CENTRE FOR ECONOMETRIC ANALYSIS
CEA@Cass

http://www.cass.city.ac.uk/cea/index.html

Cass Business School
Faculty of Finance
106 Bunhill Row
London EC1Y 8TZ

Locating Structural Change in Regression with Strongly Dependent Processes

Stepana Lazarova

CEA@Cass Working Paper Series

WP–CEA–08-2005

Locating structural change in regression with strongly dependent processes

Štěpána Lazarová*

Department of Economics, Queen Mary, University of London
Mile End Road, London E1 4NS, UK

October 18, 2005

Abstract

This paper considers estimation and testing of the time of the structural break in a regression with possibly strongly dependent data. Statistical properties of the estimator depend on the assumption made on the size of the break. Specifically, when the size of break is fixed, the asymptotic distribution depends on the entire joint distribution of the regressors and the error term. When the size of break is shrinking, the asymptotic distribution of breakpoint depends only on the second moments of data. When the break is weak, that is when its magnitude is shrinking too fast, the time of break is not estimable. For the shrinking break, hypotheses on the time of break can be carried out on the basis of the asymptotic distribution. However, since asymptotic tests do not always approximate finite sample tests well, we propose a valid bootstrap approximation of the asymptotic tests and demonstrate its performance in small samples by Monte Carlo simulation.

Keywords: Structural change, long memory, bootstrap.
JEL classification: C12, C13, C22

*This paper is a part of the PhD dissertation I am writing at London School of Economics. I gratefully acknowledge research support from the Denis Sargan Memorial Fund and from the Economic and Social Research Council through grant R000238212. I would like to thank my supervisor Javier Hidalgo, Liudas Giraitis and participants of the First Italian Congress of Econometrics and Empirical Economics in Venice in 2005 for helpful comments and suggestions. E-mail address: s.lazarova@qmul.ac.uk. Tel.: +44-20-7882-5086, fax: +44-20-8983-3580.

1 Introduction

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. On the other hand, if a presence of 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. 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 Perron (2005).

In the last decades, however, it has been recognised 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 paper 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 paper 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. 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, the paper shows that the asymptotic distribution of our breakpoint estimator is indeed invariant to the distribution of data.

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 paper is organized in the following way. Section 2 introduces a linear regression 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 3. Section 4 discusses the cases of weak break and no break. Section 5 comments on the difference in testing hypotheses about the time of break under fixed and shrinking break. In Section 6, a bootstrap approximation to the asymptotic test procedure is proposed. Section 7 reports the results of a small Monte Carlo simulation conducted to investigate small sample properties of the proposed bootstrap procedure. Section 8 concludes. The proofs are collected in Section 9 which refers to Section 10 for intermediate results.

In what follows, B denotes a p-dimensional vector of independent standard Brownian motion processes on $[0,1]$ and C stands for a generic constant. For a set S and a constant a, $S \cdot a = \{xa : x \in S\}$. For any real numbers a, b, $a \wedge b = \min\{a,b\}$. 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 $C > 0$. Notation $[\cdot]$ signifies integer part, $\mathbb{I}(\cdot)$ is the indicator function, "$\Longrightarrow$" denotes weak convergence in the space $D(S)$ of p-vectors of right-continuous functions on $S \subset \mathbb{R}$ with left-hand limits, endowed with the uniform metric $\rho(x,y) = \sup_{\tau \in S} \|x(\tau) - y(\tau)\|$ for $x, y \in D(S)$. Notation $\overset{d}{=}$ stands for equality of distribution. 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 a, b and c, we write $a = b \mod c$ if $a - b$ is divisible by c.

2 Linear regression with break

Consider the following linear regression with a break in the slope parameter:

$$y_t = \alpha + \beta' x_t + \delta_T' z_t + u_t \tag{1}$$

for $t = 1, \ldots, T$, where

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

and where k_0 is an unknown date of break, y_t is the observed dependent variable, α is an unknown intercept, β and δ_T are p-dimensional vectors of unknown parameters with $\delta_T \neq 0$, x_t is a p-dimensional vector of observations on the explanatory variables and u_t is an unobserved stochastic disturbance. 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, that is $\delta_T = \delta$.

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

$$\text{H}_0\text{: } k_0 = k_H$$

for some constant k_H against the alternative

$$\text{H}_1\text{: } k_0 \neq k_H.$$

In this paper, 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 (1) can be written in the matrix form as

$$y = \alpha\iota + X\beta + Z_0\delta_T + u \tag{2}$$

where $y = (y_1, \ldots, y_T)'$, $\iota = (1, \ldots, 1)'$, $X = (x_1, \ldots, x_T)'$ and where $Z_k = (x_1, \ldots, x_k, 0, \ldots, 0)'$ 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 $u = (u_1, \ldots, u_T)'$.

We estimate the parameters of the model by the least squares method. For $k = 1, \ldots, T-1$, let $\hat{u}(k)$ be the vector of residuals from the least squares regression of y on X and Z_k,

$$\hat{u}(k) = M_{\iota,X,Z_k} y,$$

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

$$\begin{pmatrix} \hat{\beta}_k \\ \hat{\delta}_k \end{pmatrix} = \left(W_k' M_\iota W_k\right)^{-1} W_k' M_\iota y,$$

where $W_k = (X, Z_k)$, $M_{A,B} = I - P_{A,B}$ where $P_{A,B} = P_C = C\left(C'C\right)^{-1} C'$ is the matrix of orthogonal projection on the column space of a matrix $C = (A, B)$. The least squares estimator $\hat{k}$ of the breakpoint k_0 is obtained by minimising the objective function

$$S_T(k) = ||\hat{u}(k)||^2, \tag{3}$$

that is

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

where $\Lambda \cdot T = \{k\text{: } k/T \in \Lambda\}$. If the point of minimum is not unique, we define $\hat{k} = \min\{k : S_T(k) = \min_{l \in \Lambda \cdot T} S_T(l)\}$. 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) 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 time to the frequency domain. In the frequency domain, model (1) is given by

$$w_y(\lambda_j) = \beta' w_x(\lambda_j) + \delta_T' w_{z(k_0)}(\lambda_j) + w_u(\lambda_j), \tag{4}$$

$j = 1, \ldots, T-1$, where

$$w_d(\lambda) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} d_t e^{it\lambda}$$

is the discrete Fourier transform of a sequence of p-dimensional vectors $d_1, \ldots, d_T$ and $\lambda_j = 2\pi j/T$ are the Fourier frequencies. Omission of the frequency zero in (4) permits the researcher to avoid estimating the unknown intercept α. As the discrete Fourier transform is invariant to location shift for $1 \leq j \leq T-1$, the regression (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 $\lambda_j = 2\pi j/T$,

$$F_{jk} = \frac{1}{\sqrt{2\pi T}} e^{ij\lambda_k}$$

for $j = 1, \ldots, T-1$ and $k = 1, \ldots, T$, model (4) can be written in the matrix form as

$$Fy = FX\beta + FZ_0\delta_T + Fu. \tag{5}$$

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

$$\widehat{Fu}(k) = M_{FX,FZ_k} Fy = M_{FW_k} Fy \tag{6}$$

be the vector of residuals and

$$\begin{pmatrix} \tilde{\beta}_k \\ \tilde{\delta}_k \end{pmatrix} = \left(W_k' \overline{F}' F W_k \right)^{-1} W_k' \overline{F}' F y,$$

be the estimators of the slope coefficients where now $P_A = A \left(\overline{A}' A \right)^{-1} \overline{A}'$ where $\overline{A}'$ 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

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

This implies that

$$\widehat{Fu}(k) = F\hat{u}(k),$$
$$||\widehat{Fu}(k)||^2 = \frac{1}{2\pi}||\hat{u}(k)||^2$$

for $k = 1, \ldots, T-1$. Therefore for the purpose of estimating of the time of break and 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 1-6:

Condition 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, and let $\mathcal{F}_t \vee \mathcal{G}_s$ be the union of $\mathcal{F}_t$ and $\mathcal{G}_s$, that is the smallest σ-algebra containing all elements of $\mathcal{F}_t$ and $\mathcal{G}_s$.

Condition 2 *$\{\xi_t\}$ is a stochastic process that satisfies*

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

(b) $E(\xi_t\xi_t'|\mathcal{F}_{t-1} \vee \mathcal{G}_t) = E(\xi_t\xi_t') = \Xi$ *a.s., and*

(c) the joint fourth cumulants of $\xi_{t_i j_i}$, $j_i = 1, \ldots p$ *and* $i = 1, \ldots 4$, *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 $|\kappa_\xi| = \max_{j_i=1,\ldots p, i=1,\ldots 4} |\kappa_{\xi, j_1, j_2, j_3, j_4}| < \infty$.

Condition 3 *$\{\varepsilon_t\}$ is a stochastic process that satisfies*

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

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

(c) the joint fourth cumulant of ε_{t_i}, $i = 1, \ldots 4$ *satisfies*

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

with $|\kappa| < \infty$.

Define functions A and B as

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

Condition 4 *The functions A and B satisfy the following assumptions:*

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

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

(c) $||A(\lambda)|| > 0$ *and* $|B(\lambda)| > 0$ *for* $\lambda \in (0, \pi]$.

Condition 5

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

(b) there exists $\lambda > 0$ *such that for every* $\varepsilon > 0$, *there exists* l_0 *such that* $P\left(\lambda_l^+ < \lambda\right) < \varepsilon$ *and* $P\left(\lambda_l^- < \lambda\right) < \varepsilon$ *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 6

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

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

Conditions 1-6 are similar to those of Robinson (1995a,b), Hidalgo (2003) and Lazarová (2004) and similar remarks apply. In particular, Conditions 1-3 require that the processes x_t and u_t are covariance stationary with constant fourth but not third moments. The homoskedasticity of regressors and errors implied by Conditions 1-3 is not a necessary restriction and a certain degree of heterogeneity could be allowed. Another property following from Conditions 1-3, $E(x_t x_s u_t u_s) = E(x_t x_s) E(u_t u_s)$ for all t and s, makes it possible to estimate the quantity $2\pi \int_{-\pi}^{\pi} f_{xx}(\lambda) f_{uu}(\lambda) d\lambda$ consistently by $(4\pi^2/T) \sum_{j=1}^{T-1} I_{xx}(\lambda_j) I_{uu}(\lambda_j)$ as proposed by Robinson (1998). Results in the literature, for example Taniguchi (1982) and Keenan (1987), allow to relax the condition $E(x_t x_s u_t u_s) = E(x_t x_s) E(u_t u_s)$ for short memory processes but similar results are not yet available for processes satisfying Conditions 1-4, in particular for processes that may exhibit long memory processes.

Condition 4 admits a singularity of the spectral density function at the zero frequency, but the results of this paper could be generalized to allow for a singularity at a nonzero frequency or for more than one singularity, as long as Condition 6 is satisfied. Examples of models satisfying Condition 4 include FARIMA model of Granger and Joyeux (1980) and Hosking (1981), fractional Gaussian noise of Mandelbrot and van Ness (1968) and fractional exponential model of Bloomfield (1973). An example of a model with singularities at nonzero frequencies is the Gegenbauer model of Gray, Zhang and Woodward (1989).

Condition 5 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 $\{x_t x_t'\}$.

Condition 6 places a restriction on the collective memory of regressors and errors. Robinson (1994) remarks that if x_t and u_t collectively exhibit sufficiently strong memory, the least squares estimates of the slope coefficients are not asymptotically normal. The restriction can 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 6 as it stands.

3 Asymptotic properties of the breakpoint and slope estimator

To establish 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 1 *Assume Conditions 1-6 are satisfied. If* $\delta_T = \delta \neq 0$ *or if* $\delta_T \to 0$ *and* $T \left|\left|\delta_T\right|\right|^2 \to \infty$*, then as* $T \to \infty$*,*

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

Whether the size of break δ is fixed or shrinking, the consistency of the estimator $\hat{\tau}$ of the relative time of break τ_0 is guaranteed by Proposition 1. It is interesting to note that Proposition 1 implies only that $\hat{k} - k_0 = o_p(T)$ and therefore does not assert the consistency of $\hat{k}$. In fact, it will be seen later that the estimator $\hat{k}$ of time of break is not consistent. At best, under the assumption of fixed size of break, the estimator $\hat{k}$ is bounded in probability.

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 as the sample size decreases at a moderate speed.

Proposition 2 *Assuming Conditions 1-6, if* $T \left|\left|\delta_T\right|\right|^2 \to \infty$*, then as* $T \to \infty$*,*

(a)

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

(b)

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

where

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

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

The limiting distribution of the slope estimators $\hat{\beta}$, $\hat{\delta}$ in the case of an 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.

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

$$W_0(s) = \delta' \sum_t^{0,s} x_t x_t' \delta - 2\delta' \sum_t^{0,s} x_t u_t \operatorname{sgn} s,$$

that is

$$W_0(s) = \begin{cases} \delta' \sum_{t=1}^{s} x_t x_t' \delta - 2\delta' \sum_{t=1}^{s} x_t u_t & s \geq 1, \\ 0 & s = 0 \\ \delta' \sum_{t=s+1}^{0} x_t x_t' \delta + 2\delta' \sum_{t=s+1}^{0} x_t u_t & s \leq -1. \end{cases}$$

The following proposition gives the asymptotic distribution of the breakpoint estimator for the case of a 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 1-6 to strict stationarity.

Proposition 3 *Assume that Conditions 1-6 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, $\delta_T = \delta$, then as $T \to \infty$,*

$$\hat{k} - k_0 \xrightarrow{d} \arg\min_s W_0(s).$$

The asymptotic distribution of the breakpoint estimator with break of fixed size therefore depends not only on the nuisance parameter δ but also on the distribution of x_t and u_t. While the size of jump δ can be consistently estimated by Proposition 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 beyond the scope of this paper.

It is worth noting that the distribution of 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 an 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. 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 assumptions on the size of the break. Consider the case of a diminishing magnitude of break. Define the process W as

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

where W_1, W_2 are independent standard Brownian motion processes defined on $[0, \infty)$.

Proposition 4 *Assume that Conditions 1-6 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 as* $T \to \infty$,

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

An example of consistent estimator of Σ is

$$\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^{T} x_t x_t', \tag{7}$$

whose consistency follows from Conditions 1 and 2. Let $I_{uv}(\lambda_j) = w_u(\lambda_j)\overline{w}_v(\lambda_j)$ be the cross-periodogram matrix of processes u_t and v_t. 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 5 *Assume that Conditions 1-6 hold and that $T^{1/2}\delta_T \to \delta \neq 0$ or $T^{1/2}\|\delta_T\| \to \infty$. Then as $T \to \infty$,*

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

The distribution of $\arg\min_\rho W(\rho)$ is not only free of nuisance parameters, but also the explicit form of the distribution function is known, see for example Yao (1987):

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

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.

