

SEQUENTIAL MONITORING FOR DISTRIBUTIONAL CHANGEPOINT USING DEGENERATE U -STATISTICS

B. COOPER BONIECE^{1*}, LAJOS HORVÁTH², AND LORENZO TRAPANI³

ABSTRACT. We investigate the online detection of changepoints in the distribution of a sequence of observations using degenerate U -statistic-type processes. We study weighted versions of: an ordinary, CUSUM-type scheme, a Page-CUSUM-type scheme, and an entirely novel approach based on “recycling” past observations into the training sample. With an emphasis on completeness, we consider open-ended and closed-ended schemes, in the latter case considering both short- and long-running monitoring schemes. We study the asymptotics under the null in all cases, also proposing a consistent, Monte-Carlo based approximation of critical values; and we derive the limiting distribution of the detection delays under early and late occurring changes under the alternative, thus enabling to quantify the expected delay associated with each procedure. As a crucial technical contribution, we derive all our asymptotics under the assumption that the kernels associated with our U -statistics are square summable, instead of requiring the typical absolute summability, which makes our assumption naturally easier to check. Our simulations show that our procedures work well in all cases considered, having excellent power versus several types of distributional changes, and appearing to be particularly suited to the analysis of multivariate data.

1. INTRODUCTION

We study the online, real-time detection of changepoints in the distribution of a (possibly multivariate) sequence of random variables $\{\mathbf{X}_i, i \geq 1\}$. To formalise the question, suppose that the observations have the same distribution (say, $\mathbf{X}_i \sim F$) over a “training” period $1 \leq i \leq m$; we study procedures to sequentially test for the null that, as new data come in,

¹DEPARTMENT OF MATHEMATICS, DREXEL UNIVERSITY, PHILADELPHIA, PA 19104 USA

²DEPARTMENT OF MATHEMATICS, UNIVERSITY OF UTAH, SALT LAKE CITY, UT 84112–0090 USA

³DEPARTMENT OF ECONOMICS AND MANAGEMENT, UNIVERSITY OF PAVIA, PAVIA, ITALY; DEPARTMENT OF ECONOMICS, FINANCE AND ACCOUNTING; SCHOOL OF BUSINESS AND ECONOMICS, UNIVERSITY OF LEICESTER, LEICESTER, U.K.

*RESEARCH SUPPORTED IN PART UNDER NSF GRANT DMS-2413558.

E-mail addresses: cooper.boniece@drexel.edu, horvath@math.utah.edu,

lorenzo.trapani@unipv.it.

2020 *Mathematics Subject Classification.* 60F17.

no changes occur in the distribution F - that is, we test for the null that $\mathbf{X}_{m+k} \sim F$, for each $k \geq 1$ over a “monitoring” horizon.

Detecting the possible presence of structural instability is arguably of great importance in all applied sciences. Examples include economics and finance, where instability has obvious implications for forecasting and decision-making (see e.g. [Smith and Timmermann, 2021](#)); engineering, where the safety and serviceability of engineering structures requires continuous monitoring (see e.g. [Sun et al., 2020](#), and [Malekloo et al., 2022](#)); and the analysis of biomedical time series data ([Fiecas et al., 2024](#)). Indeed, many applications require the use of the whole distribution, as opposed to specific moments such as the mean or the variance, whence the importance of testing for *distributional* changes: [Fu et al. \(2023\)](#), *inter alia*, discuss several examples in economics and finance, including density forecast and the detection of changes in the tail risk of financial variables. A comprehensive analysis of the literature on the changepoint problem in general - and on online detection in particular - is beyond the scope of this paper, and we refer to [Aue and Kirch \(2024\)](#) and [Horváth and Rice \(2024\)](#) for reviews. However, within this literature, contributions on the detection of distributional changepoints are rare, and the vast majority of papers deals with breaks in a specific moment such as, typically, the mean or the variance. Some exceptions are the papers on retrospective, offline detection by [Inoue \(2001\)](#), who uses the empirical distribution function, and [Hušková and Meintanis \(2006\)](#) and [Boniece et al. \(2025\)](#) who, *inter alia*, use the empirical characteristic function. On the other hand, online, real-time detection of changes in the distribution is relatively underexplored (see [Horváth et al., 2021](#), for an exception). In this paper, we study several online changepoint detection schemes based on weighted functionals of degenerate U -statistics, considering both open-ended schemes (where monitoring goes on for an indefinite amount of time) and closed-ended schemes (where monitoring goes on until a pre-specified time, after which it stops). The use of U -statistics-type processes in the context of changepoint detection was firstly proposed in [Csörgő and Horváth \(1989\)](#),

and subsequently studied in several contributions - examples include [Matteson and James \(2014\)](#), [Biau et al. \(2016\)](#) and [Dehling et al. \(2022\)](#) for retrospective changepoint, and [Kirch and Stoehr \(2022b\)](#) for online detection. [Boniece et al. \(2025\)](#) study retrospective distributional changepoint detection for functional data using a special case of a degenerate U -statistic, using the (generalised) energy distance ([Székely and Rizzo, 2005](#) and [Székely and Rizzo, 2013](#)). Developing a general theory for the use of degenerate U -statistics for online detection allows to consider a wide variety of measures of distributional distance, including those that are rotation invariant (such as the energy distance), which naturally lend themselves to the study of changes in multivariate data - as opposed to the use of statistics based on e.g. Cramer's distance, such as the ones employed in [Inoue \(2001\)](#).

Hypotheses of interest and the main contributions of this paper

Given a training sample $\{\mathbf{X}_i, 1 \leq i \leq m\}$ with common distribution F , as new data come in we test for the null of no distributional change at each $k \geq 1$, viz.

$$(1.1) \quad H_0 : \mathbf{X}_{m+k} \sim F, \quad k \geq 1,$$

versus the alternative that, at some point in time k_* , the distribution changes. At each point in time k during the monitoring horizon, we compute a measure of distance (defined naturally as a U -statistic) between the observations in the training sample, and the ones recorded thereafter up to k . Under the null, at each k it can be expected that such a distance will fluctuate around zero. Conversely, in the presence of a changepoint at time k_* , the distance will drift away from zero. Indeed, for the specific case of detecting changes retrospectively using a particular instance of (generalised) energy distance, [Boniece et al. \(2025\)](#) show that this behaves, modulo an asymptotically negligible term, like the square of a CUSUM process. In this contribution, we show analogous result in the online setting for a broader class of degenerate U -statistics.

The asymptotic proximity between U -statistics and CUSUM processes is a very important result, which offers the possibility of extending the basic U -statistic process. In particular, in this paper we propose *weighted* versions of degenerate U -statistics, using weights designed in order to reduce the delay in detecting a changepoint occurring close to the beginning of the monitoring horizon; further, to enhance the ability to detect changepoints occurring later on, we also study a Page-CUSUM scheme (Aue and Kirch, 2024). Indeed, by exploiting the asymptotic approximation with a series of squared CUSUM processes, we are also able to propose an entirely novel monitoring scheme based on “recycling” past observations, which have already been tested for, in the monitoring horizon. We study the limiting behaviour of our statistics under the null, showing that the asymptotic distributions of our test statistics converge to the suprema of weighted, infinite sums of centered, squared standard Wiener processes, analogous to an infinite weighted χ^2 representation in the “classical” case (Serfling, 2009), with weights given by the eigenvalues of the integral operator associated with the kernel employed in the construction of the U -statistic. Therefrom, asymptotic critical values can be calculated by standard Monte Carlo techniques. In addition to the “testing” side, we also investigate the “estimation” side of the monitoring problem: under the alternative, the point in time at which the procedure marks a break is a natural estimate of the break date. We derive the limiting distribution of the detection delay, defined as the discrepancy between the estimated break date and the actual one, under various monitoring schemes (open-ended, closed-ended with a long monitoring horizon, and closed-ended with a short monitoring horizon), and various locations of the changepoint (early or late).

We make at least five main contributions to the extant literature. Firstly, as a theoretical contribution, all the asymptotic theory is derived by requiring only the *square* summability of the eigenvalues associated to the kernels of the U -statistics. This is a marked difference compared to the usual requirement of the *absolute* summability of the eigenvalues (see

Biau et al., 2016, as a prime example). Given that the square summability of the eigenvalues is a natural consequence of the existence of the second moment of the kernel defining the U -statistic, this is an easily testable assumption, which makes our results useful also from an applied viewpoint. This requirement allows for a great deal of flexibility; in particular, it paves the way to using an entire suite of distance-based kernels - e.g., those which arise from spaces of strong negative type (Lyons, 2013), which are theoretically consistent against all distributional alternatives. Secondly, as a methodological contribution, with an emphasis on completeness we study CUSUM-type and Page-CUSUM-type detection schemes, and both open-ended and closed-ended monitoring schemes - in the latter case, considering both “short-horizon” monitoring schemes (where the procedure is carried out across a monitoring horizon whose length is negligible compared to the length of the training sample), and “long-horizon” monitoring schemes (where the monitoring horizon is comparably as large as - or larger than - the training sample). Thirdly, in addition to the CUSUM and Page-CUSUM schemes, we propose a novel detection scheme, based on expanding the training sample with observations taken from the monitoring horizon after a sufficient number of these have been shown to not have undergone any changes. This results in a richer, more informative training sample on the one hand, and, on the other hand, in a monitoring horizon where - if and when a break occurs - past observations do not “water down” the impact of the change and can lead to significantly improved detection times. Fourthly, as an advance to the “estimation” side of the problem (i.e., to the study of our test statistics under the alternative hypothesis), for all our proposed statistics we study the (distribution of the) detection delay - in the presence of a changepoint - in the cases of an early change, and also of a late change. In particular, the literature typically considers only the former (see, *inter alia*, Aue and Horváth, 2004; see however Kirch and Stoeckl, 2022a, for an exception), usually requiring a constraint on the location of the break date to be “close” to the beginning of the monitoring period (or, in other words, to happen after a period of time whose length

is asymptotically negligible compared to the length of the historical training period). Our results extend and complement the existing ones, thus helping the applied user to form an expectation on the possible detection delay, irrespective of the location of the break date k_* . Finally, we also develop an approach to test *retrospectively* for the presence of breaks based on U -statistics, along similar lines as [Matteson and James \(2014\)](#), [Biau et al. \(2016\)](#) and [Dehling et al. \(2022\)](#). This is important for our purposes, because a typical assumption in the monitoring literature is the presence of a training sample during which no break occurred, and our proposed test makes this assumption testable. Furthermore, this is also a direct technical advance on the current literature (see e.g. [Biau et al., 2016](#)), since - even in this case - we are able to derive the asymptotics under the assumption of square summability of the eigenvalues associated to the kernels of the U -statistics.

