C(K)}\left|\hat{u}'\overline{F}'\overline{H}'\left(M_{FW_k}-M_{FW_{\hat{k}}}\right)HF\hat{u}\right| \geq \frac{\lambda K}{4}\right) \\
& +P^*\left(\left\|\hat{\delta}\right\| \leq \frac{1}{2}\|\delta_T\|\right).
\end{aligned} \tag{2.65}$$

By Lemma 2.9, the first term on the right of (2.65) is smaller than $\eta/2$ with probability at least $1-\varepsilon/2$ for large K and T. Expectation of the the second term of (2.65) is $P\left(\left\|\hat{\delta}\right\| \leq \frac{1}{2}\|\delta_T\|\right)$ which is smaller than $\varepsilon/2$ for large K by Proposition 2.2.

Collecting the results, we conclude that for arbitrary $\varepsilon > 0$ and $\eta > 0$, there exists K such that

$$P\left(P^*\left(\sup_{\hat{N}^C(K)} \frac{|R_T^*(k)|}{\left|k-\hat{k}\right|} \geq \lambda \left\|\hat{\delta}\right\|^2\right) > \eta\right) < \varepsilon$$

for large T. This together with (2.57) and with the Markov inequality imply that

$$P\left(P^*\left(\left|\hat{k}^* - \hat{k}\right| > K\|\delta_T\|^{-2}\right) > \eta\right) < \varepsilon$$

for large K and T as required. ■

Lemma 2.13 *If $\delta_T \to 0$ and $T\|\delta_T\|^2 \to \infty$, then*

$$S_T^*(k) - S_T^*\left(\hat{k}\right) = \delta_T'\hat{Z}_\Delta'\overline{F}'F\hat{Z}_\Delta\delta_T - 2\delta_T'\hat{Z}_\Delta'\overline{F}'HF\hat{u}\,\mathrm{sgn}\left(k-\hat{k}\right) + o_{p^*}(1),$$

where $o_{p^*}(1)$ is uniform on $\hat{N}(K)$.

Proof. Write

$$Q_T^*(k) = \hat{\delta}'\hat{Z}_\Delta'\overline{F}'F\hat{Z}_\Delta\hat{\delta} - \hat{\delta}'\hat{Z}_\Delta'\overline{F}'FW_k\left(W_k'\overline{F}'FW_k\right)^{-1}W_k'\overline{F}'F\hat{Z}_\Delta\hat{\delta}$$

and

$$\begin{aligned} R_T^*(k) &= -2\hat{\delta}'\hat{Z}_\Delta'\overline{F}'HF\hat{u}\,\mathrm{sgn}\left(k-\hat{k}\right) \\ &\quad +2\hat{\delta}'\hat{Z}_\Delta'\overline{F}'FW_k\left(W_k'\overline{F}'FW_k\right)^{-1}W_k'\overline{F}'HF\hat{u}\,\mathrm{sgn}\left(k-\hat{k}\right) \\ &\quad +\hat{u}'\overline{F}'\overline{H}'\left(M_{FW_k} - M_{FW_{\hat{k}}}\right)HF\hat{u}. \end{aligned}$$

By Lemma 2.8, $\hat{Z}_\Delta'\overline{F}'FW_k = O_p\left(\|\delta_T\|^{-2}\right)$ uniformly on $\hat{N}(K)$ and by Lemma 2.2, $\left(W_k'\overline{F}'FW_k\right)^{-1} = O_p(T^{-1})$ uniformly on $k \in \Lambda \cdot T$. Further, by Proposition 2.2, $\hat{\delta} = O_p(\|\delta_T\|)$. Also, by Lemma 2.8, $W_k'\overline{F}'HF\hat{u} = O_{p^*}\left(T^{1/2}\right)$

uniformly on $1 \leq k \leq T$ and by Lemma 2.9, $\hat{u}'\overline{F}'\overline{H}'\left(M_{FW_k} - M_{FW_{\hat{k}}}\right) HF\hat{u} = o_{p^*}(1)$ uniformly on $\hat{N}(K)$. It follows that

$$Q_T^*(k) = \hat{\delta}'\hat{Z}_\Delta'\overline{F}'F\hat{Z}_\Delta\hat{\delta} + o_p(1)$$

and

$$R_T^*(k) = -2\hat{\delta}'\hat{Z}_\Delta'\overline{F}'HF\hat{u}\operatorname{sgn}\left(k - \hat{k}\right) + o_{p^*}(1)$$

uniformly on $\hat{N}(K)$. By (2.45),

$$S_T^*(k) - S_T^*\left(\hat{k}\right) = \hat{\delta}'\hat{Z}_\Delta'\overline{F}'F\hat{Z}_\Delta\hat{\delta} - 2\hat{\delta}'\hat{Z}_\Delta'\overline{F}'HF\hat{u}\operatorname{sgn}\left(k - \hat{k}\right) + o_{p^*}(1)$$

uniformly on $\hat{N}(K)$. ∎

Lemma 2.14 *Let v_T^2 be defined as in Lemma 2.7. If the conditions of Proposition 2.9 are satisfied, then for any $K > 0$,*

$$\arg\min_{|\rho|\leq K} S_T^*\left(\hat{k} + \left[\rho v_T^{-2}\right]\right) \xrightarrow{d^*} \arg\min_{|\rho|\leq K} \bar{W}(\rho).$$

Proof. Write $\hat{k} + \left[\rho v_T^{-2}\right] = k$. From Lemma 2.13,

$$S_T^*\left(\hat{k} + \left[\rho v_T^{-2}\right]\right) - S_T^*\left(\hat{k}\right) = \delta_T'\hat{Z}_\Delta'\overline{F}'F\hat{Z}_\Delta\delta_T - 2\delta_T'\hat{Z}_\Delta'\overline{F}'HF\hat{u}\operatorname{sgn}\left(k - \hat{k}\right) + o_{p^*}(1).$$

By Lemma 2.2,

$$\frac{1}{T}Z'_{\Delta([\sigma T],[(\sigma+\rho)T])}M_\iota Z_{\Delta([\sigma T],[(\sigma+\rho)T])} \xrightarrow{p} |\rho|\,\Sigma$$

uniformly on $\{(\sigma,\rho) : 0 \leq \sigma, \sigma + \rho \leq 1\}$. Hence

$$v_T^2\hat{Z}_\Delta'\overline{F}'F\hat{Z}_\Delta \xrightarrow{p} \frac{|\rho|}{2\pi}\Sigma$$

uniformly over $|\rho| \leq K$. Lemma 2.8 implies that

$$\frac{1}{\sqrt{T}}Z'_{\Delta([\sigma T],[(\sigma+\rho)T])}\overline{F}'HF\hat{u} \overset{p}{\Longrightarrow} \begin{cases} \frac{1}{2\pi}\Omega^{\frac{1}{2}}\left(W(\sigma+\rho) - W(\sigma)\right) & \rho \geq 0, \\ \frac{1}{2\pi}\Omega^{\frac{1}{2}}\left(W(\sigma) - W(\sigma+\rho)\right) & \rho \leq 0, \end{cases}$$

on $\{(\rho,\sigma): 0 \leq \sigma, \sigma+\rho \leq 1\}$ from which it follows that

$$v_T \hat{Z}_\Delta' \overline{F}' H F \hat{u} \operatorname{sgn}\left(k-\hat{k}\right) \stackrel{p}{\Longrightarrow} \begin{cases} \frac{1}{2\pi}\Omega^{\frac{1}{2}} W_1(\rho) & \rho \geq 0, \\ \frac{1}{2\pi}\Omega^{\frac{1}{2}} W_2(-\rho) & \rho < 0, \end{cases}$$

where W_1, W_2 are independent p-vectors of independent standard Brownian motion processes. Proceeding as in the proof of Lemma 2.7, we deduce that

$$S_T^*\left(\hat{k}+\left[\rho v_T^{-2}\right]\right) - S_T^*\left(\hat{k}\right) \stackrel{d^*}{=} \bar{W}(\rho) + o_{p^*}(1)$$

on $\rho \in [-K, K]$ where "$\stackrel{d^*}{=}$" stands for the equality of distribution conditional on data. The lemma now follows from the continuous mapping theorem. ■

Lemma 2.15 *As* $T \to \infty$,

$$\hat{\Omega}^* \xrightarrow{p^*} \Omega.$$

Proof. As in the proof of Proposition 2.5, assume that process $\{x_t\}$ is scalar. By definition, matrix $\hat{\Omega}^*$ is equal to

$$\frac{4\pi^2}{T}\sum_{j=1}^{T-1} I_{xx,j} I_{\hat{u}\hat{u},j} \left|\eta_j^*\right|^2 + \frac{4\pi^2}{T}\sum_{j=1}^{T-1} I_{xx,j}\left(I_{\hat{u}^*\hat{u}^*,j} - I_{\hat{u}\hat{u},j}\left|\eta_j^*\right|^2\right). \tag{2.66}$$

Let $\hat{z}_t^* = z_t(\hat{k}^*)$. Writing

$$w_{\hat{u}^*,j} = \left(\hat{\beta}-\hat{\beta}^*\right) w_{x,j} + \hat{\delta}\left(w_{\hat{z},j} - w_{\hat{z}^*,j}\right) + \left(\hat{\delta}-\hat{\delta}^*\right) w_{\hat{z}^*,j} + w_{\hat{u},j}\eta_j^*$$

and proceeding as in the proof of Proposition 2.5, it can be seen that, up to a multiplicative constant, the second term of (2.66) is bounded in absolute value by

$$\begin{aligned} &\frac{1}{T}\sum_{j=1}^{T-1} \left|w_{x,j}\right|^2 \left|w_{\hat{u}^*,j} - w_{\hat{u},j}\eta_j^*\right|^2 \\ &+\frac{2}{T}\sum_{j=1}^{T-1} \left|w_{x,j}\right|^2 \left|w_{\hat{u}^*,j} - w_{\hat{u},j}\eta_j^*\right| \left|\eta_j^* w_{\hat{u},j}\right| \end{aligned} \tag{2.67}$$

and that the first term of (2.67) is bounded by

$$\frac{3}{T}\left(\hat{\beta}-\hat{\beta}^{*}\right)^{2}\sum_{j=1}^{T-1}\left|w_{x,j}\right|^{4}+\frac{3}{T}\hat{\delta}^{2}\sum_{j=1}^{T-1}\left|w_{x,j}\right|^{2}\left|w_{\hat{z},j}-w_{\hat{z}^{*},j}\right|^{2}$$
$$+\frac{3}{T}\left(\hat{\delta}-\hat{\delta}^{*}\right)^{2}\sum_{j=1}^{T-1}\left|w_{x,j}\right|^{2}\left|w_{\hat{z}^{*},j}\right|^{2}. \tag{2.68}$$

By Proposition 2.8, $\hat{\beta}^{*}-\hat{\beta}=O_{p^{*}}\left(T^{-1/2}\right)$, and by Lemma 1.11, $T^{-1}\sum_{j=1}^{T-1}\left|w_{x,j}\right|^{4}=o_{p}\left(T\right)$, therefore the first term of (2.68) is $o_{p^{*}}\left(1\right)$. By the Cauchy-Schwarz inequality, the second term of (2.68) is bounded by

$$\frac{3}{T}\hat{\delta}^{2}\left(\sum_{j=1}^{T-1}\left|w_{x,j}\right|^{4}\right)^{\frac{1}{2}}\left(\sum_{j=1}^{T-1}\left|w_{\hat{z},j}-w_{\hat{z}^{*},j}\right|^{4}\right)^{\frac{1}{2}}.$$

The sum in the second bracket of the last displayed expression is bounded by

$$\frac{1}{(2\pi T)^{2}}\left(\sum_{t}^{\hat{k},\hat{k}^{*}}\left|x_{t}\right|\right)^{4}+\frac{1}{(2\pi)^{2}T}\sum_{t}^{\hat{k},\hat{k}^{*}}\left|x_{t}\right|\sum_{s}^{\hat{k},\hat{k}^{*}}\left|x_{s}\right|\left(\sum_{r}^{\hat{k},\hat{k}^{*}}\left|x_{r}\right|^{2}\sum_{v}^{\hat{k},\hat{k}^{*}}\left|x_{t-s+v}\right|^{2}\right)^{\frac{1}{2}}.$$

For any $K>0$ and $M>0$,

$$P^{*}\left(\sum_{t}^{\hat{k},\hat{k}^{*}}\left|x_{t}\right|\geq M\left\|\delta_{T}\right\|^{-2}\right)\leq P^{*}\left(\sum_{t=k_{0}-2K\|\delta_{T}\|^{-2}}^{k_{0}+2K\|\delta_{T}\|^{-2}}\left|x_{t}\right|\geq M\left\|\delta_{T}\right\|^{-2}\right)$$
$$+P^{*}\left(\left|\hat{k}-k_{0}\right|>K\left\|\delta_{T}\right\|^{-2}\right)+P^{*}\left(\left|\hat{k}^{*}-\hat{k}\right|>K\left\|\delta_{T}\right\|^{-2}\right). \tag{2.69}$$

Proposition 2.1 implies that expectation of the second term on the right of (2.69) is bounded by $\varepsilon/3$ for large K. By the Markov inequality and Conditions 1.1 and 1.2, expectation of the first term on the right of (2.69) is bounded by CK/M which is bounded by $\varepsilon/3$ for large M. For any $\eta>0$, the last term on the right of (2.69) is smaller than η for large K and T with probability at least $1-\varepsilon/3$. This means that $\sum_{t}^{\hat{k},\hat{k}^{*}}\left|x_{t}\right|=O_{p^{*}}\left(\left\|\delta_{T}\right\|^{-2}\right)$. By similar arguments, $\sum_{t}^{\hat{k},\hat{k}^{*}}\left|x_{t}\right|^{2}=O_{p^{*}}\left(\left\|\delta_{T}\right\|^{-2}\right)$. Hence, because $\hat{\delta}=\delta_{T}+O_{p}\left(T^{-1/2}\right)=O_{p}\left(\left\|\delta_{T}\right\|\right)$

by Proposition 2.2 and because

$$\sup_{t,s\in N(2K)} \sum_{v=k_0-2K\|\delta_T\|^{-2}}^{k_0+2K\|\delta_T\|^{-2}} \left|x_{t-s+v}\right|^2 \leq \sum_{v=k_0-6K\|\delta_T\|^{-2}}^{k_0+6K\|\delta_T\|^{-2}} \left|x_v\right|^2,$$

we conclude that the second term of (2.68) is $o_{p^*}(1)$.

Next, the sum in the third term of (2.68) is bounded by

$$\sum_{j=1}^{T-1} \left|w_{x,j}\right|^2 \left|w_{\hat{z},j}\right|^2 + \sum_{j=1}^{T-1} \left|w_{x,j}\right|^2 \left|w_{\hat{z}^*,j} - w_{\hat{z},j}\right|^2. \tag{2.70}$$

The first term of (2.70) is $o_{p^*}(T^2)$ by the reasons given in the discussion of (2.25) and (2.27). The second term of (2.70) is $O_{p^*}\left(T^{-1/2}\left\|\delta_T\right\|^{-3}\right)$ by the reasons discussed above. Since $\hat{\delta} - \hat{\delta}^* = O_{p^*}\left(T^{-1/2}\right)$ by Proposition 2.8, the third term of (2.68) is $o_{p^*}(1)$. It follows that the first term of (2.67) is $o_{p^*}(1)$.

The second term of (2.67) is bounded by

$$2\left(\frac{1}{T}\sum_{j=1}^{T-1} \left|w_{x,j}\right|^2 \left|w_{\hat{u}^*,j} - \eta_j^* w_{\hat{u},j}\right|^2\right)^{\frac{1}{2}} \left(\frac{1}{T}\sum_{j=1}^{T-1} \left|w_{x,j}\right|^2 \left|w_{\hat{u},j}\right|^2 \left|\eta_j^*\right|^2\right)^{\frac{1}{2}} \tag{2.71}$$

by the Cauchy-Schwarz inequality. The first bracket of (2.71) has just been shown to be $o_{p^*}(1)$. The conditional expectation of the expression in the second bracket of (2.71) is $\hat{\Omega}$ which is $O_p(1)$ by Proposition 2.5. Thus the second term of (2.67), and consequently the second term of (2.66), is $o_{p^*}(1)$.

Further, the first term of (2.66) is equal to

$$\frac{4\pi^2}{T}\sum_{j=1}^{T-1} I_{xx,j} I_{uu,j} + \frac{4\pi^2}{T}\sum_{j=1}^{T-1} I_{xx,j} I_{uu,j}\left(\left|\eta_j^*\right|^2 - 1\right)$$
$$+\frac{4\pi^2}{T}\sum_{j=1}^{T-1} I_{xx,j}\left(I_{\hat{u}\hat{u},j} - I_{uu,j}\right)\left|\eta_j^*\right|^2. \tag{2.72}$$

By the Theorem 1 of Robinson (1998), the first term of (2.72) converges to Ω in probability. Further, the conditional second moment of the second term of

(2.72) is bounded by

$$\frac{C}{T^2}\sum_{j=1}^{T-1} I_{xx,j}^2 I_{uu,j}^2 = \frac{C}{T^2}\sum_{j=1}^{T-1} f_{xx,j}^2 f_{uu,j}^2 \left(\frac{I_{xx,j}^2}{f_{xx,j}^2}\frac{I_{uu,j}^2}{f_{uu,j}^2}\right).$$

By a routine extension of the proof of bound (4.8) of Robinson (1995b), it can be shown that factors $I_{xx,j}^2/f_{xx,j}^2$ and $I_{uu,j}^2/f_{uu,j}^2$ are $O_p(1)$ uniformly in $1 \leq j \leq T-1$. An application of Lemma 1.6 to $g(\lambda) = f_{xx}^{1/2}(\lambda) f_{uu}^{1/2}(\lambda)$ leads to the conclusion that the last displayed expression is $o_p(1)$. By the Markov inequality, the second term of (2.72) is $o_{p^*}(1)$. Finally, by the proof of Theorem 1.3 the conditional expectation of the third term of (2.72) is $o_p(1)$. Combining results, we have

$$\hat{\Omega}^* = \Omega + o_{p^*}(1).$$

■

Chapter 3

Inference on the time of break in regression with long memory processes

3.1 Introduction

When making inference about the location of the breakpoint in a linear regression model, a problem that we encounter is that the limiting distribution of the location estimator depends not only on unknown parameters of the model but, more importantly, on the distribution of the regressors and the error term. The limiting distribution is therefore data dependent and unknown in general and thus it is intractable for the purposes of statistical inference.

The problem of intractability of the limiting distribution of the breakpoint estimator has been approached in several ways. One approach to this problem has been to assume that the distribution of the data is known. This is the approach followed by Hinkley (1970), who considered a regression model with deterministic regressors and independent identical Gaussian distributed error term. He found an analytic solution for the limiting distribution but, in absence of a close form for the solution, he had to rely on numerical approximations to obtain critical values. However, with nonnormal errors or nondeterministic regressors, this approach seems to be difficult or impossible to implement. A more recent method would be to approximate the limiting distribution by Monte Carlo simulation. However, since the exact form of the underlying distribution is a crucial determinant of the form of the limiting distribution, the assumption that the underlying distribution of the data is known is untenable.

Another direction that has been pursued in the literature is to assume that as the sample size increases, the magnitude of the change shrinks to zero. If the size of break diminishes at an appropriate rate, the limiting distribution of the breakpoint estimator is invariant to the form of distribution of the regressors and the error term and depends only on the first two moments of these series, as we have seen in Chapter 2. The motivation behind the assumption of the shrinking break is that in finite samples the distribution of the location of the break for small breaks can be used as an approximation for large breaks as the sample size increases. One of the first authors to consider asymptotics with break magnitude local to zero has been Picard (1985) in the context of a Gaussian autoregressive process. The assumption of shrinking break has since become standard and has been adopted or discussed in various settings. The assumption of shrinking break has been also examined in Chapter 2.

However, obtaining distribution-free asymptotics comes at a price. As the sample size increases, information contained in the sample is sufficient to detect changes that are tending to zero but the increase in information is not fast enough to maintain the precision of the location estimator. The dispersion of the distribution of the estimator grows. As a result, tests of hypotheses about the date of change against fixed alternatives lose power as the sample size increases. In other words, for tests where the date of break under the null and alternative hypothesis is a given number of periods apart, the percentage of rejections of the null hypothesis converges toward zero. This may be seen as an unacceptable consequence of modelling the break as diminishing.

Given the power loss under the shrinking break, it appears reasonable to model the size of break as fixed. Under the fixed size of change and unknown distribution of data, a possibility for making inference feasible is to estimate the joint distribution of data. A possible estimation technique is the bootstrap. When data are assumed to be independently and identically distributed, bootstrap estimation can be carried out in a relatively straightforward manner. Antoch et al. (1995) bootstrap residuals from regression with fixed break and independent identically distributed errors. When data are serially correlated, however, the basic assumption of independence, essential to the validity of bootstrap, is violated. The bootstrap procedure needs to be modified to reflect the dependence structure of the data. In our case, the dependence in the data is possibly of long range.

A bootstrap procedure that is valid for short as well as long memory time series is the frequency bootstrap of Hidalgo (2003a), employed with modifications in Chapters 1 and 2 of this thesis. This procedure asymptotically matches the covariance structure of the underlying process. The ability to estimate the second order dependence structure is sufficient for approximating distributions that are entirely described by the second-order structure, for example the Gaussian distribution. However, it does not suffice in our current scenario where a general joint distribution of a process needs to be approximated. To our knowledge, there is currently no bootstrap procedure available that approximates the joint distribution of data while allowing for strong serial dependence. It is nevertheless worth mentioning that Bühlmann (1997) has observed that the sieve bootstrap offers a valid approximation of the finite dimensional joint distribution for weakly dependent processes.

In this chapter, we wish to maintain the assumption that the magnitude of break is fixed for all sample sizes and that the underlying distribution of data is unknown. We consider nondeterministic regressors and we allow for strong temporal dependence in both regressors and errors. In order to obtain valid inference procedures, we propose a bootstrap method for estimating the joint distribution of weakly or strongly dependent processes. To accommodate the potential presence of long memory, the memory parameter is estimated explicitly and information about memory is incorporated in the model in a way that essentially amounts to fractional differencing of data in the frequency domain. Apart from modelling the memory parameter, no further assumptions are made about the structure of the model for data, and the procedure can be viewed as semiparametric within a class of linear processes.

In what follows, Section 3.2 proposes a new bootstrap procedure for estimating the distribution of the breakpoint estimator under the assumption that the magnitude of break is fixed. Asymptotic properties of the proposed bootstrap procedure are examined in Section 3.3. Section 3.4 concludes and suggests possible extensions of the bootstrap procedure. The proofs of all results are collected in Section 3.A which refers to auxiliary results in Section 3.B.

3.2 Bootstrap under fixed break

We consider the linear regression model (2.1) with a break in the slope parameter,

$$y_t = \alpha + \beta' x_t + \delta' z_t + u_t, \qquad t = 1, \ldots, T, \tag{3.1}$$

where the magnitude δ of break is fixed and different from zero. We are again interested in making inference on the parameters of the model and in particular in testing the null hypothesis $k_0 = k_H$, where k_H is a constant, against the alternative hypothesis $k_0 \neq k_H$. We estimate parameters k, β and δ by the least squares procedure discussed in Chapter 2, that is,

$$\hat{k} = \arg \min_{k \in \Lambda \cdot T} \| M_{\iota, W_k} y \|^2$$

and

$$\begin{pmatrix} \hat{\beta} \\ \hat{\delta} \end{pmatrix} = \left(W_{\hat{k}}' M_\iota W_{\hat{k}} \right)^{-1} W_{\hat{k}}' M_\iota y.$$

When processes $\{x_t\}$ and $\{u_t\}$ are strictly stationary and mutually independent and $(\delta' x_t)^2 \pm 2\delta' x_t u_t$ has a continuous distribution, it has been shown in Chapter 2 that under appropriate regularity conditions,

$$\hat{k} - k_0 \xrightarrow{d} \arg \min_s W^0(s) \tag{3.2}$$

and

$$\sqrt{T} \begin{pmatrix} \hat{\beta} - \beta \\ \hat{\delta} - \delta_T \end{pmatrix} \xrightarrow{d} N(0, V), \tag{3.3}$$

where the process W^0 and the covariance matrix V are defined in (2.8) and Proposition 2.2, respectively.

The result in (3.2) indicates that the asymptotic distribution of $\hat{k}$ depends on the underlying distribution of $\{x_t\}$ and $\{u_t\}$. When the distribution of data is unknown, it needs to be estimated in an appropriate way. The main purpose of this section is to propose a bootstrap method of approximating the limiting distribution of the time of the break.

The strategy is to obtain an estimate k^* of the time of break k_0 from the regression

$$y_t^* = \hat{\alpha} + \hat{\beta} x_t + \hat{\delta} z_t + u_t^*, \qquad t = 1, \ldots, T,$$

where $u_1^*, \ldots u_T^*$ is a bootstrap sample obtained by resampling the residuals $\hat{u}_t$. Given the nature of the limiting distribution (3.2) of the breakpoint estimator $\hat{k}$, the distribution of u_t^* needs to approximated not only the marginal distribution of u_t but also the finite-dimensional joint distributions of u_t.

For covariance stationary processes with an AR(∞) representation, Bühlmann (1998) proposes to approximate the dependence structure by AR(p) models where the order $p = p_T$ increases at a certain rate with the sample size. Bühlmann's bootstrap procedure based on resampling the fitted AR(p) residuals delivers a valid approximation of the finite-dimensional joint distributions of the underlying process provided that the coefficients ρ_l of the AR(∞) representation satisfy $\sum_{l=0}^{\infty} l \, |\rho_l| < \infty$, the process has finite fourth moments and $p_T = o\left((T/\log T)^{1/4}\right)$.

The assumption of summability of $l \, |\rho_l|$ does not admit processes where the correlation between increasingly distant observations decays slowly and where the AR(∞) coefficients are absolutely summable but the series $\sum_l l \, |\rho_l|$ diverges. To accommodate a stronger degree of dependence, the AR(∞) process would need to be approximated by AR(p) model with a faster increase in p_T. However, with increasing ratio p_T/T, the variance of estimators of the autoregressive coefficients increases and it is not immediately evident that the estimators would remain consistent. Similar observations apply to the tradeoff between the block length or the subsample length and the precision of estimators in the block bootstrap of Carlstein (1986) and Künsch (1989) or the subsampling bootstrap of Politis and Romano (1992), respectively.

We propose a bootstrap procedure based on prewhitening of the process

u_t. Suppose that u_t is a covariance stationary linear process given by

$$u_t = \sum_{l=0}^{\infty} b_l \varepsilon_{t-l} = B(L)\,\varepsilon_t, \qquad t \in \mathbb{Z},$$

where $b_0 = 1$, $\sum_{l=0}^{\infty} b_l^2 < \infty$, $B(z) = \sum_{p=0}^{\infty} b_p z^p$, L denotes the lag operator, $L\varepsilon_t = \varepsilon_{t-1}$, and where $\{\varepsilon_t\}$ is a serially uncorrelated process with $E\varepsilon_t = 0$. If u_t has long memory, the spectral density f_{uu} of u_t has a pole of order $-2d$ at the zero frequency,

$$f_{uu}(\lambda) \sim C\lambda^{-2d} \qquad \text{as } \lambda \to 0+.$$

In this case, process u_t can be conveniently represented as

$$u_t = (1-L)^{-d}\,\Psi(L)\,\varepsilon_t, \tag{3.4}$$

where $\Psi(L) = (1-L)^d B(L)$. The corresponding representation of the spectral function is

$$f_{uu}(\lambda) = g(\lambda)\,h(\lambda), \tag{3.5}$$

where $g(\lambda) = \left|1 - e^{i\lambda}\right|^{-2d}$ and $h(\lambda) = \left|\Psi\left(e^{i\lambda}\right)\right|^2 \sigma_\varepsilon^2/(2\pi)$. When $d > 0$, function g dominates over h around the zero frequency, therefore the behaviour of the spectral density of the process at low frequencies is described by g. Correspondingly, we refer to g and h as long- and short-memory part of the spectral density, respectively.