4 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 so that $T\|\delta_T\|^2 = O(1)$.

For $\tau \in (0,1)$, define

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

where for $\delta \neq 0$ the function m is defined as

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

When $\delta = 0$, the function m can be left undefined. Further, 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 6 *Assume that Conditions 1-6 hold and that* $T^{1/2}\delta_T \to \delta$ *where* $0 \leq \|\delta\| < \infty$*. Then, as* $T \to \infty$*,*

$$\hat{\tau} = \frac{\hat{k}}{T} \xrightarrow{d} L.$$

Proposition 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 limit δ is equal to zero, Proposition 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 7 *Assume that Conditions 1-6 hold and that* $T^{1/2}\delta_T \to \delta$ *where* $0 \leq \|\delta\| < \infty$*. Then as* $T \to \infty$*,*

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

The random variable $B(L)$ 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 $P(B(L) \leq b)$ 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 $B(L) \leq b$ is to be taken componentwise.

Proposition 7 together with Proposition 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 continuous to hold under a weak break, the asymptotic normality does not. 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 faster 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 4 in Section 9, if $T^{1/2}\delta_T \to \delta$,

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

Therefore it can be seen that while tests based on continuous functionals of $S_T([\tau T])$ have nontrivial power against the alternatives with $T^{1/2}\delta_T \to \delta \neq 0$, the test have no power against the alternatives with $\delta_T = o_p\left(T^{-1/2}\right)$ 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 breaks with $T\left\|\delta_T\right\|^2 \to \infty$ can be detected and their location can be estimated, breaks with $T\left\|\delta_T\right\|^2 \to C$ are detectable but there location is not estimable, and breaks with $T\left\|\delta_T\right\|^2 \to 0$ cannot be detected.

5 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 breaks is assumed fixed, Proposition 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

$$H_1 : k_0 = k_H + \Delta \tag{8}$$

where $\Delta \neq 0$ is a constant, 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 (8). 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 $T\left\|\delta_T\right\|^2 \to \infty$, Proposition 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 4 then gives the asymptotic distribution of Z_T under the null. Under the alternative hypothesis (8),

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

The test based on Z_T has therefore asymptotically no power against the alternative that $k_0 = k_H + \Delta$.

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

$$H_1\colon\ k_0 = k_H + \Delta_T, \tag{9}$$

the test has asymptotic local power against such alternatives if $\Delta_T^{-1} = O\left(\left\|\delta_T\right\|^2\right)$. When $\Delta_T^{-1} = o\left(\left\|\delta_T\right\|^2\right)$, the test is consistent, or has a global power, against the alternative hypothesis (9). For example, if we consider an alternative hypothesis to be $H_1\colon\ k_0 = k_H + T\cdot\Delta$, which corresponds to the alternative fixed in terms of the relative time of the break, $H_1\colon\ \tau_0 = \tau_H + \Delta$, the test has global power since $\Delta_T^{-1} = T^{-1}\Delta^{-1} = o\left(||\delta_T||^2\right)$.

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 propose 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.

Since the independence assumption does not hold for time series, the basic bootstrap procedure of Efron (1979) cannot be employed without a modification. There exists a vast literature on bootstrapping time series. The methods of tackling the dependence of data may be broadly categorized as parametric or nonparametric. Parametric methods have been examined by Efron and Tibshirani (1993) among others. Examples of nonparametric bootstrap procedures can be found in articles of Carlstein (1986), Künsch (1989), Politis and Romano (1992) or Bühlmann (1997, 1998).

Bootstrap methods that resample from data transformed into the frequency domain can also be subsumed under the heading of nonparametric methods. Among these methods, mention can be made of the work of Franke and Härdle (1992), Dahlhaus and Janas (1996), Paparoditis and Politis (2000) and Hidalgo (2003). However, only the approach of Hidalgo (2003) has been shown to be valid for strongly dependent data. The method of Hidalgo (2003) has an added advantage that it does not require a user-chosen parameter such as the block length in the block bootstrap of Carlstein (1996) or a lag length in the sieve bootstrap of Bühlmann (1997, 1998).

To approximate the distribution of the breakpoint estimator, we propose to use a method similar to that employed by Hidalgo (2003) and Lazarová (2004). 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 (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

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

where $\{w_{\hat{z}}\left(\lambda_j\right), 1 \leq j \leq T\}$ is the discrete Fourier transform of the sequence $\{\hat{z}_t, 1 \leq t \leq T\}$. 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 $\hat{u}^*\left(k\right)$ be the vector of residuals from the least squares regression of Fy^* on FX and FZ_k. Let

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

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

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

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 (7) and where $\widehat{\Omega}^* = \frac{4\pi^2}{T}\sum_{j=1}^{T-1} I_{xx}\left(\lambda_j\right) I_{\hat{v}\hat{v}}\left(\lambda_j\right)$. Alternatively, compute a nonpivotal statistic

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

Before discussing the bootstrap procedure, it is useful to introduce some notation. Denote P^* to be 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, denote E^*, var^* and cov^* 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, let the statement

$$X_T \xrightarrow{d^*} X \tag{10}$$

be 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 variable X in (10) is a constant, write $X_T \xrightarrow{p^*} X$. Further, for a stochastic process Y and a sequence of stochastic processes $\{Y_T\}$, let "$Y_T \xRightarrow{p} Y$" stand for the weak convergence in probability as defined by Giné and Zinn (1990).

It is also useful to define stochastic orders of magnitude o_{p^*}, O_{p^*}. Let $\{g_T\}$ be a sequence of positive finite numbers. We say that $X_T = O_{p^*}(g_T)$ 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 $T \geq T_0$,

$$P(P^*(|X_T| > Mg_T) > \eta) < \varepsilon.$$

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

$$P^*(|X_T| > \eta g_T) = o_p(1).$$

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 $o_{p^*}(1) + o_p(1) = o_{p^*}(1)$.

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

Proposition 8 *Assume that Conditions 1-6 hold and that $\delta_T \to 0$ and $T\|\delta_T\|^2 \to \infty$. Then as $T \to \infty$,*

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

where V is defined in Proposition 2.

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

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

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

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

This observation is exploited in the following proposition asserting the consistency of bootstrap.

Proposition 9 *Assume that Conditions 1-6 hold and that $\delta_T \rightarrow 0$ and $T\left\|\delta_T\right\|^2 \rightarrow \infty$. Then, as $T \rightarrow \infty$,*

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

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, $P\left(Z_T \in C_\alpha\right) = \alpha$, is replaced by a critical region C_α^* based on the bootstrap distribution where C_α^* satisfies $P^*\left(Z_T^* \in C_{\alpha*}\right) = \alpha$. Proposition 9 guarantees that the bootstrap test has asymptotically correct size.

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 (1) are generated as scalar ARFIMA$(0, d, 0)$ processes where d is the long 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_u are considered in admissible combinations such that $0 \leq d_x + d_u < 1/2$. Samples of size $T = 64$, 128, 256 and 512 are generated by the algorithm of Davis 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 $\left(\hat{\delta}^{*\prime}\hat{\Sigma}\hat{\delta}^*\right)^2 / \left(\hat{\delta}^{*\prime}\hat{\Omega}^*\hat{\delta}^*\right)\left(\hat{k}^* - \hat{k}\right)$, and an asymptotic test based on the limiting distribution of the test statistic $\left(\hat{\delta}'\hat{\Sigma}\hat{\delta}\right)^2 / \left(\hat{\delta}'\hat{\Omega}\hat{\delta}\right)\left(\hat{k} - k_0\right)$ 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 1 reports the size 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. Power of the tests is given in Table 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.

8 Conclusions

In this paper, statistical properties of estimators of location of a structural change are examined in the context of a linear regression under mild conditions on regressors and errors. 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 a 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 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 paper 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 article 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,

		Nonpivotal bootstrap test			Pivotal bootstrap test			Asymptotic test		
d_x	d_u	**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 1: Size of the bootstrap and asymptotic tests

		Nonpivotal bootstrap test			Pivotal bootstrap test			Asymptotic test		
d_x	d_u	**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: Power of the bootstrap and asymptotic tests

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). Finally, for the case of a fixed magnitude of break, it may be 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. These topics are left for possible future research.

9 Proofs of Propositions

This section contains the proofs of the results in the main body of text. Throughout this section, it is assumed that Conditions 1-6 hold.

Let $N(K) = \left\{k\colon\ |k - k_0| \le K\,||\delta_T||^{-2}\right\}$ and $N^c(K) = \Lambda \cdot T - N(K)$. For integers l, m define

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

and denote $Z_\Delta = Z_{\Delta(k_0,k)}$. 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 $\iota = \iota_T$. Denote $\iota_0 = \iota_{k_0}$ and $\iota_\Delta = (\iota_k - \iota_0)\,\mathrm{sgn}\,(k - k_0)$. Further, 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

$$S_T(k) - S_T(k_0) = Q_T(k) + R_T(k)\,. \tag{11}$$

For a generic function f, we abbreviate $f(\lambda_j)$ by f_j.

Proof of Proposition 1. For any $\lambda > 0$,

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

$$\le P\left(\inf_{N^c(K)} \frac{Q_T(k)}{|k - k_0|} \le \lambda\,||\delta_T||^2\right) + P\left(\sup_{N^c(K)} \left|\frac{R_T(k)}{k - k_0}\right| \ge \lambda\,||\delta_T||^2\right). \tag{12}$$

Lemma 4 implies that λ can be chosen such that the first term on the right of (12) is smaller than $\varepsilon/2$ for large K. We now show that the second term on the right of (12) is smaller than $\varepsilon/2$ for large K. To that end, write

$$\begin{aligned} R_T(k) &= -2\delta_T' Z_\Delta' M_\iota u \,\mathrm{sgn}(k-k_0) \\ &\quad +2\delta_T' Z_\Delta' M_\iota W_k (W_k' M_\iota W_k)^{-1} W_k' M_\iota u \,\mathrm{sgn}(k-k_0) \\ &\quad +u'(M_{\iota,X,Z_k} - M_{\iota,X,Z_0}) u. \end{aligned} \tag{13}$$

The contribution of the first term of (13) to (12) 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 5. Regarding the second term of (13), $\sup_{N^c(K)}\|Z_\Delta' M_\iota W_k/(k-k_0)\| = O_p(1)$, $(W_k' M_\iota W_k)^{-1} = O_p(T^{-1})$ and $W_k' M_\iota u = O_p\left(T^{\frac{1}{2}}\right)$ uniformly on $\Lambda\cdot T$ by Lemma 2, and $T^{-1/2}\|\delta_T\| = o\left(\|\delta_T\|^2\right)$. Therefore the contribution of this term to (12) is bounded by $\varepsilon/6$ for large K and T. The contribution of the third term of (13) to (12) 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 3. This concludes the proof that for large K the second term on the right of (12) is bounded by $\varepsilon/2$. The proposition is proved. ■

Proof of Proposition 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,

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

and

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

where B is a p-vector of independent standard Brownian motion processes on $[0,1]$. 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 $M>0$,

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

By Proposition 1, the second term on the right side 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. It follows that $\left(Z_{\hat{k}}-Z_0\right)'M_\iota\left(Z_{\hat{k}}-Z_0\right)=O_p\left(\left|\left|\delta_T\right|\right|^{-2}\right)$. 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 $O_p\left(\left|\left|\delta_T\right|\right|^{-2}\right)$. Therefore

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

and by Lemma 2,

$$\left(\frac{1}{T}W_{\hat{k}}'M_\iota W_{\hat{k}}\right)^{-1}\xrightarrow{p}\begin{pmatrix}1&\tau_0\\ \tau_0&\tau_0\end{pmatrix}^{-1}\otimes\Sigma^{-1}. \tag{14}$$

By the same arguments,

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

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

$$\frac{1}{\sqrt{T}}W_{\hat{k}}'M_\iota u=\frac{1}{\sqrt{T}}W_0'M_\iota u+\frac{1}{\sqrt{T}}\left(W_{\hat{k}}-W_0\right)'M_\iota u \tag{15}$$

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

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

The terms on the right of the last display are bounded by $\varepsilon/2$ for large M and K by Lemma 2 and Proposition 1, respectively. Therefore $T^{-1/2}(Z_{\hat{k}} - Z_0)' M_\iota u$ is $O_p\left(T^{-1/2}\|\delta_T\|^{-1}\right) = o_p(1)$ and by (15) and Lemma 2,

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

Summarising,

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

as maintained by the proposition. ■

Proof of Proposition 3. Let $\tilde{k} = \arg\min_{N(K)} S_T(k)$, $\hat{m} = \arg\min_m W_0(m)$ and $\tilde{m} = \arg\min_{|m| \leq K} W_0(m)$. For any $K > 0$ and for any integer j,

$$
P\left(\hat{k} - k_0 = j\right) = P\left(\hat{k} - k_0 = j, \left|\hat{k} - k_0\right| \leq K\right) + P\left(\hat{k} - k_0 = j, \left|\hat{k} - k_0\right| > K\right). \tag{16}
$$

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 (16) is equal to

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

By similar arguments $P(\hat{m} = j) = P(\tilde{m} = j) P(|\hat{m}| \leq K) + P(\hat{m} = j, |\hat{m}| > K)$. Therefore

$$
\begin{aligned}
\left|P\left(\hat{k} - k_0 = j\right) - P(\hat{m} = j)\right| \leq\; & \left|P\left(\tilde{k} - k_0 = j\right) - P(\tilde{m} = j)\right| \\
& + 2P\left(\left|\hat{k} - k_0\right| > K\right) + 2P(|\hat{m}| > K).
\end{aligned} \tag{17}
$$

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 6 implies that

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

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 (17) 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 (17) is smaller than $\varepsilon/3$ for large K by Proposition 1. Since by Conditions 1 and 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 (17) is smaller than $\varepsilon/3$ for large K. It follows that for each j, $P\left(\hat{k} - k_0 = j\right) - P(\hat{m} = j) \to 0$. ■