In addition to the five main contributions mentioned above, we also report an extensive discussion (in Section [4.3](#)) on possible kernel functions, and study how to “generate” kernels which are distribution-determining - that is, which give monitoring schemes that have non-trivial power in the presence of *any* distributional changes. Whilst most of the results in Section [4.3](#) are already in the extant literature, the result in Theorem [4.3](#) is, to the best of our knowledge, novel. Further, in Section [4.4](#), we study the validity of a Monte Carlo approximation to the asymptotic critical values, also offering practical guidelines. Finally, our simulations show that our procedures work particularly well with multivariate data with finite dimension. We view this as an important practical feature of our work. In the literature there are several methodologies to detect distributional changes in *univariate* data (see the review in [Horváth and Rice, 2024](#)), and some contributions for high dimensional data ([Chakraborty and Zhang, 2021](#); [Drikvandi and Modarres, 2025](#)).¹ The case of multivariate, but not high-dimensional, data is underexplored, and the available techniques usually yield

¹Indeed, our procedures, in principle, can be applied even in the case of infinite-dimensional metric spaces.

mixed results, either due to computational and scalability problems, or to a general poor performance in controlled settings (see e.g. [Holmes et al., 2024](#)).

The remainder of the paper is organised as follows. In Section 2, we spell out the null and alternative hypotheses, and the relevant assumptions; we also define the detectors employed in our monitoring schemes. In Section 3, we report the asymptotic theory under the null (Section 3.1) and the alternative, including the limiting distribution of the detection delays (Section 3.2). In Section 4, we: introduce a novel detection scheme (Section 4.1); develop a test for the assumption that the observations during the training sample have the same distribution (Section 4.2); study various examples of kernel functions, also advancing on the extant literature (Section 4.3); and propose a consistent Monte-Carlo approximation of critical values (Section 4.4). A comprehensive simulation exercise (wherein we also discuss the practical implementation of our methodology), and empirical applications, are in Section 5. Section 6 concludes. All lemmas and proofs, and further Monte Carlo and empirical results, are relegated to the Supplement.

NOTATION. Throughout, for positive sequences a_m, b_m , we write $a_m \sim b_m$ if $a_m/b_m \rightarrow 1$ as $m \rightarrow \infty$. We denote $a_m \ll b_m$ to mean $a_m = o(b_m)$ and similarly $a_m \gg b_m$ means $b_m = o(a_m)$ as $m \rightarrow \infty$. Convergence in distribution is denoted as $\xrightarrow{\mathcal{D}}$. We denote binomial coefficients as $\binom{p}{q}$. Other, relevant notation is introduced later on in the paper. We often use the shorthand notation $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$. For any interval $I \subseteq [0, \infty)$, we write $\mathbf{D}(I)$ the space of càdlàg functions endowed with with the Skorokhod topology.

2. ASSUMPTIONS AND MONITORING SCHEMES

Let $\mathbf{X}_1, \mathbf{X}_2, \dots$ be a sequence of random elements taking values in a separable metric space $(\mathcal{X}, \rho)$. We assume that there exists a historical training period $\{\mathbf{X}_i, 1 \leq i \leq m\}$ during which no change took place. Letting F denote the distribution of $\mathbf{X}_1$, we make the following

Assumption 2.1. It holds that $\mathbf{X}_i \sim F$ for all $i = 1, \dots, m$.

Assumption 2.1 is typical in this literature, where it is also known as the *noncontamination assumption* (Chu et al., 1996). In the spirit of making assumptions that are testable, as mentioned in the introduction, in Section 4.2 we construct a test (based on the same approach as discussed herein for online monitoring) to check retrospectively for no changepoint in the distribution of $\{\mathbf{X}_i, 1 \leq i \leq m\}$.

After the training period, incoming observations $\mathbf{X}_{m+k}$ are monitored, where $k \geq 1$ denotes the “current” monitoring time; we test for the null hypothesis of distributional stability versus the alternative hypothesis that a change occurs in the distribution at some point in time k_* :

$$(2.1) \quad H_A : \mathbf{X}_{m+k} \sim \begin{cases} F & k = 1, 2, \dots, k_*, \\ F_* & k = k_* + 1, k_* + 2, \dots \end{cases}$$

where $k_* \geq 1$, and $F_* \neq F$ is an unspecified distribution on $\mathcal{X}$.

Throughout this work we use the following assumption.

Assumption 2.2. It holds that $\{\mathbf{X}_i, i \geq 1\}$ is an independent sequence.

We now present the monitoring schemes, starting with a preview of how they work. At each point during the monitoring horizon, k , we construct a “detector” $\mathcal{D}_m(k)$, based on comparing the observations in the historical training sample $\{\mathbf{X}_i, 1 \leq i \leq m\}$ against the observations available in the monitoring sample up until k $\{\mathbf{X}_{m+i}, 1 \leq i \leq k\}$. As mentioned in the introduction, such a detector (heuristically) is constructed as a partial sum process of quantities which, under the null of no break, have mean zero; consequently, as k increases, under the null $\mathcal{D}_m(k)$ should range within a “boundary (function)” which evolves with k , say $g_m(k)$. As soon as such boundary is crossed, the null is rejected and a changepoint is marked; formally, H_0 is rejected as soon as

$$(2.2) \quad \mathcal{D}_m(k) > c g_m(k),$$

where the constant $c > 0$ is a critical value chosen in conjunction with the historical sample to control the asymptotic Type I Error rate.

We now introduce our detectors. Following [Matteson and James \(2014\)](#), [Biau et al. \(2016\)](#) and [Dehling et al. \(2022\)](#), our detectors $\mathcal{D}_m(k)$ are based on *degenerate* U -statistics (see e.g. [Van der Vaart, 2000](#), for a general treatment). Let $h : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ be any function satisfying

Assumption 2.3. It holds that $h(\mathbf{x}, \mathbf{y}) = h(\mathbf{y}, \mathbf{x})$; for *i.i.d.* elements $\mathbf{X}, \mathbf{Y} \sim F$, it holds that

$$(2.3) \quad \mathbb{E}h^2(\mathbf{X}, \mathbf{Y}) = \iint h^2(\mathbf{x}, \mathbf{y})dF(\mathbf{x})dF(\mathbf{y}) < \infty.$$

Assumption 2.3 requires the second moment of $h(\mathbf{X}, \mathbf{Y})$ to be finite. Heuristically, our statistics are based on sums of $h(\mathbf{X}_i, \mathbf{X}_j)$, and therefore assuming that the second moment thereof is a natural requirement to derive the asymptotics. As mentioned in the introduction, this part of the assumption is testable: given a (user-chosen) kernel $h(\cdot, \cdot)$, it can be checked whether its second moment is finite or not based e.g. on the procedures discussed in [Trapani \(2016\)](#) and [Degiannakis et al. \(2023\)](#). Indeed, the assumption is “constructive”: after determining how many moments are admitted by the data, a $h(\cdot, \cdot)$ can be chosen, by the applied user, so as to satisfy the assumption.

Note, importantly, that the assumption on the finiteness of the second moment is for the kernel $h(\mathbf{X}, \mathbf{Y})$, and *not for the data* $\mathbf{X}$: hence, $\mathbf{X}$ need not even admit any finite polynomial moment *per se*, as long as an appropriate kernel is chosen.

Given a kernel $h(\mathbf{x}, \mathbf{y})$ satisfying Assumption 2.3, for each m and $k \geq 2$, let

$$(2.4) \quad U_m(h; k) = \frac{2}{km} \sum_{i=1}^m \sum_{j=m+1}^{m+k} h(\mathbf{X}_i, \mathbf{X}_j) - \binom{m}{2}^{-1} \sum_{1 \leq i < j \leq m} h(\mathbf{X}_i, \mathbf{X}_j) \\ - \binom{k}{2}^{-1} \sum_{m < i < j \leq m+k} h(\mathbf{X}_i, \mathbf{X}_j).$$

We then define the detector

$$(2.5) \quad \mathcal{D}_m^{(1)}(k) = m^{-1} k^2 |U_m(h; k)|,$$

and its Page-type counterpart (see [Page, 1954](#); [Fremdt, 2015](#); and [Aue and Kirch, 2024](#))

$$(2.6) \quad \mathcal{D}_m^{(2)}(k) = m^{-1} \max_{0 \leq r < k} (k - r)^2 |U_m(h; r, k)|,$$

where for each $m, k \geq 2$, $0 \leq r < k - 1$,

$$(2.7) \quad U_m(h; r, k) = \frac{2}{(k - r)m} \sum_{i=1}^m \sum_{j=m+r+1}^{m+k} h(\mathbf{X}_i, \mathbf{X}_j) \\ - \binom{m}{2}^{-1} \sum_{1 \leq i < j \leq m} h(\mathbf{X}_i, \mathbf{X}_j) - \binom{k-r}{2}^{-1} \sum_{m+r < i < j \leq m+k} h(\mathbf{X}_i, \mathbf{X}_j).$$

We use the following family of *weighted* boundary functions

$$(2.8) \quad g_m(k) = \left(\frac{k/m}{1 + k/m} \right)^\beta \left(1 + \frac{k}{m} \right)^2 = g \left(\frac{k}{m} \right).$$

As is typical in this literature, the boundary functions defined in (2.8) depend on a user-chosen weight $0 \leq \beta < 1$, which determines the weights assigned to the fluctuations of $U_m(h; r, k)$: as β increases, the weight also increases, and therefore higher power/faster detection under the alternative may be expected. [Horváth et al. \(2004\)](#), [Horváth et al. \(2007\)](#) and [Ghezzi et al. \(2024\)](#) study online changepoint detection based on the CUSUM process with various values of β ; [Horváth and Trapani \(2025\)](#) and [Horváth et al. \(2025\)](#) study a weighted version of the Page-CUSUM process.

For a chosen detector $\mathcal{D}_m(k)$, we consider two separate types of monitoring schemes. First,

an “open-ended” (or indefinite monitoring horizon) scheme, based on the stopping rule

$$(2.9) \quad \tau_m = \tau_m(\mathfrak{c}) = \begin{cases} \min\{k \geq 2 : \mathcal{D}_m(k) > \mathfrak{c}g_m(k)\} \\ \infty, & \text{if } \mathcal{D}_m(k) \leq \mathfrak{c}g_m(k) \text{ for all } k \geq 2. \end{cases}$$

The procedure goes on forever, until it rejects H_0 - corresponding to having $\tau_m < \infty$.