Let

$$a_t = \sqrt{\frac{2\pi}{T}} \sum_{j=1}^{T} w_a(\lambda_j)\, e^{-it\lambda_j}, \qquad t = 1, \ldots, T,$$

be the inverse discrete Fourier transform of a generic sequence $\{w_a(\lambda_j)\}_{j=1}^{T}$. Representation (3.4) suggests to approximate the discrete Fourier transform of u_t as

$$w_u(\lambda_j) \approx B\left(e^{i\lambda_j}\right) w_\varepsilon(\lambda_j) = \left(1 - e^{i\lambda_j}\right)^{-d} \Psi\left(e^{i\lambda_j}\right) w_\varepsilon(\lambda_j), \tag{3.6}$$

and hence to approximate u_t as

$$\begin{aligned} u_t &\approx \sqrt{\frac{2\pi}{T}}\sum_{j=1}^{T} e^{-it\lambda_j} B\left(e^{i\lambda_j}\right) w_\varepsilon\left(\lambda_j\right) \\ &= \sqrt{\frac{2\pi}{T}}\sum_{j=1}^{T} e^{-it\lambda_j} \left(1-e^{i\lambda_j}\right)^{-d} \Psi\left(e^{i\lambda_j}\right) w_\varepsilon\left(\lambda_j\right), \end{aligned} \tag{3.7}$$

where "$\approx$" should be read as "approximately". The existence of the pole of function $\left(1-e^{i\lambda_j}\right)^{-d}$ at zero may prompt doubts about the reliability of the Bartlett approximation (3.6) of $w_u\left(\lambda_j\right)$ for frequencies near zero, but the results in Section 3.3 below indicate that the approximation is asymptotically valid provided B is replaced by a modification of B based on a truncated Fourier series of B.

If on the right-hand side of (3.7) the quantities d, $\Psi\left(e^{i\lambda_j}\right)$ and $w_\varepsilon\left(\lambda_j\right)$ are replaced by consistent estimators $\hat{d}$, $\hat{\Psi}\left(e^{i\lambda_j}\right)$ and $w_{\hat{\varepsilon}}\left(\lambda_j\right)$, respectively, the problem of obtaining a bootstrap sample $u_1^*, \ldots u_T^*$ becomes a problem of designing a valid bootstrap algorithm for the discrete Fourier transform $w_{\hat{\varepsilon}}\left(\lambda_j\right)$, $j=1,...,T$. These arguments lead us to propose the following bootstrap algorithm for estimation of the distribution of the breakpoint estimator.

Step 1 Obtain the centered least squares residuals

$$\hat{u}_t = \left(y_t-\bar{y}\right) - \hat{\beta}'\left(x_t-\bar{x}\right) - \hat{\delta}'\left(\hat{z}_t-\bar{\hat{z}}\right), \qquad t=1,...,T,$$

where $\bar{y} = T^{-1}\sum_{t=1}^{T} y_t$, $\bar{x} = T^{-1}\sum_{t=1}^{T} x_t$, $\hat{z}_t = x_t \mathbb{I}\left(t \le \hat{k}\right)$ and $\bar{\hat{z}} = T^{-1}\sum_{t=1}^{T} \hat{z}_t$ and where $\mathbb{I}$ denotes the indicator function. By definition, $\sum_{t=1}^{T} \hat{u}_t = 0$.

Step 2 Estimate d by the local Whittle estimator $\hat{d}$ proposed by Robinson (1995b),

$$\hat{d} = \arg\min_{a\in[0,\Delta]} H(a), \tag{3.8}$$

where $0<\Delta<1/2$,

$$H(a) = \log\left(\frac{1}{m}\sum_{j=1}^{m}\lambda_j^{2a} I_{\hat{u}\hat{u}}\left(\lambda_j\right)\right) - \frac{2a}{m}\sum_{j=1}^{m}\log\lambda_j$$

for an integer $m \in [1, [T/2])$, where the bandwidth m satisfies condition $1/m + m/T \to 0$.

Let

$$\hat{h}(\lambda) = \frac{1}{2m+1} \sum_{j=-m}^{m} \left|1 - e^{i(\lambda+\lambda_j)}\right|^{2\hat{d}} I_{\hat{u}\hat{u}}(\lambda + \lambda_j)$$

be our estimator of $h(\lambda)$ in (3.5).

Step 3 Let $M = [T/(4m)]$ and estimate innovations $\hat{\varepsilon}_t$ as

$$\hat{\varepsilon}_t = \frac{1}{\sqrt{2\pi T}} \sum_{j=1}^{T} e^{-it\lambda_j} \widehat{\hat{R}}_j w_{\hat{u}}(\lambda_j), \qquad t = 1, ..., T,$$

where

$$\begin{aligned}
\widehat{\hat{R}}_j &= \sum_{l=0}^{M} \hat{\rho}_l e^{il\lambda_j}, \qquad j = 1, ..., T, \\
\hat{\rho}_l &= \frac{1}{T} \sum_{j=1}^{T-1} \hat{R}_j e^{-il\lambda_j}, \qquad l = 0, ..., M, \\
\hat{R}_j &= \left(1 - e^{i\lambda_j}\right)^{\hat{d}} \hat{\Psi}^{-1}\left(e^{i\lambda_j}\right), \qquad j = 1, ..., T-1,
\end{aligned}$$

and where

$$\hat{\Psi}\left(e^{i\lambda}\right) = \exp\left\{-\sum_{r=1}^{M} \hat{c}_r e^{ir\lambda}\right\}, \qquad \lambda \in [0, \pi],$$

and

$$\hat{c}_r = \frac{1}{T} \sum_{l=m+1}^{[T/2]} \log \hat{h}_l e^{-ir\lambda_l} = \frac{1}{T} \sum_{l=m+1}^{[T/2]} \log \hat{h}(\lambda_l) \cos r\lambda_l, \qquad r = 1, ..., M.$$

Step 4 Draw a random sample $\varepsilon^* = (\varepsilon_1^*, \varepsilon_2^*, ..., \varepsilon_T^*)'$ with replacement from the empirical distribution of the residuals $\hat{\varepsilon}_t$, $P^*(\varepsilon_i^* = \hat{\varepsilon}_j) = T^{-1}$ for $i, j = 1, \ldots, T$, and compute the discrete Fourier transform w_{ε^*} of ε^*, that is

$$w_{\varepsilon^*}(\lambda_j) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} \varepsilon_t^* e^{it\lambda_j}, \qquad j = 1, \ldots, T.$$

Step 5 Compute

$$u_t^* = \sqrt{\frac{2\pi}{T}} \sum_{j=1}^{T} e^{-it\lambda_j} \widehat{\hat{B}}_j w_{\varepsilon^*}(\lambda_j), \qquad t = 1, ..., T, \tag{3.9}$$

where

$$\begin{aligned} \widehat{\hat{B}}_j &= \sum_{l=0}^{M} \hat{b}_l e^{il\lambda_j}, \quad j = 1, \ldots, T, \\ \hat{b}_l &= \frac{1}{T} \sum_{j=1}^{T-1} \hat{B}_j e^{-il\lambda_j}, \qquad l = 0, \ldots, M, \\ \hat{B}_j &= \left(1 - e^{i\lambda_j}\right)^{-\hat{d}} \hat{\Psi}\left(e^{i\lambda_j}\right), \qquad j = 1, \ldots, T-1. \end{aligned}$$

Step 6 Construct bootstrap sample y_t^*,

$$y_t^* = \hat{\beta}' x_t + \hat{\delta}' \hat{z}_t + u_t^*, \qquad t = 1, ..., T. \tag{3.10}$$

The regression intercept is set to zero because only the slope coefficients are of interest. Compute the bootstrap counterparts $\hat{u}^*(k)$ and $\hat{k}^*$ of estimates $\hat{u}(k)$ and $\hat{k}$, that is $\hat{u}^*(k) = M_{\iota, W_k} y^*$ and

$$\hat{k}^* = \arg\min_{k \in \Lambda \cdot T} S_T^*(k) = \min\left\{k : S_T^*(k) = \min_{l \in \Lambda \cdot T} S_T^*(l)\right\},$$

where $S_T^*(k) = ||\hat{u}^*(k)||^2 - ||\hat{u}^*(\hat{k})||^2$. Finally, compute the bootstrap counterparts $\hat{\beta}^*$ and $\hat{\delta}^*$ of estimators $\hat{\beta}$ and $\hat{\delta}$ as

$$\begin{pmatrix} \hat{\beta}^* \\ \hat{\delta}^* \end{pmatrix} = \left(W_{\hat{k}^*}' M_\iota W_{\hat{k}^*}\right)^{-1} W_{\hat{k}^*}' M_\iota y^*.$$

Step 2 of the bootstrap procedure requires a consistent estimator of the memory parameter d. In our procedure, we use the local Whittle estimator of

Robinson (1995b) but other estimators may prove suitable for the purposes of bootstrapping the date of break.

The estimator $\hat{h}$ of the short-memory part of the spectrum in Step 3 is obtained by reweighting the periodogram of residuals by the estimated long-memory part $\hat{g} = \left|1 - e^{i\lambda}\right|^{-2\hat{d}}$ of the spectrum and then smoothing. Normalizing periodogram $I_{\hat{u}\hat{u}}$ by $\hat{g}$ approximately corresponds to fractional differencing of the data. It is possible to estimate h by reversing the operations of smoothing and weighting, that is by reweighting the smoothed periodogram by $\hat{g}$. The latter method, however, is likely to produce a higher variance estimator. Reweighting periodogram prior to smoothing may be likened to applying logarithm to data in order to equalize variance across the sample. The estimator $\hat{\Psi}$ in Step 3 is obtained by the canonical spectral decomposition of $\hat{h}$, see for example Brillinger (1981, page 78-79).

It is interesting to compare Step 5 of our bootstrap to analogous Step 4 of the bootstrap procedure of Hidalgo (2005), where the function $\left(1 - e^{i\lambda}\right)^{-\hat{d}}$ is replaced by the square root of a partial sum of the Fourier series of $\left|1 - e^{i\lambda}\right|^{-2\hat{d}}$, that is by $\left|\sum_{l=-T+1}^{T-1} \hat{c}_l e^{il\lambda}\right|^{1/2}$ with

$$\hat{c}_l = \frac{(-1)^l \Gamma\left(1 - 2\hat{d}\right)}{\Gamma\left(l - \hat{d} + 1\right) \Gamma\left(1 - l - \hat{d}\right)}.$$

The replacement is motivated by a concern about the adequacy of the Bartlett approximation (3.7) for frequencies around zero because, in contrast to our case, the asymptotic distribution of the test statistic of Hidalgo (2005) depends only on the behaviour of the spectral density function for frequencies in a shrinking neighbourhood of zero. It would be informative to evaluate the performance of our bootstrap procedure under the two choices of estimator of $\left(1 - e^{i\lambda}\right)^{-d}$.

Definition (3.9) in Step 5 implies that

$$u_t^* = \sum_{l=0}^{M} \hat{b}_l \sum_{s=1}^{T} \varepsilon_s^* \mathbb{I}\left(s = t - l \mod T\right), \quad t = 1, ..., T. \tag{3.11}$$

Apart from the circularity of the bootstrap innovations, (3.11) could be viewed as a description of a time domain bootstrap for u_t. By Condition 3.4 below, the lag order M in (3.11) is required to satisfy $M = o\left((T/\log T)^{1/3}\right)$ and $M^{-1} = o\left(T^{-1/4}\right)$. A comparison with the lag order $p_T = o\left((T/\log T)^{1/4}\right)$ of Bühlmann (1997) leads to the observation that our bootstrap calls for a greater number of lags. This is to be expected as our procedure allows for strong dependence of data.

3.3 Asymptotic properties of the bootstrap procedure

The bootstrap procedure is discussed under the assumption the that $\{x_t\}$ and $\{u_t\}$ are stochastic processes that satisfy Conditions 1.1-1.5 and 2.1 with Conditions 1.2, 1.3 and 1.4 strengthened by Conditions 3.1, 3.2 and 3.3 below, respectively.

Condition 3.1 *$\{\xi_t\}$ is an independent identically distributed stochastic process with $E\xi_t = 0$, $E\left(\xi_t \xi_t'\right) = \Xi > 0$, and with finite fourth moments.*

Condition 3.2 *$\{\varepsilon_t\}$ is an independent identically distributed stochastic process that is independent of $\{\xi_t\}$ and that satisfies $E\varepsilon_t = 0$, $E\varepsilon_t^2 = \sigma_\varepsilon^2$, and $E\varepsilon_t^8 < \infty$.*

Condition 3.3 *The function $B\left(e^{i\lambda}\right)\left(1 - e^{i\lambda}\right)^d$ is twice continuously differentiable on $(0, \pi)$ and has one-sided second derivatives at 0 and π.*

In addition, we need to impose the following conditions.

Condition 3.4 *As $T \to \infty$,*

$$\frac{m^4}{T^3} + \frac{T^2}{m^3}\log T \to 0.$$

Condition 3.5 *When $0 < d < 1/2$, the coefficients b_l satisfy*

$$|b_l - b_{l+1}| \leq \frac{D\,|b_l|}{l} \qquad \text{for all } l > L \text{ and some } L < \infty,\ D < \infty,$$

and

$$b_l = O\left(l^{d-1}\right) \qquad as\ l \to \infty.$$

Conditions 3.1 and 3.2 are strong but necessary for the validity of the bootstrap procedure because shift invariance and mutual independence of the joint distributions of x_t and u_t is required. While finiteness of the moments up to fourth order only is imposed on the innovations of x_t, the innovations of u_t are required to have finite eighths moments. The stronger condition is introduced to guarantee convergence of the estimator of the short memory part of the spectral density of u_t.

Conditions 1.4 and 3.3 imply that the spectral density f_{uu} of the process u_t can be written as

$$f_{uu}(\lambda) = g(\lambda) h(\lambda), \qquad \lambda \in [0, \pi], \tag{3.12}$$

where $0 \leq d < \frac{1}{2}$, $g(\lambda) = \left|1 - e^{i\lambda}\right|^{-2d}$ and where the function h is positive, symmetric around zero, twice continuously differentiable on $(0, \pi)$ with one-sided second derivatives at 0 and π.

Condition 3.4 gives upper and lower bounds on the rate of increase to infinity of the smoothing parameter m. For example, a bandwidth in the form of $m = T^{\alpha}$ would satisfy Condition 3.4 for $\alpha \in (2/3, 3/4)$. It is worth noting that an identical bandwidth is used in the estimation of d in Step 2, in the smoothed estimate of prewhitened periodogram in Step 3, and through the parameter M also in the truncated sums in Steps 3 and 5 of the bootstrap procedure.

Condition 3.5 corresponds to Conditions (4.1) and (4.2) in Assumption 7 of Robinson (1994b). The condition is slightly stronger than quasi-monotonic convergence of b_l to zero which requires that $b_l \to 0$ as $l \to \infty$ and that $b_{l+1} \leq b_l (1 + C/l)$ for some $C < \infty$ and for all sufficiently large l. Condition 3.5 implies that b_l have bounded variation, that is $\sum_{l=0}^{\infty} |b_l - b_{l+1}| < \infty$. The condition is introduced to ensure that residual Fourier sums of the transfer function B are sufficiently small.

Condition 3.5 together with Conditions 1.1, 1.4, 3.2 and 3.3 are nearly equivalent to Condition C1 of Hidalgo (2005). The difference is that imposition of $g(\lambda) = \left|1 - e^{i\lambda}\right|^{-2d}$ on the factorization $f_{uu}(\lambda) = g(\lambda) h(\lambda)$ in our case implies the necessity to allow for functions h that have an angular point at 0 and π and are smooth everywhere else. Hidalgo (2005), on the other hand, imposes smoothness of h and h' at 0 and π, which means that the function g in factorization $f_{uu} = gh$ in general will not be equal to $\left|1 - e^{i\lambda}\right|^{-2d}$.

Conditions 3.1-3.3 are stronger than those conditions that are sufficient for obtaining the rate of convergence of the breakpoint estimator $\hat{k}$ and asymptotic normality of the slope coefficient estimators $\hat{\beta}$ and $\hat{\delta}$. To obtain these results, only a weaker form of Conditions 3.1-3.3 is needed. More specifically, Conditions 3.1 and 3.2 can be weakened to a requirement that ξ_t and ε_t are homoskedastic martingale difference processes with finite fourth cumulants with an allowance for a certain degree of cross-dependence, see Chapter 2 for details. Moreover, a smaller degree of smoothness is required, that is, the functions A and B only need to be once differentiable and so Condition 3.3 is not required. The limiting distribution of $\hat{k}$ under the assumption of diminishing size of break can also be obtained under these weaker conditions. The asymptotic distribution of location estimator $\hat{k}$ under fixed break obtains without a need for imposing Condition 3.3, but it is necessary to impose strict stationarity, finite fourths moments and mutual independence of regressors and errors.

We can now describe the statistical properties of the estimators employed in the bootstrap procedure. The first stage of the bootstrap procedure is to perform approximate fractional differencing with the difference parameter d. The following proposition affirms consistency of the local Whittle estimator $\hat{d}$ of d.

Proposition 3.1 *Under Conditions 1.1-1.5, 2.1 and 3.1-3.4,*

$$\hat{d} - d = O_p\left(m^{-1/2}\right).$$

Robinson (1995b) shows that the local Whittle estimator $\hat{d}$ is an $m^{1/2}$-consistent estimator of the memory parameter d of a linear process. Proposition 3.1 implies that $\hat{d}$ remains $m^{1/2}$-consistent when the process u_t is replaced by the regression residuals $\hat{u}_t$. We select the local Whittle estimator to estimate d but any consistent estimator of d can be utilized provided its rate of convergence is at least $m^{1/2}$. Since the spectral density is smooth away from the pole, an alternative candidate for the d estimator is the broadband estimator of Moulines and Soulier (1999).

The next prerequisite for the success of the bootstrap procedure is a valid approximation of the joint distribution of finite stretches of data.

Proposition 3.2 *Assume Conditions 1.1-1.5, 2.1 and 3.1-3.4 hold. Then for any finite K,*

$$\left(u^*_{k_0-K}, \ldots, u^*_{k_0+K}\right) \xrightarrow{d^*} \left(u_{k_0-K}, \ldots, u_{k_0+K}\right).$$

Proposition 3.2 deals with continuous blocks of data in a neighbourhood of the true date of break, but the proposition can be generalized for any finite-dimensional joint distribution of the process. The result is used in subsequent steps of our bootstrap procedure, but is itself of interest and could be adapted for other purposes.

Proposition 3.3 *Assume Conditions 1.1-1.5, 2.1 and 3.1-3.5 are satisfied. Then*

$$\hat{k}^* - \hat{k} = O_{p^*}(1).$$

Proposition 3.3 mirrors the corresponding result (2.7) for the rate of convergence of $\hat{k}$. The rate of convergence of $\hat{k}^*$ can be regarded as a preliminary result in the analysis of the limiting bootstrap distribution of $\hat{k}^*$.

The following theorem is the core result of the chapter. The theorem asserts consistency of the proposed bootstrap procedure and gives the asymptotic distribution of the bootstrap estimator of the date of the break.

Theorem 3.1 *Assume $(\delta' x_t)^2 \pm 2\delta' x_t u_t$ has a continuous distribution. Then under Conditions 1.1-1.5, 2.1 and 3.1-3.5,*

$$\hat{k}^* - \hat{k} \xrightarrow{d^*} \arg\min_k W^0(k),$$

where the process W^0 is defined in (2.8).

The assumption of continuity of the distribution of $(\delta' x_t)^2 \pm 2\delta' x_t u_t$ identifies the bootstrap breakpoint estimator $\hat{k}^*$ by ensuring that the process W^0 has a unique minimum.

A bootstrap approximation of the asymptotic test can be constructed on the basis of the conditional distribution of $\hat{k}^* - \hat{k}$. For the test of the null hypothesis H_0: $k_0 = k_H$, the bootstrap rejection region C_α^* at a level of significance α is constructed in such a way that $P^*\left(\hat{k}^* - \hat{k} \in C_\alpha^*\right) = \alpha$. The bootstrap test rejects when $\hat{k} - k_H \in C_\alpha^*$. By Theorem 3.1, the bootstrap rejection region C_α^* consistently estimates the asymptotic rejection region C_α where C_α is such that $P\left(\arg\min_k W^0(k) \in C_\alpha\right) = \alpha$.

While the examination of the asymptotic distribution of the bootstrap breakpoint estimator is the main focus of our analysis, the statistical properties of the slope estimator are also of interest. The following theorem characterizes the asymptotic distribution of the bootstrap counterparts of $\hat{\beta}$ and $\hat{\delta}$.

Theorem 3.2 *Assume Conditions 1.1-1.5, 2.1 and 3.1-3.5 hold. Then*

$$\sqrt{T}\begin{pmatrix} \hat{\beta}^* - \hat{\beta} \\ \hat{\delta}^* - \hat{\delta} \end{pmatrix} \xrightarrow{d^*} N(0, V),$$

where V is defined in Proposition 2.2.

Theorem 3.2 states that the asymptotic normality of slope coefficient estimators is preserved under fixed breaks when employing the proposed bootstrap procedure.

3.4 Conclusions

This chapter examines the problem of obtaining valid inference for the date of break under the assumption that the size of break does not change when the sample size increases. The problem of unknown distribution of underlying data is dealt with by devising a bootstrap procedure which approximates the distribution of innovations of the linear process of errors. The method is based on prewhitening procedure which delivers estimates of the innovations of a linear process. The deconvolution of the residual process is carried out in two stages. First, the degree of memory of the process is brought down by filtering out the high amplitude at low frequencies. The second stage is similar to the sieve bootstrap of Bühlmann (1997). It seems convenient to perform the prewhitening in the frequency domain.

The bootstrap values of innovations are used to generate a bootstrap sample of the left-hand side variable y_t for the purpose of obtaining a bootstrap distribution of the date of break, but the bootstrap procedure for u_t or its estimate $\hat{u}_t$ is itself of interest and can be adapted for other purposes.

There are several ways in which the techniques proposed in this chapter can be extended. The bootstrap procedure could be modified for estimation of the distribution of the date of break in a nonlinear regression

$$y_t = f(x_t, \delta_t) + u_t, \qquad t = 1, \ldots, T, \tag{3.13}$$

where $\delta_t = \delta_0$ for $t \leq \tau_0 T$ and $\delta_t = \delta_1$ for $t > \tau_0 T$ with $\delta_0 \neq \delta_1$. The procedure could also be adapted to allow for heteroskedastic errors in regression (3.1), that is, for errors of the form $\tilde{u}_t = \sigma(x_t, \delta_t) u_t$, where δ_t is as in (3.13), u_t satisfies conditions given in Section 3.3 and σ is a function known up to a finite number of parameters. In this case, the least squares residuals in Step 1 of the bootstrap procedure could be standardized by $\sigma\left(x_t, \hat{\delta}_t\right)$ where $\hat{\delta}_t$ is a $T^{1/2}$-consistent estimator of δ_t.

While several steps of the proposed bootstrap procedure, and in particular the fractional differencing, are executed in the frequency domain, it is likely that the underlying ideas could be realized also in the time domain. Steps 3-5 of the proposed bootstrap procedure could be modified in the following manner.

Step 3' Let $\hat{\alpha}_j$ be the coefficients of the binomial representation of $(1-x)^d$ at $d = \hat{d}$, that is,

$$\hat{\alpha}_j = \frac{\Gamma\left(j - \hat{d}\right)}{\Gamma(j+1)\,\Gamma\left(-\hat{d}\right)}, \qquad j = 0, 1, \ldots,$$

where Γ denotes the gamma function. Compute

$$\eta_t = \sum_{j=0}^{T} \hat{\alpha}_j \hat{u}_{t-j}, \qquad t = 1, \ldots, T,$$

with $\hat{u}_t = 0$ for $t \leq 0$. Let $\eta_t = 0$ for $t \leq 0$ and compute

$$\hat{\varepsilon}_t = \eta_t - \sum_{p=1}^{p_T} \hat{\varphi}_p \eta_{t-p}, \qquad t = 1, \ldots, T,$$

where $\hat{\varphi}_1, \ldots, \hat{\varphi}_{p_T}$ are estimated coefficients of an AR(p_T) model for η_t and where p_T increases with the sample size T at an appropriate rate.

Step 4' Draw a random sample $\varepsilon_1^*, \ldots, \varepsilon_{2T}^*$ with replacement from the centered residuals

$$\tilde{\varepsilon}_t = \hat{\varepsilon}_t - \frac{1}{T} \sum_{t=1}^{T} \hat{\varepsilon}_t, \qquad t = 1, \ldots, T.$$

Step 5' Compute

$$\eta_t^* = \varepsilon_t^* + \sum_{p=1}^{p_T} \hat{\varphi}_p \varepsilon_{t-p}^*, \qquad t = 1, \ldots, 2T$$

with the initial condition $\varepsilon_t^* = 0$ for $t \leq 0$. Let $\hat{\beta}_j$ be the coefficients of the binomial representation of $(1-x)^{-d}$ at $d = \hat{d}$,

$$\hat{\beta}_j = \frac{\Gamma\left(j + \hat{d}\right)}{\Gamma(j+1)\,\Gamma\left(\hat{d}\right)}, \qquad j = 0, 1, \ldots$$

and obtain

$$u_t^* = \eta_{t+T}^* + \sum_{j=1}^{T} \hat{\beta}_j \eta_{t+T-j}^*, \qquad t = 1, \ldots, T.$$

Details and validity of the proposed extensions would need to be examined in future research.

3.A Proofs

In what follows, we assume for simplicity that process $\{x_t\}_{t\in\mathbb{Z}}$ is scalar and denote $\operatorname{var} x_t = \sigma_x^2$.

Proof of Proposition 3.1. On account of Theorems 1 and 2 of Robinson (1995b), it suffices to show that

$$\sup_{a\in\Theta}\left|\sum_{j=1}^{m}\lambda_j^{2a}\left(I_{\hat{u}\hat{u},j}-I_{uu,j}\right)\right| = o_p\left(m^{1/2}\right). \tag{3.14}$$

Define $\tilde{u}_t = u_t - \hat{u}_t$. By the definition of $\hat{u}_t$, we have $w_{\hat{u},j} = w_{u,j} - w_{\tilde{u},j}$ for $j = 1,\ldots,T-1$, where

$$w_{\tilde{u},j} = \left(\hat{\beta}-\beta\right)w_{x,j} + \left(\hat{\delta}-\delta\right)w_{\hat{z},j} + \delta w_{\Delta\hat{z},j} \tag{3.15}$$

and $\{\Delta\hat{z}_t\}_{t=1}^{T} = \{\hat{z}_t - z_t\}_{t=1}^{T} = \{z_t(\hat{k}) - z_t(k_0)\}_{t=1}^{T}$. Hence

$$I_{\hat{u}\hat{u},j} - I_{uu,j} = I_{\tilde{u}\tilde{u},j} + 2\operatorname{Re} I_{\tilde{u}u,j}$$

and the left-hand side of (3.14) is bounded by

$$\begin{aligned}\sum_{j=1}^{m}\left|I_{\hat{u}\hat{u},j}-I_{uu,j}\right| &\le \sum_{j=1}^{m}\left(I_{\tilde{u}\tilde{u},j}+2\left|I_{\tilde{u}u,j}\right|\right)\\ &\le \sum_{j=1}^{m}I_{\tilde{u}\tilde{u},j} + 2\left(\sum_{j=1}^{m}I_{\tilde{u}\tilde{u},j}\right)^{\frac{1}{2}}\left(\sum_{j=1}^{m}I_{uu,j}\right)^{\frac{1}{2}} \qquad (3.16)\end{aligned}$$

due to the triangle and Cauchy-Schwarz inequalities. By the Cauchy-Schwarz inequality and (3.15),

$$I_{\tilde{u}\tilde{u},j} \le 3\left(\hat{\beta}-\beta\right)^2\left|w_{x,j}\right|^2 + 3\left(\hat{\delta}-\delta\right)^2\left|w_{\hat{z},j}\right|^2 + 3\delta\left|w_{\Delta\hat{z},j}\right|^2.$$

Further, Lemma 1.5 implies that $E\left(\left|w_{z([\tau T]),j}\right|^2/\left|A_j\right|^2\right) \le C$ uniformly over integers $1 \le j \le [T/2]$ and over $\tau \in [0,1]$ for some C, and Condition 1.4

implies that $|A_j|^2 \lambda_j^{2d_x} \leq C$ uniformly over integers $1 \leq j \leq [T/2]$. Therefore

$$E\sum_{j=1}^{m} I_{zz,j} \leq C\sum_{j=1}^{m} \lambda_j^{-2d_x} = O\left(m\left(\frac{T}{m}\right)^{2d_x}\right).$$

By Lemma 1.7, $\delta - \hat{\delta} = O_p\left(T^{-1/2}\right)$, therefore the contribution of $\left(\hat{\delta}-\delta\right)^2 I_{zz,j}$ into the first term of (3.16) is $O_p\left((T/m)^{2d_x-1}\right)$ which is $o_p\left(m^{1/2}\right)$ by Condition 3.4. Likewise, the contribution into the first term of (3.16) due to $\left(\hat{\beta}-\beta\right)^2 I_{xx,j}$ is $O_p\left((T/m)^{2d_x-1}\right)$. Further, we show that

$$\sup_{1\leq j\leq [T/2]} I_{\Delta\hat{z}\Delta\hat{z},j} = O_p\left(T^{-1}\right). \tag{3.17}$$

For any $D > 0$ and any finite K,

$$P\left(\sup_{1\leq j\leq [T/2]} I_{\Delta\hat{z}\Delta\hat{z},j} > \frac{D}{T}\right) \leq P\left(K\sum_{t=k_0-K}^{k_0+K} x_t^2 \geq D\right) + P\left(\left|\hat{k}-k_0\right| > K\right). \tag{3.18}$$

By the Markov inequality, the first term on the right of (3.18) is bounded by $\sigma_x^2 K^2/D < \varepsilon/2$ for sufficiently large D. The second term on the right of (3.18) bounded by $\varepsilon/2$ for large K by (3.2). It follows that (3.17) holds. This implies that the contribution due to $I_{\tilde{u}\tilde{u},j}$ into the first term of (3.16) is $o_p\left(m^{1/2}\right)$. We have thus shown that the first term of (3.16) is $O_p\left((T/m)^{2d_x-1}\right)$.