Proof of the Proposition 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,

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

where v_T^2 has been defined in Lemma 7. Let $\tilde{k} = \arg\min_{|k-k_0|\leq K v_T^{-2}} S_T(k)$, $\hat{\rho} = \arg\min_{\rho\in\mathbb{R}} W(\rho)$ and $\tilde{\rho} = \arg\min_{\rho\in[-K,K]} W(\rho)$. Reasoning as in the proof of Proposition 3, we obtain

$$\begin{aligned} \left|P\left(v_T^2\left(\hat{k} - k_0\right) \leq x\right) - P(\hat{\rho} \leq x)\right| &\leq \left|P\left(v_T^2\left(\tilde{k} - k_0\right) \leq x\right) - P(\tilde{\rho} \leq x)\right| \\ &\quad + 2P\left(\left|\hat{k} - k_0\right| > K v_T^{-2}\right) + 2P(|\hat{\rho}| > K). \end{aligned} \tag{18}$$

Since $v_T^2\left(\tilde{k} - k_0\right) = \arg\min_{|\rho|\leq\left[K v_T^{-2}\right]/v_T^{-2}} S_T\left(k_0 + \rho v_T^{-2}\right)$, Lemma 7 implies that $v_T^2\left(\tilde{k} - k_0\right) \xrightarrow{d} \arg\min_{|\rho|\leq K} W(\rho)$. Therefore the first term of (18) 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 (18) are smaller than $\varepsilon/3$ for large K and T. Inequality (18) then implies that

$$\frac{\left(\delta_T' \Sigma \delta_T\right)^2}{\delta_T' \Omega \delta_T}\left(\hat{k} - k_0\right) \xrightarrow{d} \arg\min_{\rho} W(\rho). \tag{19}$$

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 (19) remains valid with the quantities δ_T, Σ and Ω replaced by their estimators $\hat{\delta}$, $\hat{\Sigma}$ and $\hat{\Omega}$. ■

Proof of Proposition 5. Assume for simplicity that $\{x_t\}$ is a scalar process. By Theorem 1 of Robinson (1998),

$$\frac{4\pi^2}{T}\sum_{j=1}^{T-1} I_{xx,j}I_{uu,j} \xrightarrow{p} \Omega. \tag{20}$$

It is therefore sufficient to prove that

$$\frac{1}{T}\sum_{j=1}^{T-1} I_{xx,j}\left(I_{\hat{u}\hat{u},j} - I_{uu,j}\right) \xrightarrow{p} 0. \tag{21}$$

Using the fact that $I_{\hat{u}\hat{u},j} - I_{uu,j} = \left|w_{\hat{u},j} - w_{u,j}\right|^2 - 2\operatorname{Re}\left(w_{\hat{u},j} - w_{u,j}\right)\overline{w}_{u,j}$, we obtain that the left side of (21) is bounded on absolute value by

$$\frac{1}{T}\sum_{j=1}^{T-1} \left|w_{x,j}\right|^2 \left|w_{\hat{u},j} - w_{u,j}\right|^2 + \frac{2}{T}\sum_{j=1}^{T-1} \left|w_{x,j}\right|^2 \left|w_{\hat{u},j} - w_{u,j}\right| \left|w_{u,j}\right|. \tag{22}$$

From (5) and (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 (22) can be bounded by

$$\begin{aligned} &\frac{3}{T}\left(\beta - \hat{\beta}\right)^2 \sum_{j=1}^{T-1} \left|w_{x,j}\right|^4 + \frac{3}{T}\delta_T^2 \sum_{j=1}^{T-1} \left|w_{x,j}\right|^2 \left|w_{z_0,j} - w_{z_{\hat{k}},j}\right|^2 \\ &+\frac{3}{T}\left(\delta_T - \hat{\delta}\right)^2 \sum_{j=1}^{T-1} \left|w_{x,j}\right|^2 \left|w_{z_{\hat{k}},j}\right|^2. \end{aligned} \tag{23}$$

By Propositions 2 and 7, $\beta - \hat{\beta} = O_p\left(T^{-1/2}\right)$, and by Proposition 6 of Lazarová (2004), $T^{-1}\sum_{j=1}^{T-1}\left|w_{x,j}\right|^4 = o_p(T)$, therefore the first term of (23) is $o_p(1)$. The second term of (23) 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 Schwarz inequality. The expression in the last bracket is bounded by

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

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

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

$$\begin{aligned} P\left(\sum_t^{k_0,\hat{k}}|x_t| \geq M\|\delta_T\|^{-2}\right) &\leq P\left(\sum_{t=k_0-K\|\delta_T\|^{-2}}^{k_0+K\|\delta_T\|^{-2}}|x_t| \geq M\|\delta_T\|^{-2}\right) \\ &\quad + P\left(\left|\hat{k}-k_0\right| > K\|\delta_T\|^{-2}\right). \end{aligned} \tag{24}$$

By Proposition 1, the second term on the right of (24) is smaller than $\varepsilon/2$ for large K. By the Markov inequality and by Conditions 1 and 2, the first term on the right of (24) 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 (23) 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 (23) is bounded by

$$\sum_{j=1}^{T-1}|w_{x,j}|^2|w_{z_0,j}|^2 + \sum_{j=1}^{T-1}|w_{x,j}|^2\left|w_{z_0,j}-w_{z_{\hat{k}},j}\right|^2 \tag{25}$$

where the first term is $o_p(T^2)$ by Proposition 6 of Lazarová (2004) 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 and 7, we conclude that the third term of (23), and hence the first term of (22), is $o_p(1)$.

The second term of (22) is bounded by

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

since the first bracket is $o_p(1)$ as has been just shown, and where the second bracket is $O_p(1)$ by (20). ■

Proof of Proposition 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

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

$$\begin{aligned}= & \left(\sqrt{T}\delta'_T\frac{1}{T}Z'_0M_{\iota,X}Z_{[\tau T]} + \frac{1}{\sqrt{T}}u'M_{\iota,X}Z_{[\tau T]}\right)\left(\frac{1}{T}Z'_{[\tau T]}M_{\iota,X}Z_{[\tau T]}\right)^{-1} \\ & \times\left(\frac{1}{T}Z'_{[\tau T]}M_{\iota,X}Z_0\sqrt{T}\delta_T + \frac{1}{\sqrt{T}}Z'_{[\tau T]}M_{\iota,X}u\right) \qquad (26)\end{aligned}$$

where δ_T may be equal to zero. By Lemma 2,

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

uniformly in $\tau\in[0,1]$ and similarly $\frac{1}{T}Z'_{[\tau T]}M_{\iota,X}Z_{[\tau T]} \xrightarrow{p} \tau(1-\tau)\Sigma$ uniformly in $\tau\in[0,1]$. Further, by Lemma 2,

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

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 7. Define

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

and

$$Y(\tau) = \frac{1}{\tau(1-\tau)} \begin{pmatrix} \Sigma^{-1}\Omega^{\frac{1}{2}}(\tau B(1) - \tau B(\tau)) \\ \Sigma^{-1}\Omega^{\frac{1}{2}}(B(1) - \tau B(1)) \end{pmatrix} + \frac{1}{\tau(1-\tau)} \begin{pmatrix} \tau(\tau_0 - \tau)\mathbb{I}(\tau < \tau_0) \\ (\tau_0 - \tau)(\mathbb{I}(\tau_0 \le \tau) - \tau) \end{pmatrix} \otimes \delta.$$

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

$$Y_T(\tau) = \left(\frac{1}{T} W'_{[\tau T]} M_\iota W_{[\tau T]}\right)^{-1} \times \left(\frac{1}{\sqrt{T}} W'_{[\tau T]} M_\iota u + \frac{1}{T} W'_{[\tau T]} M_\iota \left(Z_0 - Z_{[\tau T]}\right) \sqrt{T}\delta_T\right). \quad (27)$$

Expressions $y'M_{\iota,X}y - y'M_{\iota,X,Z_{[\tau T]}}y$ in (26) and $Y_T(\tau)$ in (27) are continuous functions of matrices $\frac{1}{\sqrt{T}}Z'_{[\tau T]}M_\iota u$, $\frac{1}{T}Z'_{[\tau T]}M_\iota Z_{[\tau T]}$, $\frac{1}{T}X'M_\iota Z_{[\tau T]}$ and $\frac{1}{T}Z'_0 M_\iota Z_{[\tau T]}$ on $\tau \in \Lambda$, therefore we need to study the joint convergence of

$$\left(\frac{1}{\sqrt{T}} Z'_{[\tau T]} M_\iota u, \frac{1}{T} Z'_{[\tau T]} M_\iota Z_{[\tau T]}, \frac{1}{T} X' M_\iota Z_{[\tau T]}, \frac{1}{T} Z'_0 M_\iota Z_{[\tau T]}\right). \quad (28)$$

But by Lemma 2, $\frac{1}{T}Z'_{[\tau T]}M_\iota Z_{[\tau T]} \xrightarrow{p} \tau\Sigma$ and $\frac{1}{T}X'M_\iota Z_{[\tau T]} \xrightarrow{p} \tau\Sigma$ and $\frac{1}{T}Z'_0 M_\iota Z_{[\tau T]} \xrightarrow{p} (\tau_0 \wedge \tau)\Sigma$ uniformly in $\tau \in [0,1]$. Also by Lemma 2, $\frac{1}{\sqrt{T}}Z'_{[\tau T]}M_\iota u \Longrightarrow \Omega^{\frac{1}{2}}B(\tau)$ on $[0,1]$. Hence (28) converges weakly to

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

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 6 and the assumption $T^{1/2}\delta_T \to \delta$ imply that

$$(Y_T(\tau), \hat{\tau}) \Longrightarrow (Y(\tau), L) \quad (29)$$

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 $\sup_{1 \le i \le v} |\tau_i - \tau_{i-1}| < \eta$. Then for arbitrary $\rho > 0$ and any x,

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

By (29) and by the portmanteau lemma (see for example Theorem 2.1 of Billingsley (1999), page 16), the right-hand side of the last displayed inequality converges to

$$
\begin{aligned}
&\leq \sum_{i=1}^{v} P\left(Y\left(\tau_{i-1}\right) \leq x+\rho, L \in\left[\tau_{i-1}, \tau_i\right) \cap \Lambda, \sup_{t, s \in\left[\tau_{i-1}, \tau_i\right) \cap \Lambda}|Y(t)-Y(s)|<\rho\right) \\
&\quad +\sum_{i=1}^{v} P\left(L \in\left[\tau_{i-1}, \tau_i\right) \cap \Lambda, \sup_{t, s \in\left[\tau_{i-1}, \tau_i\right) \cap \Lambda}|Y(t)-Y(s)| \geq \rho\right) \\
&\leq P(Y(L) \leq x+2 \rho)+\sum_{i=1}^{v} P\left(\sup_{1 \leq i \leq v} \sup_{t, s \in\left[\tau_{i-1}, \tau_i\right) \cap \Lambda}|Y(t)-Y(s)| \geq \rho\right)
\end{aligned}
$$

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

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

for small η. Proceeding similarly, we obtain that

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

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

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

as desired. ■

Proof of Proposition 8. We have

$$
\begin{aligned}
F y^* & =F X \hat{\beta}+F Z_{\hat{k}} \hat{\delta}+H F \hat{u} \\
& =F X \hat{\beta}+F Z_{\hat{k}^*} \hat{\delta}+\tilde{u}
\end{aligned}
$$

where $\tilde{u}=H F \hat{u}+F\left(Z_{\hat{k}}-Z_{\hat{k}^*}\right) \hat{\delta}$. Therefore

$$
\begin{aligned}
\sqrt{T}\begin{pmatrix}
\hat{\beta}^*-\hat{\beta} \\
\hat{\delta}^*-\hat{\delta}
\end{pmatrix} = & \left(\frac{1}{T} W_{\hat{k}^*}^{\prime} \overline{F}^{\prime} F W_{\hat{k}^*}\right)^{-1}\left(\frac{1}{\sqrt{T}} W_{\hat{k}^*}^{\prime} \overline{F}^{\prime} H F \hat{u}\right. \\
& \left.+\frac{1}{\sqrt{T}} W_{\hat{k}^*}^{\prime} \overline{F}^{\prime} F\left(Z_{\hat{k}}-Z_{\hat{k}^*}\right) \hat{\delta}\right) .
\end{aligned}
$$

Write

$$
\begin{aligned}
Z_{\hat{k}^*}' \overline{F}' F Z_{\hat{k}^*} &= Z_0' \overline{F}' F Z_0 + \left(Z_{\hat{k}^*} - Z_0\right)' \overline{F}' F \left(Z_{\hat{k}^*} - Z_0\right) \\
&\quad + \left(Z_{\hat{k}^*} - Z_0\right)' \overline{F}' F Z_0 + Z_0' \overline{F}' F \left(Z_{\hat{k}^*} - Z_0\right)
\end{aligned}
$$

and

$$
X' \overline{F}' F Z_{\hat{k}^*} = X' \overline{F}' F Z_0 + X' \overline{F}' F \left(Z_{\hat{k}^*} - Z_0\right).
$$

For any $K > 0$, expression $P^* \left(\left\| \left(Z_{\hat{k}^*} - Z_0\right)' \overline{F}' F \left(Z_{\hat{k}^*} - Z_0\right) \right\| > M \left|\left|\delta_T\right|\right|^{-2} \right)$ is bounded by

$$
P^* \left(\sup_{N(K)} \left\| Z_\Delta' \overline{F}' F Z_\Delta \right\| > M \left|\left|\delta_T\right|\right|^{-2} \right) + P^* \left(\left| \hat{k}^* - k_0 \right| > K \left|\left|\delta_T\right|\right|^{-2} \right). \quad (30)
$$

Fix $\varepsilon, \eta > 0$. Expectation of the first term of (30) is smaller than $\varepsilon/3$ for large M and T by Lemma 2. The second term of (30) is bounded by

$$
P^* \left(\left| \hat{k}^* - \hat{k} \right| > \frac{K}{2} \left|\left|\delta_T\right|\right|^{-2} \right) + P^* \left(\left| \hat{k} - k_0 \right| > \frac{K}{2} \left|\left|\delta_T\right|\right|^{-2} \right)
$$

where the first term exceeds η with probability smaller than $\varepsilon/3$ for large K by Lemma 12 and where expectation of the second term is smaller than $\varepsilon/3$ for large K by Proposition 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, $T^{-1} X' \overline{F}' F X$ and $T^{-1} Z_0' \overline{F}' F Z_0$ are $O_p(1)$. The Schwarz inequality implies that the terms $T^{-1} \left(Z_{\hat{k}^*} - Z_0\right)' \overline{F}' F Z_0$ and $T^{-1} \left(Z_{\hat{k}^*} - Z_0\right)' \overline{F}' F X$ are $o_{p*}(1)$. It follows that

$$
\frac{1}{T} W_{\hat{k}^*}' \overline{F}' F W_{\hat{k}^*} = \frac{1}{T} W_0' \overline{F}' F W_0 + o_{p^*}(1)
$$

and by Lemma 2,

$$
\left(\frac{1}{T} W_{\hat{k}^*}' \overline{F}' F W_{\hat{k}^*} \right)^{-1} \xrightarrow{p^*} 2\pi \begin{pmatrix} 1 & \tau_0 \\ \tau_0 & \tau_0 \end{pmatrix}^{-1} \otimes \Sigma^{-1}.
$$