However, by definition, monitoring based on τ_m may *never* terminate, which may not be suitable in some applications. Thus, we also consider finite horizon (or “closed”) monitoring schemes, which are based on the stopping rule

$$(2.10) \quad \tau_{m,M} = \tau_{m,M}(\mathfrak{c}) = \begin{cases} \min\{2 \leq k \leq M-1 : \mathcal{D}_m(k) > \mathfrak{c}g_m(k)\} \\ M, & \text{if } \mathcal{D}_m(k) \leq \mathfrak{c}g_m(k) \text{ for all } 2 \leq k < M, \end{cases}$$

where $M \geq 2$ is a user-specified monitoring horizon.²

3. MAIN RESULTS

We report results under the null and under the alternatives for the “classical” monitoring schemes based on the detectors $\mathcal{D}_m^{(1)}(k)$ and $\mathcal{D}_m^{(2)}(k)$; a novel scheme is introduced in Section 4.1. From hereon, we assume that Assumptions 2.1-2.3 are in force, and thus we omit them from the statements of our results.

Let $\mathbf{X}, \mathbf{Y} \stackrel{iid}{\sim} F$. For a given h satisfying Assumption 2.3, we define its degenerate counterpart

$$(3.1) \quad \bar{h}(\mathbf{x}, \mathbf{y}) = h(\mathbf{x}, \mathbf{y}) - \mathbb{E}h(\mathbf{X}, \mathbf{y}) - \mathbb{E}h(\mathbf{x}, \mathbf{Y}) + \mathbb{E}h(\mathbf{X}, \mathbf{Y}).$$

To the function h , we associate the integral operator $A : \mathcal{L}^2(F) \rightarrow \mathcal{L}^2(F)$, defined by $Ag(\mathbf{x}) = \mathbb{E}h(\mathbf{x}, \mathbf{Y})g(\mathbf{Y}) = \int \bar{h}(\mathbf{x}, \mathbf{y})g(\mathbf{y})dF(\mathbf{y})$. Under Assumption (2.3), the spectral theorem (e.g. [Riesz and Sz.-Nagy, 1990](#)) yields that there exists an orthonormal basis $\{\phi_k\}_{k \geq 1}$

²Formally, monitoring based on $\tau_{m,M}$ rejects H_0 if $\tau_{m,M} < M$

of $\mathcal{L}^2(F)$ such that $A\phi_\ell = \lambda_\ell\phi_\ell$, $\ell \geq 1$, where $\lambda_\ell \in \mathbb{R}$ for all $\ell \geq 1$,³ such that

$$(3.2) \quad \sum_{\ell=1}^{\infty} \lambda_\ell^2 < \infty.$$

Let $\{W_1(u), u \geq 0\}$, $\{W_2(u), u \geq 0\}$, $\dots$ be independent Wiener processes, and define

$$(3.3) \quad \Gamma(u) = \sum_{\ell=1}^{\infty} \lambda_\ell (W_\ell^2(u) - u);$$

the process defined in (3.3) is typically found when studying the limiting distribution of degenerate U -statistics (e.g. [Serfling, 2009](#)).

3.1. Monitoring under H_0 . Based on the stopping rules defined in (2.9) and (2.10), the case of no detection taking place corresponds to the events $\{\tau_m = \infty\}$ and $\{\tau_{m,M} = \infty\}$ respectively. In this section, we study the probability of such events under the null hypothesis H_0 - and, therefore, the asymptotic distribution of our statistics.

We begin by presenting the limiting behaviour of CUSUM-type schemes based on using the detector $\mathcal{D}_m^{(1)}(k)$ defined in (2.5).

Theorem 3.1. *Assume H_0 holds, and consider the detector $\mathcal{D}_m(k) = \mathcal{D}_m^{(1)}(k)$. Let g_m be as in (2.8). As $m \rightarrow \infty$,*

$$(3.4) \quad P\{\tau_m = \infty\} \rightarrow P\left\{\sup_{0 < u < 1} u^{-\beta} |\Gamma(u)| > c\right\}.$$

Suppose $M = M_m \rightarrow \infty$ such that $M/m \rightarrow a_0 \in (0, \infty]$, and let $u_0 = a_0/(1 + a_0)$. Then

$$(3.5) \quad P\{\tau_{m,M} = \infty\} \rightarrow P\left\{\sup_{0 < u < u_0} u^{-\beta} |\Gamma(u)| > c\right\}.$$

³With no loss of generality, we assume they are ordered as $|\lambda_1| \geq |\lambda_2| \geq \dots$

Suppose $M = M_m \rightarrow \infty$ such that $M/m \rightarrow 0$, and let the boundary function g_m be given by $g_m(k) = (M/m) (k/M)^\beta$. Then

$$(3.6) \quad P\{\tau_{m,M} = \infty\} \rightarrow P\left\{\sup_{0 < u < 1} u^{-\beta} |\Gamma(u)| > \mathfrak{c}\right\}.$$

Theorem 3.1 contains the limiting distribution of the test statistics in various cases. Part (i) of the theorem refers to an open-ended, indefinite-horizon monitoring scheme; asymptotic control of the Type I error rate under the null is guaranteed by choosing $\mathfrak{c} = \mathfrak{c}_\alpha$ such that $P\left\{\sup_{0 < u < 1} u^{-\beta} |\Gamma(u)| > \mathfrak{c}_\alpha\right\} = 1 - \alpha$. Parts (ii) and (iii) provide analogous statements in the finite-horizon monitoring setting. In particular, part (ii) corresponds to a “long-horizon” monitoring, in the sense the monitoring horizon M is either comparable or much larger than the length of the historical sample m . The limiting distribution in both cases is given by the supremum of the weighted version of $|\Gamma(u)|$; the only difference is in the interval over which the supremum is taken. From a practical point of view, the relevant case is always (ii) - that is, critical values should be always computed from the supremum taken over the interval $(0, u_0)$, and case (i) can be viewed as an always more conservative asymptotic approximation. Finally, part (iii) corresponds to “short-horizon” monitoring, where the length of the monitoring horizon is effectively negligible compared to the length of the training period. In all cases, the critical values $\mathfrak{c}_\alpha$ can be derived by simulations, based on the definition of $\Gamma(u)$ in (3.3) - see Section 4.4.

We now study the limiting behavior of Page-type monitoring scheme, based on $\mathcal{D}_m^{(2)}(k)$ defined in (2.6). Define the two parameter process

$$(3.7) \quad G(u, v) = \sum_{\ell=1}^{\infty} \lambda_\ell \left[\left(W_\ell(u) - \frac{1-u}{1-v} W_\ell(v) \right)^2 - \left(u - v \left(\frac{1-u}{1-v} \right) \right) \left(1 - v \left(\frac{1-u}{1-v} \right) \right) \right],$$

for each $0 \leq u, v \leq 1$, with $\{W_\ell(u), u \geq 0\}$ as in (3.3), and let

$$(3.8) \quad \bar{\Gamma}(u) = \sup_{0 < v \leq u} |G(u, v)|, \quad 0 \leq v \leq 1.$$

Theorem 3.2. *Assume H_0 holds, and consider the Page-type detector $\mathcal{D}_m(k) = \mathcal{D}_m^{(2)}(k)$. Then the statements of Theorem 3.1 hold with $\bar{\Gamma}$ replacing Γ .*

3.2. Monitoring under the alternative. Consider the following notation. Let $F_* = \theta G + (1 - \theta)F$, where $0 < \theta < 1$, and $G(\mathbf{x})$ is a distribution function which, under the alternative, “contaminates” F . Define

$$(3.9) \quad h_1(\mathbf{x}) = \int h(\mathbf{x}, \mathbf{y}) dF(\mathbf{y}), \quad h_2(\mathbf{x}) = \int h(\mathbf{x}, \mathbf{y}) dF_*(\mathbf{y}),$$

$$(3.10) \quad v(\mathbf{x}) = \int h(\mathbf{x}, \mathbf{y}) d(F(\mathbf{y}) - G(\mathbf{y})) = \theta^{-1} (h_1(\mathbf{x}) - h_2(\mathbf{x})),$$

$$(3.11) \quad \nu_1 = \int v(\mathbf{x}) dF(\mathbf{x}), \quad \nu_2 = \int v(\mathbf{x}) dF_*(\mathbf{x}).$$

Assumption 3.1. As $m \rightarrow \infty$, $m\theta^2 |\mathfrak{D}_h(F, G)| \rightarrow \infty$, where $\mathfrak{D}_h(F, G)$ is defined in (4.6).

Assumption 3.1 states that the change can be “small”, but not “too small”, in order for it to be detected. In particular, whenever $|\mathfrak{D}_h(F, G)| \neq 0$, the “degree of contamination” θ is required to be larger than $O(m^{-1/2})$, but it can drift to zero, corresponding to the case of a “vanishing break”. By (3.11), $\theta^{-1}(\nu_1 - \nu_2) = \int h(\mathbf{x}, \mathbf{y}) d(F - G)^2(\mathbf{x}, \mathbf{y}) = \mathfrak{D}_h(F, G)$; hence Assumption 3.1 can be equivalently written as $m\theta |\nu_1 - \nu_2| \rightarrow \infty$, which is used extensively in the proofs.

Theorem 3.3. *Under Assumption 3.1, when either $\mathcal{D}_m(k) = \mathcal{D}_m^{(1)}(k)$ or $\mathcal{D}_m^{(2)}(k)$, it holds that $\lim_{m \rightarrow \infty} P(\tau_m < \infty) = 1$.*

Whenever $(\mathfrak{X}, \rho)$ has strong negative type (see Example 4.3 below), then under the choice $h(\mathbf{x}, \mathbf{y}) = \rho(\mathbf{x}, \mathbf{y})$, Theorem 3.3 states that our procedure is consistent against *all* distributional change alternatives, as long as $m\theta^2 \rightarrow \infty$.

The results derived thus far refer to “testing”. We now report several results, under H_A , concerning “estimation”, by studying the *detection delay* associated with our procedures $\kappa_m - k_*$, where

$$(3.12) \quad \kappa_m = \begin{cases} \min\{k > k_* : \mathfrak{D}_m(k) > c g_m(k)\} \\ \infty, & \text{if } \mathfrak{D}_m(k) \leq c g_m(k) \text{ for all } k > k_*. \end{cases}$$

We focus on two distinct settings: an “early change”, when $k_* \leq C$ for some unknown fixed constant $C > 0$, and a “late change”, wherein $k_* = \lfloor c_* m \rfloor$ for some $c_* > 0$.