Proceeding in a similar way, we obtain that $\sum_{j=1}^{m} I_{uu,j} = O_p\left(m\left(\frac{T}{m}\right)^{2d}\right)$ and hence that (3.16) is

$$O_p\left(\left(\frac{T}{m}\right)^{2d_x-1}\right) + O_p\left(m^{\frac{1}{2}}\left(\frac{T}{m}\right)^{d_x+d-\frac{1}{2}}\right) = o_p\left(m^{\frac{1}{2}}\right)$$

by Condition 3.4, because $d_x + d < 1/2$. ∎

Proof of Proposition 3.2. Using the Cramér-Wold device, we need to prove that for any finite constants $\alpha_{k_0-K}, \ldots, \alpha_{k_0+K}$,

$$\sum_{t=k_0-K}^{k_0+K} \alpha_t u_t^* \xrightarrow{d^*} \sum_{t=k_0-K}^{k_0+K} \alpha_t u_t.$$

From the definition of u_t^* in Step 5 of the bootstrap, we obtain

$$u_t^* = \sum_{l=0}^{M} \hat{b}_l \sum_{s=1}^{T} \varepsilon_s^* \mathbb{I}\left(s = t - l \bmod T\right).$$

For large enough T, $k_0 - K \geq M + 1$ and we can write $u_t^* = \sum_{l=0}^{M} \hat{b}_l \varepsilon_{t-l}^*$ and

$$\sum_{t=k_0-K}^{k_0+K} \alpha_t u_t^* = \sum_{l=0}^{M} \hat{b}_l \sum_{t=k_0-K}^{k_0+K} \alpha_t \varepsilon_{t-l}^*.$$

Let $\omega > 0$ be such that θ, $\theta + 3\omega$ and $\theta - 3\omega$ are continuity points of the distribution function of u_t. Then

$$\begin{aligned} P^*\left(\sum_{t=k_0-K}^{k_0+K} \alpha_t u_t^* \leq \theta\right) &\leq P^*\left(\sum_{l=0}^{M} b_l \sum_{t=k_0-K}^{k_0+K} \alpha_t \varepsilon_{t-l}^* \leq \theta + \omega\right) \\ &\quad + P^*\left(\left|\sum_{l=0}^{M} \left(\hat{b}_l - b_l\right) \sum_{t=k_0-K}^{k_0+K} \alpha_t \varepsilon_{t-l}^*\right| > \omega\right). \end{aligned}$$

By the Markov and Cauchy-Schwarz inequalities, the second term on the right of the last displayed inequality is bounded by

$$\frac{1}{\omega}\left|\sum_{l=0}^{M} \left(\hat{b}_l - b_l\right)\right| \sum_{t=k_0-K}^{k_0+K} |\alpha_t| E^* \left|\varepsilon_{t-l}^*\right| \leq \frac{K}{\omega}\left(M \sum_{l=0}^{M} \left(\hat{b}_l - b_l\right)^2\right)^{\frac{1}{2}} \hat{\sigma}_\varepsilon$$

which is $K\omega^{-1} o_p(1)$ by Lemmas 3.13 and 3.15. Further, for any $N > 0$,

$$\begin{aligned} & P^*\left(\sum_{l=0}^{M} b_l \sum_{t=k_0-K}^{k_0+K} \alpha_t \varepsilon_{t-l}^* \leq \theta + \omega\right) \\ \leq\ & P^*\left(\sum_{l=0}^{N} b_l \sum_{t=k_0-K}^{k_0+K} \alpha_t \varepsilon_{t-l}^* \leq \theta + 2\omega\right) \\ & + P^*\left(\left|\sum_{l=N+1}^{M} b_l \sum_{t=k_0-K}^{k_0+K} \alpha_t \varepsilon_{t-l}^*\right| > \omega\right). \end{aligned} \tag{3.19}$$

By the Markov inequality, the second term on the right of (3.19) is bounded by

$$\frac{1}{\omega^2}E^*\left|\sum_{l=N+1}^{M} b_l \sum_{t=k_0-K}^{k_0+K} \alpha_t \varepsilon_{t-l}^*\right|^2 = CK^2\frac{\hat{\sigma}_\varepsilon^2}{\omega^2}\sum_{l=N+1}^{M} b_l^2.$$

Due to the square summability of b_l, the sum on the right of the last displayed equality tends to 0 as $N \to \infty$. The first term on the right of (3.19) is bounded by

$$P\left(\sum_{l=0}^{N} b_l \sum_{t=k_0-K}^{k_0+K} \alpha_t \varepsilon_{t-l} \le \theta + 2\omega\right)$$
$$+\left|P^*\left(\sum_{l=0}^{N} b_l \sum_{t=k_0-K}^{k_0+K} \alpha_t \varepsilon_{t-l}^* \le \theta + 2\omega\right) - P\left(\sum_{l=0}^{N} b_l \sum_{t=k_0-K}^{k_0+K} \alpha_t \varepsilon_{t-l} \le \theta + 2\omega\right)\right|,$$

where the first term is bounded by

$$P\left(\sum_{l=0}^{\infty} b_l \sum_{t=k_0-K}^{k_0+K} \alpha_t \varepsilon_{t-l} \le \theta + 3\omega\right) + P\left(\left|\sum_{l=N+1}^{\infty} b_l \sum_{t=k_0-K}^{k_0+K} \alpha_t \varepsilon_{t-l}\right| > \omega\right). \tag{3.20}$$

The first term of (3.20) is bounded by $\sigma_\varepsilon^2\omega^{-2}CK^2\sum_{l=N+1}^{\infty} b_l^2$ by the Markov and Cauchy-Schwarz inequalities.

Fix $\varepsilon > 0$. Collecting results, we have that for any $\eta > 0$.

$$P^*\left(\sum_{t=k_0-K}^{k_0+K} \alpha_t u_t^* \le \theta\right) \le P\left(\sum_{t=k_0-K}^{k_0+K} \alpha_t u_t \le \theta + 3\omega\right)$$

$$+\left|P^*\left(\sum_{l=0}^{N} b_l \sum_{t=k_0-K}^{k_0+K} \alpha_t \varepsilon_{t-l}^* \le \theta + 2\omega\right) - P\left(\sum_{l=0}^{N} b_l \sum_{t=k_0-K}^{k_0+K} \alpha_t \varepsilon_{t-l} \le \theta + 2\omega\right)\right|$$
$$+\frac{\eta}{2} \tag{3.21}$$

with probability at least $1-\varepsilon/2$ for large T and N. Lemma 3.16 and Cramér-Wold device imply that for any θ and for any fixed N, the last term on the right of (3.21) is smaller than $\eta/2$ with probability no smaller than $1 - \varepsilon/2$

for large T. This means that

$$P^*\left(\sum_{t=k_0-K}^{k_0+K}\alpha_t u_t^* \le \theta\right) \le P\left(\sum_{t=k_0-K}^{k_0+K}\alpha_t u_t \le \theta + 3\omega\right) + \eta$$

with probability at least $1-\varepsilon$ for large T and N.

In a similar way, it can be shown that

$$P\left(\sum_{t=k_0-K}^{k_0+K}\alpha_t u_t \le \theta - 3\omega\right) - \eta \le P^*\left(\sum_{t=k_0-K}^{k_0+K}\alpha_t u_t^* \le \theta\right)$$

with probability greater than $1-\varepsilon$ for large T and N. The lemma now follows because ω, ε and η are arbitrarily small. ■

Define $\hat{Z}_\Delta = (Z_k - Z_{\hat{k}})\,\mathrm{sgn}(k-\hat{k})$.

Proof of Proposition 3.3. Let $u^* = (u_1^*, \ldots, u_T^*)'$ and $\check{u}^* = (\check{u}_1^*, \ldots, \check{u}_T^*)'$. By definition, $S_T^*(k) = ||\hat{u}^*(k)||^2 - \left\|\hat{u}^*\left(\hat{k}\right)\right\|^2 = Q_T^*(k) + R_T^*(k)$, where

$$\begin{aligned} Q_T^*(k) &= \hat{\delta}' Z_{\hat{k}}' M_{\iota,W_k} Z_{\hat{k}}\hat{\delta} \quad \text{and} \\ R_T^*(k) &= 2\hat{\delta}' Z_{\hat{k}}' M_{\iota,W_k} u^* + u^{*\prime}(M_{\iota,W_k} - M_{\iota,W_{\hat{k}}})u^*. \end{aligned}$$

By standard arguments, for any $K > 0$,

$$P^*\left(\left|\hat{k}^* - \hat{k}\right| > K\right) \le P^*\left(\inf_{\hat{N}^C(K)} \frac{Q_T^*(k)}{\left|k-\hat{k}\right|} \le \lambda\right) + P^*\left(\sup_{\hat{N}^C(K)} \frac{|R_T^*(k)|}{\left|k-\hat{k}\right|} \ge \lambda\right), \tag{3.22}$$

where $\hat{N}^C(K)$ is defined in (2.47). Fix $\varepsilon, \eta > 0$. Expectation of the first term on the right of (3.22) is equal to

$$P\left(\inf_{\hat{N}^C(K)} \frac{\hat{\delta}' Z_{\hat{k}}' M_{\iota,W_k} Z_{\hat{k}}\hat{\delta}}{\left|k-\hat{k}\right|} \le \lambda\right).$$

By Lemma 2.10, there exists $\lambda > 0$ such that for every $\varepsilon > 0$, there exists $K < \infty$ such that the last displayed expression is smaller than $\varepsilon/2$ for large T. Select such λ.

Write

$$R_T^*(k) = R_{1T}^*(k) + R_{2T}^*(k) + R_{3T}^*(k) + R_{4T}^*(k),$$

where

$$\begin{aligned}
R_{1T}^*(k) &= 2\hat{\delta}'\hat{Z}_\Delta' M_\iota^* \breve{u} \operatorname{sgn}\left(\hat{k} - k\right), \\
R_{2T}^*(k) &= 2\hat{\delta}'\hat{Z}_\Delta' M_\iota (u^* - \breve{u}^*) \operatorname{sgn}\left(\hat{k} - k\right), \\
R_{3T}^*(k) &= -2\hat{\delta}'\hat{Z}_\Delta' M_\iota W_k \left(W_k' M_\iota W_k\right)^{-1} W_k' M_\iota u^* \operatorname{sgn}\left(\hat{k} - k\right), \\
R_{4T}^*(k) &= u^{*\prime}(M_{\iota,W_k} - M_{\iota,W_{\hat{k}}})u^*.
\end{aligned}$$

The contribution due to $R_{1T}^*(k)$ into the second term of (3.22) is bounded by

$$P^*\left(\sup_{\hat{N}^C(K)} \frac{\left\|\hat{Z}_\Delta' M_\iota \breve{u}^*\right\|}{\left|k - \hat{k}\right|} \geq \frac{\lambda}{16\|\delta\|}\right) + P^*\left(\left\|\hat{\delta}\right\| > 2\|\delta\|\right). \tag{3.23}$$

By Lemma 3.23, the first term of (3.23) is smaller than $\eta/4$ with probability at least $1 - \varepsilon/4$ for large K and T. Since $\hat{\delta} = \delta + O_p\left(T^{-1/2}\right)$ by (3.3), the expectation of the second term of (3.23) is $o_p(1)$.

We now turn to the term $R_{2T}^*(k)$. Write

$$u_t^* - \breve{u}_t^* = \sum_{s=1}^{T} \varepsilon_s^* \sum_{l=0}^{M} \left(\hat{b}_l - b_l\right) \mathbb{I}\left(s = t - l \mod T\right)$$

and define

$$r_t^* = \sum_{s=1}^{T} \varepsilon_s^* \sum_{l=0}^{M} \left(\hat{b}_l - b_l\right) \mathbb{I}\left(s = t - l \mod T\right) - \left(\frac{1}{T}\sum_{s=1}^{T} \varepsilon_s^*\right) \sum_{l=0}^{M} \left(\hat{b}_l - b_l\right).$$

We have

$$\sup_{\hat{N}^C(K)} \frac{1}{\left|k-\hat{k}\right|} \left\| \hat{Z}'_{\Delta} M_{\iota} \left(u^{*} - \breve{u}_t^{*}\right) \right\| \leq \left(\sup_{\hat{N}^C(K)} \frac{1}{\left|k-\hat{k}\right|} \sum_{t}^{\hat{k},k} \left\|x_t\right\|^2 \right)^{\frac{1}{2}} \times \left(\sup_{\hat{N}^C(K)} \frac{1}{\left|k-\hat{k}\right|} \sum_{t}^{\hat{k},k} r_t^{*2} \right)^{\frac{1}{2}}.$$

Since $(a-b)^2 \leq 2\left(a^2+b^2\right)$, we obtain

$$\begin{aligned} r_t^{*2} &\leq 2\left(\sum_{s=1}^{T} \varepsilon_s^{*} \sum_{l=0}^{M} \left(\hat{b}_l - b_l\right) \mathbb{I}\left(s = t-l \mod T\right)\right)^2 \\ &\quad +2\left(\frac{1}{T}\sum_{s=1}^{T}\varepsilon_s^{*}\right)^2 \left(\sum_{l=0}^{M}\left(\hat{b}_l - b_l\right)\right)^2. \end{aligned} \tag{3.24}$$

By the Cauchy-Schwarz inequality, the first term on the right of (3.24) is bounded by

$$\sum_{s=1}^{T} \varepsilon_s^{*2} \sum_{l=0}^{M} \mathbb{I}\left(s = t-l \mod T\right) \sum_{p=0}^{M} \left(\hat{b}_p - b_p\right)^2.$$

For sufficiently large T,

$$\begin{aligned} & \sup_{\hat{N}^c(K)} \frac{1}{\left|k-\hat{k}\right|} \sum_{t}^{\hat{k},k} \sum_{s=1}^{T} \varepsilon_s^{*2} \sum_{l=0}^{M} \mathbb{I}\left(s = t-l \mod T\right) \\ = & \sup_{\hat{N}^c(K)} \frac{1}{\left|k-\hat{k}\right|} \sum_{r=0}^{M} \sum_{t}^{\hat{k},k} \varepsilon_{t-r}^{*2} \leq \sum_{r=0}^{M} \sup_{\hat{N}^c(K)} \frac{1}{\left|k-\hat{k}\right|} \sum_{t}^{\hat{k},k} \varepsilon_{t-r}^{*2} \\ = & \ M O_{p^*}(1). \end{aligned}$$

Since $M\sum_{l=0}^{M}\left(\hat{b}_l - b_l\right)^2$ is $O_p\left(Mm^{-2}T\log T\right) = o_p(1)$ by Lemma 3.13 and

Condition 3.4, the second term on the right of (3.24) is bounded by

$$2\left(\frac{1}{T}\sum_{s=1}^{T}\varepsilon_s^*\right)^2 M\sum_{l=0}^{M}\left(\hat{b}_l-b_l\right)^2=o_{p^*}(1).$$

Therefore

$$\sup_{\hat{N}^C(K)}\frac{1}{\left|k-\hat{k}\right|}\sum_t^{\hat{k},k}r_t^{*2}=o_{p^*}(1)$$

and also

$$\sup_{\hat{N}^C(K)}\frac{1}{\left|k-\hat{k}\right|}\left\|\hat{Z}_\Delta' M_\iota\left(u^*-\check{u}^*\right)\right\|=o_{p^*}(1)$$

because $\sup_{\hat{N}^C(K)}\left|k-\hat{k}\right|^{-1}\sum_t^{\hat{k},k}\|x_t\|^2$ is $O_p(1)$. It follows that

$$\begin{aligned}P^*\left(\sup_{\hat{N}^C(K)}\frac{\left|R_{2T}^*(k)\right|}{\left|k-\hat{k}\right|}\geq\frac{\lambda}{4}\right) &\leq P^*\left(\sup_{\hat{N}^C(K)}\frac{\left\|\hat{Z}_\Delta' M_\iota\left(u^*-\check{u}^*\right)\right\|}{\left|k-\hat{k}\right|}\geq\frac{\lambda}{16\,\|\delta\|}\right)\\ &\quad+P^*\left(\left\|\hat{\delta}\right\|>2\,\|\delta\|\right),\end{aligned}$$

where the first term on the right is smaller than $\eta/4$ with probability equal or larger than $1-\varepsilon/4$ for large K and T.

Regarding $R_{3T}^*(k)$, Lemma 2.8 implies that $\hat{Z}_\Delta' M_\iota W_k/\left|k-\hat{k}\right|$ is $O_p(1)$ uniformly on $\hat{N}^C(K)$ and $(W_k' M_\iota W_k)^{-1}$ is $O_p(T^{-1})$ uniformly on $\hat{N}^C(K)$. Further, $W_k' M_\iota u^*$ is $O_{p^*}\left(T^{1/2}\right)$ uniformly over $1\leq k\leq T$ by Lemma 3.22. It follows that

$$P^*\left(\sup_{\hat{N}^C(K)}\frac{\left|R_{3T}^*(k)\right|}{\left|k-\hat{k}\right|}\geq\frac{\lambda}{4}\right)<\frac{\eta}{4}$$

with probability no less than $1-\varepsilon/4$ for large T. Finally, since R_{4T}^* is $O_{p^*}(1)$ uniformly on $\hat{N}^C(K)$ by Lemma 3.24, we have

$$P^*\left(\sup_{\hat{N}^C(K)}\frac{\left|R_{4T}^*(k)\right|}{\left|k-\hat{k}\right|}\geq\frac{\lambda}{4}\right)\leq P^*\left(\sup_{\hat{N}^C(K)}\left|R_{4T}^*(k)\right|\geq\frac{\lambda K}{4}\right)<\frac{\eta}{4}$$

with probability larger than $1-\varepsilon/4$ for large K and T. Collecting the results, the bound $\hat{k}^* - \hat{k} = O_{p^*}(1)$ is established. ■

Proof of Proposition 3.1. Write $\hat{k}^* = \hat{k} + \arg\min_{r \in T \cdot \Lambda - \hat{k}} S_T^*(\hat{k}+r)$. Fix $K > 0$ and let

$$\tilde{k}^* = \arg\min_{k \in \hat{N}(K)} S_T^*(k) = \hat{k} + \arg\min_{|r| \leq K} S_T^*\left(\hat{k}+r\right),$$

where $\hat{N}(K)$ is defined in (2.46). We have

$$\begin{aligned} P^*\left(\hat{k}^* - \hat{k} = j\right) &= P^*\left(\hat{k}^* - \hat{k} = j, \left|\hat{k}^* - \hat{k}\right| \leq K\right) \\ &\quad + P^*\left(\hat{k}^* - \hat{k} = j, \left|\hat{k}^* - \hat{k}\right| > K\right). \qquad (3.25) \end{aligned}$$

Since conditionally on data, the event $\left\{\left|\hat{k}^* - \hat{k}\right| \leq K\right\}$ is equivalent to the event $\left\{\hat{k}^* = \tilde{k}^*\right\}$, the first term on the right of (3.25) is equal to

$$P^*\left(\tilde{k}^* - \hat{k} = j\right) - P^*\left(\tilde{k}^* - \hat{k} = j, \left|\hat{k}^* - \hat{k}\right| > K\right).$$

Let $\hat{m} = \arg\min_{m \in N} W^0(m)$ and $\tilde{m} = \arg\min_{|m| \leq K} W^0(m)$. Arguing as above, $P(\hat{m} = j) = P(\tilde{m} = j) - P(\tilde{m} = j, |\hat{m}| > K) + P(\hat{m} = j, |\hat{m}| > K)$. Therefore

$$\begin{aligned} &\left|P^*\left(\hat{k}^* - \hat{k} = j\right) - P(\hat{m} = j)\right| \leq \left|P^*\left(\tilde{k}^* - \hat{k} = j\right) - P(\tilde{m} = j)\right| \\ &+2P^*\left(\left|\hat{k}^* - \hat{k}\right| > K\right) + 2P(|\hat{m}| > K). \qquad (3.26) \end{aligned}$$

By strict stationarity of x_t and u_t, the conditional joint distribution of $\left(Z_{\hat{k}+r} - Z_{\hat{k}}\right)' u^*$ is equal to the conditional joint distribution of $(Z_{k_0+r} - Z_{k_0})' u^*$ for $|r| \leq K$, and the finite dimensional joint distributions of distribution of $\left(Z_{\hat{k}+r} - Z_{\hat{k}}\right)' \left(Z_{\hat{k}+r} - Z_{\hat{k}}\right)$ is equal to the finite dimensional joint distributions of $(Z_{k_0+r} - Z_{k_0})'(Z_{k_0+r} - Z_{k_0})$ for $r \in \mathbb{Z}$. Therefore by Lemma 3.25 and Proposition 3.2,

$$S_T^*\left(\hat{k}+r\right) \xrightarrow{d^*} \delta'(Z_{k_0+r} - Z_{k_0})'(Z_{k_0+r} - Z_{k_0})\delta - 2\delta'^{(Z_{k_0+r} - Z_{k_0})'}u$$

on $|r| \leq K$. The distribution of the right-hand side of the last displayed expression is the same as distribution of the process $W^0(r)$. By assumption, $(\delta' x_t)^2 \pm 2\delta' x_t u_t$ has a continuous distribution. Therefore the process W^0 has a unique minimum,

$$\tilde{k}^* - \hat{k} = \arg\min_{|r|\le K} S_T^*\left(\hat{k}+r\right) \xrightarrow{d^*} \arg\min_{|r|\le K} W^0(r),$$

and the first term on the right of (3.26) is smaller than $\eta/3$ with probability at least $1-\varepsilon/3$ for large T. The second term on the right of (3.26) is smaller than $\eta/3$ with probability no smaller than $1-\varepsilon/3$ for large enough K and T by Proposition 3.3. The third term on the right of (3.26) is bounded by $\eta/3$ for large K because $\hat{m} = O_p(1)$. In sum, $P^*\left(\hat{k}^* - \hat{k} = j\right) \xrightarrow{p} P(\hat{m} = j)$ for each j and so $\hat{k}^* - \hat{k} \xrightarrow{d^*} \arg\min_m W^0(m)$ as required. ■

Proof of Theorem 3.2.

From the bootstrap model (3.10), we have

$$y^* = X\hat{\beta} + Z_{\hat{k}}\hat{\delta} + u^* = W_{\hat{k}^*}' M_\iota \begin{pmatrix} \hat{\beta} \\ \hat{\delta} \end{pmatrix} + \left(Z_{\hat{k}} - Z_{\hat{k}^*}\right)\hat{\delta} + u^*,$$

where $y^* = (y_1, \ldots, y_T)'$ and $u^* = (u_1^*, \ldots, u_T^*)'$. Therefore

$$\begin{aligned} \sqrt{T}\begin{pmatrix} \hat{\beta}^* - \hat{\beta} \\ \hat{\delta}^* - \hat{\delta} \end{pmatrix} &= \left(\frac{1}{T} W_{\hat{k}^*}' M_\iota W_{\hat{k}^*}\right)^{-1} \frac{1}{\sqrt{T}} W_{\hat{k}^*}' M_\iota \left(Z_{\hat{k}} - Z_{\hat{k}^*}\right)\hat{\delta} \\ &\quad + \left(\frac{1}{T} W_{\hat{k}^*}' M_\iota W_{\hat{k}^*}\right)^{-1} \frac{1}{\sqrt{T}} W_{\hat{k}^*}' M_\iota u^*. \end{aligned}$$

Proceeding as in the proof of Proposition 2.8, we obtain

$$\left(\frac{1}{T} W_{\hat{k}^*}' M_\iota W_{\hat{k}^*}\right)^{-1} \xrightarrow{p^*} \begin{pmatrix} 1 & \tau_0 \\ \tau_0 & \tau_0 \end{pmatrix}^{-1} \otimes \Sigma^{-1}$$

and

$$\frac{1}{\sqrt{T}} W_{\hat{k}^*}' M_\iota \left(Z_{\hat{k}} - Z_{\hat{k}^*}\right) = O_{p^*}\left(T^{-\frac{1}{2}}\right).$$

Lemma 3.21 implies that

$$\frac{1}{\sqrt{T}} W_{\hat{k}^*}' M_\iota u^* \xrightarrow{p} \begin{pmatrix} \Omega^{\frac{1}{2}} W(1) \\ \Omega^{\frac{1}{2}} W(\tau) \end{pmatrix}.$$

By Proposition 2.2, $\hat{\delta} = \delta + O_p\left(T^{-1/2}\right)$. The theorem follows from these results and from the continuous mapping theorem. ∎

3.B Lemmas

Let us introduce the following notation:

$$\begin{aligned} \check{h}_l &= \frac{1}{2m+1} \sum_{j=-m}^{m} \left| 2\sin\left(\frac{\lambda_{l+j}}{2}\right) \right|^{2\hat{d}} I_{uu,l+j}, \\ \check{h}_l &= \frac{1}{2m+1} \sum_{j=-m}^{m} \left| 2\sin\left(\frac{\lambda_{l+j}}{2}\right) \right|^{2d} I_{uu,l+j} \\ \tilde{h}_l &= \frac{1}{2m+1} \sum_{j=-m}^{m} \left| 2\sin\left(\frac{\lambda_{l+j}}{2}\right) \right|^{2d} f_{uu,l+j} \end{aligned}$$

for $l = m+1, \ldots, [T/2]$. In what follows, when supremum is taken over values of l, the range is $l = m+1, \ldots, [T/2]$ unless stated otherwise.

Lemma 3.1 *Assume Conditions 1.1-1.5, 2.1 and 3.1-3.4 hold. Then*

$$E\left| \frac{1}{T} \sum_{l=m+1}^{[T/2]} \frac{\check{h}_l - \tilde{h}_l}{h_l} \cos\left(r\lambda_l\right) \right|^2 = O\left(\frac{1}{T}\right).$$

Proof. The proof is a standard extension of Theorem 1 of Hidalgo and Yajima (2002), and it is omitted. ∎

Lemma 3.2 *Assume Conditions 1.1-1.5, 2.1 and 3.1-3.4 hold. Then*
(a) uniformly in $1 \leq r \leq M$,

$$\frac{1}{T} \sum_{l=m+1}^{[T/2]} \frac{\check{h}_l - \tilde{h}_l}{\tilde{h}_l} \cos r\lambda_l = O_p\left(T^{-\frac{1}{2}}\right) \tag{3.27}$$

(b)

$$\frac{1}{T}\sum_{l=m+1}^{[T/2]}\left(\check{h}_l-\tilde{h}_l\right)^2\cos r\lambda_l=O_p\left(m^{-1}\right),$$

and

(c)

$$\sup_l\left|\check{h}_l-\tilde{h}_l\right|=o_p\left(1\right).$$

Proof. (a) This bound follows from Lemma 3.1 and the Markov inequality.

(b) By an obvious extension of Theorem 2.1 of Hidalgo and Robinson (2002), $E\left(\check{h}_l-\tilde{h}_l\right)^2=O\left(m^{-1}\right)$ uniformly over $m+1\leq l\leq[T/2]$, and so $T^{-1}\sum_{l=m+1}^{[T/2]}\left(\check{h}_l-\tilde{h}_l\right)^2=O_p\left(m^{-1}\right)$.