Second, write

$$
\frac{1}{\sqrt{T}} W_{\hat{k}^*}' \overline{F}' H F \hat{u} = \frac{1}{\sqrt{T}} W_0' \overline{F}' H F \hat{u} + \left(\frac{1}{\sqrt{T}} \left(W_{\hat{k}^*} - W_0\right) \overline{F}' H F \hat{u} \right). \quad (31)
$$

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

$$P^*\left(\sup_{N(K)}\left\|Z_\Delta'\overline{F}'HF\hat{u}\right\|>M\left\|\delta_T\right\|^{-1}\right)+P^*\left(\left|\hat{k}^*-k_0\right|>K\left\|\delta_T\right\|^{-2}\right). \quad (32)$$

By Lemma 8, the first term of (32) is smaller than η with probability at least $1-\varepsilon/2$ for large M. Arguing as in (30), the second term of (32) 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 (31) and Lemma 8,

$$\frac{1}{\sqrt{T}}W_{\hat{k}^*}'\overline{F}'HF\hat{u}\xrightarrow{d^*}\frac{1}{2\pi}\Omega^{\frac{1}{2}}\begin{pmatrix}B(1)\\B(\tau_0)\end{pmatrix}.$$

Finally, for any $M>0$,

$$\begin{aligned}&P^*\left(\left\|\left(Z_{\hat{k}^*}-Z_{\hat{k}}\right)'\overline{F}'FW_{\hat{k}^*}\right\|>M\left\|\delta_T\right\|^{-2}\right)\\ \leq\ &P^*\left(4\sup_{N(K/2)}\sup_{1\leq l\leq T}\left\|Z_\Delta'\overline{F}'FW_l\right\|>M\left\|\delta_T\right\|^{-2}\right)\\ &+P^*\left(\left|\hat{k}^*-k_0\right|>K\left\|\delta_T\right\|^{-2}\right)+P^*\left(\left|\hat{k}-k_0\right|>K\left\|\delta_T\right\|^{-2}\right). \quad (33)\end{aligned}$$

By Lemma 2, expectation of the first term of (33) 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. By the continuous mapping theorem,

$$\sqrt{T}\begin{pmatrix}\hat{\beta}^*-\hat{\beta}\\\hat{\delta}^*-\hat{\delta}\end{pmatrix}\xrightarrow{d^*}2\pi\left(\begin{pmatrix}1&\tau_0\\\tau_0&\tau_0\end{pmatrix}^{-1}\otimes\Sigma^{-1}\right)\frac{1}{2\pi}\begin{pmatrix}\Omega^{\frac{1}{2}}B(1)\\\Omega^{\frac{1}{2}}B(\tau_0)\end{pmatrix}.$$

This implies the validity of the proposition. ■

Proof of Proposition 9. For any $0<K<\infty$, let $\hat{\rho}$, $\tilde{\rho}$ and v_T^2 be as in the proof of Proposition 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 4, write

$$\left|P^*\left(v_T^2\left(\hat{k}^*-\hat{k}\right)\leq x\right)-P\left(\hat{\rho}\leq x\right)\right|\leq\left|P^*\left(v_T^2\left(\tilde{k}^*-\hat{k}\right)\leq x\right)-P\left(\tilde{\rho}\leq x\right)\right|$$

$$+2P^{*}\left(\left|\hat{k}^{*}-\hat{k}\right|>Kv_{T}^{-2}\right)+2P\left(|\hat{\rho}|>K\right). \tag{34}$$

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 14 implies that

$$v_{T}^{2}\left(\tilde{k}^{*}-\hat{k}\right)\overset{d^{*}}{\rightarrow}\arg\min_{|\rho|\leq K}W\left(\rho\right)=\tilde{\rho}$$

and so the first term on the right of (34) is smaller than $\eta/3$ with probability at least $1-\varepsilon/3$ for large T. By Lemma 12, the second term on the right of (34) 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 (34) is bounded by $\eta/3$ for large K because $\hat{\rho}=O_{p}\left(1\right)$. Therefore the right-hand side of (34) is $o_{p}\left(1\right)$ and

$$\frac{\left(\delta_{T}'\Sigma\delta_{T}\right)^{2}}{\delta_{T}'\Omega\delta_{T}}\left(\hat{k}^{*}-\hat{k}\right)\overset{d^{*}}{\rightarrow}\arg\min_{\rho}W\left(\rho\right).$$

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 7 and 8, $\hat{\Sigma}\overset{p}{\rightarrow}\Sigma$ by Conditions 1 and 2, and $\hat{\Omega}^{*}\overset{p^{*}}{\rightarrow}\Omega$ by Lemma 15, it follows that

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

■

10 Lemmas

This section contains the some auxiliary results and their proofs. Throughout this section, it is assumed that Conditions 1-6 hold.

Lemma 1 *For any matrix*

$$\left(\begin{array}{cc}A_{1} & B_{1}\\ A_{2} & B_{2}\end{array}\right)$$

and $A=\left(\begin{array}{c}A_{1}\\ A_{2}\end{array}\right)$ *and* $B=\left(\begin{array}{c}B_{1}\\ B_{2}\end{array}\right)$,

$$A'M_{B}A\geq A_{1}'M_{B_{1}}A_{1}.$$

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, Perron (1998) or Lemma 2 of Brown et al. (1975). ■

Lemma 2 *For any $0 < K < \infty$, as $T \to \infty$,*

(a) $\frac{1}{T} Z_{[\tau T]} M_\iota Z_{[\sigma T]} \xrightarrow{p} (\tau \wedge \sigma) \Sigma$ *uniformly on* $(\tau, \sigma) \in [0,1]^2$,

(b) $\frac{1}{\sqrt{T}} Z'_{[\tau T]} M_\iota u \Longrightarrow \Omega^{\frac{1}{2}} B(\tau)$ *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) $\sup_{1 \le k \le T} \| W'_k M_\iota u \| = O_p\left(T^{\frac{1}{2}}\right)$,

(e) $\sup_{1 \le k, l \le T} \| Z'_k M_\iota Z_l \| = O_p(T)$,

(f) $\sup_{k \in N(K)} \sup_{1 \le l \le T} \| Z'_\Delta M_\iota Z_l \| = O_p\left(\|\delta_T\|^{-2}\right)$,

(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) $\sup_{k \in N^c(K)} \sup_{1 \le l \le T} x \left\| \frac{Z'_\Delta M_\iota Z_l}{k - k_0} \right\| = O_p(1)$.

Proof. Parts (a) and (b) follow from Propositions 4 and 2 of Lazarová (2004), respectively, after noting that

$$\frac{1}{T} Z'_{[\tau T]} M_\iota Z_{[\sigma T]} = \frac{2\pi}{T} Z'_{[\tau T]} \overline{F}' F Z_{[\sigma T]} = \frac{2\pi}{T} \sum_{j=1}^{T-1} w_{z([\tau T]),j} \overline{w}_{z([\sigma T]),j}$$

and

$$\frac{1}{\sqrt{T}} Z'_{[\tau T]} M_\iota u = \frac{2\pi}{\sqrt{T}} Z'_{[\tau T]} \overline{F}' F u = \frac{2\pi}{\sqrt{T}} \sum_{j=1}^{T-1} w_{z([\tau T]),j} \overline{w}_{u,j}.$$

Parts (c), (d) and (e) are implied by parts (a) and (b). Part (f) follows from the bound

$$\| Z'_\Delta M_\iota Z_l \| = \left\| Z'_\Delta Z_l - \frac{1}{T} Z'_\Delta \iota_\Delta \iota'_l Z_l \right\| \le \|Z_\Delta\|^2 + |k - k_0|^{\frac{1}{2}} \|Z_\Delta\| \frac{l^{\frac{1}{2}}}{T} \|Z_l\| \quad (35)$$

and from Conditions 1, 2 and 5. 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

$$\frac{1}{\sqrt{T}} Z'_{\Delta([\tau_0 T],[(\tau_0+\rho)T])} M_\iota u \Longrightarrow \begin{cases} \Omega^{\frac{1}{2}} \left(B(\tau_0 + \rho) - B(\tau_0)\right) & \rho \ge 0, \\ \Omega^{\frac{1}{2}} \left(B(\tau_0) - B(\tau_0 + \rho)\right) & \rho \le 0, \end{cases}$$

from which it follows that

$$\|\delta_T\| Z'_\Delta M_\iota u \Longrightarrow \begin{cases} \Omega^{\frac{1}{2}} B_1(\rho) & \rho \ge 0, \\ \Omega^{\frac{1}{2}} B_2(-\rho) & \rho < 0 \end{cases}$$

where B_1, B_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 (35) after noting that Condition 5 implies that $\sup_{k \in N^c(K)} |k - k_0|^{-1} \|Z_\Delta\|^2 = O_p(1)$. ■

Lemma 3 *As* $T \to \infty$,

(a) $u'(M_{\iota,X,Z_k} - M_{\iota,X,Z_0})\, u = O_p(1)$ *uniformly on* $k \in \Lambda \cdot T$,

(b) *if* $T \|\delta_T\|^2 \to \infty$ *then* $u'(M_{\iota,X,Z_k} - M_{\iota,X,Z_0})\, u = o_p(1)$ *uniformly on* $N(K)$.

Proof. Denote $W_\Delta = (W_k - W_0)\,\mathrm{sgn}(k - k_0) = (0, Z_\Delta)$. For any non-singular matrices A and $A + B$,

$$P_{A+B} - P_A = B\left(\overline{A}'A\right)^{-1}\overline{A}' + A\left(\overline{A}'A\right)^{-1}\overline{B}' + B\left(\overline{A}'A\right)^{-1}\overline{B}'$$

$$-(A+B)\left(\left(\overline{A}+\overline{B}\right)'(A+B)\right)^{-1}\left(\overline{A}'B + \overline{B}'A + \overline{B}'B\right)\left(\overline{A}'A\right)^{-1}\left(\overline{A}+\overline{B}\right)'. \tag{36}$$

Let $A = M_\iota W_0$ and $B = M_\iota W_\Delta\,\mathrm{sgn}(k - k_0)$. Then

$$u'(M_{\iota,X,Z_0} - M_{\iota,X,Z_k})\, u = (M_\iota u)'(P_{M_\iota W_k} - P_{M_\iota W_0})\, M_\iota u$$

$$\begin{aligned}
= \;& 2u'M_\iota W_\Delta (W_0'M_\iota W_0)^{-1} W_0'M_\iota u\,\mathrm{sgn}(k - k_0) \\
& + u'M_\iota W_\Delta (W_0'M_\iota W_0)^{-1} W_\Delta' M_\iota u \\
& - u'M_\iota W_k (W_k'M_\iota W_k)^{-1} \{(W_0'M_\iota W_\Delta + W_\Delta'M_\iota W_0)\,\mathrm{sgn}(k - k_0) \\
& + W_\Delta'M_\iota W_\Delta\} (W_0'M_\iota W_0)^{-1} W_k'M_\iota u.
\end{aligned}$$

The bounds in part (a) and (b) are implied by Lemma 2. ∎

Lemma 4 *There exists* $\lambda > 0$ *such that for every* $\varepsilon > 0$, *there exists* $K < \infty$ *such that*

$$P\left(\inf_{N^c(K)} \frac{Q_T(k)}{|k - k_0|} \geq \lambda \,||\delta_T||^2\right) \geq 1 - \varepsilon.$$

Proof. By the definition of $Q_T(k)$, the left side of the last displayed inequality is bounded from below by

$$P\left(\inf_{N^c(K)} \lambda_{\min}\left(\frac{Z_\Delta' M_{\iota,X,Z_k} Z_\Delta}{|k - k_0|}\right) \geq \lambda\right). \tag{37}$$

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 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

$$\lambda_{\min}\left(\frac{Z_\Delta' M_{\iota,X,Z_k} Z_\Delta}{|k-k_0|}\right) \geq \lambda_{\min}\left\{\frac{Z_\Delta' M_{\iota_\Delta} Z_\Delta}{|k-k_0|}\left(\frac{1}{T} Z_k' M_{\iota_0,\iota_\Delta} Z_k\right)^{-1} \frac{1}{T} Z_0' M_{\iota_0} Z_0\right\}. \tag{38}$$

Similar inequality is obtained in the case $k < k_0$,

$$\begin{aligned}\lambda_{\min}\left(\frac{Z_\Delta' M_{\iota,X,Z_k} Z_\Delta}{|k-k_0|}\right) \quad \geq \quad & \lambda_{\min}\left\{\frac{Z_\Delta' M_{\iota_\Delta} Z_\Delta}{|k-k_0|}\left(\frac{1}{T}(X-Z_k)' M_{\iota_\Delta,\iota-\iota_0}(X-Z_k)\right)^{-1}\right. \\ & \left. \times \frac{1}{T}(X-Z_0)' M_{\iota-\iota_0}(X-Z_0)\right\}. \end{aligned} \tag{39}$$

By Conditions 1 and 2, as $l \to \infty$ and $l \to -\infty$, $\frac{1}{l}\sum_t^{k_0,k_0+l} x_t \xrightarrow{p} 0$ and $\frac{1}{l}\sum_t^{k_0,k_0+l} x_t x_t' \xrightarrow{p} \Sigma$ where $\Sigma > 0$, 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 $\lambda > 0$ with probability at least $1 - \varepsilon/3$ for large K. Similarly, since $\frac{1}{T}\sum_{t=1}^T x_t \xrightarrow{p} 0$ and $\frac{1}{T}\sum_{t=1}^T x_t x_t' \xrightarrow{p} \Sigma$ as $T \to \infty$, 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 (38) and (39). 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 $\lambda > 0$ such that (37) is greater than $1 - \varepsilon$, and the lemma is established. ∎

The following lemma extends Hájek-Rényi inequality to the cross-product of two mean-adjusted series possibly exhibiting long-memory dependence.