We introduce some further notation. For $\mathbf{X} \sim F$ and $\mathbf{X}^* \sim F_*$, we set $\sigma^2 = \text{var}(v(\mathbf{X}))$, and $\sigma_*^2 = \text{var}(v(\mathbf{X}^*))$; note under Assumption 3.1, σ_* and σ may also drift to zero. In order to simplify some asymptotic expressions, we make the following

Assumption 3.2. As $m \rightarrow \infty$, $\sigma(\mathfrak{D}_h(F, G))^{-1/2} \rightarrow \zeta$, and $\sigma_*(\mathfrak{D}_h(F, G))^{-1/2} \rightarrow \zeta_*$.

The next two theorems provide the limiting distribution of $\kappa_m - k_*$ when both $\mathfrak{D}_m(k) = \mathfrak{D}_m^{(1)}(k)$ and $\mathfrak{D}_m^{(2)}(k)$. Let

$$(3.13) \quad \rho = \frac{1 - \beta}{2 - \beta}, \quad w = \left(\frac{c}{\theta |\nu_1 - \nu_2|} \right)^{1/(2-\beta)}, \quad v_m = \frac{2\sigma_*}{(2 - \beta) |\nu_1 - \nu_2|} (wm^\eta)^{1/2},$$

$$(3.14) \quad v'_m = \frac{m^{1/2}}{\theta |\mathfrak{D}_h(F, G)|^{1/2}}.$$

Theorem 3.4. Assume H_A holds. Let κ_m be as in (3.12) with $\mathfrak{D}_m(k) = \mathfrak{D}_m^{(1)}(k)$ and g_m as in (2.8). If $k_* \leq C$ with some $C > 0$, and Assumptions 3.1 and 3.2 hold, then

$$(3.15) \quad \frac{\kappa_m - k_* - wm^\rho}{v_m} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

If $k_* = c_* m$ for some $c_* > 0$, and Assumptions 3.1 and 3.2 hold, then

$$(3.16) \quad \frac{\kappa_m - k_*}{v'_m} \xrightarrow{\mathcal{D}} \mathcal{H}_{c_*}(c),$$

where $\mathcal{H}_{c_*}$ is defined in (3.16) in the Supplement.

Theorem 3.5. Assume H_A holds. Let κ_m be as in (3.12) based on the detector $\mathcal{D}_m(k) = \mathcal{D}_m^{(2)}(k)$, with g_m as in (2.8). If $k_* \leq C$ with some $C > 0$, and Assumptions 3.1 and 3.2 hold, and further

$$(3.17) \quad \sigma_* \theta(wm^\eta)^{3/2-\beta} \rightarrow \infty,$$

where w, η are given in (3.13), then the limit (3.15) holds.

If $k_* = c_* m$ for some $c_* > 0$, and Assumption 3.1 and 3.2 hold, then

$$(3.18) \quad \frac{\kappa_m - k_*}{v'_m} \xrightarrow{\mathcal{D}} \tilde{\mathcal{H}}_{c_*}(\mathbf{c}),$$

where $\tilde{\mathcal{H}}_{c_*}$ is defined in (3.18) in the Supplement, and v'_m is as in (3.14).

Theorems 3.4 and 3.5 describe the delay time under both monitoring schemes (CUSUM and Page-CUSUM, respectively). The theorems (roughly) state that - in the early change regime where k_* occurs a finite number of periods after the start of the monitoring horizon - the expected delay is given by $w m^\rho$ - that is, roughly $w m^\rho$ observations after the change-point are needed prior to detecting a change. Since ρ approaches 0 as β approaches 1, choosing values of β close to 1 can shorten detection times; this is also observed in Aue and Horváth (2004). Considering the late change regime, as mentioned in the theorems, the (lengthy) definition of the limit variables $\mathcal{H}_{c_*}$ and $\tilde{\mathcal{H}}_{c_*}$ is relegated to equations (D.51) and (D.77) in the Supplement, for ease of exposition. We remark, however, that both $\mathcal{H}_{c_*}(\mathbf{c})$ and $\tilde{\mathcal{H}}_{c_*}(\mathbf{c})$ are non-Gaussian, strictly positive, and for the same fixed $\mathbf{c}$, the variable $\tilde{\mathcal{H}}_{c_*}(\mathbf{c})$ can be seen to be stochastically smaller than $\mathcal{H}_{c_*}(\mathbf{c})$, reflecting a well-documented advantage of shorter delay times in Page-type detection procedures under late changes (c.f. Fremdt, 2015). Seeing as both $\mathcal{H}_{c_*}$ and $\tilde{\mathcal{H}}_{c_*}$ are well-defined random variables, the theorems entail that, in the late change regime, the number of observations needed in order to detect a change is proportional to v'_m . When

$0 < \theta < 1$ (i.e., when the size of the break is fixed), this entails that the detection delay is proportional to $m^{1/2}$; seeing as the breakdate k_* is proportional to m , this means that detection is relatively quick. On the other hand, when $\theta \rightarrow 0$ (corresponding to a break of vanishing size), this inflates v'_m and, therefore, the detection delay.

Finally, as can be expected, in all cases small values of $|\mathfrak{D}_h(F, G)|$ yield larger delay times.

4. COMPLEMENTS AND EXTENSIONS

4.1. A “repurposing” approach. Consider $U_m(h; k)$ defined in (2.4). In essence, at each time k , $U_m(h; k)$ functions as a two-sample test statistic that compares the historical sample $\mathbf{X}_1, \dots, \mathbf{X}_m$ with the entire monitored sequence $\mathbf{X}_{m+1}, \dots, \mathbf{X}_{m+k}$. The sequential detection schemes considered above focus their attention on the *monitoring* sample. In the classical CUSUM-based detector, if $k > k_*$, the “second sample” is comprised of both pre- and post-change observations, which may adversely affect test performance under H_A , especially for late changes. On the other hand, the Page monitoring scheme, based only on the truncated sequences $\mathbf{X}_{m+r+1}, \dots, \mathbf{X}_{m+k}$, in effect “tosses away” earliest observations.

In this section, we propose a novel, alternative sequential detection scheme, which focuses primarily on the *historical* sample. The rationale is as follows: if no rejection of H_0 occurs after a suitable number of monitoring periods, in principle “recycling” a portion of the earliest-monitored observations back into the training data may serve the dual purpose of increasing information about the historical baseline period and - simultaneously - reducing any contamination in the monitored sample following a changepoint, potentially leading to improved power and/or faster detection time. Naturally, such an approach can be expected to lead to improved power on finite time horizons, with - for changes that are “small” in magnitude - some possible risk of adding post-change observations into the training sample.

Hence, in order to complement our approaches above, we also consider a moving-window

counterpart, based on a user-specified monitoring window size $w \geq 2$, viz.

$$(4.1) \quad \mathcal{D}_m^{(3)}(k, w) = m^{-1}(k \wedge w)^2 \left| \tilde{U}_m(h; w, k) \right|,$$

where $\tilde{U}_m(h, w; k) = U_m(h; k)$ for $2 \leq k \leq w$, and for $k > w$,

$$(4.2) \quad \begin{aligned} \tilde{U}_m(h, w; k) = & \frac{2}{w(m+r)} \sum_{i=1}^{m+r} \sum_{j=m+r+1}^{m+k} h(\mathbf{X}_i, \mathbf{X}_j) \\ & - \binom{m+r}{2}^{-1} \sum_{1 \leq i < j \leq m+r} h(\mathbf{X}_i, \mathbf{X}_j) - \binom{w}{2}^{-1} \sum_{m+r < i < j \leq m+k} h(\mathbf{X}_i, \mathbf{X}_j), \end{aligned}$$

where $r = k - w$. Thus, in $\tilde{U}_m(h, w; k)$, once the monitoring period k exceeds the prespecified window length w , the earliest monitored observations $\mathbf{X}_{m+1}, \dots, \mathbf{X}_{m+k-w}$ are recycled back into the training sample. The detector $\mathcal{D}_m^{(1)}(k)$ can be viewed as $\mathcal{D}_m^{(3)}(k)$ with $w = k$, and for values of $k < w$, we have $\mathcal{D}_m^{(3)}(k, w) = \mathcal{D}_m^{(1)}(k)$.

There are many possible choices for w ; for illustration and flexibility, we consider the case where the moving monitoring window length is given by

$$(4.3) \quad w = w(k, m) = \lfloor c_w m + b_w(k - c_w m) \vee 0 \rfloor,$$

for some and $0 \leq b_w \leq 1$ and $c_w \geq 0$. Here, c_w represents a pre-specified minimum window size before any repurposing begins, whereas b_w represents the proportion of the monitored data retained in the moving window after repurposing starts - e.g. when $c_w = 1, b_w = 1/2$, the monitoring window grows until it reaches the length of the historical sample, after which an observation is repurposed back into the training sample every two new monitoring periods.⁴

⁴Alternative choices made in related monitoring procedures include considering all possible two-sample segmentations at every time k , as in ?, among others (e.g., [Aue and Kirch, 2024](#)).

Let $f(u, c_w) = u/(1 - u) - c_w$

$$(4.4) \quad \mathbf{y}(u) = \begin{cases} 0 & 0 \leq u \leq c_w/(1 + c_w), \\ \frac{f(u, c_w)(1 - b_w)}{1 + f(u, c_w)(1 - b_w)} & c_w/(1 + c_w) < u \leq 1, \end{cases}$$

and define $\Gamma(u, b_w, c_w) = \sum_{\ell=1}^{\infty} \lambda_{\ell} [(W_{\ell}(u) - W_{\ell}(\mathbf{y}(u)))^2 - (u - \mathbf{y}(u))]$, with W_{ℓ} as in (3.3).