(c) Write

$$\sup_l\left|\check{h}_l-\tilde{h}_l\right|\leq\sup_l\left|h_{1l}\right|+\sup_l\left|h_{2l}\right|+\sup_l\left|h_{3l}\right|,$$

where

$$\begin{aligned}
h_{1l} &= \frac{1}{2m+1}\sum_{j=-m}^{m}h_{l+j}\left(\frac{I_{uu,l+j}}{f_{uu,l+j}}-\frac{2\pi}{\sigma_\varepsilon^2}I_{\varepsilon\varepsilon,l+j}\right),\\
h_{2l} &= \frac{h_l}{2m+1}\sum_{j=-m}^{m}\left(\frac{2\pi}{\sigma_\varepsilon^2}I_{\varepsilon\varepsilon,l+j}-1\right),\\
h_{3l} &= \frac{1}{2m+1}\sum_{j=-m}^{m}\left(h_{l+j}-h_l\right)\left(\frac{2\pi}{\sigma_\varepsilon^2}I_{\varepsilon\varepsilon,l+j}-1\right).
\end{aligned}$$

Noting that $(\sup_l|a_l|)^p\leq\sum_l a_l^p$ for $p\geq0$, we have

$$\left(\sup_l\left|h_{1l}\right|\right)^2\leq\sum_{l=m+1}^{[T/2]}\left(\frac{1}{2m+1}\sum_{j=-m}^{m}h_{l+j}\left(\frac{I_{uu,l+j}}{f_{uu,l+j}}-\frac{2\pi}{\sigma_\varepsilon^2}I_{\varepsilon\varepsilon,l+j}\right)\right)^2. \quad (3.28)$$

Proceeding as with the proof of (4.8) of Robinson (1995b), the expectation of the right-hand side of (3.28) is bounded by

$$\frac{Cm}{(2m+1)^2}\sum_{l=m+1}^{[T/2]}\frac{\log^2(l+m)}{l-m}\leq\frac{C\log^3 T}{m},$$

so that by the Markov inequality and Condition 3.4,

$$\sup_l |h_{1l}| = O_p\left(m^{-1/2}\log^{3/2}T\right) = o_p(1).$$

Next,

$$\left(\sup_l |h_{2l}|\right)^4 \leq \sum_{l=m+1}^{[T/2]} h_l^4\left(\frac{1}{2m+1}\sum_{j=-m}^{m}\left(\frac{2\pi}{\sigma_\varepsilon^2}I_{\varepsilon\varepsilon,l+j}-1\right)\right)^4. \tag{3.29}$$

By the arguments in the proof of Theorem 7.7.4 of Brillinger (1981), the expectation of the right-hand side of (3.29) is bounded by

$$C\sum_{l=m+1}^{[T/2]} h_l^4 m^{-2}\leq CTm^{-2}=o(1)$$

because h is a bounded function. By the Markov inequality, $\sup_l |h_{2l}| = o_p(1)$. Further, by the Hölder inequality,

$$\begin{aligned}\left(\sup_l |h_{3l}|\right)^4 &\leq \sup_{\substack{m+1\leq l\leq[T/2]\\ -m\leq j\leq m}}(h_{l+j}-h_l)^4\sum_{l=m+1}^{[T/2]}\frac{1}{2m+1}\sum_{j=-m}^{m}\left(\frac{2\pi}{\sigma_\varepsilon^2}I_{\varepsilon\varepsilon,l+j}-1\right)^4\\ &= O\left(\left(\frac{m}{T}\right)^4\right)O_p(T)=O_p\left(\frac{m^4}{T^3}\right)=o_p(1)\end{aligned}$$

because $h_{l+j}=h_l+O(m/T)$ uniformly in l and j and because ε_t is an i.i.d. sequence with finite eighth moments. In sum,

$$\sup_l\left|\check{h}_l-\tilde{h}_l\right|=o_p(1)$$

as required. ■

Lemma 3.3 *For* $l = m+1, \ldots, [T/2]$,

$$\frac{1}{2m+1} \sum_{j=-m}^{m} \lambda_{l+j}^{\delta} \leq \begin{cases} \lambda_l^{\delta} & 0 \leq \delta < 1, \\ \lambda_{l-m}^{\delta} & \delta < 0. \end{cases}$$

Proof. When $0 \leq \delta < 1$, the function λ^{δ} concave and so $\left(\lambda_{l-j}^{\delta} + \lambda_{l+j}^{\delta}\right)/2 \leq \lambda_l^{\delta}$ for $-m \leq j \leq m$. The bound easily follows. When $\delta < 0$, convexity of the function λ^{δ} implies that

$$\frac{1}{2m+1} \sum_{j=-m}^{m} \lambda_{l+j}^{\delta} \leq \frac{1}{2}\left(\lambda_{l-m}^{\delta} + \lambda_{l+m}^{\delta}\right) \leq \lambda_{l-m}^{\delta}.$$

■

Lemma 3.4 *Assume Conditions 1.1-1.5, 2.1 and 3.1-3.4 hold. Then*
(a) uniformly in $1 \leq r \leq M$,

$$\frac{1}{T} \sum_{l=m+1}^{[T/2]} \frac{\breve{h}_l - \check{h}_l}{\check{h}_l} \cos r\lambda_l = O_p\left(m^{-\frac{1}{2}} \frac{\log r}{r}\right) + O_p\left(m^{\frac{3}{2}} \frac{\log^2 T}{T^2}\right), \tag{3.30}$$

(b)

$$\frac{1}{T} \sum_{l=m+1}^{[T/2]} \left(\breve{h}_l - \check{h}_l\right)^2 = O_p\left(m^{-1}\right) \tag{3.31}$$

and

(c)

$$\sup_l \left|\breve{h}_l - \check{h}_l\right| = o_p(1).$$

Proof. Let us define

$$\phi_j = \left(\hat{d} - d\right) \log \left|2 \sin\left(\frac{\lambda_j}{2}\right)\right|^2$$

and

$$q_{pl} = \frac{2\pi}{\sigma_\varepsilon^2} \frac{1}{2m+1} \sum_{j=-m}^{m} \frac{\phi_{l+j}^p}{p!} \frac{h_{l+j}}{h_l} \left(I_{\varepsilon\varepsilon, l+j} - \frac{\sigma_\varepsilon^2}{2\pi}\right).$$

for $p \geq 1$. We first observe that $\tilde{h}_l = (2m+1)^{-1} \sum_{j=-m}^{m} h_{l+j}$, so that by the Taylor theorem, $\tilde{h}_l = h_l + O_p(M^{-2})$ uniformly in $m+1 \leq l \leq [T/2]$.

(a) By the Taylor theorem,

$$\begin{aligned}
& \breve{h}_l - \check{h}_l \\
= & \frac{1}{2m+1} \sum_{j=-m}^{m} \left| 2\sin\left(\frac{\lambda_{l+j}}{2}\right) \right|^{2d} \left(\phi_{l+j} + \frac{1}{2}\phi_{l+j}^2 \right) I_{uu,l+j} \\
& + \frac{8\left(\hat{d}-d\right)^3}{2m+1} \sum_{j=-m}^{m} \left| 2\sin\left(\frac{\lambda_{l+j}}{2}\right) \right|^{2\tilde{d}} \log^3 \left| 2\sin\left(\frac{\lambda_{l+j}}{2}\right) \right| I_{uu,l+j}, \quad (3.32)
\end{aligned}$$

where $\tilde{d}$ is an intermediate point between d and $\hat{d}$. The first term on the right of (3.32) is

$$\begin{aligned}
& h_l \left(q_{1l} + q_{2l}\right) + \frac{1}{2m+1} \sum_{j=-m}^{m} \left(\phi_{l+j} + \frac{1}{2}\phi_{l+j}^2 \right) h_{l+j} \left(\frac{I_{uu,l+j}}{f_{uu,l+j}} - \frac{2\pi}{\sigma_\varepsilon^2} I_{\varepsilon\varepsilon,l+j} \right) \\
& + \frac{1}{2m+1} \sum_{j=-m}^{m} \left(\phi_{l+j} + \frac{1}{2}\phi_{l+j}^2 \right) h_{l+j}. \qquad (3.33)
\end{aligned}$$

The contribution of the first term of (3.33) into the left-hand side of (3.30) is bounded in absolute value by

$$\left(\sup_l \left| \frac{h_l}{\tilde{h}_l} \right| \right) \frac{1}{T} \sum_{l=m+1}^{[T/2]} \left| q_{1l} + q_{2l} \right| = O_p\left(m^{-1}\right)$$

because h is bounded and bounded away from zero, $\tilde{h}_l = h_l + O\left(M^{-2}\right)$ uniformly in l, $\hat{d} - d = O_p\left(m^{-1/2}\right)$ and because

$$\begin{aligned}
E\left(\frac{q_{pl}^2}{\left(\hat{d}-d\right)^{2p}} \right) & \leq \frac{C}{m^2} E \left| \sum_{j=-m}^{m} \log^p \left(2\sin\left(\frac{\lambda_j}{2}\right) \right) \frac{h_{l+j}}{h_l} \left(I_{\varepsilon\varepsilon,l+j} - \frac{\sigma_\varepsilon^2}{2\pi} \right) \right|^2 \\
& \leq Cm^{1/2}.
\end{aligned}$$

By the Cauchy-Schwarz inequality, the square of the contribution of the second term of (3.33) into the left-hand side of (3.30) is bounded by

$$\frac{C}{T}\sum_{l=m+1}^{[T/2]}\left|\frac{1}{2m+1}\sum_{j=-m}^{m}\left(\phi_{l+j}+\frac{1}{2}\phi_{l+j}^2\right)h_{l+j}\left(\frac{I_{uu,l+j}}{f_{uu,l+j}}-\frac{2\pi}{\sigma_\varepsilon^2}I_{\varepsilon\varepsilon,l+j}\right)\right|^2. \quad (3.34)$$

Now for $p \geq 1$,

$$\begin{aligned} & E\left(\left(\hat{d}-d\right)^{-2p}\left|\frac{1}{2m+1}\sum_{j=-m}^{m}\frac{\phi_{l+j}^p}{p!}h_{l+j}\left(\frac{I_{uu,l+j}}{f_{uu,l+j}}-\frac{2\pi}{\sigma_\varepsilon^2}I_{\varepsilon\varepsilon,l+j}\right)\right|^2\right) \\ \leq\ & \frac{C}{2m+1}\sum_{j=-m}^{m}\log^{2p}\left(2\sin\frac{\lambda_{l+j}}{2}\right)E\left(\frac{I_{uu,l+j}}{f_{uu,l+j}}-\frac{2\pi}{\sigma_\varepsilon^2}I_{\varepsilon\varepsilon,l+j}\right)^2. \quad (3.35)\end{aligned}$$

Proceeding as with the proof of expression (4.8) of Robinson (1995b), we obtain

$$E\left(\frac{I_{uu,l+j}}{f_{uu,l+j}}-\frac{2\pi}{\sigma_\varepsilon^2}I_{\varepsilon\varepsilon,l+j}\right)^2 \leq C\frac{\log(l+j)}{l+j},$$

where the constant C does not depend on l and j. Since $\log^{2p}\left(2\sin\frac{\lambda_{l+j}}{2}\right) \leq C\log^{2p}T$, the right-hand side of (3.35) is bounded by $C(l-m)^{-1}\log(l-m) \times \log^{2p}T$. By the Cauchy-Schwarz and Markov inequalities, (3.34) is

$$\begin{aligned} & \sum_{p=1}^{2}\left(\hat{d}-d\right)^{2p}O_p\left(\frac{\log^{2p}T}{T}\sum_{l=m+1}^{[T/2]}\frac{\log(l-m)}{l-m}\right) \\ =\ & O_p\left(m^{-1}\frac{\log^4 T}{T}+m^{-2}\frac{\log^6 T}{T}\right)=O_p\left(m^{-1}\frac{\log^4 T}{T}\right)\end{aligned}$$

and the contribution of the second term of (3.33) into the left-hand side of (3.30) is $O_p\left(m^{-1/2}T^{-1/2}\log^2 T\right)$.

Regarding the third term of (3.33), we have by standard arguments that

$$\begin{aligned}\frac{1}{2m+1}\sum_{j=-m}^{m}\log^p\left|2\sin\left(\frac{\lambda_{l+j}}{2}\right)\right|h_{l+j} &= \log^p\left|2\sin\left(\frac{\lambda_l}{2}\right)\right|h_l \\ &\quad +O\left(\frac{m^2}{l-m}\frac{\log^{p-1}T}{T}\right)\end{aligned}$$

uniformly in $m+1 \leq l \leq [T/2]$ for $p \geq 1$, and so the contribution of the third term of (3.33) into the left-hand side of (3.30) is

$$\frac{1}{T}\sum_{l=m+1}^{[T/2]}\frac{h_l}{\tilde{h}_l}\left(\phi_l+\frac{\phi_l^2}{2}\right)\cos r\lambda_l$$
$$+\left(\left(\hat{d}-d\right)+\left(\hat{d}-d\right)^2\right)O\left(\frac{1}{T}\sum_{l=m+1}^{[T/2]}\frac{m^2}{l-m}\frac{\log T}{T}\right). \quad (3.36)$$

The first term of (3.36) can be written as

$$\frac{1}{T}\sum_{l=m+1}^{[T/2]}\left(\phi_l+\frac{\phi_l^2}{2}\right)\cos r\lambda_l+\frac{1}{T}\sum_{l=m+1}^{[T/2]}\left(\phi_l+\frac{\phi_l^2}{2}\right)\left(\frac{h_l}{\tilde{h}_l}-1\right)\cos r\lambda_l. \quad (3.37)$$

Since for $p \geq 1$,

$$\frac{2}{T}\sum_{l=1}^{T/2}\log^p\left|2\sin\left(\frac{\lambda_l}{2}\right)\right|\cos\left(r\lambda_l\right)=o\left(r^{-1}\log r\right)$$

by Theorem III-23 of Yong (1974), the first term of (3.37) is $O_p\left(m^{-1/2}\right)$ $o\left(r^{-1}\log r\right)$. Also, since $\log^p\left|2\sin\left(\frac{\lambda_l}{2}\right)\right|$ is absolutely integrable for $p \geq 1$ and $\sup_l\left|h_l/\tilde{h}_l-1\right| = O_p\left(M^{-2}\right)$, the second term of (3.37) is $O_p\left(m^{-1/2}M^{-2}\right)$ which is $o_p\left(m^{-1}\right)$ by Condition 3.4. The second term of (3.36) is $O_p\left(m^{-1/2}m^2T^{-2}\log^2 T\right)$ and so the total contribution of the third term of (3.33) into the left-hand side of (3.30) is

$$\begin{aligned} & o_p\left(m^{-1/2}r^{-1}\log r\right)+O_p\left(m^{-1/2}M^{-2}\right)+O_p\left(m^{3/2}\frac{\log^2 T}{T^2}\right) \\ = \; & o_p\left(m^{-1/2}r^{-1}\log r\right)+O_p\left(m^{3/2}\frac{\log^2 T}{T^2}\right) \end{aligned}$$

by Condition 3.4.

Finally, the contribution of the second term on the right of (3.32) into the left-hand side of (3.30) is bounded in absolute value by

$$C\left|\hat{d}-d\right|^3 \frac{\log^3 T}{T}\sum_{l=m+1}^{[T/2]}\left(\frac{1}{2m+1}\sum_{j=-m}^{m} I_{uu,l+j}\right). \tag{3.38}$$

Employing the arguments from the proof of Proposition 3.1 and applying Lemma 3.3, we obtain

$$E\frac{1}{2m+1}\sum_{j=-m}^{m} I_{uu,l+j} \leq \frac{D}{2m+1}\sum_{j=-m}^{m}\lambda_{l+j}^{-2d} \leq D\left(\frac{l-m}{T}\right)^{-2d}.$$

By the Markov inequality, expression (3.38) is

$$O_p\left(m^{-3/2}\log^3 T\right) O_p\left(\frac{1}{T}\sum_{l=m+1}^{[T/2]}\left(\frac{l-m}{T}\right)^{-2d}\right) = O_p\left(m^{-3/2}\log^3 T\right).$$

Collecting the results and applying Condition 3.4, we arrive at the bound

$$\frac{1}{T}\sum_{l=m+1}^{[T/2]}\frac{\check{h}_l-\tilde{h}_l}{\tilde{h}_l}\cos r\lambda_l = O_p\left(m^{-1/2}\frac{\log^2 r}{r}\right) + O_p\left(m^{3/2}\frac{\log^2 T}{T^2}\right).$$

(b) By (3.32), (3.33) and the Cauchy-Schwarz inequality, the left-hand side of (3.31) is bounded by

$$\begin{aligned}
&\frac{2}{T}\sum_{l=m+1}^{[T/2]}\left(q_{1l}^2+q_{2l}^2\right)\\
&+\frac{1}{T}\sum_{l=m+1}^{[T/2]}\left(\frac{1}{2m+1}\sum_{j=-m}^{m}\left(\phi_l+\frac{\phi_l^2}{2}\right)h_{l+j}\left(\frac{I_{uu,l+j}}{f_{uu,l+j}}-\frac{2\pi}{\sigma_\varepsilon^2}I_{\varepsilon\varepsilon,l+j}\right)^2\right)\\
&+\frac{C}{T}\sum_{l=m+1}^{[T/2]}\frac{1}{2m+1}\sum_{j=-m}^{m}\left(\phi_l^2+\phi_l^4\right)\\
&+C\left(\hat{d}-d\right)^6\log^6 T\frac{1}{T}\sum_{l=m+1}^{[T/2]}\frac{1}{2m+1}\sum_{j=-m}^{m} I_{uu,l+j}^2.
\end{aligned} \tag{3.39}$$

The first two terms of (3.39) are $O_p(m^{-2})$ and $O_p(m^{-1}T^{-1}\log^4 T)$, respectively, by the arguments employed in the proof of part (a). The third term of (3.39) is bounded by $CT^{-1}\sum_{l=m+1}^{[T/2]}(\phi_j^2+\phi_j^4)$ which is $O_p(m^{-1})$ by Proposition 3.1 and by the integrability of the function $\log^{2p}|2\sin(\lambda_l/2)|$ for $p \geq 1$. We have

$$\frac{1}{2m+1}\sum_{j=-m}^{m} I_{uu,l+j}^2 = \frac{1}{2m+1}\sum_{j=-m}^{m} f_{uu,l+j}^2 \frac{I_{uu,l+j}^2}{f_{uu,l+j}^2}.$$

By the arguments in the proof of expression (4.8) of Robinson (1995b), the expectation of the factor $I_{uux,l+j}^2/f_{uu,l+j}^2$ is $O(1)$ uniformly in integers $1 \leq j \leq [T/2]$. Moreover, $f_{uu}(\lambda) \leq C\lambda^{-2d}$ for $0 < \lambda \leq \pi$ by Condition 1.4, therefore by Lemma 3.3,

$$\begin{aligned} E\frac{1}{2m+1}\sum_{j=-m}^{m} I_{uu,l+j}^2 &= \frac{1}{2m+1}\sum_{j=-m}^{m} f_{uu,l+j}^2 E\frac{I_{uu,l+j}^2}{f_{uu,l+j}^2} \\ &\leq \frac{C}{2m+1}\sum_{j=-m}^{m} f_{uu,l+j}^2 \\ &\leq \frac{C}{2m+1}\sum_{j=-m}^{m} \lambda_{l+j}^{-4d} \leq C\lambda_{l-m}^{-4d} \qquad (3.40) \end{aligned}$$

and so $(2m+1)^{-1}E\sum_{j=-m}^{m} I_{uu,l+j}^2$ is $O\left(((l-m)/T)^{-4d}\right)$ uniformly over $l = m+1,\ldots,[T/2]$. Discussion of the sum $T^{-1}\sum_{l=m+1}^{[T/2]}((l-m)/T)^{-4d}$ for values $d \in [0,1/4)$, $d = 1/4$ and $d \in (1/4,1/2)$ together with the Markov inequality leads to the conclusion that the fourth term of (3.39) is $O_p(m^{-3}\log^7 T)$. Employing Condition 3.4, we conclude that (3.39) is $O_p(m^{-1})$.

(c) Noting that $\sup_l|a_l| \leq \sum_l |a_l|$ and $\sup_l|a_l| = \left(\sup_l|a_l|^2\right)^{1/2}$ and using (3.32), (3.33) and the triangle and Cauchy-Schwarz inequalities, we obtain that $\sup_l \left|\breve{h}_l - \check{h}_l\right|$ is bounded by

$$C\left(\sum_{l=m+1}^{[T/2]}\left(q_{1l}^2+q_{2l}^2\right)\right)^{\frac{1}{2}} + C\sup_l\left|\phi_l+\phi_l^2\right|$$
$$+\left(\sum_{l=m+1}^{[T/2]}\left(\frac{1}{2m+1}\sum_{j=-m}^{m}\left(\phi_l+\frac{\phi_l^2}{2}\right)h_{l+j}\left(\frac{I_{uu,l+j}}{f_{uu,l+j}}-\frac{2\pi}{\sigma_\varepsilon^2}I_{\varepsilon\varepsilon,l+j}\right)\right)^2\right)^{\frac{1}{2}}$$
$$+C\left(\left(\hat{d}-d\right)^6\log^6 T\sum_{l=m+1}^{[T/2]}\frac{1}{2m+1}\sum_{j=-m}^{m}I_{uu,l+j}^2\right)^{\frac{1}{2}}. \tag{3.41}$$

By the arguments employed in parts (a) and (b), the first, third and fourth term of (3.41) is $O_p\left(T^{1/2}m^{-1}\right)$, $O_p\left(m^{-1/2}\log^2 T\right)$ and $O_p\left(T^{1/2}m^{-3/2}\log^{7/2}T\right)$, respectively. Further, it is obvious that by Proposition 3.1, $\sup_l\left|\phi_l+\phi_l^2\right| = O_p\left(m^{-\frac{1}{2}}\log T\right)$. Condition 3.4 implies that $\sup_l\left|\breve{h}_l-\check{h}_l\right| = o_p(1)$. ■

Lemma 3.5 *Assume Conditions 1.1-1.5, 2.1 and 3.1-3.4 hold. Then for $l = m+1,\ldots,[T/2]$,*
(a)
$$\frac{1}{2m+1}\sum_{j=-m}^{m}\left|I_{\hat{u}\hat{u},l+j}-I_{uu,l+j}\right| = O_p\left(T^{-\frac{1}{2}}\right)r_l,$$

(b)
$$\frac{1}{2m+1}\sum_{j=-m}^{m}\left(I_{\hat{u}\hat{u},l+j}-I_{uu,l+j}\right)^2 = O_p\left(T^{-1}\right)s_l,$$

where the O_p terms are uniform over $l = m+1,\ldots,[T/2]$ and where

$$E\left|r_l\right| \le D\left(\frac{l-m}{T}\right)^{-(d_x+d)} \quad \text{and} \quad E\left|s_l\right| \le D\left(\frac{l-m}{T}\right)^{-2(d_x+d)},$$

with a constant D that does not depend on l and T.

Proof. (a) By the arguments employed in the proof of Proposition 3.1, it is sufficient to examine the behaviour of

$$\frac{1}{2m+1}\sum_{j=-m}^{m} I_{\tilde{u}\tilde{u},l+j} + \left(\frac{1}{2m+1}\sum_{j=-m}^{m} I_{\tilde{u}\tilde{u},l+j}\right)^{\frac{1}{2}}\left(\frac{1}{2m+1}\sum_{j=-m}^{m} I_{uu,l+j}\right)^{\frac{1}{2}}. \tag{3.42}$$

By (3.15) and the Cauchy-Schwarz inequality, the first term of (3.42) is bounded by

$$\frac{3}{2m+1}\sum_{j=-m}^{m}\left(\left(\beta-\hat{\beta}\right)^2 I_{xx,l+j} + \left(\delta-\hat{\delta}\right)^2 I_{\hat{z}\hat{z},l+j} + \delta^2 I_{\Delta\hat{z}\Delta\hat{z},l+j}\right).$$

Employing the arguments from the proof of Lemma 3.4, we obtain

$$E\frac{1}{2m+1}\sum_{j=-m}^{m} I_{\hat{z}\hat{z},l+j} \leq \frac{D}{2m+1}\sum_{j=-m}^{m}\lambda_{l+j}^{-2d_x} \leq D\left(\frac{l-m}{T}\right)^{-2d_x}$$

The same bound applies for $E\left(2m+1\right)^{-1}\sum_{j=-m}^{m} I_{xx,l+j}$. In a similar way, $E\left(2m+1\right)^{-1}\sum_{j=-m}^{m} I_{uu,l+j} \leq D\left(\left(l-m\right)/T\right)^{-2d}$. The fact that $\left(\beta-\hat{\beta}\right)^2 = O_p\left(T^{-1}\right)$ and $\left(\delta-\hat{\delta}\right)^2 = O_p\left(T^{-1}\right)$ together with bound (3.17) and the Cauchy-Schwarz inequality now imply that (3.42) is bounded by $O_p\left(T^{-\frac{1}{2}}\right) r_l$ where $E\left|r_l\right| \leq D\left(\left(l-m\right)/T\right)^{-d_x-d}$.

(b) The Cauchy-Schwarz inequality indicates that we need to investigate the stochastic magnitude of

$$\frac{1}{2m+1}\sum_{j=-m}^{m} I^2_{\tilde{u}\tilde{u},l+j} + \left(\frac{1}{2m+1}\sum_{j=-m}^{m} I^2_{\tilde{u}\tilde{u},l+j}\right)^{\frac{1}{2}}\left(\frac{1}{2m+1}\sum_{j=-m}^{m} I^2_{uu,l+j}\right)^{\frac{1}{2}}. \tag{3.43}$$

Another application of the Cauchy-Schwarz inequality bounds the first term of (3.43) by

$$\frac{C\left(\hat{\beta}-\beta\right)^4}{2m+1}\sum_{j=-m}^{m} I^2_{xx,l+j} + \frac{C\left(\hat{\delta}-\delta\right)^4}{2m+1}\sum_{j=-m}^{m} I^2_{\hat{z}\hat{z},l+j} + \frac{C\delta^4}{2m+1}\sum_{j=-m}^{m} I^2_{\Delta\hat{z}\Delta\hat{z},l+j}.$$

From (3.40), we have $(2m+1)^{-1} E \sum_{j=-m}^{m} I_{uu,l+j}^2 \leq C\lambda_{l-m}^{-4d}$. In a similar way, it can be shown that $(2m+1)^{-1} E \sum_{j=-m}^{m} I_{xx,l+j}^2 \leq C\lambda_{l-m}^{-4d_x}$. Furthermore, the arguments used in the proof of Lemma 1.11 indicate that for $z_t(k) = z_t \mathbb{I}(t \leq k)$, $E\left(I_{zz,j}^2 / f_{xx,j}^2\right) = O(1)$ uniformly in $1 \leq j \leq [T/2]$ and $1 \leq k \leq T$. Therefore we have also

$$E \frac{1}{2m+1} \sum_{j=-m}^{m} I_{\hat{z}\hat{z},l+j}^2 \leq C\lambda_{l-m}^{-4d_x}.$$

The conclusion now follows in the same way as in the part (a). ■

Lemma 3.6 *Assume Conditions 1.1-1.5, 2.1 and 3.1-3.4 hold. Then (a) uniformly in* $1 \leq r \leq M$*,*

$$\frac{1}{T} \sum_{l=m+1}^{[T/2]} \frac{\hat{h}_l - \check{h}_l}{\tilde{h}_l} \cos r\lambda_l = O_p\left(T^{-\frac{1}{2}}\right),$$

(b)

$$\frac{1}{T} \sum_{l=m+1}^{[T/2]} \left(\hat{h}_l - \check{h}_l\right)^2 = O_p\left(T^{-1}\right),$$

(c)

$$\sup_l \left|\hat{h}_l - \check{h}_l\right| = o_p(1).$$

Proof. (a) Since $\sup_l \left|\tilde{h}_l\right|^{-1} = O(1)$, we have by standard inequalities that

$$\left|\frac{1}{T} \sum_{l=m+1}^{[T/2]} \frac{\hat{h}_l - \check{h}_l}{\tilde{h}_l} \cos r\lambda_l\right| \leq \frac{C}{T} \sum_{l=m+1}^{[T/2]} \frac{1}{2m+1} \sum_{j=-m}^{m} \left|I_{\hat{u}\hat{u},l+j} - I_{uu,l+j}\right|.$$

By Lemma 3.5 part (a), the right-hand side of the last displayed inequality is $O_p\left(T^{-\frac{1}{2}}\right) T^{-1} \sum_{l=m+1}^{[T/2]} r_l$, where

$$E \frac{1}{T} \sum_{l=m+1}^{[T/2]} r_l \leq \frac{D}{T} \sum_{l=m+1}^{[T/2]} \left(\frac{l-m}{T}\right)^{-(d_x+d)} \leq C.$$

The conclusion follows from the Markov inequality.