Lemma 5 *Let* $\overline{u} = T^{-1} \sum_{t=1}^{T} u_t$*. Then for any* $\alpha > 0$ *and for any integers* m*,* T *such that* $m < T$*,*

$$P\left(\max_{m\leq k\leq T} \frac{1}{k}\left\|\sum_{t=1}^{k} x_t\left(u_t-\overline{u}\right)\right\| > \alpha\right) < \frac{D}{\alpha^2 m} \tag{40}$$

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

$$A_k = \left\{\frac{1}{k}\left|S_k\right| > \alpha, \frac{1}{j}\left|S_j\right| \leq \alpha \text{ for } m \leq j < k\right\}.$$

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

$$\begin{aligned}
& P\left(\max_{m\leq k\leq T} \frac{1}{k}\left|\sum_{t=1}^{k} x_t\left(u_t-\overline{u}\right)\right| > \alpha\right) \\
\leq\ & \frac{1}{\alpha^2}\left(\frac{1}{m^2} ES_m^2 + \sum_{k=m+1}^{T} \frac{1}{k^2} E\left(S_k^2 - S_{k-1}^2\right)\left(1-\mathbb{I}\left(A_{k-1}\right)-\ldots-\mathbb{I}\left(A_m\right)\right)\right) \\
\leq\ & \frac{1}{\alpha^2}\left(\frac{1}{m^2} ES_m^2 + \sum_{k=m+1}^{T} \frac{1}{k^2}\left(Ex_k^2\left(u_k-\overline{u}\right)^2 + 2\left|Ex_k\left(u_k-\overline{u}\right)S_{k-1}\right|\right)\right). \quad (41)
\end{aligned}$$

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

$$\left|Ex_t x_s u_r u_v\right| \leq C\varphi_{|t-s|}\psi_{|r-v|}$$

where $\varphi_k = \sum_{j=0}^{\infty}\left|a_j\right|\left|a_{j+k}\right|$ and $\psi_k = \sum_{j=0}^{\infty}\left|b_j\right|\left|b_{j+k}\right|$, 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,

$$\begin{aligned} \left|Ex_k \left(u_k - \overline{u}\right) S_{k-1}\right| &\leq \sum_{t=1}^{k-1} \varphi_{k-t}\psi_{k-t} + \frac{C}{T}\sum_{t=1}^{k-1}\sum_{s=1}^{T} \varphi_{k-t}\left(\psi_{|k-s|} + \psi_{|t-s|}\right) \\ &\quad + \frac{C}{T^2}\sum_{t=1}^{k-1}\sum_{s,r=1}^{T} \varphi_{k-t}\psi_{|s-r|} \\ &\leq C \end{aligned}$$

uniformly in k by Lemma 2 of Robinson (1998). Therefore the second term in the bracket on the right of (41) is bounded by

$$C\sum_{k=m+1}^{T} \frac{1}{k^2} \leq \frac{C}{m}$$

and thus we conclude that (40) holds true. ■

Lemma 6 *If $T\left\|\delta_T\right\|^2 \to \infty$, then for any $K < \infty$, as $T \to \infty$,*

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

where the $o_p(1)$ term is uniform on $N(K)$.

Proof. Write

$$Q_T(k) = \delta_T' Z_\Delta' M_\iota Z_\Delta \delta_T - \delta_T' Z_\Delta' M_\iota W_k \left(W_k' M_\iota W_k\right)^{-1} W_k M_\iota Z_\Delta \delta_T.$$

Since by Lemma 2, the terms and $Z_\Delta' M_\iota W_k$ and $\left(W_k' M_\iota W_k\right)^{-1}$ are $O_p\left(||\delta_T||^{-2}\right)$ and $O_p\left(T^{-1}\right)$ 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 6 and $u'\left(M_{\iota,X,Z_k} - M_{\iota,X,Z_0}\right)u$ is uniformly on $N(K)$ by Lemma 3, the decomposition of $R_T(k)$ in (13) implies that

$$\begin{aligned} R_T(k) &= -2\delta_T' Z_\Delta' M_\iota u \,\mathrm{sgn}(k-k_0) + O_p\left(\|\delta_T\|\,\|\delta_T\|^{-2}\,T^{-1}T^{1/2}\right) + o_p(1) \\ &= -2\delta_T' Z_\Delta' M_\iota u \,\mathrm{sgn}(k-k_0) + o_p(1)\,. \end{aligned}$$

The lemma now follows from (11). ■

Lemma 7 *For $\delta_T \neq 0$, let $v_T^2 = \left(\delta_T'\Sigma\delta_T\right)^2/\delta_T'\Omega\delta_T$. If the conditions of Proposition 4 are satisfied then for any $K > 0$, as $T \to \infty$,*

$$\arg\min_{|\rho|\le K} S_T\left(k_0 + \left[\rho v_T^{-2}\right]\right) \xrightarrow{d} \arg\min_{|\rho|\le K} W(\rho)\,.$$

Proof. For any given K and $\rho \in [-K, K]$, write $k = k_0 + \left[\rho v_T^{-2}\right]$. By Lemma 6,

$$S_T\left(k_0 + \left[\rho v_T^{-2}\right]\right) - S_T(k_0) = \delta_T' Z_\Delta' M_\iota Z_\Delta \delta_T - 2\delta_T' Z_\Delta' M_\iota u \,\mathrm{sgn}\left[\rho v_T^{-2}\right] + o_p(1) \tag{42}$$

where the $o_p(1)$ term is uniform on $N(K)$. Lemma 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 $\{(\sigma,\rho) : 0 \le \sigma, \sigma+\rho \le 1\}$, 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, it can be seen that

$$v_T Z_\Delta' M_\iota u \overset{p}{\Longrightarrow} \begin{cases} \Omega^{\frac{1}{2}} B_1(\rho) & \rho \ge 0, \\ \Omega^{\frac{1}{2}} B_2(-\rho) & \rho < 0 \end{cases}$$

where B_1, B_2 are independent p-vectors of independent standard Brownian motion processes. Since $-2\frac{\delta_T'}{v_T}\Omega^{\frac{1}{2}}B_1(\rho)$ and $2\frac{\delta_T'}{v_T}\Omega^{\frac{1}{2}}B_2(\rho)$ are equal in distribution to $2\frac{\left(\delta_T'\Omega\delta_T\right)^{\frac{1}{2}}}{v_T}W_1(\rho)$ and $2\frac{\left(\delta_T'\Omega\delta_T\right)^{\frac{1}{2}}}{v_T}W_2(\rho)$, respectively, the left-hand side of (42) is equal in distribution to

$$\begin{aligned} & |\rho|\,\frac{\delta_T'\Sigma\delta_T}{v_T^2} + 2\frac{\left(\delta_T'\Omega\delta_T\right)^{\frac{1}{2}}}{v_T}\left(W_1(\rho)\,\mathbb{I}(\rho \ge 0) + W_2(-\rho)\,\mathbb{I}(\rho < 0)\right) + o_p(1) \\ =\; & 2\frac{\delta_T'\Omega\delta_T}{\delta_T'\Sigma\delta_T} W(\rho) + o_p(1) \end{aligned}$$

uniformly on $\rho \in [-K, K]$. Now observing that

$$\arg\min_{|\rho|\leq K} 2\frac{\delta_T' \Omega \delta_T}{\delta_T' \Sigma \delta_T} W(\rho) = \arg\min_{|\rho|\leq K} W(\rho),$$

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

$$R_T^*(k) = 2\hat{\delta}' Z_{\hat{k}}' \overline{F}' M_{FW_k} HF\hat{u} + \hat{u}' \overline{F}' \overline{H}' \left(M_{FW_k} - M_{FW_{\hat{k}}} \right) HF\hat{u},$$

so that

$$Q_T^*(k) + R_T^*(k) = \|\hat{v}_k\|^2 - \|\hat{v}_{\hat{k}}\|^2 = S_T^*(k) - S_T^*\left(\hat{k}\right). \tag{43}$$

Let $\widehat{N}(K) = \left\{ k\text{: } \left| k - \hat{k} \right| \leq K \left|\left|\delta_T\right|\right|^{-2} \right\}$, $\widehat{N^c}(K) = \Lambda \cdot T - \widehat{N}(K)$ and denote $\widehat{Z}_\Delta = Z_{\Delta\left(\hat{k}, k\right)}$.

Lemma 8 *As $T \to \infty$,*

(a) $\frac{2\pi}{\sqrt{T}} Z_{[\tau T]}' \overline{F}' HF\hat{u} \overset{p}{\Longrightarrow} \Omega^{\frac{1}{2}} B(\tau)$,

(b) for any $K > 0$, $\sup_{k \in N(K)} \sup_{1 \leq l \leq T} \left\| Z_\Delta' \overline{F}' HFZ_l \right\| = O_{p^*}\left(\|\delta_T\|^{-2} \right)$,

(c) for any $K > 0$, $\sup_{N(K)} \left\| Z_\Delta' \overline{F}' HF\hat{u} \right\| = O_{p^*}\left(\|\delta_T\|^{-1} \right)$,

(d) for any $K > 0$, $\sup_{k \in \hat{N}(K)} \sup_{1 \leq l \leq T} \left\| \hat{Z}_\Delta' \overline{F}' FZ_l \right\| = O_p\left(\|\delta_T\|^{-1} \right)$,

(e) $W_k' \overline{F}' HF\hat{u} = O_{p^*}\left(T^{1/2} \right)$ *uniformly over* $1 \leq k \leq T$,

(f) $\left(\frac{1}{T} W_{\hat{k}}' \overline{F}' FW_{\hat{k}} \right)^{-1} \overset{p}{\to} 2\pi \begin{pmatrix} 1 & \tau_0 \\ \tau_0 & \tau_0 \end{pmatrix}^{-1} \otimes \Sigma^{-1}$,

(g) for every $\varepsilon > 0$ *there exist* $K, M > 0$ *such that*

$$P\left(\sup_{k \in \widehat{N^c}(K)} \sup_{1 \leq l \leq T} \frac{1}{\left| k - \hat{k} \right|} \left\| \hat{Z}_\Delta' \overline{F}' FZ_l \right\| > M \right) < \varepsilon,$$

(h) for every $\varepsilon, \eta > 0$ *there exist* $K, M > 0$ *such that*

$$P\left(P^*\left(\sup_{k \in \widehat{N^c}(K)} \sup_{1 \leq l \leq T} \frac{1}{\left| k - \hat{k} \right|} \left\| \hat{Z}_\Delta' \overline{F}' HFZ_l \right\| > M \right) > \eta \right) < \varepsilon.$$

Proof. Part (a) follows from Proposition 7 of Lazarová (2004) and from the remark at the end of its proof because

$$Z'_{[\tau T]}\overline{F}'HF\hat{u}=\sum_{j=1}^{T-1}\overline{w}_{z(\tau),j}w_{\hat{u},j}\eta_j^*.$$

To show the validity of part (b), define matrix G as $G=\overline{F}'HF$. By the definition of matrices F and H, matrix G is a real circulant matrix with elements $g_{t,s}=g_{s-t}$, $1\leq t,s\leq T$, where $g_t=\frac{1}{2\pi T}\sum_{j=1}^{T-1}\eta_j^*e^{i\lambda_j t}$. 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 Schwarz inequality,

$$\left\|Z'_\Delta\overline{F}'HFZ_l\right\|=\left\|\sum_t^{k_0,k}x_tg^tZ_l\right\|\leq\left(\sum_t^{k_0,k}\|x_t\|^2\right)^{\frac{1}{2}}\left(\sum_t^{k_0,k}\left\|g^tZ_l\right\|^2\right)^{\frac{1}{2}}.$$

Therefore

$$\begin{aligned}\sup_{k\in N(K)}\sup_{1\leq l\leq T}\left\|Z'_\Delta\overline{F}'HFZ_l\right\| \;\leq\; & \left(\sum_{t=k_0-K\|\delta_T\|^{-2}}^{k_0+K\|\delta_T\|^{-2}}\|x_t\|^2\right)^{\frac{1}{2}}\\ & \times\left(\sum_{t=k_0-K\|\delta_T\|^{-2}}^{k_0+K\|\delta_T\|^{-2}}\sup_{1\leq l\leq T}\left\|g^tZ_l\right\|^2\right)^{\frac{1}{2}}. \quad (44)\end{aligned}$$

The expression in the first bracket on the right of (44) is $O_p\left(\|\delta_T\|^{-2}\right)$ by Conditions 1 and 2. Further,

$$\left\|g^tZ_l\right\|^2=g^tZ_lZ_l'\left(g^t\right)'\leq g^tXX'\left(g^t\right)'=\operatorname{tr}\left(g^t\right)'g^tXX'$$

and so $E^*\sup_{1\leq l\leq T}\left\|g^tZ_l\right\|^2\leq\operatorname{tr}E^*\left(g^t\right)'g^tXX'$. For any t and s, $E^*g_tg_s=\frac{1}{4\pi^2}\left(-\frac{1}{T^2}+\frac{1}{T}I\left(t=s\right)\right)$, therefore $E^*\left(g^t\right)'g^t=\frac{1}{4\pi^2T}M_\iota$ and

$$E^*\sup_{1\leq l\leq T}\left\|g^tZ_l\right\|^2\leq\operatorname{tr}\frac{1}{4\pi^2T}M_\iota XX'=\frac{1}{4\pi^2}\operatorname{tr}\frac{1}{T}X'M_\iota X=O_p\left(1\right) \qquad (45)$$

by Lemma 2. By the Markov inequality, the expression in the second bracket of (44) 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 follows from part (b) of Lemma 2. To prove part (d), write

$$2\pi\left\|\hat{Z}'_\Delta\overline{F}'FZ_l\right\|\leq\left\|\hat{Z}_\Delta\right\|^2+\left|k-\hat{k}\right|\left\|\hat{Z}_\Delta\right\|\frac{l^{\frac{1}{2}}}{T}\left\|Z_l\right\|.$$

For any $M>0$,

$$P\left(\sup_{k\in\widehat{N}(K)}\left\|\hat{Z}_{\Delta}\right\|^{2}>M\left\|\delta_{T}\right\|^{-2}\right)\leq P\left(\sum_{t=\hat{k}-K\|\delta_{T}\|^{-2}}^{\hat{k}+K\|\delta_{T}\|^{-2}}\left\|x_{t}\right\|^{2}>M\left\|\delta_{T}\right\|^{-2}\right)$$

which is bounded by CK/M by the Markov inequality and Conditions 1 and 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. Part (e) follows from part (a) and part (f) follows from (14). In part (g), we have

$$\sup_{\substack{k\in\widehat{N^{c}}(K)\\ 1\leq l\leq T}}2\pi\left\|\frac{\hat{Z}_{\Delta}'\overline{F}'FZ_{l}}{k-\hat{k}}\right\|\leq\sup_{k\in\widehat{N^{c}}(K)}\frac{\left\|\hat{Z}_{\Delta}\right\|^{2}}{\left|k-\hat{k}\right|}+\sup_{k\in\widehat{N^{c}}(K)}\frac{\left\|\hat{Z}_{\Delta}\right\|}{\left|k-\hat{k}\right|^{\frac{1}{2}}}\sup_{1\leq l\leq T}\frac{l^{\frac{1}{2}}}{T}\left\|Z_{l}\right\|.$$