Theorem 4.1. *Assume H_0 holds, and consider the detector $\mathcal{D}_m(k) = \mathcal{D}_m^{(3)}(k, w)$, with $w = w(k, m)$ as in (4.3), and let g_m be given by (2.8). As $m \rightarrow \infty$, the statements of Theorem 3.1(i)-(ii) hold, with $\Gamma(u, b_w, c_w)$ in place of $\Gamma(u)$.*

4.2. Testing for the stability of the training sample. Assumption 2.1 requires that the training sample $\mathbf{X}_1, \dots, \mathbf{X}_m$ is stable - that it, it undergoes no breaks. As mentioned above, this is a typical, and testable, assumption. We now (briefly) discuss a U -statistic based approach to test retrospectively for the null hypothesis of no distributional changes in the training sample. We use the sequence

$$\begin{aligned} \mathfrak{R}(k) = & \frac{2}{k(m-k)} \sum_{i=1}^k \sum_{j=k+1}^m h(\mathbf{X}_i, \mathbf{X}_j) - \binom{k}{2}^{-1} \sum_{1 \leq i < j \leq k} h(\mathbf{X}_i, \mathbf{X}_j) \\ & - \binom{m-k}{2}^{-1} \sum_{k+1 \leq i < j \leq m} h(\mathbf{X}_i, \mathbf{X}_j), \end{aligned}$$

for $2 \leq k \leq m-2$, and define the corresponding process

$$\mathfrak{r}_m(t) = \begin{cases} 0, & t \notin [2/m, 1 - 2/m] \\ mt^2(1-t)^2 \mathfrak{R}(mt), & 2/m \leq t \leq 1 - 2/m. \end{cases}$$

As is typical in this literature (Horváth and Rice, 2024), we consider a *weighted* version of $\mathfrak{r}_m(t)$, in order to enhance the power of our test in the presence of changes occurring close

to the beginning/end of the sample; we propose the following family of weight functions

$$(4.5) \quad \mathbf{q}(t) = (t(1-t))^\zeta, \text{ for some } \zeta < 1.$$

A “natural” choice to detect the presence of a possible change is to use the sup-norm of the weighted version of $\mathbf{r}_m(t)$, viz. $\sup_{0 < t < 1} |\mathbf{r}_m(t)| / \mathbf{q}(t)$.

Theorem 4.2. *If Assumptions 2.1–2.3 hold, then*

$$\sup_{0 < t < 1} \frac{|\mathbf{r}_m(t)|}{\mathbf{q}(t)} \xrightarrow{\mathcal{D}} \sup_{0 \leq t \leq 1} \frac{1}{\mathbf{q}(t)} \left| \sum_{\ell=1}^{\infty} \lambda_\ell (B_\ell^2(t) - t(1-t)) \right|,$$

where $\{B_\ell(t), 0 \leq t \leq 1\}$, $\ell = 1, 2, \dots$ are independent Brownian bridges.

Theorem 4.2 contains the limit of the maximally selected weighted version of $\mathbf{r}_m(t)$. Several further results such as power versus the alternative, and a consistent estimator of the break date, could be readily derived by extending the theory in Horváth and Rice (2024). The same result - for the case $\mathbf{q}(t) = 1$ - was proven by Biau et al. (2016), under the more restrictive condition $\sum_{\ell=1}^{\infty} |\lambda_\ell| < \infty$. Hence, similarly to the other results above, Theorem 4.2 improves on the current literature by requiring the milder condition $\sum_{\ell=1}^{\infty} \lambda_\ell^2 < \infty$.

4.3. Examples of kernel functions. We discuss some examples of possible kernel functions $h(\cdot, \cdot)$, and a methodology (plus an example) to construct “distribution-determining” kernels $h(\cdot, \cdot)$ - that is, functions $h(\cdot, \cdot)$ which can discriminate *any* change in distribution.

Example 4.1. Suppose $\mathcal{X} = \mathbb{R}^d$, and let $\eta \in (0, 2)$. The kernel $h(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|^\eta$ is connected with the energy distance between two independent vectors $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^n$, defined as $\mathcal{E}_\eta(\mathbf{X}, \mathbf{Y}) = 2\mathbb{E} \|\mathbf{X} - \mathbf{Y}\|^\eta - \mathbb{E} \|\mathbf{X} - \mathbf{X}'\|^\eta - \mathbb{E} \|\mathbf{Y} - \mathbf{Y}'\|^\eta$, where $\mathbf{X}', \mathbf{Y}'$ are independent copies of $\mathbf{X}$ and $\mathbf{Y}$ respectively. Székely and Rizzo (2005) show that $\mathcal{E}_\eta(\mathbf{X}, \mathbf{Y}) \geq 0$, with equality if and only if $\mathbf{X} \stackrel{\mathcal{D}}{=} \mathbf{Y}$. As also argued in Biau et al. (2016) and Boniece et al. (2025), $U_m(h; k)$ in (2.4) is the empirical counterpart to $\mathcal{E}_\eta$, evaluating the distance between

the distribution of the training sample and that of the monitored sequence up to time k . When trying to detect changepoint in possibly multivariate time series, the energy distance is particularly advantageous due to its rotational invariance (Székely and Rizzo, 2013).⁵ In the case of using $h(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|^\eta$, it is immediate to see that Assumption 2.3 holds as long as $\mathbb{E} \|\mathbf{X}\|^{2\eta} < \infty$. In turn, this suggests that η can e.g. be chosen *a posteriori* by the applied user after checking how many moments the data admit.

Example 4.2. Chen et al. (2025) propose the so-called *Grothendieck divergence*, defined as $\mathcal{G}_\eta(\mathbf{X}, \mathbf{Y}) = 2\mathbb{E}\psi(\mathbf{X}, \mathbf{Y}) - \mathbb{E}\psi(\mathbf{X}, \mathbf{X}') - \mathbb{E}\psi(\mathbf{Y}, \mathbf{Y}')$, where

$$\psi(\mathbf{x}, \mathbf{y}) = \arccos \left[\frac{1 + \langle \mathbf{x}, \mathbf{y} \rangle}{\sqrt{(1 + \langle \mathbf{x}, \mathbf{x} \rangle)(1 + \langle \mathbf{y}, \mathbf{y} \rangle)}} \right],$$

satisfying Assumption 2.3, with no moment requirements on $\mathbf{X}$ or $\mathbf{Y}$. By Proposition 1 in Chen et al. (2025), the Grothendieck divergence is distribution determining - that is, it is nonzero if and only if the distributions of $\mathbf{X}$ and $\mathbf{Y}$ differ.

Example 4.3. Consider a separable metric space $(\mathcal{X}, \rho)$ with finite first moment. Then, $(\mathcal{X}, \rho)$ is said to have *negative type* (Lyons, 2013), if it holds that

$$(4.6) \quad \mathfrak{D}_\rho(G_1, G_2) = \int \rho(\mathbf{x}, \mathbf{y}) d(G_1 - G_2)^2(\mathbf{x}, \mathbf{y}) \leq 0.$$

The space $(\mathcal{X}, \rho)$ is said to have *strong negative type* if (4.6) is satisfied with the additional property that equality holds if and only if $G_1 = G_2$. Hence, taking $h(\mathbf{x}, \mathbf{y}) = \rho(\mathbf{x}, \mathbf{y})$ when $(\mathcal{X}, \rho)$ has strong negative type yields an *omnibus* test for changes in the distribution. Examples of spaces with strong negative type include $\mathbb{R}^d$ (the energy distance in Example 4.1 is a special case of (4.6)), or more generally all separable Hilbert spaces. Notably, from Lyons (2013), if $(\mathcal{X}, \rho)$ has negative type, then for any $0 < r < 1$, $(\mathcal{X}, \rho^r)$ has *strong* negative

⁵As mentioned in the introduction, statistics based on other distances, such as Cramér's distance or the Cramér-von Mises-Smirnov distance do not share this property.

type. In particular, from Meckes (2013), if $1 \leq p \leq 2$ and $\mathcal{X} = \mathcal{L}^p[0, 1]$ is the space of real-valued p -integrable functions and ρ its usual metric, then $(\mathcal{X}, \rho^r)$ has strong negative type for any $0 < r < 1$. In the case of using the kernel $h(\mathbf{x}, \mathbf{y}) = \rho(\mathbf{x}, \mathbf{y})$, it is immediate to see that Assumption 2.3 holds as long as $\mathbb{E}[\rho^2(\mathbf{x}, \mathbf{y})] < \infty$. Then, similarly to Example 4.1, the definition of $\rho(\mathbf{x}, \mathbf{y})$ is “constructive”, in that either it can be chosen based on how many moments the data admit (as long as (4.6) holds); or, given a metric $\rho(\mathbf{x}, \mathbf{y})$ and a dataset, it can be tested whether Assumption 2.3 holds by testing whether $\mathbb{E}[\rho^2(\mathbf{x}, \mathbf{y})] < \infty$.

Example 4.4. Arlot et al. (2019) study multiple changepoint detection (retrospectively) based on positive semidefinite kernel, providing several examples of possible kernel functions suitable to various data types (e.g. vector-valued data, multinomial data, text or graph-valued data; see their Section 3.2); their paper also contains a comprehensive set of references on the literature on kernel functions. Of particular interest is the family of *characteristic kernels* (Fukumizu et al., 2007; Sriperumbudur et al., 2010; Sriperumbudur et al., 2011), whose “mean” changes whenever there is a change in the distribution of the underlying observations $\mathbf{X}_i$ - thus being able to pick up any distributional change, or, as mentioned above, being “distribution-determining”.⁶ A possible example of a characteristic kernel (see Fukumizu et al., 2003) is the Gaussian kernel $h(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|_2^2 / (2a^2))$, where $a > 0$ is a bandwidth parameter. By Corollary 16 in Sejdinovic et al. (2013), there is a one to one correspondence between characteristic kernels and (semi)metrics of the strong negative type.

Examples 4.3 and 4.4 suggest that it is possible to choose $h(\mathbf{x}, \mathbf{y})$ so as to be “distribution-determining” - essentially, producing kernels by means of kernels. Indeed, consider the user-chosen function $K(\mathbf{x}, \mathbf{y}) : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$, such that $K(\mathbf{x}, \mathbf{y})$ is symmetric, positive semidefinite,⁷ and non-degenerate - that is, the map $\mathbf{x} \rightarrow K(\cdot, \mathbf{x})$ is injective. Given such a kernel, define the semimetric $\delta(\mathbf{x}, \mathbf{y}) = K(\mathbf{x}, \mathbf{x}) + K(\mathbf{y}, \mathbf{y}) - 2K(\mathbf{x}, \mathbf{y})$. Sejdinovic et al. (2013)

⁶The mean function is defined as the function $\mu_i(\cdot)$ such that, for all $g(\cdot)$, $\langle \mu_i(\cdot), g(\cdot) \rangle = \mathbb{E} \langle h(\mathbf{X}_i, \cdot), g(\cdot) \rangle$.