(b) Using part (b) of Lemma 3.5, the proof follows by arguments similar to part (a).

(c) For $l = m+1, \ldots, 2m$, we have

$$\left|\hat{h}_l - \breve{h}_l\right| \leq \frac{1}{2m+1} \sum_{j=-m}^{m} \left|I_{\hat{u}\hat{u},l+j} - I_{uu,l+j}\right| \leq \frac{1}{2m+1} \sum_{j=1}^{3m} \left|I_{\hat{u}\hat{u},j} - I_{uu,j}\right|.$$

From (3.14), we obtain that $\sup_{m+1 \leq l \leq 2m} \left|\hat{h}_l - \breve{h}_l\right| = o_p\left(m^{-1/2}\right)$.

For $l = 2m+1, \ldots, [T/2]$, write $l = 2mp + k$ for some $1 \leq p \leq M$ and $1 \leq k \leq 2m$. We have

$$\left|\hat{h}_{2mp+k} - \breve{h}_{2mp+k}\right| \leq \frac{1}{2m+1} \sum_{j=-m}^{m} \left|I_{\hat{u}\hat{u},2mp+k+j} - I_{uu,2mp+k+j}\right|$$

$$\begin{aligned} &= \frac{1}{2m+1} \sum_{j=-m+k}^{m} \left|I_{\hat{u}\hat{u},2mp+j} - I_{uu,2mp+j}\right| \\ &\quad + \frac{1}{2m+1} \sum_{j=m+1}^{m+k} \left|I_{\hat{u}\hat{u},2mp+j} - I_{uu,2mp+j}\right| \end{aligned}$$

$$\begin{aligned} &\leq \frac{1}{2m+1} \sum_{j=-m}^{m} \left|I_{\hat{u}\hat{u},2mp+j} - I_{uu,2mp+j}\right| \\ &\quad + \frac{1}{2m+1} \sum_{j=-m}^{m} \left|I_{\hat{u}\hat{u},2m(p+1)+j} - I_{uu,2m(p+1)+j}\right|. \end{aligned}$$

Therefore for $1 \leq p \leq M$,

$$\begin{aligned} \sup_{1 \leq k \leq 2m} \left|\hat{h}_{2mp+k} - \breve{h}_{2mp+k}\right| &\leq \frac{1}{2m+1} \sum_{j=-m}^{m} \left|I_{\hat{u}\hat{u},2mp+j} - I_{uu,2mp+j}\right| \\ &\quad + \frac{1}{2m+1} \sum_{j=-m}^{m} \left|I_{\hat{u}\hat{u},2m(p+1)+j} - I_{uu,2m(p+1)+j}\right| \end{aligned}$$

and

$$\begin{aligned}
\sup_{2m+1\le l\le[T/2]}\left|\hat{h}_l-\check{h}_l\right| &\le \sup_{1\le p\le M}\sup_{1\le k\le 2m}\left|\hat{h}_{2mp+k}-\check{h}_{2mp+k}\right| \\
&\le \sum_{p=1}^{M}\sup_{1\le k\le 2m}\left|\hat{h}_{2mp+k}-\check{h}_{2mp+k}\right| \\
&\le 2\sum_{p=1}^{M+1}\frac{1}{2m+1}\sum_{j=-m}^{m}\left|I_{\hat{u}\hat{u},2mp+j}-I_{uu,2mp+j}\right|
\end{aligned}$$

which by Lemma 3.5 is $O_p\left(T^{-1/2}\right)\sum_{p=1}^{M} r_{2mp}$ where

$$E\left|\sum_{p=1}^{M} r_{2mp}\right| \le D\sum_{p=1}^{M}\left(\frac{2mp-m}{T}\right)^{-(d_x+d)} = O\left(Tm^{-1}\right)$$

so that by the Markov inequality and Condition 3.4,

$$\sup_l\left|\hat{h}_l-\check{h}_l\right| = o_p\left(m^{-1/2}\right)+O_p\left(T^{-1/2}\right)O_p\left(Tm^{-1}\right) = O_p\left(T^{1/2}m^{-1}\right) = o_p\left(1\right).$$

■

Lemma 3.7 *Let φ be a piecewise twice continuously differentiable function on $[0,2\pi]$ with $\varphi(0)=\varphi(2\pi)$. Then uniformly in $s\neq 0$,*

(a)

$$\int_0^{2\pi}\varphi(\lambda)\cos s\lambda d\lambda = \begin{cases} O\left(s^{-1}\right) & s \text{ real,} \\ O\left(s^{-2}\right) & s \text{ integer,}\end{cases}$$

(b)

$$\int_0^{2\pi}\varphi(\lambda)\sin s\lambda d\lambda = \begin{cases} O\left(s^{-1}\right) & s \text{ real,} \\ O\left(s^{-2}\right) & s \text{ integer.}\end{cases}$$

Proof. Let $t_1<\ldots<t_{p-1}$, where $t_1>t_0=0$ and $t_{p-1}<t_p=2\pi$, be the points of discontinuity of φ.

(a) Integrating twice by parts, we obtain

$$\begin{aligned}\int_0^{2\pi} \varphi(\lambda)\cos s\lambda d\lambda &= \sum_{j=1}^{p} \int_{t_{j-1}}^{t_j} \varphi(\lambda)\cos s\lambda d\lambda \\ &= \sum_{j=1}^{p} \left(\frac{1}{s} [\varphi(\lambda)\sin s\lambda]_{t_{j-1}}^{t_j} - \frac{1}{s}\int_{t_{j-1}}^{t_j} \varphi'(\lambda)\sin s\lambda d\lambda \right)\end{aligned}$$

$$\begin{aligned} &= \frac{1}{s}(\varphi(2\pi)\sin 2\pi s - \varphi(0)\sin 0) \\ &\quad -\frac{1}{s}\sum_{j=1}^{p}\left(\frac{1}{s}[-\varphi'(\lambda)\cos s\lambda]_{t_{j-1}}^{t_j} + \frac{1}{s}\int_{t_{j-1}}^{t_j}\varphi''(\lambda)\cos s\lambda d\lambda\right) \\ &= \frac{1}{s}(\varphi(2\pi)\sin 2\pi s - \varphi(0)\sin 0) + \frac{1}{s^2}(\varphi'(2\pi)\cos 2\pi s - \varphi'(0)\cos 0) \\ &\quad -\frac{1}{s^2}\int_0^{2\pi}\varphi''(\lambda)\cos s\lambda d\lambda.\end{aligned}$$

From here the bound follows easily.

(b) The proof is similar to the proof of part (a) and is therefore omitted.

■

Lemma 3.8 *Let φ be a function on $[-\pi,\pi]$ that is symmetric around zero and twice continuously differentiable on $(0,\pi)$ with one-sided second derivatives at 0 and π. Then its Fourier coefficients $v_r = \frac{1}{2\pi}\int_{-\pi}^{\pi}\varphi(\lambda)e^{-i\lambda r}d\lambda$ satisfy*

$$r^2 |v_r| \le C$$

for $r \ge 0$.

Proof. Let v_r'' be the Fourier coefficient of φ'', that is

$$v_r'' = \frac{1}{2\pi}\int_{-\pi}^{\pi}\varphi''(\lambda)e^{-ir\lambda}d\lambda.$$

A simple algebra gives

$$v_r'' = \frac{1}{\pi}\varphi'(\pi)\cos(r\pi) - r^2 v_r. \tag{3.44}$$

By the Bessel inequality, see for example Zygmund (2002, p. 13),

$$\sum_{r=0}^{\infty} |v_r''|^2 \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} |\varphi''(\lambda)|^2 \, d\lambda < \infty$$

because φ'' is bounded and piecewise continuous. Therefore $v_r'' \to 0$ and (3.44) implies that $|r^2 v_r| \leq C$. ■

Let us define

$$\begin{aligned} \tilde{c}_r &= \frac{1}{T} \sum_{l=m+1}^{[T/2]} \log \tilde{h}_l \cos r\lambda_l, \qquad c_{r,T} = \frac{1}{T} \sum_{l=m+1}^{[T/2]} \log h_l \cos r\lambda_l \quad \text{and} \\ c_r &= \frac{1}{4\pi} \int_{-\pi}^{\pi} \log h(\lambda) \, e^{-ir\lambda} d\lambda = \frac{1}{2\pi} \int_{0}^{\pi} \log h(\lambda) \cos r\lambda d\lambda. \end{aligned}$$

Lemma 3.9 *Under Conditions 1.1-1.5, 2.1 and 3.1-3.4, uniformly in* $1 \leq r \leq M$,
(a) $\hat{c}_r - \tilde{c}_r = O_p\left(T^{-\frac{1}{2}} + m^{-\frac{1}{2}} r^{-1} \log r\right)$,
(b) $\tilde{c}_r - c_{r,T} = O_p(M^{-2})$,
(c) $c_{r,T} - c_r = O_p(T^{-1} r^{-1})$,
(d) $\sum_{r=M+1}^{\infty} c_r e^{ir\lambda_j} = o(M^{-1})$.

Proof. (a) By the Taylor theorem,

$$\begin{aligned} \hat{c}_r - \tilde{c}_r &= \frac{1}{T} \sum_{l=m+1}^{[T/2]} \left(\log \hat{h}_l - \log \tilde{h}_l\right) \cos r\lambda_l \\ &= \frac{1}{T} \sum_{l=m+1}^{[T/2]} \frac{\hat{h}_l - \tilde{h}_l}{\tilde{h}_l} \cos r\lambda_l + \frac{1}{T} \sum_{l=m+1}^{[T/2]} \frac{\left(\hat{h}_l - \tilde{h}_l\right)^2}{\eta_l^2} \cos r\lambda_l, \end{aligned}$$

where η_l is an intermediate point between $\hat{h}_l$ and $\tilde{h}_l$. The right-hand side of the displayed equality is equal to

$$\begin{aligned} &\frac{1}{T} \sum_{l=m+1}^{[T/2]} \frac{\hat{h}_l - \breve{h}_l}{\tilde{h}_l} \cos r\lambda_l + \frac{1}{T} \sum_{l=m+1}^{[T/2]} \frac{\breve{h}_l - \check{h}_l}{\tilde{h}_l} \cos r\lambda_l \\ &+ \frac{1}{T} \sum_{l=m+1}^{[T/2]} \frac{\check{h}_l - \tilde{h}_l}{\tilde{h}_l} \cos r\lambda_l + \frac{1}{T} \sum_{l=m+1}^{[T/2]} \frac{\left(\hat{h}_l - \tilde{h}_l\right)^2}{\eta_l^2} \cos r\lambda_l. \end{aligned} \tag{3.45}$$

By part (a) of Lemmas 3.2, 3.4 and 3.6, the first three terms of (3.45) are $O_p\left(T^{-1/2}+m^{-1}+m^{-1/2}r^{-1}\log r+m^{3/2}T^{-2}\log^2 T\right)$ which by Condition 3.4 is $O_p\left(T^{-\frac{1}{2}}+m^{-\frac{1}{2}}r^{-1}\log^2 r\right)$. By the Cauchy-Schwarz inequality, the fourth term of (3.45) is bounded by

$$\left(\sup_l \eta_l^{-2}\right)\frac{C}{T}\sum_{l=m+1}^{[T/2]}\left(\left(\hat{h}_l-\check{h}_l\right)^2+\left(\check{h}_l-\check{h}_l\right)^2+\left(\check{h}_l-\tilde{h}_l\right)^2\right).$$

By part (c) of Lemmas 3.2, 3.4 and 3.6,

$$\sup_l\left|\hat{h}_l-\tilde{h}_l\right|\leq\sup_l\left|\hat{h}_l-\check{h}_l\right|+\sup_l\left|\check{h}_l-\check{h}_l\right|+\sup_l\left|\check{h}_l-\tilde{h}_l\right|=o_p(1).$$

Since h is bounded and bounded away from zero, this implies that $\sup_l \eta_l^{-2}=O_p(1)$. Further, by part (b) of Lemmas 3.2, 3.4 and 3.6,

$$\frac{C}{T}\sum_{l=m+1}^{[T/2]}\left(\left(\hat{h}_l-\check{h}_l\right)^2+\left(\check{h}_l-\check{h}_l\right)^2+\left(\check{h}_l-\tilde{h}_l\right)^2\right)=O_p\left(m^{-1}+T^{-1}\right)=O_p\left(m^{-1}\right).$$

From here the conclusion of part (a) is obvious.

(b) By the mean value theorem,

$$|\tilde{c}_r-c_{r,T}|\leq\frac{1}{T}\sum_{l=m+1}^{[T/2]}\left|\log\tilde{h}_l-\log h_l\right|=\frac{1}{T}\sum_{l=m+1}^{[T/2]}\frac{1}{\eta_l}\left|\tilde{h}_l-h_l\right|,$$

where η_l is an intermediate point between h_l and $\tilde{h}_l$. Since $\tilde{h}_l=h_l+O\left(M^{-2}\right)$ uniformly in l and therefore also $\sup_l|\eta_l|^{-1}=O(1)$, the last displayed expression is $O_p\left(M^{-2}\right)=o_p\left(T^{-\frac{1}{2}}\right)$ uniformly in $m+1\leq l\leq[T/2]$.

(c) By symmetry, we have

$$c_{r,T}-c_r=\frac{1}{2T}\sum_{l=m+1}^{T-m-1}\log h_l\cos r\lambda_l-\frac{1}{4\pi}\int_{-\pi}^{\pi}\log h(\lambda)\cos r\lambda d\lambda$$

which by the formula of Brillinger (1981, p. 15) is equal to

$$\begin{aligned}&\frac{1}{T}\int_0^{2\pi}\left(T\frac{\lambda}{2\pi}-\left[T\frac{\lambda}{2\pi}\right]-\frac{1}{2}\right)\frac{h'(\lambda)}{h(\lambda)}\cos(r\lambda)\,d\lambda\\&-\frac{r}{T}\int_0^{2\pi}\left(T\frac{\lambda}{2\pi}-\left[T\frac{\lambda}{2\pi}\right]-\frac{1}{2}\right)\log h(\lambda)\sin(r\lambda)\,d\lambda. \qquad (3.46)\end{aligned}$$

Since the function h is piecewise twice continuously differentiable, the first term of (3.46) is $O(r^{-2}T^{-1})$ by Lemma 3.7 and the second term of (3.46) is $O(r^{-1}T^{-1})$ uniformly in r. Therefore $c_{r,T}-c_r=O(r^{-1}T^{-1})$.

(d) Since function $\log h(\lambda)$ is piecewise twice continuously differentiable on $[0,\pi]$, Lemma 3.8 implies that

$$\left|\sum_{r=M+1}^{\infty} c_r e^{ir\lambda_j}\right|\le\sum_{r=M+1}^{\infty}|c_r|\le\sum_{r=M+1}^{\infty}\frac{C}{r^2}=O\left(M^{-1}\right).$$

■

Let us denote $\Psi_j=\Psi\left(e^{i\lambda_j}\right)$ and $\hat{\Psi}_j=\hat{\Psi}\left(e^{i\lambda_j}\right)$.

Lemma 3.10 *Under Conditions 1.1-1.5, 2.1 and 3.1-3.4,*

$$\sup_{1\le j\le T}\left|\hat{\Psi}_j-\Psi_j\right|=O_p\left(m^{-1}T^{\frac{1}{2}}\right).$$

Proof. By the Taylor theorem,

$$\begin{aligned}\hat{\Psi}_j-\Psi_j &= \Psi_j\left(\frac{\hat{\Psi}_j}{\Psi_j}-1\right)=\Psi_j\left(\exp\left\{\sum_{r=1}^{\infty}c_r e^{ir\lambda_j}-\sum_{r=1}^{M}\hat{c}_r e^{ir\lambda_j}\right\}-1\right)\\ &= \Psi_j e^{\eta_j}\left(\sum_{r=1}^{\infty}c_r e^{ir\lambda_j}-\sum_{r=1}^{M}\hat{c}_r e^{ir\lambda_j}\right),\end{aligned}$$

where $0<\left|\eta_j\right|<\left|\sum_{r=1}^{\infty}c_r e^{ir\lambda_j}-\sum_{r=1}^{M}\hat{c}_r e^{ir\lambda_j}\right|$. The last display is equal to

$$\begin{aligned}\Psi_j e^{\eta_j}&\left(\sum_{r=M+1}^{\infty}c_r e^{ir\lambda_j}+\sum_{r=1}^{M}\left(c_r-c_{r,T}\right)e^{ir\lambda_j}+\sum_{r=1}^{M}\left(c_{r,T}-\tilde{c}_r\right)e^{ir\lambda_j}\right.\\&\left.+\sum_{r=1}^{M}\left(\tilde{c}_r-\hat{c}_r\right)e^{ir\lambda_j}\right)\end{aligned}$$

By Lemma 3.9, the bracket of the last displayed expression is $o(M^{-1}) + O(T^{-1}\log T) + O_p(M^{-1}) + O_p\left(MT^{-\frac{1}{2}} + m^{-\frac{1}{2}}\log^2 T\right)$ which is $O_p\left(m^{-1}T^{\frac{1}{2}}\right) = o_p(1)$ by Condition 3.4. This also implies that $\sup_{1\le j\le T}\left|\eta_j\right| = o_p(1)$ and therefore $\sup_{1\le j\le T}\left|e^{\eta_j}\right| = O_p(1)$. Since $\sup_{1\le j\le T}\left|\Psi_j\right| \le C$ by (3.5) and Condition 3.3, we obtain $\sup_{1\le j\le T}\left|\hat{\Psi}_j - \Psi_j\right| = O_p\left(m^{-1}T^{\frac{1}{2}}\right)$ as required. ∎

Let us denote $\hat{\gamma}_j = \left(1-e^{i\lambda_j}\right)^{-\hat{d}}$ and $\gamma_j = \left(1-e^{i\lambda_j}\right)^{-d}$.

Lemma 3.11 *Under Conditions 1.1-1.5, 2.1 and 3.1-3.4,*

(a)

$$\sup_{1\le j\le[T/2]}\left|\frac{\hat{\gamma}_j}{\gamma_j} - 1\right| = O_p\left(m^{-\frac{1}{2}}\log T\right),$$

(b)

$$\sup_{1\le j\le[T/2]}\left|\frac{\gamma_j}{\hat{\gamma}_j} - 1\right| = O_p\left(m^{-\frac{1}{2}}\log T\right).$$

Proof. (a) For $1 \le j \le [T/2]$, write

$$\frac{\hat{\gamma}_j}{\gamma_j} = \left(1-e^{i\lambda_j}\right)^{d-\hat{d}} = \left(2\sin\frac{\lambda_j}{2}\right)^{d-\hat{d}} e^{i\left(d-\hat{d}\right)\theta_j},$$

where $0 \le \theta_j < 2\pi$. By the mean value theorem,

$$\left|\left(2\sin\frac{\lambda_j}{2}\right)^{d-\hat{d}} - 1\right| \le \left|\ln\left(2\sin\frac{\lambda_j}{2}\right)\right|\left(2\sin\frac{\lambda_j}{2}\right)^{\delta}\left|d-\hat{d}\right|,$$

where $0 < |\delta| < \left|d-\hat{d}\right|$. Since for $0 \le \lambda \le \pi$, $2\sin\frac{\lambda}{2} \ge \frac{2}{\pi}\lambda$, we have

$$\sup_{1\le j\le[T/2]}\left|\ln\left(2\sin\frac{\lambda_j}{2}\right)\right| \le \left|\ln\left(2\sin\frac{\lambda_1}{2}\right)\right| \le \left|\ln\left(\frac{2}{\pi}\lambda_1\right)\right| = O(\ln T).$$

Further,

$$\begin{aligned}\sup_{1\le j\le[T/2]}\left(2\sin\frac{\lambda_j}{2}\right)^{\delta} &\le \left(2\sin\frac{\lambda_1}{2}\right)^{-\left|d-\hat{d}\right|} \le \left(\frac{2}{\pi}\lambda_1\right)^{-\left|d-\hat{d}\right|} \\ &= \left(\frac{\pi}{2}\frac{1}{\lambda_1}\right)^{\left|d-\hat{d}\right|} = \left(\frac{T}{4}\right)^{\left|d-\hat{d}\right|} = O_p(1)\end{aligned}$$

because by Proposition 3.1, $d - \hat{d} = O_p\left(m^{-1/2}\right)$, and because by Condition 3.4, $T^{m^{-1/2}} = e^{m^{-1/2}\log T} \to 1$. It follows that $(2\sin(\lambda_j/2))^{d-\hat{d}} = 1 + O_p\left(m^{-1/2}\log T\right)$ uniformly over $1 \le j \le [T/2]$.

Further, by the Taylor theorem, $e^{i\left(d-\hat{d}\right)\theta} = 1 + O\left(d - \hat{d}\right)$ uniformly in θ, and therefore

$$\sup_{1\le j\le [T/2]}\left|\frac{\hat{\gamma}_j}{\gamma_j} - 1\right| = O_p\left(m^{-\frac{1}{2}}\log T\right).$$

(b) Part (a) implies that

$$\frac{\gamma_j}{\hat{\gamma}_j} = \frac{1}{1 + O_p\left(m^{-\frac{1}{2}}\log T\right)} = 1 + O_p\left(m^{-\frac{1}{2}}\log T\right)$$

uniformly over $1 \le j \le [T/2]$. ■

Lemma 3.12 *Let φ be a complex function satisfying the following conditions: There exist constants $0 < C_\varphi < \infty$ and d such that $|\varphi(\lambda)| \sim C_\varphi \lambda^{-d}$ as $\lambda \to 0+$, $\varphi(\lambda)$ is differentiable on $(0, \pi]$, $\left|\frac{d\varphi(\lambda)}{d\lambda}\right| = O\left(\frac{|\varphi(\lambda)|}{\lambda}\right)$ uniformly over $(0, \pi]$ as $\lambda \to 0+$, $|\varphi(\lambda)| > 0$ for $\lambda \in (0, \pi]$ and $\varphi(2\pi - \lambda) = \overline{\varphi(\lambda)}$ for $\lambda \in (0, \pi]$. Then for $r = 0, \ldots, M$,*
(a)

$$\frac{1}{T}\sum_{j=1}^{T-1}\varphi(\lambda_j)\, e^{ir\lambda_j} - \frac{1}{2\pi}\int_{-\pi}^{\pi}\varphi(\lambda)\, e^{ir\lambda}d\lambda = \begin{cases} O\left(rT^{-1} + T^{-1}\right) & -1 < d < 0, \\ O\left(rT^{-1} + T^{-1}\log T\right) & d = 0, \\ O\left(rT^{-1} + T^{d-1}\right) & 0 < d < 1, \end{cases}$$

(b)

$$\frac{1}{T}\sum_{j=1}^{T-1}\varphi(\lambda_j)\, e^{ir\lambda_j} = \begin{cases} O(\log T) & d = 1, \\ O\left(T^{d-1}\right) & d > 1. \end{cases}$$

Proof. (a) The assumptions of the lemma imply that $|\varphi(\lambda)| \le C\lambda^{-d}$ for $0 < \lambda \le \pi$ for some C. We have

$$\left|\frac{1}{T}\sum_{j=1}^{T-1}\varphi(\lambda_j)\, e^{ir\lambda_j} - \frac{1}{2\pi}\int_{-\pi}^{\pi}\varphi(\lambda)\, e^{ir\lambda}d\lambda\right| \le \frac{2}{T}|\varphi(\lambda_1)| + \frac{2}{2\pi}\int_0^{\lambda_1}|\varphi(\lambda)|\, d\lambda$$

$$+\left|\frac{2}{2\pi}\sum_{j=2}^{[T/2]}\left(\int_{\lambda_{j-1}}^{\lambda_j}\left(\varphi(\lambda)e^{ir\lambda}-\varphi(\lambda_j)e^{ir\lambda_j}\right)d\lambda\right)\right|. \tag{3.47}$$

The first term on the right of (3.47) is bounded by $CT^{-1}\lambda_1^{-d}=O\left(T^{d-1}\right)$. The second term on the right of (3.47) is bounded by $C\int_0^{\lambda_1}\lambda^{-d}d\lambda=O\left(T^{d-1}\right)$ for $-1<d<1$. Regarding the third term on the right of (3.47), we obtain

$$\begin{aligned}
&\left|\int_{\lambda_{j-1}}^{\lambda_j}\left(\varphi(\lambda)e^{ir\lambda}-\varphi(\lambda_j)e^{ir\lambda_j}\right)d\lambda\right| \\
\leq\ & \max_{\lambda_{j-1}\leq\lambda\leq\lambda_j}\left|\left(\varphi(\lambda)e^{ir\lambda}-\varphi(\lambda_j)e^{ir\lambda_j}\right)\right|\int_{\lambda_{j-1}}^{\lambda_j}d\lambda \\
\leq\ & \frac{C}{T^2}\max_{\lambda_{j-1}\leq\lambda\leq\lambda_j}\left(\left|\varphi'(\lambda)\right|+r\left|\varphi(\lambda)\right|\right)\leq\frac{C}{T^2}\max_{\lambda_{j-1}\leq\lambda\leq\lambda_j}\left(\frac{|\varphi(\lambda)|}{\lambda}+r\left|\varphi(\lambda)\right|\right) \\
\leq\ & C\left(T^{d-1}j^{-d-1}+rT^{d-2}j^{-d}\right)
\end{aligned}$$

for $j=2,\ldots,[T/2]$, where the second inequality follows from the mean value theorem. Therefore the third term on the right of (3.47) is bounded by

$$CT^{d-2}r\sum_{j=2}^{[T/2]}j^{-d}+CT^{d-1}\sum_{j=2}^{[T/2]}j^{-d-1}=O\left(\frac{r}{T}\right)+\begin{cases}O\left(T^{-1}\right) & -1<d<0,\\ O\left(T^{-1}\log T\right) & d=0,\\ O\left(T^{d-1}\right) & 0<d<1.\end{cases}$$

Collecting the results, we obtain the bound in part (a).

(b) The expression $T^{-1}\sum_{j=1}^{T-1}\varphi(\lambda_j)e^{ir\lambda_j}$ is bounded in absolute value by

$$\frac{2}{T}\left|\varphi(\lambda_1)\right|+\left(\frac{2}{T}\sum_{j=2}^{T-1}\varphi(\lambda_j)e^{ir\lambda_j}-\frac{2}{2\pi}\int_{\lambda_1}^{\pi}\varphi(\lambda)e^{ir\lambda}d\lambda\right)+\frac{2}{2\pi}\int_{\lambda_1}^{\pi}\left|\varphi(\lambda)\right|d\lambda. \tag{3.48}$$

The first term of (3.48) is bounded by $CT^{-1}\lambda_1^{-d}=O\left(T^{d-1}\right)$. By the same arguments as in the proof of part (a), the second term of (3.48) is bounded by

$$CT^{d-2}r\sum_{j=2}^{[T/2]}j^{-d}+CT^{d-1}\sum_{j=2}^{[T/2]}j^{-d-1}=\begin{cases}O\left(rT^{-1}\log T+1\right) & d=1,\\ O\left(T^{d-1}\right) & d>1.\end{cases}$$

The third term of (3.48) is bounded by

$$C\int_{\lambda_1}^{\pi}\lambda^{-d}d\lambda=\begin{cases}O(\log T) & d=1,\\ O\left(T^{d-1}\right) & d>1.\end{cases}$$

These results yield the bound in part (b). ■

Let us denote $B_j=B\left(e^{i\lambda_j}\right)$ and $R_j=B_j^{-1}$. Conditions 1.1 and 1.4 imply that ε_t has a representation

$$\varepsilon_t=\sum_{l=0}^{\infty}\rho_l u_{t-l},$$

where ρ_l are square summable and

$$\rho_l=\frac{1}{2\pi}\int_{-\pi}^{\pi}R\left(e^{i\lambda}\right)e^{-il\lambda}d\lambda$$

with $R\left(e^{i\lambda}\right)=B^{-1}\left(e^{i\lambda}\right)$.