Now

$$\begin{aligned}\sup_{k\in\widehat{N^{c}}(K)}\frac{1}{\left|k-\hat{k}\right|}\left\|\hat{Z}_{\Delta}\right\|^{2}&\leq 2\sup_{k\in\widehat{N^{c}}(K)}\frac{|k-k_{0}|}{\left|k-\hat{k}\right|}\sup_{k\in\widehat{N^{c}}(K)}\frac{1}{|k-k_{0}|}\left\|Z_{\Delta}\right\|^{2}\\&\quad+2\sup_{k\in\widehat{N^{c}}(K)}\frac{1}{\left|k-\hat{k}\right|}\left\|Z_{\hat{k}}-Z_{0}\right\|^{2}.\end{aligned}$$

The factor $\sup_{k\in\widehat{N^{c}}(K)}\frac{|k-k_{0}|}{\left|k-\hat{k}\right|}$ 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 $\sup_{k\in\widehat{N^{c}}(K)}\frac{1}{|k-k_{0}|}\left\|Z_{\Delta}\right\|^{2}$ is greater than M is bounded by

$$P\left(\sup_{|k-k_{0}|\geq K/2}\frac{1}{|k-k_{0}|}\left\|Z_{\Delta}\right\|^{2}>M\right)+P\left(\left|\hat{k}-k_{0}\right|>\frac{K}{2}\left\|\delta_{T}\right\|^{-2}\right).\quad(46)$$

The second term of (46) is bounded by $\varepsilon/2$ for large K by Proposition 1 and the first term of (46) is bounded by $\varepsilon/2$ for large M by Condition 5. Further, for any $N>0$,

$$P\left(\sup_{k\in\widehat{N^{c}}(K)}\frac{1}{\left|k-\hat{k}\right|}\left\|Z_{\hat{k}}-Z_{0}\right\|^{2}>M\right)\leq P\left(\left\|Z_{\hat{k}}-Z_{0}\right\|^{2}>MK\left\|\delta_{T}\right\|^{-2}\right)$$

$$\leq P\left(\sum_{t=k_{0}-N\|\delta_{T}\|^{-2}}^{k_{0}+N\|\delta_{T}\|^{-2}}\left\|x_{t}\right\|^{2}>MK\left\|\delta_{T}\right\|^{-2}\right)+P\left(\left|\hat{k}-k_{0}\right|>N\left\|\delta_{T}\right\|^{-2}\right).\quad(47)$$

By the Markov inequality the first term on the right of (47) is bounded by CN/MK. Proposition 1 then implies that both terms on the right of (47) can be bounded by $\varepsilon/2$ for large M and N for any $K > 0$. Gathering the results and recalling that the factor $\sup_{1\le l\le T} l^{\frac{1}{2}} T^{-1} \|Z_l\|$ is $O_p(1)$ by Condition 5, we can conclude that part (g) holds.

Finally, to prove part (h), we follow the steps in part (b) and write

$$\begin{aligned}\sup_{\substack{k\in\widehat{N^c}(K)\\ 1\le l\le T}} \frac{1}{\left|k-\hat{k}\right|}\left\|\hat{Z}_\Delta'\overline{F}'HFZ_l\right\| &\le \left(\sup_{\widehat{N^c}(K)} \frac{1}{\left|k-\hat{k}\right|}\sum_t^{\hat{k},k}\|x_t\|^2\right)^{\frac{1}{2}} \\ &\quad\times\left(\sup_{\widehat{N^c}(K)} \frac{1}{\left|k-\hat{k}\right|}\sum_t^{\hat{k},k}\sup_{1\le l\le T}\left\|g^t Z_l\right\|^2\right)^{\frac{1}{2}}. \quad (48)\end{aligned}$$

Proceeding as in the proof of part (g), it can be seen that the expression

$$EP^*\left(\sup_{\widehat{N^c}(K)} \frac{1}{\left|k-\hat{k}\right|}\sum_t^{\hat{k},k}\|x_t\|^2 > M\right) = P\left(\sup_{\widehat{N^c}(K)} \frac{1}{\left|k-\hat{k}\right|}\sum_t^{\hat{k},k}\|x_t\|^2 > M\right)$$

is smaller than $\varepsilon/2$ for large M and K. The expression in the second bracket on the right of (48) is equal to $\sup_{1\le l\le T}\|g^t Z_l\|^2$. Part (g) is then implied by (45) and by the Markov inequality. ■

Lemma 9 *Assume Conditions 1-6 hold. Then as $T\to\infty$,*
(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 $\hat{W}_\Delta = (W_k - W_{\hat{k}})\,\mathrm{sgn}\left(k-\hat{k}\right) = \left(0,\hat{Z}_\Delta\right)$. Proceeding as in Lemma 3 and taking $A = FW_{\hat{k}}$ and $B = F\hat{W}_\Delta\,\mathrm{sgn}\left(k-\hat{k}\right)$ in (36), we obtain that

$$\hat{u}'\overline{F}'\overline{H}'\left(M_{FW_{\hat{k}}}-M_{FW_k}\right)HF\hat{u} = \hat{u}'\overline{F}'\overline{H}'\left(P_{FW_k}-P_{FW_{\hat{k}}}\right)HF\hat{u}$$

$$
\begin{aligned}
= \;& 2\operatorname{Re}\hat{u}'\overline{F}'\overline{H}'F\hat{W}_\Delta\left(W_{\hat{k}}'\overline{F}'FW_{\hat{k}}\right)^{-1}W_{\hat{k}}'\overline{F}'HF\hat{u}\operatorname{sgn}\left(k-\hat{k}\right)\\
&+\hat{u}'\overline{F}'\overline{H}'F\hat{W}_\Delta\left(W_{\hat{k}}'\overline{F}'FW_{\hat{k}}\right)^{-1}\hat{W}_\Delta'\overline{F}'HF\hat{u}\\
&-\hat{u}'\overline{F}'\overline{H}'FW_k\left(W_k'\overline{F}'FW_k\right)^{-1}\left\{\left(W_{\hat{k}}'\overline{F}'F\hat{W}_\Delta+\hat{W}_\Delta'\overline{F}'FW_{\hat{k}}\right)\operatorname{sgn}\left(k-\hat{k}\right)\right.\\
&\left.+\hat{W}_\Delta'\overline{F}'F\hat{W}_\Delta\right\}\left(W_{\hat{k}}'\overline{F}'FW_{\hat{k}}\right)^{-1}W_k'\overline{F}'HF\hat{u}.
\end{aligned}
$$

Therefore the lemma holds by Lemmas 2 and 8. ■

Lemma 10 *There exists $\lambda > 0$ such that for every $\varepsilon > 0$, there exists $K < \infty$ and $T_0 < \infty$ such that for all $T \geq T_0$,*

$$
EP^*\left(\inf_{\widehat{N^c}(K)}\frac{Q_T^*(k)}{\left|k-\hat{k}\right|}\geq\lambda\left\|\hat{\delta}\right\|^2\right)\geq 1-\varepsilon.
$$

Proof. If an event A does not depend on η_j^*, 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 hypothesised inequality is bounded from below by

$$
P\left(\inf_{\widehat{N^c}(K)}\lambda_{\min}\left(\frac{\hat{Z}_\Delta'M_{\iota,X,Z_k}\hat{Z}_\Delta}{\left|k-\hat{k}\right|}\right)\geq 2\pi\lambda\right).
$$

Denote $\hat{\iota}_\Delta = (\iota_k - \iota_{\hat{k}})\operatorname{sgn}\left(k-\hat{k}\right)$. Proceeding as in the proof of Lemma 4, we obtain inequality

$$
\lambda_{\min}\left(\frac{\hat{Z}_\Delta'M_{\iota,X,Z_k}\hat{Z}_\Delta}{\left|k-\hat{k}\right|}\right)\geq\lambda_{\min}\left\{\frac{\hat{Z}_\Delta'M_{\hat{\iota}_\Delta}\hat{Z}_\Delta}{\left|k-\hat{k}\right|}\left(\frac{1}{T}Z_k'M_{\hat{\iota}_\Delta,\iota_{\hat{k}}}Z_k\right)^{-1}\frac{1}{T}Z_{\hat{k}}'M_{\iota_{\hat{k}}}Z_{\hat{k}}\right\} \tag{49}
$$

for $k > \hat{k}$ and

$$
\begin{aligned}
\lambda_{\min}\left(\frac{\hat{Z}_\Delta'M_{\iota,X,Z_k}\hat{Z}_\Delta}{\left|k-\hat{k}\right|}\right)\;\geq\;&\lambda_{\min}\left\{\frac{\hat{Z}_\Delta'M_{\hat{\iota}_\Delta}\hat{Z}_\Delta}{\left|k-\hat{k}\right|}\left(\frac{1}{T}(X-Z_k)'M_{\hat{\iota}_\Delta,\iota-\iota_{\hat{k}}}(X-Z_k)\right)^{-1}\right.\\
&\left.\times\frac{1}{T}(X-Z_{\hat{k}})'M_{\iota-\iota_{\hat{k}}}(X-Z_{\hat{k}})\right\}.
\end{aligned} \tag{50}
$$

for $k < \hat{k}$. Write the first factor in the curly bracket in (49) 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 $\delta>0$,

$$P\left(\sup_{\widehat{N^c}(K)}\left|\frac{k-k_0}{k-\hat{k}}-1\right|>\delta\right)=P\left(\sup_{\widehat{N^c}(K)}\left|\frac{\hat{k}-k_0}{k-\hat{k}}\right|>\delta\right)\leq P\left(\left|\hat{k}-k_0\right|>\delta K\left\|\delta_T\right\|^{-2}\right)$$

which is smaller than $\varepsilon/2$ for large K by Proposition 1. Also, for any $\delta>0$ and any $M>0$,

$$P\left(\sup_{\widehat{N^c}(K)}\frac{\left\|Z_{\hat{k}}-Z_0\right\|^2}{\left|k-\hat{k}\right|}>\delta\right)\leq P\left(\sum_t^{k_0,\hat{k}}\left\|x_t\right\|^2>\delta K\left\|\delta_T\right\|^{-2}\right)$$

$$\leq P\left(\sum_{t=k_0-M\|\delta_T\|^{-2}}^{k_0+M\|\delta_T\|^{-2}}\left\|x_t\right\|^2>\delta K\left\|\delta_T\right\|^{-2}\right)+P\left(\left|\hat{k}-k_0\right|>M\left\|\delta_T\right\|^{-2}\right).\quad(51)$$

The second term of (51) is smaller than $\varepsilon/2$ for large M by Proposition 1. By the Markov inequality, the first term of (51) 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 and 2, $\lim_{l\to\pm\infty}|l|^{-1}\sum_t^{k_0,k_0+l}x_tx_t'\xrightarrow{p}\Sigma$, 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 (49) can be written as

$$\frac{1}{T}Z_{\hat{k}}'M_{\iota_{\hat{k}}}Z_{\hat{k}}=\frac{1}{T}\sum_{t=1}^{\hat{k}}x_tx_t'-\left(\frac{1}{T}\sum_{t=1}^{\hat{k}}x_t\right)\left(\frac{1}{\hat{k}}\sum_{t=1}^{\hat{k}}x_t'\right).$$

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 4. Similarly, the eigenvalues of the remaining factors of (49) and (50) are bounded from below by a positive number. Concluding as in the proof of Lemma 4, the current lemma is established. ■

Lemma 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 $m < T$,*

$$EP^*\left(\max_{m\le k\le T}\frac{1}{k}\left|\left|Z_k'\overline{F}'HFu\right|\right| > \alpha\right) \le \frac{C}{\alpha^2 m}$$

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 $g_t = \frac{1}{2\pi T}\sum_{j=1}^{T-1}\eta_j^* e^{i\lambda_j t}$. Arguing as in the proof of Lemma 5,

$$\begin{aligned} & P^*\left(\max_{m\le k\le T}\frac{1}{k}\left|Z_k'\overline{F}'HFu\right| > \alpha\right) \\ \le\; & \frac{1}{\alpha^2}\left(\frac{1}{m^2}E^*S_m^{*2} + \sum_{k=m+1}^{T}\frac{1}{k^2}\left(E^*x_k^2 d_k^2 + 2\left|E^*x_k d_k S_{k-1}^*\right|\right)\right). \end{aligned} \quad (52)$$

Because $E^*g_t g_s = \frac{1}{4\pi^2}\left(-\frac{1}{T^2} + \frac{1}{T}\mathbb{I}\left(t = s\right)\right)$, we have

$$E^*S_k^{*2} = \frac{1}{4\pi^2 T}\sum_{t,s=1}^{k}\sum_{r,v=1}^{T} x_t x_s u_r u_v \mathbb{I}\left(t - s = r - v\right) - \frac{1}{4\pi^2 T^2}\sum_{t,s=1}^{k}\sum_{r,v=1}^{T} x_t x_s u_r u_v.$$

In a similar way, $E^*x_k^2 d_k^2 = \left(4\pi^2 T\right)^{-1} x_k^2 \sum_{r=1}^{T} u_r^2 - \left(4\pi^2 T^2\right)^{-1} x_k^2 \sum_{r,v=1}^{T} u_r u_v$ and $E^*x_k d_k S_{k-1}^* = \left(4\pi^2 T\right)^{-1} x_k \sum_{t=1}^{k-1} x_t \sum_{r,v=1}^{T} u_r u_v \mathbb{I}\left(k - t = r - v\right)$ $- \left(4\pi^2 T^2\right)^{-1} x_k \sum_{t=1}^{k-1} x_t \sum_{r,v=1}^{T} u_r u_v$. Proceeding as in the proof of Lemma 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 side of (52) 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 12 *If* $T\,||\delta_T||^2 \to \infty$ *then as* $T \to \infty$,

$$\hat{k}^* - \hat{k} = O_{p^*}\left(||\delta_T||^{-2}\right).$$

Proof. Fix $\varepsilon, \eta > 0$. For any $K < \infty$,

$$\begin{aligned} & P^*\left(\left|\hat{k}^* - \hat{k}\right| > K\,||\delta_T||^{-2}\right) \\ \leq\ & P^*\left(\inf_{\widehat{N^c}(K)} \frac{Q_T^*(k)}{\left|k-\hat{k}\right|} \leq \lambda \left\|\hat{\delta}\right\|^2\right) + P^*\left(\sup_{\widehat{N^c}(K)} \frac{\left|R_T^*(k)\right|}{\left|k-\hat{k}\right|} \geq \lambda \left\|\hat{\delta}\right\|^2\right). \end{aligned}$$