⁷That is, for each tuple $\{x_1, \dots, x_n\}$, the matrix $\{K(x_i, x_j)\}_{1 \leq i \leq j \leq n}$ is positive semidefinite.

show that $\delta(\mathbf{x}, \mathbf{y})$ is a *semimetric* of negative type on $\mathcal{X}$. In turn, by Proposition 3 in [Sejdinovic et al. \(2013\)](#), this entails that there are a Hilbert space $\mathcal{H}$ and an injective map $\varphi(\cdot)$ such that $\delta(\mathbf{x}, \mathbf{y}) = \|\varphi(\mathbf{x}) - \varphi(\mathbf{y})\|_{\mathcal{H}}^2$; therefore, $\delta^{1/2}(\mathbf{x}, \mathbf{y})$ is a metric of negative type on $\mathcal{X}$. Then, based on Remark 3.19 in [Lyons \(2013\)](#), $\delta^s(\mathbf{x}, \mathbf{y})$ is - for any $s \in (0, 1/2)$ - a metric of *strong* negative type. Thus, revisiting Example 4.3, given a nondegenerate kernel $K(\mathbf{x}, \mathbf{y})$, the family of functions $h(\mathbf{x}, \mathbf{y}) = [K(\mathbf{x}, \mathbf{x}) + K(\mathbf{y}, \mathbf{y}) - 2K(\mathbf{x}, \mathbf{y})]^{s/2}$ defines a family of “distribution-determining” kernels for any $s \in (0, 1/2)$. Indeed, in the following theorem we extend Remark 3.19 in [Lyons \(2013\)](#), showing that even $\delta^{1/2}(\mathbf{x}, \mathbf{y})$ is a distribution-determining kernel.

Theorem 4.3. *Let $\mathcal{X}$ be a separable, complete metric space, and $K(\mathbf{x}, \mathbf{y})$ be a continuous, non-degenerate kernel. Then $\delta^{1/2}(\mathbf{x}, \mathbf{y})$ is a metric of strong negative type.*

To the best of our knowledge, the result in Theorem 4.3 is novel in this literature. According to the theorem, the kernel $h(\mathbf{x}, \mathbf{y}) = [K(\mathbf{x}, \mathbf{x}) + K(\mathbf{y}, \mathbf{y}) - 2K(\mathbf{x}, \mathbf{y})]^{1/2}$, is characteristic, and therefore, considering Example 4.4, an *omnibus* test for distributional change can be based on it.

It is easily seen that when $K(\mathbf{x}, \mathbf{y})$ is strictly positive definite, it is non-degenerate. Thus, in order to construct a distribution determining kernel $h(\mathbf{x}, \mathbf{y})$, it suffices to follow the procedure above starting from a positive definite kernel.⁸ A leading example is based on the Gaussian kernel, discussed in the next example.

Example 4.5. Consider the Gaussian kernel $K_g(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|^2 / (2a^2))$ for some $a > 0$; this is a non-degenerate kernel (see e.g. [Arlot et al. \(2019\)](#)). Then, by the above, it is easy to see that $\delta^{1/2}(\mathbf{x}, \mathbf{y}) = [K_g(\mathbf{x}, \mathbf{x}) + K_g(\mathbf{y}, \mathbf{y}) - 2K_g(\mathbf{x}, \mathbf{y})]^{1/2}$, is a metric of negative type; further, by Theorem 4.3, it is also a metric of *strong* negative type.

⁸Other sufficient conditions can be found in [Sriperumbudur et al. \(2010\)](#) and [Sriperumbudur et al. \(2011\)](#).

4.4. On implementation. The limiting processes of our monitoring schemes under H_0 all depend on the (infinite sequence of) eigenvalues λ_i of the operator A defined above, which necessitates some approximation when obtaining critical values. A possible approach is based on estimating the eigenvalues λ_i from the historical sample via the $m \times m$ matrix A_m , where

$$(4.7) \quad \{A_m\}_{i,j} = \frac{1}{m} \left(h(\mathbf{X}_i, \mathbf{X}_j) - h_{1,i} - h_{1,j} - \binom{m}{2}^{-1} \sum_{1 \leq i' < j' < m} h(\mathbf{X}_{i'}, \mathbf{X}_{j'}) \right),$$

with $h_{1,i} = \sum_{\ell=1}^m h(\mathbf{X}_i, \mathbf{X}_\ell) \mathbf{1}_{\{\ell \neq i\}} / (m-1)$. Let $|\hat{\lambda}_{1,m}| \geq |\hat{\lambda}_{2,m}| \geq \dots \geq |\hat{\lambda}_{m,m}|$ denote the eigenvalues of the matrix A_m , define the sigma-field $\mathcal{F} = \sigma\{\mathbf{X}_\ell, \ell \geq 1\}$, and let $\{W_1(u), u \geq 0\}$, $\{W_2(u), u \geq 0\}, \dots$ be independent Wiener processes, independent of $\mathcal{F}$. The approximations to the limiting processes $\Gamma(u)$, $\bar{\Gamma}(u)$ and $\Gamma(u, b_w, c_w)$ under H_0 are constructed as follows

$$(4.8) \quad \hat{\Gamma}_m(u) = \sum_{\ell=1}^m \hat{\lambda}_{\ell,m} (W_\ell^2(u) - u),$$

$$(4.9) \quad \hat{\bar{\Gamma}}_m(u) = \sup_{0 < v \leq u} |\hat{G}_m(u, v)|,$$

$$(4.10) \quad \hat{\Gamma}_m(u, b_w, c_w) = \sum_{\ell=1}^m \hat{\lambda}_{\ell,m} [(W_\ell(u) - W_\ell(\mathbf{y}(u)))^2 - (u - \mathbf{y}(u))],$$

where $\mathbf{y}(u)$ is defined in (4.4) and

$$\hat{G}_m(u, v) = \sum_{\ell=1}^m \hat{\lambda}_{\ell,m} \left[\left(W_\ell(u) - \frac{1-u}{1-v} W_\ell(v) \right)^2 - \left(u - v \left(\frac{1-u}{1-v} \right) \right) \left(1 - v \left(\frac{1-u}{1-v} \right) \right) \right].$$

This method is proposed in [Biau et al. \(2016\)](#); hereafter, we formalise it, showing that the approximations (4.8)-(4.10) converge (a.s. conditionally on the data) to the limiting processes. Let “ $\Rightarrow_{\mathcal{F}}$ ” denote the almost sure weak convergence under $P(\cdot|\mathcal{F})$.

Theorem 4.4. *As $m \rightarrow \infty$, it holds that, for all $0 < u_0 \leq 1$ and $0 \leq \beta < 1$,*

$$\begin{aligned}
(4.11) \quad & \sup_{0 < u < u_0} u^{-\beta} |\widehat{\Gamma}_m(u)| \Rightarrow_{\mathcal{F}} \sup_{0 < u < u_0} u^{-\beta} |\Gamma(u)|, \\
& \sup_{0 < u < u_0} u^{-\beta} |\widehat{\bar{\Gamma}}_m(u)| \Rightarrow_{\mathcal{F}} \sup_{0 < u < u_0} u^{-\beta} |\bar{\Gamma}(u)|, \\
& \sup_{0 < u < u_0} |\widehat{\Gamma}_m(u, b_w, c_w)| \Rightarrow_{\mathcal{F}} \sup_{0 < u < u_0} |\Gamma(u, b_w, c_w)|.
\end{aligned}$$

The theorem requires that the number of eigenvalues employed grows with m ; in (4.8)-(4.10) all the eigenvalues of A_m are used, but employing only a fraction (e.g., $m/2$) still yields the same result.

5. SIMULATIONS AND APPLICATIONS

5.1. Simulation study. We report a set of Monte Carlo simulations to investigate the empirical rejection frequencies and the detection delays under alternatives of our procedures. We report only a set of simulations based on the case $\mathcal{X} = \mathbb{R}^5$.⁹ We use the following kernels: $h^{(1)}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_1^{1/2}$; $h^{(2)}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_2$; and $h^{(3)}(\mathbf{x}, \mathbf{y}) = [1 - \exp(-\|\mathbf{x} - \mathbf{y}\|_2^2 / (2a^2))]^{1/2}$. The kernel $h^{(2)}$ corresponds to the usual energy distance; $h^{(3)}$ is based directly on Example 4.5, with a set equal to the sample median of $\{\|\mathbf{X}_i - \mathbf{X}_j\|_2, 1 \leq i, j \leq m\}$. In all scenarios, we consider historical samples of length $m \in \{50, 100, 200\}$, and we report results for each of the detectors $\mathcal{D}_m^{(i)}$, $i = 1, 2, 3$, based on the boundary function (2.8) with $\beta \in \{0, 0.5, 0.9\}$.¹⁰ We begin by examining the performance of our procedures under H_0 ; in all cases, we generate the observations as $\mathbf{X}_i \sim i.i.d. \mathcal{N}(0, \mathbf{I}_5)$, and we set the monitoring horizon $M = 10m$. Empirical rejection frequencies are reported in Table 5.1.¹¹

Broadly speaking, size control is ensured in all cases as m increases. This can be read in conjunction with the online monitoring literature, where often detection schemes are

⁹Further simulations, which essentially confirm the results in this section, are available upon request.

¹⁰For reference, recall that: $\mathcal{D}_m^{(1)}$ is the “ordinary” detection scheme defined in (2.5); $\mathcal{D}_m^{(2)}$ is the “Page-type” scheme defined in (2.6); and $\mathcal{D}_m^{(3)}$ is the novel “repurposing” scheme introduced in (4.1).

¹¹Note that, for each empirical rejection frequency, the 95% confidence interval is $[0.04, 0.06]$.

TABLE 5.1. Empirical rejection probabilities under H_0 , nominal level 0.05

Scheme	Kernel β	$h^{(1)}$			$h^{(1)}$			$h^{(1)}$		
		$m = 50$	$m = 100$	$m = 200$	$m = 50$	$m = 100$	$m = 200$	$m = 50$	$m = 100$	$m = 200$
$\mathcal{D}_m^{(1)}$	0	0.056	0.050	0.043	0.065	0.057	0.051	0.077	0.066	0.056
	0.5	0.065	0.047	0.050	0.057	0.043	0.058	0.065	0.067	0.048
	0.9	0.057	0.056	0.047	0.059	0.049	0.051	0.054	0.066	0.053
$\mathcal{D}_m^{(2)}$	0	0.055	0.050	0.044	0.068	0.058	0.050	0.078	0.067	0.057
	0.5	0.059	0.046	0.050	0.057	0.047	0.055	0.062	0.063	0.047
	0.9	0.050	0.051	0.045	0.052	0.045	0.046	0.049	0.060	0.055
$\mathcal{D}_m^{(3)}$	0	0.047	0.056	0.046	0.060	0.048	0.050	0.074	0.063	0.047
	0.5	0.048	0.057	0.056	0.048	0.050	0.053	0.051	0.062	0.048
	0.9	0.050	0.055	0.047	0.056	0.046	0.050	0.057	0.057	0.052

found to be conservative (we refer e.g. to the simulations in [Horváth et al., 2007](#), and the comments therein). When using kernels $h^{(1)}$ and $h^{(2)}$, no oversize is observed whenever $m > 50$, and our procedures have a (mild) tendency to over-reject only in very few cases when $m = 50$. Conversely, kernel $h^{(3)}$ seems to over-reject very often, unless $m = 200$; note, however, that partnering $h^{(3)}$ with $\beta = 0.9$ results in no oversize even for m as little as 50. Hence, the results in the table offer several guidelines to the applied user as far as the choice of the kernel and of the weight β are concerned.