Lemma 3.13 *Assume Conditions 1.1-1.5, 2.1 and 3.1-3.4 hold. Then*

(a)

$$\sum_{l=0}^{M}\left(\hat{b}_l-b_l\right)^2=O_p\left(m^{-2}T\log T\right),$$

(b)

$$\sum_{l=0}^{M}\left(\hat{\rho}_l-\rho_l\right)^2=O_p\left(m^{-2}T\log T\right).$$

Proof. (a) Define

$$\check{b}_l=\frac{1}{T}\sum_{j=1}^{T-1}B_j e^{-il\lambda_j},\qquad l=0,\ldots,M$$

and write

$$\sum_{l=0}^{M}\left(\hat{b}_l-\check{b}_l\right)^2=\frac{1}{T^2}\sum_{j=1}^{T-1}\sum_{k=1}^{T-1}\left(\hat{B}_j-B_j\right)\overline{\left(\hat{B}_k-B_k\right)}\sum_{l=0}^{M}e^{-il(\lambda_j-\lambda_k)},\quad(3.49)$$

where

$$
\begin{aligned}
\left|\hat{B}_j - B_j\right| &= \left|\hat{\gamma}_j\hat{\Psi}_j - \gamma_j\Psi_j\right| \\
&\le \left|\hat{\gamma}_j - \gamma_j\right|\left|\Psi_j\right| + \left|\gamma_j\right|\left|\hat{\Psi}_j - \Psi_j\right| + \left|\hat{\gamma}_j - \gamma_j\right|\left|\hat{\Psi}_j - \Psi_j\right| \\
&= \left|\gamma_j\right| O_p\left(m^{-1}T^{\frac{1}{2}}\right)
\end{aligned}
$$

uniformly over $1 \le j \le T-1$ by Condition 3.4 and Lemmas 3.10 and 3.11. Also, simple algebra yields

$$
\sum_{l=0}^{M} e^{-il\lambda} = \begin{cases} M+1 & \lambda = 0, \\ e^{-i\frac{\lambda}{2}M}\frac{\sin\left(\frac{\lambda}{2}(M+1)\right)}{\sin\frac{\lambda}{2}} & \lambda \ne 0. \end{cases}
$$

Since $\sin\frac{\lambda}{2} > \frac{\lambda}{\pi}$ for $0 \le \lambda \le \pi$, it is

$$
\sum_{l=0}^{M} e^{-il(\lambda_j-\lambda_k)} = O\left(T\frac{1}{|j-k|_+}\right).
$$

Therefore (3.49) is

$$
O_p\left(m^{-2}T\right)\frac{1}{T}\sum_{j=1}^{T-1}\sum_{k=1}^{T-1}\left|\gamma_j\gamma_k\right|\frac{1}{|j-k|_+} = O_p\left(m^{-2}T\log T\right)
$$

by the Cauchy-Schwarz inequality and Lemma 1.6.

Further, by Lemma 3.12,

$$
\breve{b}_l - b_l = \frac{1}{T}\sum_{j=1}^{T-1} B_j e^{-il\lambda_j} - \frac{1}{2\pi}\int_{-\pi}^{\pi} B\left(e^{i\lambda}\right)e^{-il\lambda}d\lambda = O\left(T^{-1}l + T^{-1}\log T + T^{d-1}\right)
$$

uniformly in $l = 0,\ldots,M$, and so $\sum_{l=0}^{M}\left(\breve{b}_l - b_l\right)^2 = O\left(m^{-3}T + m^{-1}T^{2d-1}\right)$. Elementary inequalities imply that $\sum_{l=0}^{M}\left(\hat{b}_l - b_l\right)^2 = O_p\left(m^{-2}T\log T\right)$.
(b) Define

$$
\breve{\rho}_l = \frac{1}{T}\sum_{j=1}^{T-1} R_j e^{-il\lambda_j}, \quad l = 0,\ldots,M.
$$

Proceeding as in part (a), we obtain that $\sum_{l=0}^{M}(\hat{\rho}_l - \check{\rho}_l)^2 = O_p(m^{-2}T\log T)$. By Lemma 3.12,

$$\check{\rho}_l - \rho_l = \frac{1}{T}\sum_{j=1}^{T-1} R_j e^{-il\lambda_j} - \frac{1}{2\pi}\int_{-\pi}^{\pi} R(\lambda)\, e^{-il\lambda} d\lambda = O\left(T^{-1}r + T^{-1}\log T\right)$$

for $l = 0, \ldots, M$, and so $\sum_{l=0}^{M}(\check{\rho}_l - \rho_l)^2 = O(MT^{-2})$. This implies that $\sum_{l=0}^{M}(\hat{\rho}_l - \rho_l)^2 = O_p(m^{-2}T\log T) + O(m^{-3}T) = O_p(m^{-2}T\log T)$. ∎

Lemma 3.14 *Assume Conditions 1.1-1.5, 2.1 and 3.1-3.4 hold. Then*

$$\frac{1}{T}\sum_{t=1}^{T}(\hat{\varepsilon}_t - \varepsilon_t)^2 = o_p(1). \tag{3.50}$$

Proof. Assume for simplicity that process $\{x_t\}$ is scalar. From the definition of $\hat{\varepsilon}_t$ in Step 3 of the bootstrap procedure,

$$\begin{aligned}
\hat{\varepsilon}_t &= \frac{1}{\sqrt{T}}\sum_{j=1}^{T-1}\sum_{l=0}^{M}\hat{\rho}_l w_{\hat{u},j} e^{-i(t-l)\lambda_j} \\
&= \frac{1}{\sqrt{T}}\sum_{j=1}^{T-1}\sum_{l=0}^{M}\rho_l w_{u,j} e^{-i(t-l)\lambda_j} + \frac{1}{\sqrt{T}}\sum_{j=1}^{T-1}\sum_{l=0}^{M}(\hat{\rho}_l - \rho_l)\, w_{u,j} e^{-i(t-l)\lambda_j} \\
&\quad + \frac{1}{\sqrt{T}}\sum_{j=1}^{T-1}\sum_{l=0}^{M}\rho_l\,(w_{\hat{u},j} - w_{u,j})\, e^{-i(t-l)\lambda_j} \\
&\quad + \frac{1}{\sqrt{T}}\sum_{j=1}^{T-1}\sum_{l=0}^{M}(\hat{\rho}_l - \rho_l)\,(w_{\hat{u},j} - w_{u,j})\, e^{-i(t-l)\lambda_j},
\end{aligned}$$

where the T-th frequency in the sum over j is omitted because $w_{\hat{u},T} = T^{-1/2}\sum_{t=1}^{T}\hat{u}_t = 0$. Since $|a+b+c+d|^2 \leq 4\left(|a|^2 + |b|^2 + |c|^2 + |d|^2\right)$, we have

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}(\hat{\varepsilon}_t-\varepsilon_t)^2 \;\le\; & \frac{4}{T}\sum_{t=1}^{T}\left(\frac{1}{\sqrt{T}}\sum_{j=1}^{T-1}\sum_{l=0}^{M}\rho_l w_{u,j}e^{-i(t-l)\lambda_j}-\varepsilon_t\right)^2 \\
& +\frac{4}{T}\sum_{j=1}^{T-1}|w_{u,j}|^2\left|\sum_{l=0}^{M}(\hat{\rho}_l-\rho_l)\,e^{il\lambda_j}\right|^2 \\
& +\frac{4}{T}\sum_{j=1}^{T-1}|w_{\hat{u},j}-w_{u,j}|^2\left|\sum_{l=0}^{M}\rho_l e^{il\lambda_j}\right|^2 \\
& +\frac{4}{T}\sum_{j=1}^{T-1}|w_{\hat{u},j}-w_{u,j}|^2\left|\sum_{l=0}^{M}(\hat{\rho}_l-\rho_l)\,e^{il\lambda_j}\right|^2. \qquad (3.51)
\end{aligned}
$$

Write

$$
\frac{1}{\sqrt{T}}\sum_{j=1}^{T-1}\sum_{l=0}^{M}\rho_l w_{u,j}e^{-i(t-l)\lambda_j}=\sum_{l=0}^{M}\rho_l\sum_{s=1}^{T}u_s\mathbb{I}\left(t-l=s \mod T\right)-\sum_{l=0}^{M}\rho_l\frac{1}{T}\sum_{s=1}^{T}u_s.
$$

Since $\varepsilon_t=\sum_{l=0}^{\infty}\rho_l u_{t-j}$, we have

$$
\begin{aligned}
\varepsilon_t-\frac{1}{\sqrt{T}}\sum_{j=1}^{T-1}\sum_{l=0}^{M}\rho_l w_{u,j}e^{-i(t-l)\lambda_j} \;=\; & \sum_{l=M+1}^{\infty}\rho_l u_{t-l}-\sum_{l=0}^{M}\rho_l\frac{1}{T}\sum_{s=1}^{T}u_s \\
& +\sum_{l=t}^{M}\rho_l\left(u_{t-l+T}-u_{t-l}\right)\mathbb{I}\left(t\le M\right)
\end{aligned}
$$

and therefore by the Cauchy-Schwarz inequality, the first term on the right of (3.51) is bounded by

$$
\begin{aligned}
& \frac{C}{T}\sum_{t=1}^{T}\left(\sum_{l=M+1}^{\infty}\rho_l u_{t-l}\right)^2+C\left(\sum_{l=0}^{M}\rho_l\right)^2\left(\frac{1}{T}\sum_{s=1}^{T}u_s\right)^2 \\
& +\frac{C}{T}\sum_{t=1}^{M}\left(\sum_{l=t}^{M}\rho_l\left(u_{t-l+T}-u_{t-l}\right)\right)^2. \qquad (3.52)
\end{aligned}
$$

Expectation of the first term (3.52) is bounded by

$$C\frac{1}{T}\sum_{t=1}^{T}\sum_{l,p=M+1}^{\infty}\left|\rho_l\rho_p\right| E\left|u_{t-l}u_{t-p}\right| \le C\left(\sum_{l=M+1}^{\infty}\left|\rho_l\right|\right)^2.$$

The coefficient ρ_l is equal to $\sum_{j=0}^{l}\chi_j\phi_{l-j}$, where

$$\chi_j = (2\pi)^{-1}\int_{-\pi}^{\pi}\left(1-e^{i\lambda}\right)^d e^{i\lambda j}d\lambda.$$

When $d>0$, $\chi_j = O\left(j^{-1-d}\right)$ by the Stirling formula. Since $\phi_j = O\left(l^{-2}\right)$, ρ_l is $O\left(l^{-1-d}\right)$ and therefore it is absolutely summable. When $d=0$, $\rho_l=\phi_l$ and ρ_l is again absolutely summable. It follows that the right-hand side of the last displayed inequality is $o(1)$. The second term of (3.52) is $o_p(1)$ because $\sum_{l=0}^{M}\left|\rho_l\right|<\infty$ and $\bar{u}=o_p(1)$ by Conditions 1.1 and 3.2. Expectation of the third term (3.52) is bounded by $CT^{-1}\sum_{t=1}^{M}\left(\sum_{l=t}^{M}\left|\rho_l\right|\right)^2$. When $d>0$,

$$\frac{1}{T}\sum_{t=1}^{M}\left(\sum_{l=t}^{M}\left|\rho_l\right|\right)^2 \le \frac{C}{T}\sum_{t=1}^{M}\left(\sum_{l=t}^{M}l^{-1-d}\right)^2 \le \frac{C}{T}\sum_{t=1}^{M}t^{-2d} = O\left(T^{-1}M^{1-2d}\right)$$

while when $d=0$,

$$\frac{1}{T}\sum_{t=1}^{M}\left(\sum_{l=t}^{M}\left|\rho_l\right|\right)^2 \le \frac{C}{T}\sum_{t=1}^{M}\left(\sum_{l=t}^{M}l^{-2}\right)^2 \le \frac{C}{T}\sum_{t=1}^{M}t^{-2} = O\left(T^{-1}\right).$$

By the Markov inequality, the first term on the right of (3.51) is $o_p(1)$.

To obtain a bound for the second term on the right of (3.51), we observe that

$$\begin{aligned}\left|\sum_{l=0}^{M}\left(\hat{\rho}_l-\rho_l\right)e^{il\lambda_j}\right|^2 &\le M\sum_{l=0}^{M}\left(\hat{\rho}_l-\rho_l\right)^2 = O_p\left(T^2m^{-3}\log T\right)\\ &= o_p(1) \qquad (3.53)\end{aligned}$$

by Lemma 3.13 and Condition 3.4. Since $T^{-1}\sum_{j=1}^{T-1}\left|w_{u,j}\right|^2 = T^{-1}\sum_{t=1}^{T}\left(u_t-\bar{u}\right)^2$ is $O_p(1)$ by Conditions 1.1 and 3.1, the second term of (3.51) is $o_p(1)$.

Turning to the third term on the right of (3.51), the absolutely summability of ρ_l implies that

$$\frac{1}{T}\sum_{j=1}^{T-1}\left|w_{\hat{u},j}-w_{u,j}\right|^2\left|\sum_{l=0}^{M}\rho_l e^{il\lambda_j}\right|^2 \le \frac{C}{T}\sum_{j=1}^{T-1}\left|w_{\hat{u},j}-w_{u,j}\right|^2.$$

Arguing as in the proof of Proposition 3.1, we obtain

$$\begin{aligned}&\frac{1}{T}\sum_{j=1}^{T-1}\left|w_{\hat{u},j}-w_{u,j}\right|^2\\ \le\ & 3\left(\hat{\beta}-\beta\right)^2\frac{1}{T}\sum_{j=1}^{T-1}I_{xx,j}+3\left(\hat{\delta}-\delta\right)^2\frac{1}{T}\sum_{j=1}^{T-1}I_{\hat{z}\hat{z},j}+3\delta^2\frac{1}{T}\sum_{j=1}^{T-1}I_{\Delta\hat{z}\Delta\hat{z},j}.\end{aligned}$$

For $1\le k\le T$, $\sup_k T^{-1}\sum_{j=1}^{T-1}I_{zz,j}\le 2T^{-1}\sum_{t=1}^{T}x_t^2=O_p(1)$ by Conditions 1.1 and 3.2, and $\sup_{1\le j\le T-1}I_{\Delta\hat{z}\Delta\hat{z},j}=O_p\left(T^{-1}\right)$ by (3.17). By Lemma 1.7, $\left(\hat{\beta}-\beta\right)^2=O_p\left(T^{-1}\right)$ and $\left(\hat{\delta}-\delta\right)^2=O_p\left(T^{-1}\right)$. This means that

$$\frac{1}{T}\sum_{j=1}^{T-1}\left|w_{\hat{u},j}-w_{u,j}\right|^2=O_p\left(T^{-1}\right) \tag{3.54}$$

and that the third term of (3.51) is $o_p(1)$.

Finally, combining results (3.53) and (3.54), we can see that the fourth term of (3.51) is $o_p(1)$. ■

Lemma 3.15 *Assume Conditions 1.1-1.5, 2.1 and 3.1-3.4 are satisfied. Let* $\hat{\sigma}_\varepsilon^2=T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t^2$. *Then*

$$\hat{\sigma}_\varepsilon^2\xrightarrow{p}\sigma_\varepsilon^2.$$

Proof. Write

$$\hat{\sigma}_\varepsilon^2=\frac{1}{T}\sum_{t=1}^{T}\left(\hat{\varepsilon}_t-\varepsilon_t\right)^2+\frac{2}{T}\sum_{t=1}^{T}\left(\hat{\varepsilon}_t-\varepsilon_t\right)\varepsilon_t+\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2.$$

The first term is $o_p(1)$ by Lemma 3.14. The third term converges in probability to σ_ε^2 by Conditions 1.1 and 3.2. By the Cauchy-Schwarz inequality, the second term is $o_p(1)$. ■

Lemma 3.16 *Assume Conditions 1.1-1.5, 2.1 and 3.1-3.4 hold. Then*

$$\varepsilon_t^* \xrightarrow{d^*} \varepsilon_t.$$

Proof. Denote by $d_2(\cdot,\cdot)$ the Mallows metric as defined for example by Bickel and Freedman (1981). Let $\hat{F}_T(x) = T^{-1}\sum_{t=1}^{T}\mathbb{I}(\hat{\varepsilon}_t \le x)$, $F_T(x) = T^{-1}\sum_{t=1}^{T}\mathbb{I}(\varepsilon_t \le x)$ and $F(x) = P(\varepsilon_t \le x)$. Then

$$d_2\left(\hat{F}_T, F\right) \le d_2\left(\hat{F}_T, F_T\right) + d_2(F_T, F). \tag{3.55}$$

Let U be a random variable distributed uniformly on $\{1, 2, \ldots, T\}$. We have

$$d_2\left(\hat{F}_T, F_T\right) \le E_U(\hat{\varepsilon}_U - \varepsilon_U)^2 = \frac{1}{T}\sum_{t=1}^{T}(\hat{\varepsilon}_t - \varepsilon_t)^2.$$

By Lemma 3.14, the last expression converges to zero in probability. The second term of (3.55) converges to zero almost surely by Lemma 8.4 of Bickel and Freedman (1981). Therefore $d_2\left(\hat{F}_T, F\right) = o_p(1)$ and the lemma holds. ■

For integers $0 \le l \le T$, let $\left\{w_{a(k),j}, 1 \le j \le T\right\}$ be the discrete Fourier transform of the sequence $\{a_t\mathbb{I}(t \le l), 1 \le t \le T\}$.

Lemma 3.17 *Let $\tau \wedge \sigma = \min\{\tau, \sigma\}$. Under Conditions 1.1-1.5, 2.1 and 3.1-3.4,*

$$\frac{1}{T}\sum_{j=1}^{T-1}|A_jB_j|^2\,\bar{w}_{\xi([\tau T]),j}w_{\xi([\sigma T]),j} \Longrightarrow (\tau \wedge \sigma)\frac{\sigma_\xi^2}{2\pi}\int_{-\pi}^{\pi}\left|A\left(e^{i\lambda}\right)B\left(e^{i\lambda}\right)\right|^2 d\lambda \tag{3.56}$$

on $(\tau, \sigma) \in [0, 1]^2$.

Proof. We first prove that

$$\frac{1}{T}\sum_{j=1}^{T-1}|A_jB_j|^2\left|w_{\xi([\tau T]),j}\right|^2 \Longrightarrow \tau\frac{\sigma_\xi^2}{2\pi}\int_{-\pi}^{\pi}\left|A\left(e^{i\lambda}\right)B\left(e^{i\lambda}\right)\right|^2 d\lambda \tag{3.57}$$

on $\tau \in [0,1]$. By Lemma 1.1, the expectation of the the left-hand side of (3.57) is

$$\frac{[\tau T]}{T} \frac{\sigma_\xi^2}{T} \sum_{j=1}^{T-1} |A_j B_j|^2 \longrightarrow \tau \frac{\sigma_\xi^2}{2\pi} \int_{-\pi}^{\pi} \left| A\left(e^{i\lambda}\right) B\left(e^{i\lambda}\right) \right|^2 d\lambda \tag{3.58}$$

and its second moment is

$$\begin{aligned}
&\frac{1}{T^2} \sum_{j,k=1}^{T-1} |A_j B_j A_k B_k|^2 \operatorname{cum}\left(\bar{w}_{\xi([\tau T]),j}, w_{\xi([\tau T]),j}, w_{\xi([\tau T]),k}, \bar{w}_{\xi([\tau T]),k}\right) \\
&+\frac{1}{T^2} \sum_{j,k=1}^{T-1} |A_j B_j A_k B_k|^2 E\left|w_{\xi([\tau T]),j}\right|^2 E\left|w_{\xi([\tau T]),k}\right|^2 \\
&+\frac{1}{T^2} \sum_{j,k=1}^{T-1} |A_j B_j A_k B_k|^2 \left|E\bar{w}_{\xi([\tau T]),j} w_{\xi([\tau T]),k}\right|^2 \\
&+\frac{1}{T^2} \sum_{j,k=1}^{T-1} |A_j B_j A_k B_k|^2 \left|E w_{\xi([\tau T]),j} w_{\xi([\tau T]),k}\right|^2 .
\end{aligned} \tag{3.59}$$

We have

$$\operatorname{cum}\left(\bar{w}_{\xi([\tau T]),j}, w_{\xi([\tau T]),j}, w_{\xi([\tau T]),k}, \bar{w}_{\xi([\tau T]),k}\right) = \frac{[\tau T]}{T^2} \kappa_\xi \leq \frac{C}{T}$$

uniformly over $\tau \in [0,1]$, where $\kappa_\xi = \operatorname{cum}(\xi_t, \xi_t, \xi_t, \xi_t)$. The first term of (3.59) is bounded by $CT^{-3} \sum_{j,k=1}^{T-1} |A_j B_j A_k B_k|^2$ which is $O\left(T^{-1}\right)$ by Lemma 1.1. The second term on the right of (3.59) is equal to

$$\left(\frac{[\tau T]}{T}\right)^2 \sigma_\xi^4 \left(\frac{1}{T} \sum_{j=1}^{T-1} |A_j B_j|^2\right)^2 \to \tau^2 \left(\frac{\sigma_\xi^2}{2\pi} \int_{-\pi}^{\pi} \left| A\left(e^{i\lambda}\right) B\left(e^{i\lambda}\right) \right|^2 d\lambda\right)^2$$

by Lemma 1.1. Since

$$\left|E\bar{w}_{\xi([\tau T]),j} w_{\xi([\tau T]),k}\right| = \left| \frac{\sigma_\xi^2}{T} \sum_{t=1}^{[\tau T]} e^{-it(\lambda_j - \lambda_k)} \right| \leq \frac{C}{|j-k|_+},$$

the third term on the right of (3.59) is bounded by

$$CT^{-2}\sum_{j,k=1}^{T-1}|A_jB_jA_kB_k|^2\frac{1}{|j-k|_+^2}\leq CT^{-2}\sum_{j=1}^{T-1}|A_jB_j|^4=o\left(1\right)$$

due to the Cauchy-Schwarz inequality and Lemma 1.6. Similarly, the fourth term on the right of (3.59) is $o\left(1\right)$. It follows that

$$E\left|\frac{1}{T}\sum_{j=1}^{T-1}|A_jB_j|^2\left|w_{\xi([\tau T]),j}\right|^2\right|^2\to\tau^2\left(\frac{\sigma_\xi^2}{2\pi}\int_{-\pi}^{\pi}\left|A\left(e^{i\lambda}\right)B\left(e^{i\lambda}\right)\right|^2d\lambda\right)^2. \quad (3.60)$$

Convergence of the expectation and the second moment in (3.58) and (3.60) together with the Markov inequality implies that for each τ and σ from $[0,1]$,

$$\frac{1}{T}\sum_{j=1}^{T-1}|A_jB_j|^2\left|w_{\xi([\tau T]),j}\right|^2\xrightarrow{p}\tau\frac{\sigma_\xi^2}{2\pi}\int_{-\pi}^{\pi}\left|A\left(e^{i\lambda}\right)B\left(e^{i\lambda}\right)\right|^2d\lambda.$$

Since $T^{-1}\sum_{j=1}^{T-1}|A_jB_j|^2\,w_{\xi([\tau T]),j}^2$ is increasing in τ and the limiting function is continuous, the convergence is uniform over $\tau\in[0,1]$ by the arguments in the proof of Lemma A.10 of Hansen (2000). This proves (3.57).

Next, (3.57) and stationarity of ξ_t imply that also

$$\frac{1}{T}\sum_{j=1}^{T-1}|A_jB_j|^2\left|w_{\xi([\tau T]),j}-w_{\xi([\sigma T]),j}\right|^2\Longrightarrow|\tau-\sigma|\frac{\sigma_\xi^2}{2\pi}\int_{-\pi}^{\pi}\left|A\left(e^{i\lambda}\right)B\left(e^{i\lambda}\right)\right|^2d\lambda \quad (3.61)$$

on $(\tau,\sigma)\in[0,1]^2$. Consider $\tau\geq\sigma$. Writing

$$\begin{aligned}\left|w_{\xi([\tau T]),j}\right|^2 &= \left|w_{\xi([\sigma T]),j}\right|^2+\left|w_{\xi([\tau T]),j}-w_{\xi([\sigma T]),j}\right|^2\\ &\quad+2\,\mathrm{Re}\left(\bar{w}_{\xi([\tau T]),j}-\bar{w}_{\xi([\sigma T]),j}\right)w_{\xi([\sigma T]),j}\end{aligned}$$

and noting that $\sum_{j=1}^{T-1}|A_jB_j|^2\left(\bar{w}_{\xi([\tau T]),j}-\bar{w}_{\xi([\sigma T]),j}\right)w_{\xi([\sigma T]),j}$ is a real number due to the fact that $\bar{w}_{\xi([\tau T]),j}=w_{\xi([\tau T]),T-j}$ for all $\tau\in[0,1]$ and $j=1,\dots T-1$, we obtain

$$
\begin{aligned}
& \frac{2}{T}\sum_{j=1}^{T-1}|A_jB_j|^2\left(\bar{w}_{\xi([\tau T]),j}-\bar{w}_{\xi([\sigma T]),j}\right)w_{\xi([\sigma T]),j} \\
= \ & \frac{1}{T}\sum_{j=1}^{T-1}|A_jB_j|^2\left(\left|w_{\xi([\tau T]),j}\right|^2-\left|w_{\xi([\sigma T]),j}\right|^2-\left|w_{\xi([\tau T]),j}-w_{\xi([\sigma T]),j}\right|^2\right) \\
\Longrightarrow \ & (\tau-\sigma-|\tau-\sigma|)\frac{\sigma_\xi^2}{2\pi}\int_{-\pi}^{\pi}\left|A\left(e^{i\lambda}\right)B\left(e^{i\lambda}\right)\right|^2 d\lambda=0
\end{aligned}
$$

by (3.57) and (3.61). This means that the left-hand side of (3.56) is equal to

$$
\begin{aligned}
& \frac{1}{T}\sum_{j=1}^{T-1}|A_jB_j|^2\left(\bar{w}_{\xi([\tau T]),j}-\bar{w}_{\xi([\sigma T]),j}\right)w_{\xi([\sigma T]),j}+\frac{1}{T}\sum_{j=1}^{T-1}|A_jB_j|^2\left|w_{\xi([\sigma T]),j}\right|^2 \\
\Longrightarrow \ & \sigma\frac{\sigma_\xi^2}{2\pi}\int_{-\pi}^{\pi}\left|A\left(e^{i\lambda}\right)B\left(e^{i\lambda}\right)\right|^2 d\lambda.
\end{aligned}
$$

Noting that $\sigma=\tau\wedge\sigma$ for $\tau\geq\sigma$ and that symmetrical arguments hold for the case $\tau\leq\sigma$, we arrive at (3.56). ■

Lemma 3.18 *Assume Conditions 1.1-1.5, 2.1 and 3.1-3.5 hold. Let* $\tilde{B}_j=\sum_{l=0}^{M}b_le^{il\lambda_j}$. *Then*

(a)

$$\tilde{B}_j-B_j=O\left(\lambda_j^{-1}\left(\frac{T}{m}\right)^{d-1}\right)\qquad j=1,\ldots,[T/2],$$

and

(b)

$$\left|\frac{\tilde{B}_j}{B_j}\right|\leq C\left(\frac{j}{m}\right)^{d}\qquad j=1,\ldots,[T/2].$$

Proof. (a) When $0 < d < 1/2$, we have

$$\begin{aligned}
\left|\tilde{B}_j - B_j\right| &= \left|\sum_{l=M+1}^{\infty} b_l e^{il\lambda_j}\right| \le \sum_{l=M+1}^{\infty} |b_l - b_{l+1}| \left|\sum_{p=M+1}^{l} e^{ip\lambda_j}\right| \\
&\le \frac{2\pi}{\lambda_j} \sum_{l=M+1}^{\infty} |b_l - b_{l+1}| \le \frac{C}{\lambda_j} \sum_{l=M+1}^{\infty} \frac{|b_l|}{l} \le \frac{C}{\lambda_j} \sum_{l=M+1}^{\infty} l^{d-2} \\
&= O\left(\lambda_j^{-1} \left(\frac{T}{m}\right)^{d-1}\right),
\end{aligned}$$

where the first inequality follows from the summation by parts, the second inequality is due to the fact that $\left|\sum_{p=M+1}^{l} e^{ip\lambda}\right| \le \pi/\lambda$ for $0 < \lambda \le \pi$, and the remaining two inequalities are due to Condition 3.5.