By Lemma 10, we can choose $\lambda > 0$ and $K < \infty$ such that

$$EP^*\left(\inf_{\widehat{N^c}(K)} \frac{Q_T^*(k)}{\left|k-\hat{k}\right|} \leq \lambda \left\|\hat{\delta}\right\|^2\right) \leq \varepsilon \tag{53}$$

for large T. Write

$$\begin{aligned} R_T^*(k) &= 2\hat{\delta}'\hat{Z}_\Delta'\overline{F}'M_{FW_k}HF\hat{u} + \hat{u}'\overline{F}'\overline{H}'\left(M_{FW_k} - M_{FW_{\hat{k}}}\right)HF\hat{u} \\ &= 2\hat{\delta}'\hat{Z}_\Delta'\overline{F}'HFu - 2\hat{\delta}'\hat{Z}_\Delta'\overline{F}'FW_k\left(W_k'\overline{F}'FW_k\right)^{-1}W_k'\overline{F}'HF\hat{u} \qquad (54) \\ &\quad +2\hat{\delta}'\hat{Z}_\Delta'\overline{F}'HF(\hat{u}-u) + \hat{u}'\overline{F}'\overline{H}'\left(M_{FW_k} - M_{FW_{\hat{k}}}\right)HF\hat{u}. \qquad (55) \end{aligned}$$

Examining the first term of (54), we have

$$\begin{aligned} & P^*\left(\sup_{\widehat{N^c}(K)} \frac{\left|2\hat{\delta}'\hat{Z}_\Delta'\overline{F}'HFu\right|}{\left|k-\hat{k}\right|} \geq \lambda\left\|\hat{\delta}\right\|^2\right) \\ \leq\ & P^*\left(\sup_{\widehat{N^c}(K)} \left\|\frac{\hat{Z}_\Delta'\overline{F}'HFu}{k-\hat{k}}\right\| \geq \frac{\lambda}{4}\,\|\delta_T\|\right) \\ & +P^*\left(\left\|\hat{\delta}\right\| \leq \frac{1}{2}\,\|\delta_T\|\right). \end{aligned} \tag{56}$$

By Lemma 11 and by the second order stationarity, expectation of the first term on the right of (56) is bounded by

$$\frac{C}{\lambda^2\,||\delta_T||^2\,K\,||\delta_T||^{-2}} = \frac{C}{\lambda^2 K} \leq \frac{\varepsilon}{2}$$

for K large enough. Further, since $P^*\left(\left\|\hat{\delta}\right\| \leq \frac{1}{2}\|\delta_T\|\right) = \mathbb{I}\left(\left\|\hat{\delta}\right\| \leq \frac{1}{2}\|\delta_T\|\right)$, expectation of the second term on the right of (56) is equal to $P\left(\left\|\hat{\delta}\right\| \leq \frac{1}{2}\|\delta_T\|\right)$ which is smaller than $\varepsilon/2$ for large T because by Proposition 2, $\hat{\delta} \xrightarrow{p} \delta$.

Regarding the second term of (54), 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 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 8. Moreover, for any $M > 0$,

$$
\begin{aligned}
& P^*\left(\sup_{\widehat{N^c}(K)}\left\|\frac{2\hat{\delta}'}{\sqrt{T}}\frac{\hat{Z}_\Delta'\overline{F}'FW_k}{k-\hat{k}}\right\| \geq \lambda\left\|\hat{\delta}\right\|^2\right) \\
\leq \quad & P^*\left(\sup_{\widehat{N^c}(K)}\left\|\frac{\hat{Z}_\Delta'\overline{F}'FW_k}{k-\hat{k}}\right\| \geq M\right) + P^*\left(\left\|\hat{\delta}\right\| \leq \frac{2M}{\lambda}\frac{1}{\sqrt{T}}\right)
\end{aligned} \tag{57}
$$

Expectation of the first term on the right of (57) is bounded by $\varepsilon/2$ for large M by Lemma 8 and expectation of the second term on the right of (57) is bounded by $\varepsilon/2$ for large T since $\hat{\delta} = \delta_T + O_p\left(T^{-1/2}\right)$ by Proposition 2 and since $T^{-\frac{1}{2}}\left\|\delta_T\right\|^{-1} = o(1)$.

Turning to the first term of (55), from (5) and (6) and from the definition of $\hat{u}$,

$$
F(\hat{u}-u) = FX\left(\beta-\hat{\beta}\right) + F\left(Z_0 - Z_{\hat{k}}\right)\delta_T + FZ_{\hat{k}}\left(\delta_T - \hat{\delta}\right).
$$

Therefore

$$
\begin{aligned}
& P^*\left(\sup_{\widehat{N^c}(K)}\left|\frac{2\hat{\delta}'\hat{Z}_\Delta'\overline{F}'HF(\hat{u}-u)}{k-\hat{k}}\right| \geq \lambda\left\|\hat{\delta}\right\|^2\right) \\
\leq & P^*\left(\sup_{\widehat{N^c}(K)}\left\|\frac{\hat{Z}_\Delta'\overline{F}'HFX\left(\beta-\hat{\beta}\right)}{k-\hat{k}}\right\| \geq \frac{\lambda}{6}\left\|\hat{\delta}\right\|\right) \\
& + P^*\left(\sup_{\widehat{N^c}(K)}\left\|\frac{\hat{Z}_\Delta'\overline{F}'HF\left(Z_0-Z_{\hat{k}}\right)\delta_T}{k-\hat{k}}\right\| \geq \frac{\lambda}{6}\left\|\hat{\delta}\right\|\right) \\
& + P^*\left(\sup_{\widehat{N^c}(K)}\left\|\frac{\hat{Z}_\Delta'\overline{F}'HFZ_{\hat{k}}\left(\delta_T-\hat{\delta}\right)}{k-\hat{k}}\right\| \geq \frac{\lambda}{6}\left\|\hat{\delta}\right\|\right).
\end{aligned} \tag{58}
$$

For any $M > 0$, the first term on the right of (58) is bounded by

$$
P^*\left(\sup_{\widehat{N^c}(K)}\left\|\frac{\hat{Z}_\Delta'\overline{F}'HFX}{k-\hat{k}}\right\| \geq M\right) + P^*\left(\left\|\beta-\hat{\beta}\right\| \geq \frac{\lambda}{6M}\left\|\hat{\delta}\right\|\right) \tag{59}
$$

By Lemma 8, the first term of (59) is smaller than $\eta/6$ with probability larger than $1-\varepsilon/6$ for large K and M. Expectation of the second term of (59) is smaller than $\varepsilon/6$ for large T because by Proposition 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 (58) 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 (58) is bounded by

$$
\begin{aligned}
& P^*\left(\frac{1}{K}\|\delta_T\|^3 \sup_{1\leq l\leq T}\left\|Z_l'\overline{F}'HF\left(Z_0 - Z_{\hat{k}}\right)\right\| \geq \frac{\lambda}{12}\left\|\hat{\delta}\right\|\right) \\
\leq\ & P^*\left(\sup_{N(K)}\sup_{1\leq l\leq T}\left\|Z_\Delta'\overline{F}'HFZ_l\right\| \geq \frac{\lambda K}{24}\|\delta_T\|^{-2}\right) + P^*\left(\left\|\hat{\delta}\right\| < \frac{1}{2}\left|\left|\delta_T\right|\right|\right) \\
& + P^*\left(\left|\hat{k} - k_0\right| > K\left|\left|\delta_T\right|\right|^{-2}\right). \qquad (60)
\end{aligned}
$$

By Lemma 8, the first term on the right of (60) 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 (60) is bounded by $\varepsilon/3$ for large T and K, respectively.

Finally, for the second term of (55),

$$
\begin{aligned}
& P^*\left(\sup_{\widehat{N^c}(K)}\left|\frac{\hat{u}'\overline{F}'\overline{H}'\left(M_{FW_k} - M_{FW_{\hat{k}}}\right)HF\hat{u}}{k - \hat{k}}\right| \geq \lambda\left\|\hat{\delta}\right\|^2\right) \\
\leq\ & P^*\left(\sup_{\widehat{N^c}(K)}\left|\hat{u}'\overline{F}'\overline{H}'\left(M_{FW_k} - M_{FW_{\hat{k}}}\right)HF\hat{u}\right| \geq \frac{\lambda K}{4}\right) \\
& + P^*\left(\left\|\hat{\delta}\right\| \leq \frac{1}{2}\|\delta_T\|\right). \qquad (61)
\end{aligned}
$$

By Lemma 9, the first term on the right of (61) is smaller than $\eta/2$ with probability at least $1-\varepsilon/2$ for large K and T. Expectation of the the second term of (61) 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.

Collecting the results, we conclude that for arbitrary $\varepsilon > 0$ and $\eta > 0$, there exists K such that

$$
P\left(P^*\left(\sup_{\widehat{N^c}(K)}\frac{\left|R_T^*(k)\right|}{\left|k - \hat{k}\right|} \geq \lambda\left\|\hat{\delta}\right\|^2\right) > \eta\right) < \varepsilon
$$

for large T. This together with (53) and with the Markov inequality imply that

$$
P\left(P^*\left(\left|\hat{k}^* - \hat{k}\right| > K\left|\left|\delta_T\right|\right|^{-2}\right) > \eta\right) < \varepsilon
$$

for large K and T as required. ■

Lemma 13 *If $\delta_T \to 0$ and $T \|\delta_T\|^2 \to \infty$ then as $T \to \infty$,*

$$S_T^*(k) - S_T^*\left(\hat{k}\right) = \delta_T' \hat{Z}_\Delta' \overline{F}' F \hat{Z}_\Delta \delta_T - 2\delta_T' \hat{Z}_\Delta' \overline{F}' H F \hat{u} \operatorname{sgn}\left(k - \hat{k}\right) + o_{p^*}(1)$$

where $o_{p^*}(1)$ is uniform on $\hat{N}(K)$.

Proof. Write

$$Q_T^*(k) = \hat{\delta}' \hat{Z}_\Delta' \overline{F}' F \hat{Z}_\Delta \hat{\delta} - \hat{\delta}' \hat{Z}_\Delta' \overline{F}' F W_k \left(W_k' \overline{F}' F W_k\right)^{-1} W_k' \overline{F}' F \hat{Z}_\Delta \hat{\delta}$$

and

$$\begin{aligned} R_T^*(k) &= -2\hat{\delta}' \hat{Z}_\Delta' \overline{F}' H F \hat{u} \operatorname{sgn}\left(k - \hat{k}\right) \\ &\quad + 2\hat{\delta}' \hat{Z}_\Delta' \overline{F}' F W_k \left(W_k' \overline{F}' F W_k\right)^{-1} W_k' \overline{F}' H F \hat{u} \operatorname{sgn}\left(k - \hat{k}\right) \\ &\quad + \hat{u}' \overline{F}' \overline{H}' \left(M_{FW_k} - M_{FW_{\hat{k}}}\right) H F \hat{u}. \end{aligned}$$

By Lemma 8, $\hat{Z}_\Delta' \overline{F}' F W_k = O_p\left(\|\delta_T\|^{-2}\right)$ uniformly on $\hat{N}(K)$ and by Lemma 2, $\left(W_k' \overline{F}' F W_k\right)^{-1} = O_p\left(T^{-1}\right)$ uniformly on $k \in \Lambda \cdot T$. Further, by Proposition 2, $\hat{\delta} = O_p\left(\|\delta_T\|\right)$. Also, by Lemma 8, $W_k' \overline{F}' H F \hat{u} = O_{p^*}\left(T^{1/2}\right)$ uniformly on $1 \le k \le T$ and by Lemma 9, $\hat{u}' \overline{F}' \overline{H}' \left(M_{FW_k} - M_{FW_{\hat{k}}}\right) H F \hat{u} = o_{p^*}(1)$ 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

$$R_T^*(k) = -2\hat{\delta}' \hat{Z}_\Delta' \overline{F}' H F \hat{u} \operatorname{sgn}\left(k - \hat{k}\right) + o_{p^*}(1)$$

uniformly on $\hat{N}(K)$. By (43),

$$S_T^*(k) - S_T^*\left(\hat{k}\right) = \hat{\delta}' \hat{Z}_\Delta' \overline{F}' F \hat{Z}_\Delta \hat{\delta} - 2\hat{\delta}' \hat{Z}_\Delta' \overline{F}' H F \hat{u} \operatorname{sgn}\left(k - \hat{k}\right) + o_{p^*}(1)$$

uniformly on $\hat{N}(K)$. ■

Lemma 14 *Let v_T^2 be defined as in Lemma 7. If the conditions of Proposition 9 are satisfied, then for any $K > 0$, as $T \to \infty$,*

$$\arg\min_{|\rho| \le K} S_T^*\left(\hat{k} + \left[\rho v_T^{-2}\right]\right) \xrightarrow{d^*} \arg\min_{|\rho| \le K} W(\rho).$$

Proof. Write $\hat{k}+\left[\rho v_T^{-2}\right]=k$. From Lemma 13,

$$S_T^*\left(\hat{k}+\left[\rho v_T^{-2}\right]\right)-S_T^*\left(\hat{k}\right)=\delta_T'\hat{Z}_\Delta'\overline{F}'F\hat{Z}_\Delta\delta_T-2\delta_T'\hat{Z}_\Delta'\overline{F}'HF\hat{u}\,\mathrm{sgn}\left(k-\hat{k}\right)+o_{p^*}(1).$$