We now turn to examining the power of our procedure. We consider three main alternative hypotheses, where - in all cases - $\mathbf{X}_i \stackrel{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ for $1 \leq i \leq k_*$ and subsequently changes into:

$$(5.1) \quad H_{A,1} : \mathbf{X}_{k_*+1} \stackrel{iid}{\sim} \mathcal{N}(\boldsymbol{\mu}, \mathbf{I}_d),$$

$$(5.2) \quad H_{A,2} : \mathbf{X}_{k_*+1} \stackrel{iid}{\sim} \mathcal{N}(\mathbf{0}, \Sigma),$$

$$(5.3) \quad H_{A,3} : \mathbf{X}_{k_*+1} = (X_{k_*+1,1}, \dots, X_{k_*+1,d})^\top \text{ with } X_{k_*+1,i} \stackrel{iid}{\sim} t_\nu / \sqrt{\text{var}(t_\nu)}.$$

Equation (5.1) corresponds to a location change; (5.2) to a scale change with no change in location; and, finally, (5.3) is a tail alternative, where the distribution of the data changes into a Student's t with ν degrees of freedom. In all three cases, we consider both the case

TABLE 5.2. Empirical power and delay times, strong changes ($\beta = 0$, randomised k_*)

Scheme	Alternative Kernel	$H_{A,1}$			$H_{A,2}$			$H_{A,3}$		
		$h^{(1)}$	$h^{(2)}$	$h^{(3)}$	$h^{(1)}$	$h^{(2)}$	$h^{(3)}$	$h^{(1)}$	$h^{(2)}$	$h^{(3)}$
$\mathcal{D}_m^{(1)}$	Power	0.841	0.865	0.849	0.805	0.948	0.967	0.967	0.767	0.780
	Med. delay	127	118	122	129	115	110	95	155	148
$\mathcal{D}_m^{(2)}$	Power	0.881	0.919	0.902	0.836	0.999	1.000	0.999	0.776	0.789
	Med. delay	114	107	110	125	109	103	87	148	144
$\mathcal{D}_m^{(3)}$	Power	0.985	0.996	0.990	1.000	1.000	1.000	1.000	0.921	0.952
	Med. delay	91	87	89	94	85	83	72	105	104

of “strong” changes and “weak” ones, depending on the size of the change - “strong” changes correspond to $\boldsymbol{\mu} = (0.3, \dots, 0.3)^\top$ in (5.1), $\{\Sigma\}_{i,j} = \exp(|i - j|/10)$ in (5.2), and $\nu = 2.5$ in (5.3); “weak” changes correspond to $\boldsymbol{\mu} = (0.25, \dots, 0.25)^\top$ in (5.1), $\{\Sigma\}_{i,j} = \exp(|i - j|/5)$ in (5.2), and $\nu = 3$ in (5.3). All the powers reported hereafter are size-adjusted - that is, each procedure has been calibrated so as to ensure that the empirical rejection frequencies under the null match the nominal level (set to 0.05).

In a first set of experiments reported in Tables 5.2 and 5.3, we consider the empirical rejection frequencies and the delays for a randomised choice of k_* ,¹² in the presence of a strong change; we report results only for $\beta = 0$ in (2.8), but results with different values of β did not change the overall findings and are available upon request. This case is empirically relevant, seeing as no prior knowledge as to the location of k_* is available. As the table shows, the power is satisfactory in all cases; detection based on the scheme proposed in Section 4.1, $\mathcal{D}_m^{(3)}$, seems to offer shorter delays, improving on both $\mathcal{D}_m^{(1)}$ and $\mathcal{D}_m^{(2)}$. Interestingly, this seems to be the case for both strong and weak changes, across all alternative hypotheses $H_{A,1} - H_{A,3}$, and for each choice of kernel $h(\cdot, \cdot)$.

¹²The value of k_* , at each iteration, has been picked from $\{10, m, 5m\}$ with equal probability.

TABLE 5.3. Empirical power and delay times, weak changes ($\beta = 0$, randomised k_*)

Scheme	Alternative Kernel	$H_{A,1}$			$H_{A,2}$			$H_{A,3}$		
		$h^{(1)}$	$h^{(2)}$	$h^{(3)}$	$h^{(1)}$	$h^{(2)}$	$h^{(3)}$	$h^{(1)}$	$h^{(2)}$	$h^{(3)}$
$\mathcal{D}_m^{(1)}$	Power	0.762	0.787	0.786	0.577	0.722	0.735	0.772	0.376	0.414
	Med. delay	153	145	148	203	179	174	154	182	191
$\mathcal{D}_m^{(2)}$	Power	0.787	0.812	0.813	0.592	0.723	0.738	0.783	0.382	0.426
	Med. delay	138	131	132	189	174	166	145	176	180
$\mathcal{D}_m^{(3)}$	Power	0.848	0.911	0.899	0.251	0.608	0.746	0.899	0.141	0.156
	Med. delay	103	99	100	102	112	111	105	91	95

In order to assess more precisely the impact of the changepoint location, we now report results for the three cases of break location used above, viz.: a “very early” break corresponding to $k_* = 10$; a medium change with $k_* = m$; and a “late” break with $k_* = 5m$. We report the detection delays, under a randomised alternative,¹³ for the case of a strong change (Table 5.4) and of a weak change (Table 5.5); in Section A in the Supplement, we report the power (see Table A.1 for strong changes, and Table A.2 for weak changes). Considering the former set of results first, the performance of all detectors $\mathcal{D}_m^{(i)}$ is comparable in the presence of an early change. Results are broadly the same under a medium changepoint location, $k_* = m$, although - when using $\mathcal{D}_m^{(3)}$ - the power deteriorates as β increases, which suggests that the “repurposing” detection scheme is better employed for low values of β . As can be expected, all results worsen when the change occurs late; this is more pronounced in the case of the detector $\mathcal{D}_m^{(1)}$, which is “dragged down” by previous observations, and naturally improves when past observations are either discarded or “recycled”; note that the novel detector $\mathcal{D}_m^{(3)}$ offers a slightly higher power, and slightly better detection delays, compared to

¹³At each iteration, the alternative has been picked from the set $\{H_{A,1}, H_{A,2}, H_{A,3}\}$ with equal probability.

TABLE 5.4. Median delay times, strong changes (randomised alternative $H_{A,i}$)

Scheme	β	$k_* = 10$			$k_* = m$			$k_* = 5m$		
		$h^{(1)}$	$h^{(2)}$	$h^{(3)}$	$h^{(1)}$	$h^{(2)}$	$h^{(3)}$	$h^{(1)}$	$h^{(2)}$	$h^{(3)}$
$\mathcal{D}_m^{(1)}$	0	73	79	78	135	149	142	321	340	340
	0.5	56	63	59	131	149	142	319	353	354
	0.9	48	57	54	154	167	165	350	375	377
$\mathcal{D}_m^{(2)}$	0	70	77	76	117	133	127	314	356	343
	0.5	52	60	57	113	130	126	316	373	366
	0.9	42	53	51	135	151	148	365	392	401
$\mathcal{D}_m^{(3)}$	0	52	57	56	93	100	98	306	301	301
	0.5	45	50	48	102	110	109	370	350	341
	0.9	43	50	48	125	133	128	382	463	441

TABLE 5.5. Median delay times, weak changes (randomised alternative $H_{A,i}$)

Scheme	β	$k_* = 10$			$k_* = m$			$k_* = 5m$		
		$h^{(1)}$	$h^{(2)}$	$h^{(3)}$	$h^{(1)}$	$h^{(2)}$	$h^{(3)}$	$h^{(1)}$	$h^{(2)}$	$h^{(3)}$
$\mathcal{D}_m^{(1)}$	0	139	151	148	243	264	258	328	312	316
	0.5	112	128	126	235	272	253	332	322	318
	0.9	102	124	116	270	304	309	349	332	330
$\mathcal{D}_m^{(2)}$	0	133	146	142	214	238	231	337	323	327
	0.5	103	121	120	210	247	233	339	337	331
	0.9	89	113	107	243	279	289	354	340	335
$\mathcal{D}_m^{(3)}$	0	77	79	81	130	138	141	409	399	420
	0.5	69	74	74	140	140	145	453	431	432
	0.9	65	67	66	159	143	152	447	477	459

the Page-CUSUM scheme $\mathcal{D}_m^{(2)}$. Similar results are found in the case of a weak change (Tables 5.5 and A.2), although in such a case the detection scheme $\mathcal{D}_m^{(3)}$ worsens dramatically in the presence of a late change, especially (as also noted above) when β is large.

To summarize the findings above, the monitoring schemes $\mathcal{D}_m^{(1)}$ and $\mathcal{D}_m^{(2)}$ are essentially able to detect almost any change, irrespective of the size; however, this occurs with a possibly large delay, potentially many times longer than k_* itself. This effect occurs essentially

because $\mathcal{D}_m^{(1)}$ and $\mathcal{D}_m^{(2)}$, by construction, do not contaminate the historical sample with post-change data, so even in the presence of a small change they will eventually accumulate enough post-change data to reject H_0 . Conversely, $\mathcal{D}_m^{(3)}$ may “recycle” even post-change data, thus contaminating the historical sample, which explains the low power in the presence of small changes. On the other hand, in the presence of large breaks, $\mathcal{D}_m^{(3)}$ often offers a massive reduction in delay time.

Finally, in Section A of the Supplement we report further Monte Carlo evidence on the distribution of the detection delay (Figures A.1 and A.2), and on the comparison with the standard CUSUM detector (Tables A.3 and A.4).