When $d = 0$, the function B is piecewise twice continuously differentiable on $[0, \pi]$ and by Lemma 3.8, $b_l = O\left(l^{-2}\right)$. In this case,

$$\left|\tilde{B}_j - B_j\right| \le \sum_{l=M+1}^{\infty} |b_l| = o\left(\sum_{l=M+1}^{\infty} l^{-2}\right) = o\left(M^{-1}\right) = o\left(\lambda_j^{-1} \left(\frac{T}{m}\right)^{-1}\right).$$

(b) When $0 < d < 1/2$, Conditions 1.4 and 3.5 imply that

$$\left|\frac{\tilde{B}_j}{B_j}\right| = B_j^{-1} \left|\sum_{l=1}^{M} b_l e^{il\lambda_j}\right| \le C\lambda_j^d \sum_{l=1}^{M} |b_l| \le C\lambda_j^d \sum_{l=1}^{M} l^{-1+d} \le C\left(\frac{j}{m}\right)^d. \tag{3.62}$$

When $d = 0$, $|B_j|^{-1} \le C$ by Condition 1.4, and (3.62) imply that $\left|\tilde{B}_j / B_j\right| \le C$. ■

Lemma 3.19 *Assume Conditions 1.1-1.5, 2.1 and 3.1-3.5 hold. Then*

$$\frac{1}{T} \sum_{j=1}^{T-1} |A_j|^2 \left|\tilde{B}_j - B_j\right|^2 = o(1).$$

Proof. By Condition 1.4, there exists a constant D such that $\left|A\left(e^{i\lambda}\right)\right| < D\lambda^{-d_x}$ and $\left|B\left(e^{i\lambda}\right)\right| < D\lambda^{-d}$ for $0 < \lambda \le \pi$. By Lemma 3.18 part (b), $\left|\tilde{B}_j\right| \le C|B_j|$ for $j = 1, \ldots, m$, therefore $\left|\tilde{B}_j - B_j\right|^2 \le 2\left|\tilde{B}_j\right|^2 + 2|B_j|^2 \le C|B_j|^2$

when $1 \leq j \leq m$. By Lemma 3.18 part (a),

$$\begin{aligned}
\frac{1}{T}\sum_{j=1}^{[T/2]}|A_j|^2\left|\tilde{B}_j-B_j\right|^2 &\leq \frac{C}{T}\sum_{j=1}^{m}|A_j|^2|B_j|^2+\frac{1}{T}\sum_{j=m+1}^{[T/2]}|A_j|^2\left|\tilde{B}_j-B_j\right|^2 \\
&\leq \frac{C}{T}\sum_{j=1}^{m}\lambda_j^{-2(d_x+d)}+\frac{1}{T}\sum_{j=m+1}^{[T/2]}\lambda_j^{-2d_x-2}\left(\frac{T}{m}\right)^{2d-2} \\
&= O\left(\left(\frac{m}{T}\right)^{1-2(d_x+d)}\right)=o(1)
\end{aligned}$$

because $d_x + d < \frac{1}{2}$. Therefore

$$\frac{1}{T}\sum_{j=1}^{T-1}|A_j|^2\left|\tilde{B}_j-B_j\right|^2=\frac{2}{T}\sum_{j=1}^{[T/2]}|A_j|^2\left|\tilde{B}_j-B_j\right|^2+O\left(\frac{1}{T}\right)=o(1).$$

■

Lemma 3.20 *Under Conditions 1.1-1.5, 2.1 and 3.1-3.5,*

$$\frac{1}{\sqrt{T}}\sum_{j=1}^{T-1}\bar{A}_jB_j\bar{w}_{\xi([\tau T]),j}w_{\varepsilon^*,j}\overset{p}{\Longrightarrow}\Omega^{\frac{1}{2}}W(\tau).$$

Proof. Denote $Y_T^*(\tau)=T^{-1/2}\sum_{j=1}^{T-1}\bar{A}_jB_j\bar{w}_{\xi([\tau T]),j}w_{\varepsilon^*,j}$. To show that $Y_T^*(\tau)\overset{p}{\Longrightarrow}\Omega^{\frac{1}{2}}W(\tau)$, we need to show that the finite dimensional distributions of the process Y_T^* converge in probability to the finite dimensional distributions of the process $\Omega^{1/2}W$, and that the process Y_T^* is tight. First, $E^*Y_T^*(\tau)=0$ and

$$\operatorname{cov}^*\left(Y_T^*(\tau),Y_T^*(\sigma)\right)=\frac{\hat{\sigma}_\varepsilon^2}{T}\sum_{j=1}^{T-1}|A_jB_j|^2\,\bar{w}_{\xi([\tau T]),j}w_{\xi([\sigma T]),j}.$$

By Lemmas 3.15 and 3.17,

$$\operatorname{cov}^*\left(Y_T^*(\tau),Y_T^*(\sigma)\right)\xrightarrow{p}(\tau\wedge\sigma)\frac{\sigma_\xi^2\sigma_\varepsilon^2}{2\pi}\int_{-\pi}^{\pi}\left|A\left(e^{i\lambda}\right)B\left(e^{i\lambda}\right)\right|^2d\lambda=(\tau\wedge\sigma)\,\Omega$$

for any $(\tau,\sigma)\in[0,1]^2$. The covariance structure of Y_T^* therefore converges in probability to the covariance structure of the Gaussian process $\Omega^{\frac{1}{2}}W$, and the convergence of the finite dimensional distribution occurs as long as the Lindeberg condition is satisfied, that is,

$$\frac{1}{T}\sum_{j=1}^{T-1}\left|A_jB_j\right|^2 I_{\xi,j}E^*\left|w_{\varepsilon^*,j}\right|^2 \mathbb{I}\left(\frac{1}{T}\left|A_jB_j\right|^2 I_{\xi,j}\left|w_{\varepsilon^*,j}\right|^2>\theta\right)\xrightarrow{p}0 \quad \text{for any } \theta>0. \tag{3.63}$$

An et al. (1983) showed that

$$\sup_{1\leq j\leq\left[\frac{T}{2}\right]}\left(\frac{2\pi}{\sigma_\xi^2}\frac{1}{\log T}\left|w_{\xi,j}\right|^2\right)\leq 1 \quad \text{a.s.}$$

This implies that

$$\begin{aligned}\sup_{1\leq j\leq\left[\frac{T}{2}\right]}\frac{1}{T}\left|A_jB_j\right|^2 I_{\xi,j} &\leq \sup_{1\leq j\leq\left[\frac{T}{2}\right]}\frac{C}{T}\left|A_jB_j\right|^2\log T\leq \sup_{1\leq j\leq\left[\frac{T}{2}\right]}\frac{C}{T}\lambda_j^{-2(d_x+d)}\log T\\ &\leq CT^{2(d_x+d)-1}\log T.\end{aligned}$$

The left of (3.63) is therefore bounded by

$$\frac{1}{T}\sum_{j=1}^{T-1}\left|A_jB_j\right|^2 I_{\xi,j}E^*\left|w_{\varepsilon^*,j}\right|^2\mathbb{I}\left(\left|w_{\varepsilon^*,j}\right|^2>C\theta T^{1-2(d_x+d)}\log^{-1}T\right). \tag{3.64}$$

The expectation of $T^{-1}\sum_{j=1}^{T-1}\left|A_jB_j\right|^2 I_{\xi,j}$ is bounded because $EI_{\xi,j}=\sigma_\xi^2$ and $T^{-1}\sum_{j=1}^{T-1}\left|A_jB_j\right|^2=O(1)$ by Lemma 1.1 and Condition 1.4. The conditional expectation in (3.64) converges to zero in probability because $w_{\varepsilon^*,j}$ has finite fourth moment. This means that the Lindeberg condition (3.63) holds and therefore that the finite dimensional distributions of Y_T^* converge to those of $\Omega^{\frac{1}{2}}W$.

In order to show that Y_T^* is tight, it is sufficient to verify the moment condition (13.14) for Theorem 13.5 of Billingsley (1999). This condition is valid if for $0\leq\rho\leq\sigma\leq\tau\leq 1$,

$$E^*\left|Y_T^*(\rho)-Y_T^*(\sigma)\right|^2\left|Y_T^*(\tau)-Y_T^*(\rho)\right|^2=O_p(1)(\tau-\sigma)^2, \tag{3.65}$$

where $O_p(1)$ is uniform over ρ, σ and τ. With the purpose of proving (3.65), we examine $E^* \left|Y_T^*(\tau) - Y_T^*(\sigma)\right|^4$. Let α_j, $j = 1, \ldots, T-1$, be complex constant such that $\alpha_{T-j} = \bar{\alpha}_j$. Then

$$\begin{aligned}
& E^* \left| \frac{1}{\sqrt{T}} \sum_{j=1}^{T-1} \alpha_j w_{\varepsilon^*,j} \right|^4 \\
= & \frac{1}{T^2} \sum_{j,k,l,m=1}^{T-1} \alpha_j \bar{\alpha}_k \alpha_l \bar{\alpha}_m E^* w_{\varepsilon^*,j} \bar{w}_{\varepsilon^*,k} w_{\varepsilon^*,l} \bar{w}_{\varepsilon^*,m} \\
= & \frac{\hat{\kappa}}{T^3} \sum_{j,k,l,m=1}^{T-1} \alpha_j \bar{\alpha}_k \alpha_l \bar{\alpha}_m \mathbb{I}(j - k + l - m = 0 \mod T) \\
& + \frac{3\hat{\sigma}_\varepsilon^4}{T^2} \sum_{j,k=1}^{T-1} \left|\alpha_j \alpha_k\right|^2
\end{aligned} \tag{3.66}$$

because

$$\begin{aligned}
& E^* w_{\varepsilon^*,j} \bar{w}_{\varepsilon^*,k} w_{\varepsilon^*,l} \bar{w}_{\varepsilon^*,m} \\
= & \frac{\hat{\kappa}}{T} \mathbb{I}(j - k + l - m = 0 \mod T) + \hat{\sigma}_\varepsilon^4 \mathbb{I}(j = k, l = m \mod T) \\
& + \hat{\sigma}_\varepsilon^4 \mathbb{I}(l = T - j, m = T - k \mod T) + \hat{\sigma}_\varepsilon^4 \mathbb{I}(j = m, k = l \mod T).
\end{aligned}$$

By the Cauchy-Schwarz inequality, the first term on the right of (3.66) is bounded by

$$\frac{\hat{\kappa}}{T^3} \sum_{j,k,l,m=1}^{T-1} \left|\alpha_j \alpha_k\right|^2 \mathbb{I}(j - k + l - m = 0 \mod T) = \hat{\kappa} \frac{T-1}{T^3} \sum_{j,k,l,m=1}^{T-1} \left|\alpha_j \alpha_k\right|^2,$$

and therefore

$$E^* \left| \frac{1}{\sqrt{T}} \sum_{j=1}^{T-1} \alpha_j w_{\varepsilon^*,j} \right|^4 \leq \left(\hat{\kappa} + 3\hat{\sigma}_\varepsilon^4\right) \left(\frac{1}{T} \sum_{j=1}^{T-1} \left|\alpha_j\right|^2 \right)^2.$$

Taking $\alpha_j = \bar{A}_j B_j \bar{w}_j$ where $w_j = w_{\xi([\tau T]),j} - w_{\xi([\sigma T]),j}$ and where $0 \leq \tau, \sigma \leq 1$, we obtain

$$E^* \left| Y_T^* (\tau) - Y_T^* (\sigma) \right|^4 \leq \left(\hat{\kappa} + 3\hat{\sigma}_\varepsilon^4 \right) \left(\frac{1}{T} \sum_{j=1}^{T-1} \left| A_j B_j \right|^2 \left| w_j \right|^2 \right)^2 . \qquad (3.67)$$

The expression in the second round bracket on the right of (3.67) converges weakly to $|\tau - \sigma| \, \Omega$ by Lemma 3.17. Since $\hat{\kappa} + 3\hat{\sigma}_\varepsilon^4 = O_p(1)$, the Markov inequality implies that

$$E^* \left| Y_T^* (\tau) - Y_T^* (\sigma) \right|^4 = O_p (1) (\tau - \sigma)^2 ,$$

where $O_p(1)$ is uniform over $(\tau, \sigma) \in [0,1]^2$. By the Cauchy-Schwarz inequality, for any $\sigma \leq \rho \leq \tau$,

$$\begin{aligned} E^* \left| Y_T^* (\rho) - Y_T^* (\sigma) \right|^2 \left| Y_T^* (\tau) - Y_T^* (\rho) \right|^2 &= O_p (1) \left((\rho - \sigma)^2 (\tau - \rho)^2 \right)^{\frac{1}{2}} \\ &= O_p (1) (\tau - \sigma)^2 , \end{aligned}$$

where $O_p(1)$ is uniform over $(\tau, \sigma) \in [0,1]^2$. We conclude that (3.65) holds and that the process Y_T^* is tight. The lemma is established. ∎

Lemma 3.21 *Under Conditions 1.1-1.5, 2.1 and 3.1-3.5,*

$$\frac{1}{\sqrt{T}} Z'_{[\tau T]} M_\iota u^* \overset{p}{\Longrightarrow} \Omega^{\frac{1}{2}} W(\tau) .$$

Proof. Lemma 3.20 implies that it is sufficient to prove that

$$\frac{1}{\sqrt{T}} Z'_{[\tau T]} M_\iota u^* - \frac{1}{\sqrt{T}} \sum_{j=1}^{T-1} \bar{A}_j B_j \bar{w}_{\xi([\tau T]),j} w_{\varepsilon^*,j} \overset{p}{\Longrightarrow} 0. \qquad (3.68)$$

We prove (3.68) in three steps. In the first step, we prove that

$$\frac{1}{\sqrt{T}} Z'_{[\tau T]} M_\iota u^* - \frac{1}{\sqrt{T}} Z'_{[\tau T]} M_\iota \breve{u}^* \overset{p}{\Longrightarrow} 0, \qquad (3.69)$$

where $\breve{u}^* = (\breve{u}_1^*, \ldots, \breve{u}_T^*)'$ and

$$\breve{u}_t^* = \sum_{l=0}^{M} b_l \sum_{s=1}^{T} \varepsilon_s^* \mathbb{I}\left(s = t - l \mod T\right), \qquad t = 1, \ldots, T.$$

Denote $Y_{1T}^*(\tau) = T^{-1/2} Z'_{[\tau T]} M_\iota \left(u^* - \breve{u}^*\right)$. To prove that $Y_{1T}^* \overset{p}{\Longrightarrow} 0$, we need to show that $Y_{1T}^*(\tau) \overset{p^*}{\to} 0$ for every $\tau \in [0,1]$ and that the process Y_{1T}^* is tight. To this end, we examine $E^* \left|Y_{1T}^*(\tau) - Y_{1T}^*(\sigma)\right|^4$ for $(\tau, \sigma) \in [0,1]$. For $\sigma \leq \tau$, write

$$Y_{1T}^*(\tau) - Y_{1T}^*(\sigma) = a - b,$$

where

$$\begin{aligned} a &= \frac{1}{\sqrt{T}} \sum_{t=[\sigma T]+1}^{[\tau T]} x_t \sum_{l=0}^{M} \left(\hat{b}_l - b_l\right) \sum_{r=1}^{T} \varepsilon_r^* \mathbb{I}\left(r = t - l \mod T\right) \\ b &= \frac{1}{\sqrt{T}} \sum_{t=[\sigma T]+1}^{[\tau T]} x_t \sum_{l=0}^{M} \left(\hat{b}_l - b_l\right) \bar{\varepsilon}^* \end{aligned}$$

and where $\bar{\varepsilon}^* = T^{-1} \sum_{t=1}^{T} \varepsilon_t^*$. We have

$$E^* \varepsilon_t^* \varepsilon_s^* \varepsilon_r^* \varepsilon_v^* = \begin{cases} \hat{\kappa} & t = s = r = v, \\ \hat{\sigma}_\varepsilon^4 & t = s \neq r = v \text{ or } t = r \neq s = v \text{ or } t = v \neq s = r, \\ 0 & \text{otherwise} \end{cases}$$

and therefore

$$\begin{aligned} E^* a^4 &= \frac{1}{T^2} \sum_{t,s,t',s'=[\sigma T]+1}^{[\tau T]} x_t x_s x_{t'} x_{s'} \sum_{l,p,l',p'=0}^{M} \left(\hat{b}_l - b_l\right)\left(\hat{b}_p - b_p\right) \\ &\quad \times \left(\hat{b}_{l'} - b_{l'}\right)\left(\hat{b}_{p'} - b_{p'}\right) \\ &\quad \times \left\{\hat{\kappa} \mathbb{I}\left(t - l = s - p = t' - l' = s' - p' \mod T\right)\right. \\ &\quad \left. + 3\hat{\sigma}_\varepsilon^4 \mathbb{I}\left(t - l = s - p, t' - l' = s' - p' \mod T\right)\right\}. \qquad (3.70) \end{aligned}$$

The factor in the curly bracket on the right of (3.70) is bounded by

$$\left(\hat{\kappa} + 3\hat{\sigma}_\varepsilon^4\right) \mathbb{I}\left(t - l = s - p \mod T\right) \mathbb{I}\left(t' - l' = s' - p' \mod T\right)$$

and therefore (3.70) is bounded by

$$\begin{aligned}
&\left(\hat{\kappa}+3\hat{\sigma}_{\varepsilon}^{4}\right)\left(\frac{1}{T}\sum_{t,s=[\sigma T]+1}^{[\tau T]} x_t x_s \sum_{l,p=0}^{M}\left(\hat{b}_l-b_l\right)\left(\hat{b}_p-b_p\right)\mathbb{I}\left(t-l=s-p \bmod T\right)\right)^2 \\
\leq\ &\left(\hat{\kappa}+3\hat{\sigma}_{\varepsilon}^{4}\right)\left(\frac{1}{T}\sum_{t=[\sigma T]+1}^{[\tau T]}\sum_{l=0}^{M} x_t^2\left(\hat{b}_l-b_l\right)^2\sum_{s=[\sigma T]+1}^{[\tau T]}\sum_{p=0}^{M}\mathbb{I}\left(t-s=l-p \bmod T\right)\right)^2 \\
\leq\ &\left(\hat{\kappa}+3\hat{\sigma}_{\varepsilon}^{4}\right)\left(\frac{1}{T}\sum_{t=[\sigma T]+1}^{[\tau T]} x_t^2\right)^2\left(M\sum_{l=0}^{M}\left(\hat{b}_l-b_l\right)^2\right)^2,
\end{aligned}$$

where the first inequality is due to the Cauchy-Schwarz inequality. Further,

$$E^*\left(\bar{\varepsilon}^*\right)^4=\frac{\hat{\kappa}}{T^3}+\frac{3\hat{\sigma}_{\varepsilon}^{4}}{T^2}$$

and so by the Cauchy-Schwarz inequality,

$$E^* b^4 \leq \left(\frac{\hat{\kappa}}{T^2}+\frac{3\hat{\sigma}_{\varepsilon}^{4}}{T}\right)\left(\frac{1}{T}\sum_{t=[\sigma T]+1}^{[\tau T]} x_t^2\right)^2\left(M\sum_{l=0}^{M}\left(\hat{b}_l-b_l\right)^2\right)^2.$$

Since $(a-b)^4 \leq 8\left(a^4+b^4\right)$, we obtain

$$\begin{aligned}
E^*\left|Y_{1T}^*(\tau)-Y_{1T}^*(\sigma)\right|^4 &\leq E^* a^4+E^* b^4 \\
&= O_p(1)\left(\frac{1}{T}\sum_{t=[\sigma T]+1}^{[\tau T]} x_t^2\right)^2\left(M\sum_{l=0}^{M}\left(\hat{b}_l-b_l\right)^2\right)^2
\end{aligned}$$

because $\hat{\kappa}$ and $\hat{\sigma}_{\varepsilon}^{4}$ are $O_p(1)$. By Conditions 1.1 and 3.1, $T^{-1}\sum_{t=[\sigma T]+1}^{[\tau T]} x_t^2 \Longrightarrow (\tau-\sigma)\,\sigma_x^2$ and by Lemma 3.13 and Condition 3.4,

$$M\sum_{l=0}^{M}\left(\hat{b}_l-b_l\right)^2=O_p\left(Mm^{-2}T\log T\right)=o_p(1).$$

Therefore

$$E^*\left|Y_{1T}^*(\tau)-Y_{1T}^*(\sigma)\right|^4=(\tau-\sigma)^2\,o_p(1). \tag{3.71}$$

Since $Y_{1T}^{*}(0)=0$, the Markov inequality and (3.71) imply that $Y_{1T}^{*}(\tau)$ converges to zero in probability for every $\tau\in[0,1]$. Further, by the Cauchy-Schwarz inequality, for any $\sigma\leq\rho\leq\tau$,

$$\begin{aligned}E^{*}\left|Y_{1T}^{*}(\rho)-Y_{1T}^{*}(\sigma)\right|^{2}\left|Y_{1T}^{*}(\tau)-Y_{1T}^{*}(\rho)\right|^{2} &= o_{p}(1)\left((\rho-\sigma)^{2}(\tau-\rho)^{2}\right)^{\frac{1}{2}}\\ &= o_{p}(1)(\tau-\sigma)^{2}.\end{aligned}$$

By Theorem 13.5 and Condition (13.14) of Billingsley (1999), the process Y_{1T}^{*} is tight. We conclude that $Y_{1T}^{*}\overset{p}{\Longrightarrow}0$.

In the second step, we write $T^{-1/2}Z_{[\tau T]}'M_{\iota}\check{u}^{*}=T^{-1/2}\sum_{j=1}^{T-1}\tilde{B}_{j}\bar{w}_{x([\tau T]),j}w_{\varepsilon^{*},j}$, where $\tilde{B}_{j}$ is defined in Lemma 3.18, and show that

$$\frac{1}{\sqrt{T}}\sum_{j=1}^{T-1}\tilde{B}_{j}\bar{w}_{x([\tau T]),j}w_{\varepsilon^{*},j}-\frac{1}{\sqrt{T}}\sum_{j=1}^{T-1}\bar{A}_{j}\tilde{B}_{j}\bar{w}_{\xi([\tau T]),j}w_{\varepsilon^{*},j}\overset{p}{\Longrightarrow}0. \tag{3.72}$$

Denoting

$$Y_{2T}^{*}(\tau)=\frac{1}{\sqrt{T}}\sum_{j=1}^{T-1}\bar{A}_{j}\tilde{B}_{j}\left(\frac{\bar{w}_{x([\tau T]),j}}{\bar{A}_{j}}-\bar{w}_{\xi([\tau T]),j}\right)w_{\varepsilon^{*},j}$$

and proceeding as with expression (3.67), we obtain

$$E^{*}\left|Y_{2T}^{*}(\tau)-Y_{2T}^{*}(\sigma)\right|^{4}\leq\left(\hat{\kappa}+3\hat{\sigma}_{\varepsilon}^{4}\right)\left(\frac{1}{T}\sum_{j=1}^{T-1}\left|A_{j}\tilde{B}_{j}\right|^{2}\left|V_{j}\right|^{2}\right)^{2},$$

where $V_{j}=\left(w_{x([\tau T]),j}-w_{x([\tau T]),j}\right)/A_{j}-\left(w_{\xi([\tau T]),j}-w_{\xi([\sigma T]),j}\right)$. In the proof of Lemma 1.7, it has been shown that

$$E\left|V_{j}V_{k}\right|^{2}\leq C(\tau-\sigma)^{2}\frac{1}{jk}$$

uniformly over $(\tau,\sigma)\in[0,1]^{2}$ and $1\leq j,k\leq[T/2]$. This means that

$$E\left(\frac{1}{T}\sum_{j=1}^{T-1}\left|A_{j}B_{j}\right|^{2}\left|V_{j}\right|^{2}\right)^{2}\leq C(\tau-\sigma)^{2}\left(\frac{1}{T}\sum_{j=1}^{T-1}\frac{1}{j}\left|A_{j}\tilde{B}_{j}\right|^{2}\right)^{2}.$$

By the Cauchy-Schwarz inequality,

$$\frac{1}{T}\sum_{j=1}^{T-1}\frac{1}{j}\left|A_j\tilde{B}_j\right|^2 \leq \frac{2}{T}\sum_{j=1}^{T-1}\frac{1}{j}\left|A_jB_j\right|^2+\frac{2}{T}\sum_{j=1}^{T-1}\left|A_j\right|^2\left|\tilde{B}_j-B_j\right|^2. \tag{3.73}$$

By Condition 1.4, the first term on the right of (3.73) is bounded by

$$\frac{C}{T}\sum_{j=1}^{T-1}\lambda_j^{-2(d_x+d)}j^{-1} \leq CT^{-1+2(d_x+d)}\sum_{j=1}^{T-1}j^{-1-2(d_x+d)} \leq CT^{-1+2(d_x+d)} = o(1)$$

because $d_x+d<1/2$. By Lemma 3.19, the second term on the right of (3.73) is $o(1)$. It follows that

$$E\left(\frac{1}{T}\sum_{j=1}^{T-1}\left|A_jB_j\right|^2\left|V_j\right|^2\right)^2 = (\tau-\sigma)^2 o(1)$$

and therefore by the Markov inequality that

$$E^*\left|Y_{2T}^*(\tau)-Y_{2T}^*(\sigma)\right|^4 = (\tau-\sigma)^2 o_p(1).$$

Proceeding as with the process Y_{1T}^*, we conclude that $Y_{2T}^* \overset{p}{\Longrightarrow} 0$.

The third and final step is to prove that

$$\frac{1}{\sqrt{T}}\sum_{j=1}^{T-1}\bar{A}_j\tilde{B}_j\bar{w}_{\xi([\tau T]),j}w_{\varepsilon^*,j}-\frac{1}{\sqrt{T}}\sum_{j=1}^{T-1}\bar{A}_jB_j\bar{w}_{\xi([\tau T]),j}w_{\varepsilon^*,j} \overset{p}{\Longrightarrow} 0, \tag{3.74}$$

where $B_j=\sum_{l=0}^{\infty}b_le^{il\lambda_j}$. We define

$$Y_{3T}^*(\tau)=\frac{1}{\sqrt{T}}\sum_{j=1}^{T-1}\bar{A}_j\left(\tilde{B}_j-B_j\right)\bar{w}_{\xi([\tau T]),j}w_{\varepsilon^*,j}$$

and note as before that $E^*\left|Y_{3T}^*(\tau)-Y_{3T}^*(\sigma)\right|^4$ is bounded by

$$\left(\hat{\kappa}+3\hat{\sigma}_\varepsilon^4\right)\left(\frac{1}{T}\sum_{j=1}^{T-1}\left|A_j\right|^2\left|\tilde{B}_j-B_j\right|^2\left|\bar{w}_{\xi([\tau T]),j}-\bar{w}_{\xi([\sigma T]),j}\right|^2\right)^2.$$

From the proof of Lemma 1.7 we know that

$$E\left(\left|\bar{w}_{\xi([\tau T]),j}-\bar{w}_{\xi([\sigma T]),j}\right|^2\left|\bar{w}_{\xi([\tau T]),k}-\bar{w}_{\xi([\sigma T]),k}\right|^2\right)\leq C\left(\tau-\sigma\right)^2$$

uniformly in $(\tau,\sigma)\in[0,1]^2$ and $1\leq j,k\leq[T/2]$, therefore

$$\begin{aligned}
& E\left(\frac{1}{T}\sum_{j=1}^{T-1}|A_j|^2\left|\tilde{B}_j-B_j\right|^2\left|\bar{w}_{\xi([\tau T]),j}-\bar{w}_{\xi([\sigma T]),j}\right|^2\right)^2 \\
\leq\ & C\left(\tau-\sigma\right)^2\left(\frac{1}{T}\sum_{j=1}^{T-1}|A_j|^2\left|\tilde{B}_j-B_j\right|^2\right)^2 \\
=\ & \left(\tau-\sigma\right)^2 o\left(1\right)
\end{aligned}$$

by Lemma 3.19. In the same way as in the previous steps, we conclude that $Y^*_{3T}\stackrel{p}{\Longrightarrow}0$. ■

Lemma 3.22 *Assume Conditions 1.1-1.5, 2.1 and 3.1-3.5 hold. Then*
(a) $T^{-\frac{1}{2}}Z'_kM_\iota u^*=O_{p^*}\left(1\right)$ *uniformly over* $1\leq k\leq T$,
(b) $\hat{Z}'_\Delta M_\iota u^*=O_{p^*}\left(1\right)$ *uniformly over* $\hat{N}\left(K\right)$.