By Lemma 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 8 implies that

$$\frac{1}{\sqrt{T}}Z'_{\Delta([\sigma T],[(\sigma+\rho)T])}\overline{F}'HF\hat{u}\overset{p}{\Longrightarrow}\begin{cases}\frac{1}{2\pi}\Omega^{\frac{1}{2}}\left(B(\sigma+\rho)-B(\sigma)\right) & \rho\geq 0,\\ \frac{1}{2\pi}\Omega^{\frac{1}{2}}\left(B(\sigma)-B(\sigma+\rho)\right) & \rho\leq 0,\end{cases}$$

on $\{(\rho,\sigma):0\leq\sigma,\sigma+\rho\leq 1\}$ from which it follows that

$$v_T\hat{Z}_\Delta'\overline{F}'HF\hat{u}\,\mathrm{sgn}\left(k-\hat{k}\right)\overset{p}{\Longrightarrow}\begin{cases}\frac{1}{2\pi}\Omega^{\frac{1}{2}}B_1(\rho) & \rho\geq 0,\\ \frac{1}{2\pi}\Omega^{\frac{1}{2}}B_2(-\rho) & \rho<0\end{cases}$$

where B_1, B_2 are independent p-vectors of independent standard Brownian motion processes. Proceeding as in the proof of Lemma 7, we deduce that

$$S_T^*\left(\hat{k}+\left[\rho v_T^{-2}\right]\right)-S_T^*\left(\hat{k}\right)\overset{d^*}{=}W(\rho)+o_{p^*}(1)$$

on $\rho\in[-K,K]$. The lemma now follows from the continuous mapping theorem. ■

Lemma 15 *As $T\to\infty$,*

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

Proof. As in the proof of Proposition 5, assume that process $\{x_t\}$ is scalar. By definition, matrix $\hat{\Omega}^*$ is equal to

$$\frac{4\pi^2}{T}\sum_{j=1}^{T-1}I_{xx,j}I_{\hat{u}\hat{u},j}\left|\eta_j^*\right|^2+\frac{4\pi^2}{T}\sum_{j=1}^{T-1}I_{xx,j}\left(I_{\hat{v}\hat{v},j}-I_{\hat{u}\hat{u},j}\left|\eta_j^*\right|^2\right). \tag{62}$$

Let $\hat{z}_t^*=z_t(\hat{k}^*)$. Writing

$$w_{\hat{v},j}=\left(\hat{\beta}-\hat{\beta}^*\right)w_{x,j}+\hat{\delta}\left(w_{\hat{z},j}-w_{\hat{z}^*,j}\right)+\left(\hat{\delta}-\hat{\delta}^*\right)w_{\hat{z}^*,j}+w_{\hat{u},j}\eta_j^*$$

and proceeding as in the proof of Proposition 5, it can be seen that, up to a multiplicative constant, the second term of (62) is bounded in absolute value by

$$\frac{1}{T}\sum_{j=1}^{T-1}|w_{x,j}|^2\left|w_{\hat{v},j}-w_{\hat{u},j}\eta_j^*\right|^2+\frac{2}{T}\sum_{j=1}^{T-1}|w_{x,j}|^2\left|w_{\hat{v},j}-w_{\hat{u},j}\eta_j^*\right|\left|\eta_j^*w_{\hat{u},j}\right| \quad (63)$$

and that the first term of (63) is bounded by

$$\frac{3}{T}\left(\hat{\beta}-\hat{\beta}^*\right)^2\sum_{j=1}^{T-1}|w_{x,j}|^4+\frac{3}{T}\hat{\delta}^2\sum_{j=1}^{T-1}|w_{x,j}|^2\left|w_{\hat{z},j}-w_{\hat{z}^*,j}\right|^2$$
$$+\frac{3}{T}\left(\hat{\delta}-\hat{\delta}^*\right)^2\sum_{j=1}^{T-1}|w_{x,j}|^2\left|w_{\hat{z}^*,j}\right|^2. \quad (64)$$

By Proposition 8, $\hat{\beta}^*-\hat{\beta}=O_{p^*}\left(T^{-1/2}\right)$, and by Proposition 6 of Lazarová (2004), $T^{-1}\sum_{j=1}^{T-1}|w_{x,j}|^4=o_p(T)$, therefore the first term of (64) is $o_{p^*}(1)$. By the Schwarz inequality, the second term of (64) is bounded by

$$\frac{3}{T}\hat{\delta}^2\left(\sum_{j=1}^{T-1}|w_{x,j}|^4\right)^{\frac{1}{2}}\left(\sum_{j=1}^{T-1}\left|w_{\hat{z},j}-w_{\hat{z}^*,j}\right|^4\right)^{\frac{1}{2}}.$$

The sum in the second bracket of the last displayed expression is bounded by

$$\frac{1}{(2\pi T)^2}\left(\sum_{t}^{\hat{k},\hat{k}^*}|x_t|\right)^4+\frac{1}{(2\pi)^2T}\sum_{t}^{\hat{k},\hat{k}^*}|x_t|\sum_{s}^{\hat{k},\hat{k}^*}|x_s|\left(\sum_{r}^{\hat{k},\hat{k}^*}|x_r|^2\sum_{v}^{\hat{k},\hat{k}^*}|x_{t-s+v}|^2\right)^{\frac{1}{2}}.$$

For any $K>0$ and $M>0$,

$$P^*\left(\sum_{t}^{\hat{k},\hat{k}^*}|x_t|\geq M\|\delta_T\|^{-2}\right)\leq P^*\left(\sum_{t=k_0-2K\|\delta_T\|^{-2}}^{k_0+2K\|\delta_T\|^{-2}}|x_t|\geq M\|\delta_T\|^{-2}\right)$$
$$+P^*\left(\left|\hat{k}-k_0\right|>K\|\delta_T\|^{-2}\right)+P^*\left(\left|\hat{k}^*-\hat{k}\right|>K\|\delta_T\|^{-2}\right). \quad (65)$$

Proposition 1 implies that expectation of the second term on the right of (65) is bounded by $\varepsilon/3$ for large K by . By the Markov inequality and Conditions 1 and 2, expectation of the first term on the right of (65) 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 (65) is smaller than η for large K and T with probability at least $1-\varepsilon/3$. This means that $\sum_t^{\hat{k},\hat{k}^*}|x_t| = O_{p^*}\left(\|\delta_T\|^{-2}\right)$. By similar arguments, $\sum_t^{\hat{k},\hat{k}^*}|x_t|^2 = O_{p^*}\left(\|\delta_T\|^{-2}\right)$. Hence because $\hat{\delta} = \delta_T + O_p\left(T^{-1/2}\right) = O_p\left(\|\delta_T\|\right)$ by Proposition 2 and because

$$\sup_{t,s\in N(2K)} \sum_{v=k_0-2K\|\delta_T\|^{-2}}^{k_0+2K\|\delta_T\|^{-2}} |x_{t-s+v}|^2 \leq \sum_{v=k_0-6K\|\delta_T\|^{-2}}^{k_0+6K\|\delta_T\|^{-2}} |x_v|^2 ,$$

we conclude that the second term of (64) is $o_{p^*}(1)$.

Next, the sum in the third term of (64) is bounded by

$$\sum_{j=1}^{T-1} |w_{x,j}|^2 |w_{\hat{z},j}|^2 + \sum_{j=1}^{T-1} |w_{x,j}|^2 |w_{\hat{z}^*,j} - w_{\hat{z},j}|^2 . \tag{66}$$

The first term of (66) is $o_{p^*}\left(T^2\right)$ by the reasons given in the discussion of (23) and (25). The second term of (66) is $O_{p^*}\left(T^{-1/2}\|\delta_T\|^{-3}\right)$ by the reasons discussed above. Since $\hat{\delta}-\hat{\delta}^* = O_{p^*}\left(T^{-1/2}\right)$ by Proposition 8, the third term of (64) is $o_{p^*}(1)$. It follows that the first term of (63) is $o_{p^*}(1)$.

The second term of (63) is bounded by

$$2\left(\frac{1}{T}\sum_{j=1}^{T-1}|w_{x,j}|^2\left|w_{\hat{v},j}-\eta_j^* w_{\hat{u},j}\right|^2\right)^{\frac{1}{2}}\left(\frac{1}{T}\sum_{j=1}^{T-1}|w_{x,j}|^2|w_{\hat{u},j}|^2\left|\eta_j^*\right|^2\right)^{\frac{1}{2}} \tag{67}$$

by the Schwarz inequality. The first bracket of (67) has just been shown to be $o_{p^*}(1)$. The conditional expectation of the expression in the second bracket of (67) is $\hat{\Omega}$ which is $O_p(1)$ by Proposition 5. Thus the second term of (63), and consequently the second term of (62), is $o_{p^*}(1)$.

Further, the first term of (62) is equal to

$$\frac{4\pi^2}{T}\sum_{j=1}^{T-1}I_{xx,j}I_{uu,j}+\frac{4\pi^2}{T}\sum_{j=1}^{T-1}I_{xx,j}I_{uu,j}\left(\left|\eta_j^*\right|^2-1\right)+\frac{4\pi^2}{T}\sum_{j=1}^{T-1}I_{xx,j}\left(I_{\hat{u}\hat{u},j}-I_{uu,j}\right)\left|\eta_j^*\right|^2 . \tag{68}$$

By the Theorem 1 of Robinson (1998), the first term of (68) converges to Ω in probability. Further, the conditional second moment of the second term of (68) 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 5 of Lazarová (2004) 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 (68) is $o_{p^*}(1)$. Finally, by the proof of Theorem 3 of Lazarová (2004) the conditional expectation of the third term of (68) is $o_p(1)$. Combining results, we have

$$\hat{\Omega}^* = \Omega + o_{p^*}(1).$$

■

References

[1] Altissimo, F. and V. Corradi, 2003, Strong rules for detecting the number of breaks in a time series, Journal of Econometrics 117, 207-244.

[2] Andrews, D.W.K., 1993, Tests for parameter instability and structural change with unknown change point, Econometrica 61, 821-856.

[3] Antoch, J. and M. Hušková, 1999, Estimators of changes, in Asymptotics, Nonparametrics, and Time Series, ed. by Subir Ghosh, Marcel Dekker, 533-578.

[4] Antoch, J., M. Hušková and N. Veraverbeke, 1995, Change-point problem and bootstrap, Nonparametric Statistics 5, 123-144.

[5] 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.

[6] Bai, J., 1994, Least squares estimation of a shift in linear processes, Journal of Time Series Analysis 15, 453-472.

[7] Bai, J., 1997a, Estimation of a change point in multiple regression models, The Review of Economics and Statistics 79, 551-563.

[8] Bai, J., 1997b, Estimating multiple breaks one at a time, Econometric Theory 13, 315-352.

[9] Bai, J. and P. Perron, 1998, Estimating and testing linear models with multiple structural changes, Econometrica 66, 47-78.

[10] 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.

[11] Billingsley, P., 1999, Convergence of probability measures, second edition, Wiley.

[12] Bloomfield, P., 1973, An exponential model for the spectrum of a scalar time series, Biometrika 60, 217-226.

[13] 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.

[14] Bühlmann, P., 1997, Sieve bootstrap for time series, Bernoulli 3, 123-148.

[15] Bühlmann, P., 1998, Sieve bootstrap for smoothing in nonstationary time series, The Annals of Statistics 26, 48-83.

[16] 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.

[17] Dahlhaus, R. and D. Janas, 1996, A frequency domain bootstrap for ratio statistics in time series analysis, The Annals of Statistics 24, 1934-1963.

[18] Davis, R.B. and D.S. Harte, 1987, Tests for Hurst effect, Biometrika 74, 95-101.

[19] Efron, B., 1979, Bootstrap methods: Another look at the jackknife, The Annals of Statistics 7, 1-26.

[20] Efron, B. and R.J. Tibshirani, 1993, An introduction to the bootstrap, Chapman and Hall, New York.

[21] Fiteni, I., 2002, Robust estimation of structural break points, Econometric Theory 18, 349-386.

[22] Fiteni, I., 2004, τ-estimators of regression models with structural change of unknown location, Journal of Econometrics 119, 19-44.

[23] Franke, J., and W. Härdle, 1992, On bootstrapping kernel spectral estimates, The Annals of Statistics 20, 121-145.

[24] Giné, E. and J. Zinn, 1990, Bootstrapping general empirical measures, The Annals of Probability 18, 851-869.

[25] 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.

[26] Gray, H.L., N. Zhang and W.A. Woodward, 1989, On generalized fractional processes, Journal of Time Series Analysis 10, 233-257.

[27] Hidalgo, J., 2003, An alternative bootstrap to moving blocks for time series regression models, Journal of Econometrics 117, 369-399.

[28] Hinkley, D.V., 1970, Inference about the change-point in a sequence of random variables, Biometrika 57, 1-17.

[29] Horváth, L. and P. Kokoszka, 1997, The effect of long-range dependence on changepoint estimators, Journal of Statistical Planning 64, 57-81.

[30] Hosking, J.R.M., 1981, Fractional differencing, Biometrika 68, 165-176.

[31] Keenan, D.M., 1987, Limiting behavior of functionals of higher-order sample cumulant spectra, The Annals of Statistics 15, 134-151.

[32] Kounias, E.G. and T.-S. Weng, 1969, An inequality and almost sure convergence, The Annals of Mathematical Statistics 40, 1091-1093.

[33] Kuan, C.-M. and C.-C. Hsu, 1998, Change-point estimation of fractionally integrated processes, Journal of Time Series Analysis 19, 693-708.

[34] Künsch, H.R., 1989, The jackknife and the bootstrap for general stationary observations, The Annals of Statistics 17, 1217-1241.

[35] Lazarová, Š., 2004, Testing for structural changes in regression with long memory processes, Journal of Econometrics, forthcoming.

[36] Magnus, J.R. and H. Neudecker, 1988, Matrix differential calculus with applications in statistics and econometrics, John Wiley and Sons.

[37] Mandelbrot, B.B. and J.W. van Ness, 1968, Fractional Brownian motions, fractional noises and applications, SIAM Review 10, 422-437.

[38] 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.

[39] Perron, P., 2005, Dealing with structural breaks, manuscript, forthcoming in Palgrave Handbook of Econometrics, Vol. 1: Econometric Theory.

[40] 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.

[41] Robinson, P.M., 1994, Rates of convergence and optimal spectral bandwidth for long range dependence, Probability Theory and Related Fields 99, 443-473.

[42] Robinson, P.M., 1995a, Log-periodogram regression of time series with long range dependence, The Annals of Statistics 23, 1048-1072.

[43] Robinson, P.M., 1995b, Gaussian semiparametric estimation of long range dependence, The Annals of Statistics 23, 1630-1661.

[44] Robinson, P.M., 1998, Inference-without-smoothing in the presence of nonparametric autocorrelation, Econometrica 66, 1163-1182.

[45] Robinson, P.M. and F.J. Hidalgo, 1997, Time series regression with long-range dependence, The Annals of Statistics 25, 77-104.

[46] Taniguchi, M., 1982, On estimation of the integrals of the fourth order cumulant spectral density, Biometrika 69, 117-122.

[47] van der Vaart, A. W., J. A. Wellner, 1996, Weak convergence and empirical processes (with applications to statistics), Springer.

[48] 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.