Proof. (a) This bound follows from Lemma 3.21.
(b) We have

$$u^*_t=\sum_{l=0}^{M}\hat{b}_l\sum_{s=1}^{T}\varepsilon^*_s\mathbb{I}\left(s=t-l\mod T\right)$$

and

$$\hat{Z}'_\Delta M_\iota u^*=\sum_t^{\hat{k},k}x_t\left(u^*_t-\bar{u}^*\right).$$

Since

$$E^*u^*_tu^*_s=\hat{\sigma}^2_\varepsilon\sum_{l,p=0}^{M}\hat{b}_l\hat{b}_p\mathbb{I}\left(t-s=l-p\mod T\right),$$

it is

$$\operatorname{var}^*\sum_t^{\hat{k},k}x_tu^*_t=\hat{\sigma}^2_\varepsilon\sum_{t,s}^{\hat{k},k}x_tx_s\sum_{l,p=0}^{M}\hat{b}_l\hat{b}_p\mathbb{I}\left(t-s=l-p\mod T\right).$$

By the Cauchy-Schwarz inequality, the last display is bounded by

$$\hat{\sigma}_\varepsilon^2 \sum_{t=\hat{k}-K}^{\hat{k}+K} x_t^2 \sum_{l=0}^{M} \hat{b}_l^2 \sum_{s=\hat{k}-K}^{\hat{k}+K} \sum_{p=0}^{M} \mathbb{I}\left(t-s=l-p \mod T\right) \leq CK\hat{\sigma}_\varepsilon^2 \sum_{t=\hat{k}-K}^{\hat{k}+K} x_t^2 \sum_{l=0}^{M} \hat{b}_l^2.$$

By Lemma 3.15, $\hat{\sigma}_\varepsilon^2 = O_p(1)$. Further, $\sum_{t=\hat{k}-K}^{\hat{k}+K} x_t^2 = KO_p(1) = O_p(1)$ and

$$\sum_{l=0}^{M} \hat{b}_l^2 \leq 2\sum_{l=0}^{M} b_l^2 + 2\sum_{l=0}^{M}\left(\hat{b}_l - b_l\right)^2 = O_p(1) + o_p(1)$$

by the square summability of b_l and by Lemma 3.13, and so $\operatorname{var}^* \sum_t^{\hat{k},k} x_t u_t^*$ is $O_p(1)$.

Next,

$$\operatorname{var}^* \sum_t^{\hat{k},k} x_t \bar{u}^* = \sum_{t,s}^{\hat{k},k} x_t x_s E^* \bar{u}^{*2} = o_{p^*}(1)$$

because $\sum_t^{\hat{k},k} x_t = O_{p^*}(1)$ uniformly on $\hat{N}(K)$ and because

$$\begin{aligned}
\operatorname{var}^* \bar{u}^* &= \frac{1}{T}\left(\sum_{l=0}^{M} \hat{b}_l\right)^2 \hat{\sigma}_\varepsilon^2 \leq \frac{M}{T}\sum_{l=0}^{M} \hat{b}_l^2 \hat{\sigma}_\varepsilon^2 \\
&\leq \frac{2\hat{\sigma}_\varepsilon^2}{m}\left(\sum_{l=0}^{M} b_l^2 + \sum_{l=0}^{M}\left(\hat{b}_l - b_l\right)^2\right) = o_p(1).
\end{aligned}$$

These results imply that $\hat{Z}_\Delta' M_\iota u^* = O_{p^*}(1)$ uniformly on $\hat{N}(K)$. ■

Lemma 3.23 *Assume Conditions 1.1-1.5, 2.1 and 3.1-3.5 hold. Let* $\overline{\breve{u}}^* = T^{-1}\sum_{t=1}^{T} \breve{u}_t^*$. *Then for any* $\alpha, \varepsilon, \eta > 0$, *there exists* K *and* T_0 *such that for all* $T \geq T_0$,

$$P\left(P^*\left(\sup_{K\leq k\leq T} \frac{1}{k}\left|\sum_{t=1}^{k} x_t\left(\breve{u}_t^* - \overline{\breve{u}}^*\right)\right| > \alpha\right) > \eta\right) < \varepsilon. \tag{3.75}$$

Proof. Let $S_k^* = \sum_{t=1}^{k} x_t\left(\breve{u}_t^* - \overline{\breve{u}}^*\right)$ and let an event A_k^* be defined as

$$A_k^* = \left\{\frac{1}{k}\left|S_k^*\right| > \alpha, \frac{1}{j}\left|S_j^*\right| \leq \alpha \text{ for } m \leq j < k \middle|\, x_t, u_t\right\}.$$

Proceeding as in the proof of Theorem 1 of Kounias and Weng (1969) and in the proof of a version of the maximal inequality of Kuan and Hsu (1998), the conditional probability inside the outer bracket in (3.75) can be bounded by

$$\begin{aligned}
&\frac{1}{\alpha^2}\left(\frac{E^* S_K^{*2}}{K^2}+\sum_{t=K+1}^{T}\frac{1}{t^2}E^*\left(S_k^{*2}-S_{k-1}^{*2}\right)\left(1-\mathbb{I}\left(A_{k-1}^*\right)-\ldots-\mathbb{I}\left(A_m^*\right)\right)\right)\\
\leq\ &\frac{\hat\sigma_\varepsilon^2}{\alpha^2}D_K, \qquad (3.76)
\end{aligned}$$

where

$$\hat\sigma_\varepsilon^2 D_K=\frac{E^* S_K^{*2}}{K^2}+\sum_{t=K+1}^{T}\frac{1}{t^2}\left(E^* x_t^2\left(\breve u_t^*-\overline{\breve u}^*\right)^2+2\left(E^* x_t^2\left(\breve u_t^*-\overline{\breve u}^*\right)^2\right)^{\frac12}\left(E^* S_{t-1}^{*2}\right)^{\frac12}\right).$$

We have

$$E^*\left(\breve u_t^*-\overline{\breve u}^*\right)\left(\breve u_s^*-\overline{\breve u}^*\right)=\hat\sigma_\varepsilon^2\sum_{l,p=0}^{M}b_l b_p\mathbb{I}\left(t-s=l-p \mod T\right)-T\hat\sigma_\varepsilon^2\bar b^2,$$

where $\bar b=T^{-1}\sum_{l=0}^{M}b_l$, and so

$$E^* x_t^2\left(\breve u_t^*-\overline{\breve u}^*\right)^2=\hat\sigma_\varepsilon^2 x_t^2\left(\sum_{l=0}^{M}b_l^2-T\bar b^2\right)\leq C\hat\sigma_\varepsilon^2 x_t^2\sum_{l=0}^{M}b_l^2$$

by the Cauchy-Schwarz inequality. The square summability of b_l implies that

$$E\left|\hat\sigma_\varepsilon^{-2}E^* x_t^2\left(\breve u_t^*-\overline{\breve u}^*\right)^2\right|\leq C,$$

where the constant C does not depend on t. Further,

$$E^* S_k^{*2}=\hat\sigma_\varepsilon^2\sum_{t,s=1}^{k}x_t x_s\sum_{l,p=0}^{M}b_l b_p\mathbb{I}\left(t-s=l-p \mod T\right)-T\hat\sigma_\varepsilon^2\bar b^2\sum_{t,s=1}^{k}x_t x_s,$$

so that

$$\begin{aligned}
E\hat{\sigma}_{\varepsilon}^{-2}\left|E^{*}S_{k}^{*2}\right| &\leq \sum_{t,s=1}^{k}\sum_{j=0}^{\infty}|a_{j}|\left|a_{j+|t-s|}\right|\sum_{l,p=0}^{M}|b_{l}|\,|b_{p}|\,\mathbb{I}\left(t-s=l-p \mod T\right) \\
&\quad +C\frac{k}{m}\sum_{t=1}^{k}Ex_{t}^{2} \\
&\leq C\sum_{t,s=1}^{k}\sum_{j=0}^{\infty}\sum_{l=0}^{M}|a_{j}|\left|a_{j+|t-s|}\right|\,|b_{l}|\left|b_{l+|t-s|}\right|+C\frac{k^{2}}{m}.
\end{aligned}$$

Arguing as in the proof of Lemma 2.5, it can be shown that

$$\sum_{t,s=1}^{k}\sum_{j=0}^{\infty}\sum_{l=0}^{M}|a_{j}|\left|a_{j+|t-s|}\right|\,|b_{l}|\left|b_{l+|t-s|}\right|\leq Ck$$

where the constant C does not depend on k, and therefore that

$$E\hat{\sigma}_{\varepsilon}^{-2}\left|E^{*}S_{k}^{*2}\right|\leq C\left(k+\frac{k^{2}}{m}\right).$$

There results imply that

$$\begin{aligned}
E\left|D_{K}\right| &\leq C\frac{K+m^{-1}K^{2}}{K^{2}}+C\sum_{t=K+1}^{T}\frac{1}{t^{2}}\left(1+t^{\frac{1}{2}}+m^{-\frac{1}{2}}t\right) \\
&\leq C\left(K^{-\frac{1}{2}}+m^{-\frac{1}{2}}\log m\right).
\end{aligned}$$

By the Markov inequality, the left-hand side of (3.75) is bounded by

$$\frac{2}{\eta}\frac{\sigma_{\varepsilon}^{2}}{\alpha^{2}}E\left|D_{K}\right|+P\left(\hat{\sigma}_{\varepsilon}^{2}>2\sigma_{\varepsilon}^{2}\right).$$

By Lemma 3.15, the second term of the last displayed sum is smaller than $\varepsilon/2$ for large T. It follows that for large T and K, the left-hand side of (3.75) is smaller than ε. ■

Lemma 3.24 *Assume Conditions 1.1-1.5, 2.1 and 3.1-3.5 hold. Then*
(a) $\sup_{k\in T\cdot\Lambda}u^{*\prime}(M_{\iota,W_{\hat{k}}}-M_{\iota,W_{k}})u^{*}=O_{p^{*}}(1)$,
(b) $\sup_{\hat{N}(K)}u^{*\prime}(M_{\iota,W_{\hat{k}}}-M_{\iota,W_{k}})u^{*}=O_{p^{*}}\left(T^{-1/2}\right)$.

Proof. Define

$$\hat{W}_\Delta = (W_k - W_{\hat{k}})\,\mathrm{sgn}\left(k - \hat{k}\right) = \left(0, \hat{Z}_\Delta\right).$$

Proceeding as in the proof of Lemma 2.9, write

$$\begin{aligned}
& u^{*\prime}(M_{\iota,W_{\hat{k}}} - M_{\iota,W_k})u^* \\
= \;& u^{*\prime}M_\iota \hat{W}_\Delta \left(W_{\hat{k}}' M_\iota W_{\hat{k}}\right)^{-1} W_{\hat{k}}' M_\iota u^* + u^{*\prime} M_\iota W_{\hat{k}} \left(W_{\hat{k}}' M_\iota W_{\hat{k}}\right)^{-1} \hat{W}_\Delta' M_\iota u^* \\
& + u^{*\prime} M_\iota \hat{W}_\Delta \left(W_{\hat{k}}' M_\iota W_{\hat{k}}\right)^{-1} \hat{W}_\Delta' M_\iota u^* \\
& - u^{*\prime} M_\iota W_k \left(W_k' M_\iota W_k\right)^{-1} \left(W_{\hat{k}}' M_\iota \hat{W}_\Delta + \hat{W}_\Delta' M_\iota W_{\hat{k}} + \hat{W}_\Delta' M_\iota \hat{W}_\Delta\right) \\
& \times \left(W_{\hat{k}}' M_\iota W_{\hat{k}}\right)^{-1} W_k' M_\iota u^*.
\end{aligned}$$

By Lemma 3.22, $W_k' M_\iota u^*$ and $\hat{W}_\Delta' M_\iota u^*$ are $O_{p^*}\left(T^{1/2}\right)$ uniformly over $1 \leq k \leq T$. In addition, $\hat{W}_\Delta' M_\iota u^*$ is $O_{p^*}(1)$ uniformly on $\hat{N}(K)$. By Lemma 2.8, $W_{\hat{k}}' M_\iota \hat{W}_\Delta$ and $\hat{W}_\Delta' M_\iota \hat{W}_\Delta$ are $O_p(1)$ uniformly on $\hat{N}(K)$ and $\left(W_{\hat{k}}' M_\iota W_{\hat{k}}\right)^{-1} = O_p\left(T^{-1}\right)$. By Lemma 2.2, $W_k' M_\iota W_k = O_p(T)$ uniformly over $1 \leq k \leq T$ and $\left(W_k' M_\iota W_k\right)^{-1} = O_p\left(T^{-1}\right)$ uniformly on $T \cdot \Lambda$ and therefore also $W_{\hat{k}}' M_\iota \hat{W}_\Delta$ and $\hat{W}_\Delta' M_\iota \hat{W}_\Delta$ are $O_p(1)$ uniformly over $1 \leq k \leq T$. Stochastic magnitude of the individual factors give the stochastic magnitude of $u^{*\prime}(M_{\iota,W_{\hat{k}}} - M_{\iota,W_k})u^*$ for $k \in T \cdot \Lambda$ and for $k \in \hat{N}(K)$ in part (a) and (b) of the lemma, respectively. ■

Lemma 3.25 *Assume Conditions 1.1-1.5, 2.1 and 3.1-3.5 are satisfied. Then for any finite* $K > 0$*,*

$$S_T^*(k) = \delta' \left(Z_k - Z_{\hat{k}}\right)' \left(Z_k - Z_{\hat{k}}\right) \delta - 2\delta' \left(Z_k - Z_{\hat{k}}\right)' u^* + o_{p^*}(1)$$

uniformly on $\hat{N}(K)$*.*

Proof. From the proof of Proposition 3.3, we have

$$S_T^*(k) = \hat{\delta}' Z_{\hat{k}}' M_{\iota,W_k} Z_{\hat{k}} \hat{\delta} + 2\hat{\delta}' Z_{\hat{k}}' M_{\iota,W_k} u^* + u^{*\prime}(M_{\iota,W_k} - M_{\iota,W_{\hat{k}}})u^*.$$

Write

$$Z_{\hat{k}}' M_{\iota,W_k} Z_{\hat{k}} = \hat{Z}_\Delta' M_\iota \hat{Z}_\Delta - \hat{Z}_\Delta' M_\iota W_k \left(W_k' M_\iota W_k\right)^{-1} W_k' M_\iota \hat{Z}_\Delta. \tag{3.77}$$

By the arguments employed earlier, the second term on the right of (3.77) can be shown to be $O_p(T^{-1})$. The first term on the right of (3.77) is equal to

$$\hat{Z}'_{\Delta}\hat{Z}_{\Delta} - \frac{1}{T}\sum_{t,s}^{\hat{k},k} x_t x'_s = \hat{Z}'_{\Delta}\hat{Z}_{\Delta} + O_p(T^{-1})$$

uniformly on $\hat{N}(K)$ because $\sum_t^{\hat{k},k} x_t = O_p(1)$ on $\hat{N}(K)$.

In a similar way, write

$$Z'_{\hat{k}} M_{\iota,W_k} u^* = (Z_{\hat{k}} - Z_k)' M_\iota u^* - (Z_{\hat{k}} - Z_k)' M_\iota W_k (W'_k M_\iota W_k)^{-1} W'_k M_\iota u^*. \tag{3.78}$$

The second term on the right of (3.78) is $O_p(1)\,O_p(T^{-1})\,O_{p^*}(T^{1/2}) = O_{p^*}(T^{-1/2})$. The first term on the right of (3.78) is equal to

$$(Z_{\hat{k}} - Z_k)' u^* + \sum_t^{\hat{k},k} x_t \bar{u}^* = (Z_{\hat{k}} - Z_k)' u^* + o_{p^*}(1)$$

by the arguments employed in the proof of Lemma 3.22.

The third term of S_T^* is $O_{p^*}(T^{-1/2})$ uniformly on $\hat{N}(K)$ by Lemma 3.24 and therefore

$$S_T^*(k) = \hat{\delta}'(Z_k - Z_{\hat{k}})'(Z_k - Z_{\hat{k}})\hat{\delta} - \hat{\delta}'(Z_{\hat{k}} - Z_k)' u^* + o_{p^*}(1).$$

Since $\hat{\delta} = \delta + O_p(T^{-1/2})$ by Proposition 2.2, $\sup_{\hat{N}(K)} \|Z_k - Z_{\hat{k}}\| = O_p(1)$ and $\sup_{\hat{N}(K)} \left\|(Z_{\hat{k}} - Z_k)'\hat{\delta}\right\| = O_p(1)$, the lemma is established. ■

References

Adelman, I., 1965, Long cycles - fact or artifact? American Economic Review 55, 444-463.

Altissimo, F. and V. Corradi, 2003, Strong rules for detecting the number of breaks in a time series, Journal of Econometrics 117, 207-244.

An, H-Z., C. Zhao-Guo and E. J. Hannan, 1983, The maximum of periodogram, Journal of Multivariate Analysis 13, 383-400.

Andrews, D. W. K., 1993, Tests for parameter instability and structural change with unknown change point, Econometrica 61, 821-856.

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

Antoch, J. and M. Hušková, 1999, Estimators of changes, in Asymptotics, nonparametrics, and time series, edited by Subir Ghosh, Marcel Dekker, 533-578.

Antoch, J., M. Hušková and Z. Prášková, 1997, Effect of dependence on statistics for determination of change, Journal of Statistical Planning and Inference 60, 291-310.

Antoch, J., M. Hušková and N. Veraverbeke, 1995, Change-point problem and bootstrap, Nonparametric Statistics 5, 123-144.

Bai, J., 1994, Least squares estimation of a shift in linear processes, Journal of Time Series Analysis 15, 453-472.

Bai, J., 1997a, Estimation of a change point in multiple regression models, The Review of Economics and Statistics 79, 551-563.

Bai, J., 1997b, Estimating multiple breaks one at a time, Econometric Theory 13, 315-352.

Bai, J. and P. Perron, 1998, Estimating and testing linear models with multiple structural changes, Econometrica 66, 47-78.

Baillie, R. T., 1996, Long memory processes and fractional integration in econometrics, Journal of Econometrics 73, 5-59.

Banerjee, A. and G. Urga, 2005, Modelling structural breaks, long memory and stock market volatility: an overview, Journal of Econometrics 129, 1-34.

Beran, J., 1994, Statistics for long-memory processes, Chapman and Hall.

Berkes, I., L. Horváth, P. Kokoszka and Q.-M. Shao, 2006, On discriminating between long-range dependence and changes in mean, The Annals of Statistics 34, forthcoming.

Bhattacharya, P. K., 1987, Maximum likelihood estimation of a change-point in the distribution of independent random variables: General multiparameter case, Journal of Multivariate Analysis 23, 183-208.

Bhattacharya, R. N., V. K. Gupta and E. Waymire, 1983, The Hurst effect under trends, Journal of Applied Probability 20, 649-662.

Bickel, P. J. and D. A. Freedman, 1981, Some asymptotic theory for the bootstrap, The Annals of Statistics 9, 1196-1217.

Billingsley, P., 1999, Convergence of probability measures, second edition, Wiley.

Brillinger, D. R., 1981, Time series: Data analysis and theory, McGraw-Hill.

Brown, R. L., J. Durbin and J. M. Evans, 1975, Techniques for testing the constancy of regression relationships over time, Journal of the Royal Statistical Society, Series B, 37, 149-192.

Bühlmann, P., 1997, Sieve bootstrap for time series, Bernoulli 3, 123-148.

Bühlmann, P., 1998, Sieve bootstrap for smoothing in nonstationary time series, The Annals of Statistics 26, 48-83.

Carlstein, E., 1986, The use of subseries values for estimating the variance of a general statistic from a stationary sequence, The Annals of Statistics 14, 1171-1179.

Chu, C.-S., K. Hornik and C.-M. Kuan, 1994, MOSUM tests for parameter constancy, Biometrika 82, 603-617.

Dahlhaus, R. and D. Janas, 1996, A frequency domain bootstrap for ratio statistics in time series analysis, The Annals of Statistics 24, 1934-1963.

Davidson J. and T. Teräsvirta (editors), 2002, Long memory and non-linear time series, Journal of Econometrics 110, 105-435.

Davies, R. B. and D. S. Harte, 1987, Tests for Hurst effect, Biometrika 74, 95-101.

Diebold, F. X., L. E. Ohanian and J. Berkowitz, 1998, Dynamic equilibrium economies: A framework for comparing models and data, Review of Economic Studies 65, 433-451.

Diebold, F. X. and A. Inoue, 2001, Long memory and regime switching, Journal of Econometrics 105, 131-159.

Dufour, J.-M. and E. Ghysels (editors), 1996, Recent developments in the econometrics of structural change, Journal of Econometrics 70, 1-316.

Efron, B., 1979, Bootstrap methods: Another look at the jackknife, The Annals of Statistics 7, 1-26.

Efron, B. and R. Tibshirani, 1986, Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy, Statistical Science 1, 54-77.

Eicker, F., 1979, The asymptotic distribution of the suprema of the standardized empirical processes, The Annals of Statistics 7, 116-138.

Engle, R. F. and A. D. Smith, 1999, Stochastic permanent breaks, Review of Econometrics and Statistics 81, 553-574.

Fiteni, I., 2002, Robust estimation of structural break points, Econometric Theory 18, 349-386.

Fiteni, I., 2004, τ-estimators of regression models with structural change of unknown location, Journal of Econometrics 119, 19-44.

Franke, J., and W. Härdle, 1992, On bootstrapping kernel spectral estimates, The Annals of Statistics 20, 121-145.

Freedman, D. A., 1984, On bootstrapping two-stage least-squares estimates in stationary linear models, The Annals of Statistics 12, 827-842.

Giné, E. and J. Zinn, 1990, Bootstrapping general empirical measures, The Annals of Probability 18, 851-869.

Giraitis, L. and D. Surgailis, 1990, A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotic normality of Whittle's estimate, Probability Theory and Related Fields 86, 87-104.

Granger, C. W. J., 1966, The typical spectral shape of an economic variable, Econometrica 34, 150-161.

Granger, C. W. J. and R. Joyeux, 1980, An introduction to long-memory time series models and fractional differencing, Journal of Time Series Analysis 1, 15-29.

Granger, C. W. J. and N. Hyung, 2001, Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns, Journal of Empirical Finance 11, 399-421.

Gray, H. L., N. Zhang and W. A. Woodward, 1989, On generalized fractional processes, Journal of Time Series Analysis 10, 233-257.

Hall, P., 1979, On the rate of convergence of normal extremes, Journal of Applied Probability 16, 433-439.

Hansen, B. E., 2000, Testing for structural change in conditional models, Journal of Econometrics 97, 93-115.

Hidalgo, J., 2003a, An alternative bootstrap to moving blocks for time series regression models, Journal of Econometrics 117, 369-399.

Hidalgo, J. 2003b, Bootstrap test for breaks of a regression model with dependent data, manuscript, LSE, May.

Hidalgo, J., 2005, Uniform confidence bands for kernel regression estimates with dependent data, manuscript, LSE, November.

Hidalgo, J. and P. M. Robinson, 1996, Testing for structural change in a long-memory environment, Journal of Econometrics 70, 159-174.

Hidalgo, J. and P. M. Robinson, 2002, Adapting to unknown disturbance autocorrelation in regression with long memory, Econometrica 70, 1545-1581.

Hidalgo, J. and Y. Yajima, 2002, Prediction and signal extraction of strongly dependent processes in the frequency domain, Econometric Theory 18, 584-624.

Hinkley, D. V., 1970, Inference about the change-point in a sequence of random variables, Biometrika 57, 1-17.

Horváth, L. and P. Kokoszka, 1997, The effect of long-range dependence on changepoint estimators, Journal of Statistical Planning 64, 57-81.

Hosking, J. R. M., 1981, Fractional differencing, Biometrika 68, 165-176.

Hurst, H. E., 1951, Long-term storage capacity of reservoirs, Transactions of the American Society of Civil Engineers 116, 770-799.

Iacone, F., 2006, Local Whittle estimation of the memory parameter in presence of deterministic components, manuscript, LSE, May.

Jaeschke, D., 1979, The asymptotic distribution of the supremum of the standardized empirical distribution function on subintervals, The Annals of Statistics 7, 108-115.

Keenan, D. M., 1987, Limiting behavior of functionals of higher-order sample cumulant spectra, The Annals of Statistics 15, 134-151.

Kreiss, J.-P., 1988, Asymptotic statistical inference for a class of stochastic processes, Habilitationsschrift, Fachbereich Mathematik, University of Hamburg.

Kounias, E. G. and T.-S. Weng, 1969, An inequality and almost sure convergence, The Annals of Mathematical Statistics 40, 1091-1093.

Kuan, C.-M. and K. Hornik, 1995, The generalized fluctuation test: A unifying view, Econometric Reviews 14, 135-161.

Kuan, C.-M. and C.-C. Hsu, 1998, Change-point estimation of fractionally integrated processes, Journal of Time Series Analysis 19, 693-708.

Künsch, H., 1986, Discrimination between monotonic trends and long-range dependence, Journal on Applied Probability 23, 1025-1030.

Künsch, H. R., 1989, The jackknife and the bootstrap for general stationary observations, The Annals of Statistics 17, 1217-1241.

Magnus, J. R. and H. Neudecker, 1988, Matrix differential calculus with applications in statistics and econometrics, John Wiley and Sons.

Mandelbrot, B. B. and J. W. van Ness, 1968, Fractional Brownian motions, fractional noises and applications, SIAM Review 10, 422-437.

Moulines, E. and P. Soulier, 1999, Broadband log-periodogram regression of time series with long-range dependence, The Annals of Statistics 27, 1415-1439.

Nelson, C. R. and C. I. Plosser, 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.

Paparoditis, E. and D. N. Politis, 2000, The local bootstrap for kernel estimators under general dependence conditions, Annals of the Institute of Statistical Mathematics 52, 139-159.

Parke, W. R., 1999, What is fractional integration?, The Review of Economics and Statistics 81, 632-638.

Perron, P., 1989, The great crash, the oil price shock, and the unit root hypothesis, Econometrica 57, 1361-1401.

Perron, P., 2006, Dealing with structural breaks, Chapter 8 in Handbook of Econometrics, Vol. 1, Econometric theory, edited by T. C. Mills and K. Patterson, Palgrave Macmillan.

Picard, D., 1985, Testing and estimating change-points in time series, Advances in Applied Probability 17, 841-867.

Ploberger, W. and W. Krämer, 1990, The local power of the CUSUM of squares tests, Econometric Theory 6, 335-347.

Ploberger, W. and W. Krämer, 1992, The CUSUM test with OLS residuals, Econometrica 60, 271-285.

Ploberger, W., W. Krämer and K. Kontrus, 1989, A new test for structural stability in the linear regression model, Journal of Econometrics 40, 307-318.

Politis, D. N. and J. P. Romano, 1992, A general resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation, The Annals of Statistics 20, 1985-2007.

Quandt, R. E., 1960, Tests of the hypothesis that a linear regression system obeys two separate regimes, Journal of the American Statistical Association 55, 324-330.

Ramos, E., 1988, Resampling methods for time series, Technical Report ONR-C-2, Department of Statistics, Harvard University.

Robinson, P. M., 1994a, Time series with strong dependence, Chapter 2 in Advances in Econometrics, Vol. 1, edited by C. A. Sims, Cambridge University Press, 47-95.

Robinson, P. M., 1994b, Rates of convergence and optimal spectral bandwidth for long range dependence, Probability Theory and Related Fields 99, 443-473.

Robinson, P. M., 1995a, Log-periodogram regression of time series with long range dependence, The Annals of Statistics 23, 1048-1072.

Robinson, P. M., 1995b, Gaussian semiparametric estimation of long range dependence, The Annals of Statistics 23, 1630-1661.

Robinson, P. M., 1998, Inference-without-smoothing in the presence of non-parametric autocorrelation, Econometrica 66, 1163-1182.

Robinson, P. M., 2003, Long-memory time series, Chapter 1 in Time Series with Long Memory, edited by P. M. Robinson, Oxford University Press, 4-32.

Robinson, P. M. and F. J. Hidalgo, 1997, Time series regression with long-range dependence, The Annals of Statistics 25, 77-104.

Scott, D. J., 1973, Central limit theorems for martingales and for processes with stationary increments using a Skorokhod representation approach, Advances in Applied Probability 5, 119-137.

Sen, P. K., 1980, Asymptotic theory of some tests for a possible change in the regression slope occurring at an unknown time point, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 52, 203-218.

Shimotsu, K., 2005, Simple (but effective) tests of long memory versus structural breaks, manuscript, Queen's University, November.

Stock, J. H., 1994, Unit roots, structural breaks and trends, Chapter 46 in Handbook of Econometrics, Vol. 4, edited by R. F. Engle and D. L. McFadden, North-Holland, 2739-2841.

Taniguchi, M., 1982, On estimation of the integrals of the fourth order cumulant spectral density, Biometrika 69, 117-122.

van der Vaart, A. W., 1998, Asymptotic statistics, Cambridge University Press.

van der Vaart, A. W. and J. A. Wellner, 1996, Weak convergence and empirical processes (with applications to statistics), Springer.

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.

Yong, C. H., 1974, Asymptotic behaviour of trigonometric series, Chinese University of Hong Kong.

Zygmund, A., 2002, Trigonometric Series, third edition, Volumes I and II combined (reissue), Cambridge University Press.