5.2. Empirical illustration. We apply our methodology to the detection of changes in the heart rate (ECG) recording of an infant. We use the same dataset as in Nason et al. (2000): a series of 2,048 observations recorded in beats per minute, sampled overnight every 16 seconds from 21:17:59 to 06:27:18, from a 66 day old infant.¹⁴ We investigate the possible presence of changepoints in the logs of the original series; apart from this transformation, no further preprocessing is applied to the dataset. We use a training sample spanning between observations 975 and 1,169 (corresponding to 01:37:59 till 02:29:43, with $m = 195$), and for purposes of illustration we consider two experiments with different monitoring horizons: one until 02:51:19 (with $M = 81$) and one until 03:55:59 (with $M = 331$).¹⁵ The training sample has been selected to ensure that the non-contamination assumption is satisfied. In Tables B.1-B.3 in the Supplement, we report some descriptive statistics for the monitoring sample, and some preliminary analysis, showing that: the data undergo no distributional change during the training sample (Table B.1); the data are independent (Table B.2); and a large number of moments exist (Table B.3). Results are reported in Table 5.6. We have used -

¹⁴The data are available as part of the R package `wavethresh`, and they were originally recorded by Prof. Peter Fleming, Dr Andrew Sawczenko and Jeanine Young of the Institute of Child Health, Royal Hospital for Sick Children, Bristol.

¹⁵See Figure B.1 in the Supplement for a graphical representation of the training sample.

by way of robustness check - two kernels (a distance based one, and a Gaussian one), and various values of the weight β . Broadly speaking, both the CUSUM and the Page-CUSUM detectors $\mathcal{D}_m^{(1)}$ and $\mathcal{D}_m^{(2)}$ detect a changepoint within minutes of each other when β is small, whereas the Page-CUSUM type detector $\mathcal{D}_m^{(2)}$ exhibits a much faster detection when $\beta = 0.9$ - results are virtually the same in the case of a “short” and of a “long(er)” monitoring horizon, and for both kernels. On the other hand, the repurposing scheme underpinning $\mathcal{D}_m^{(3)}$ seems to be slower at picking up the presence of the changepoint, possibly due to the small size of the break; in fact, when expanding the monitoring horizon, $\mathcal{D}_m^{(3)}$ detects a *break*, missing the previous one. In Figures B.2 and B.3 in the Supplement, we report a graphical representation of the data, with the interquartile range of the break dates found by our procedures (for the cases of a short and long monitoring horizon) - based on this, the break appears to be found between 02:39:35 and 02:51:19 (corresponding to the end of the monitoring horizon, which indicates no detection at all) when using a short monitoring horizon, and between 02:41:11 and 02:59:19 when using a long monitoring horizon.¹⁶

6. DISCUSSION AND CONCLUSIONS

In this paper, we propose several online monitoring schemes to detect changes in the distribution of a sequence of observations. The building block of our analysis is the use of degenerate U -statistic-type processes: we study (the weighted versions of) both an ordinary, CUSUM-type scheme, and a Page-CUSUM-type scheme, considering both an open-ended and a closed-ended scheme (in the latter case, studying both a “long” and a “short” monitoring horizon). We also propose a novel monitoring scheme, based on expanding the training sample as the monitoring goes on, when no changes are found. As a by-product, we also propose a test for the offline, retrospective detection of changepoints, which is useful

¹⁶In the original article by [Nason et al. \(2000\)](#), a changepoint at a similar point in time is also found; this is interpreted to coincide with a transition from sleep to awake state, as marked by a trained expert who was watching the infant.

TABLE 5.6. Changepoint detection in the ECG pattern (logs of beats per minute)

$M = 81$											
Kernel	β	detector	$\mathcal{D}_m^{(1)}$	$\mathcal{D}_m^{(2)}$	$\mathcal{D}_m^{(3)}$	Kernel	β	detector	$\mathcal{D}_m^{(1)}$	$\mathcal{D}_m^{(2)}$	$\mathcal{D}_m^{(3)}$
$\ \mathbf{x} - \mathbf{y}\ _2$	0.0		1, 218	none	none	Gaussian	0.0		1, 208	1, 230	1, 249
	0.5		1, 216	1, 212	none		0.5		1, 202	1, 200	none
	0.9		1, 231	1, 182	none		0.9		1, 206	1, 175	none
$M = 331$											
Kernel	β	detector	$\mathcal{D}_m^{(1)}$	$\mathcal{D}_m^{(2)}$	$\mathcal{D}_m^{(3)}$	Kernel	β	detector	$\mathcal{D}_m^{(1)}$	$\mathcal{D}_m^{(2)}$	$\mathcal{D}_m^{(3)}$
$\ \mathbf{x} - \mathbf{y}\ _2$	0.0		1, 250	1, 251	1, 280	Gaussian	0.0		1, 228	1, 230	1, 249
	0.5		1, 238	1, 212	1, 283		0.5		1, 213	1, 200	1, 286
	0.9		1, 231	1, 182	1, 287		0.9		1, 206	1, 175	1, 292

The table contains the stopping times, with “none” indicating that no changepoint was detected at a nominal level of 0.05. The kernel denoted as Gaussian is defined as $h(\mathbf{x}, \mathbf{y}) = [1 - \exp(-\|\mathbf{x} - \mathbf{y}\|_2^2 / (2a^2))]^{1/2}$. The detectors indicated as $\mathcal{D}_m^{(1)}$, $\mathcal{D}_m^{(2)}$ and $\mathcal{D}_m^{(3)}$ refer to (respectively): the CUSUM-based detector defined in equation (2.5); the Page-type detector defined in (2.6); and the repurposing scheme introduced in (4.1).

when testing for the maintained assumption that - during the training sample - no changes have occurred. We study the asymptotics of our schemes both under the null of no change (thus allowing to compute asymptotic critical values, for which we also study a Monte-Carlo based approximation method), and under the alternative (deriving the limiting distribution of the detection delay in both cases of an early and a late occurring change, which in turn is useful to quantify the expected delay associated with each procedure). Importantly, all our asymptotics is derived under the assumption that the kernel functions associated with the U -statistics are square summable, as opposed to the assumption of absolute summability which is customarily made in the literature. Whilst this is a (major) technical advance, it is also of practical importance, since - given a choice of the kernel function - it can be readily verified using one of the tests available in the literature. Monte Carlo evidence shows that our procedures work well in all cases considered, and seem to be particularly suited to the analysis of multivariate data of fixed dimension, which are often neglected by the changepoint literature.

Several interesting extensions can be considered, extending the theory developed herein. As a leading example, the extension of our techniques and of our theory to the analysis of “modern” datasets (e.g., functional-valued time series, network data or non-Euclidean data) seems to be feasible as an extension of the present work - the main change required in these cases is the choice of an appropriate kernel. This extension, and others, are under current investigation by the authors.

REFERENCES

- Arlot, S., A. Celisse, and Z. Harchaoui (2019). A kernel multiple change-point algorithm via model selection. *Journal of Machine Learning Research* 20(162), 1–56.
- Aue, A. and L. Horváth (2004). Delay time in sequential detection of change. *Statistics & Probability Letters* 67(3), 221–231.
- Aue, A. and C. Kirch (2024). The state of cumulative sum sequential changepoint testing 70 years after Page. *Biometrika* 111(2), 367–391.
- Biau, G., K. Bleakley, and D. M. Mason (2016). Long signal change-point detection. *Electronic Journal of Statistics* 10(2), 2097–2123.
- Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley.
- Bogachev, V. I. and M. A. S. Ruas (2007). *Measure theory*, Volume 1. Springer.
- Boniece, B. C., L. Horváth, and L. Trapani (2025). On changepoint detection in functional data using empirical energy distance. *Journal of Econometrics*, *forthcoming*.
- Broock, W. A., J. A. Scheinkman, W. D. Dechert, and B. LeBaron (1996). A test for independence based on the correlation dimension. *Econometric reviews* 15(3), 197–235.
- Chakraborty, S. and X. Zhang (2021). High-dimensional change-point detection using generalized homogeneity metrics. *arXiv preprint arXiv:2105.08976*.
- Chen, Y., H. Lin, X. Wang, and C. Wen (2025). A minimax two-sample test for functional data via Grothendieck’s divergence. *Journal of the American Statistical Association* (just-accepted), 1–23.
- Christmann, A. and I. Steinwart (2008). *Support vector machines*. Springer.

- Chu, C.-S. J., M. Stinchcombe, and H. White (1996). Monitoring structural change. *Econometrica* 64(5), 1045–1065.
- Csörgő, M. and L. Horváth (1989). Invariance principles for changepoint problems. In *Multivariate Statistics and Probability*, pp. 151–168. Elsevier.
- Degiannakis, S., G. Filis, G. Siourounis, and L. Trapani (2023). Superkurtosis. *Journal of Money, Credit and Banking* 55(8), 2061–2091.
- Dehling, H., K. Vuk, and M. Wendler (2022). Change-point detection based on weighted two-sample U-statistics. *Electronic Journal of Statistics* 16(1), 862–891.
- Doob, J. L. (1990). *Stochastic Processes*. Wiley.
- Drikvandi, R. and R. Modarres (2025). A distribution-free method for change point detection in non-sparse high dimensional data. *Journal of Computational and Graphical Statistics* 34(1), 290–305.
- Fiecas, M. B., K. R. Cullen, and R. Killick (2024). Generalised mixed effects models for changepoint analysis of biomedical time series data. *arXiv preprint arXiv:2410.00183*.
- Fremdt, S. (2015). Page’s sequential procedure for change-point detection in time series regression. *Statistics* 49(1), 128–155.
- Fu, Z., Y. Hong, and X. Wang (2023). On multiple structural breaks in distribution: An empirical characteristic function approach. *Econometric Theory* 39(3), 534–581.
- Fukumizu, K., F. Bach, and M. Jordan (2003). Kernel dimensionality reduction for supervised learning. In *Advances in Neural Information Processing Systems*, Volume 16.
- Fukumizu, K., A. Gretton, X. Sun, and B. Schölkopf (2007). Kernel measures of conditional dependence. In *Advances in Neural Information Processing Systems*, Volume 20.
- Ghezzi, F., E. Rossi, and L. Trapani (2024). Fast online changepoint detection using heavily-weighted CUSUM and veto-based decision rules. *arXiv preprint arXiv:2402.04433*.
- Holmes, M., I. Kojadinovic, and A. Verhoijzen (2024). Multi-purpose open-end monitoring procedures for multivariate observations based on the empirical distribution function. *Journal of Time Series Analysis* 45(1), 27–56.
- Horváth, L., M. Hušková, P. Kokoszka, and J. Steinebach (2004). Monitoring changes in linear models. *Journal of Statistical Planning and Inference* 126(1), 225–251.
- Horváth, L., P. Kokoszka, and J. Steinebach (2007). On sequential detection of parameter changes in linear regression. *Statistics and Probability Letters* 80, 1806–1813.

- Horváth, L., P. Kokoszka, and S. Wang (2021). Monitoring for a change point in a sequence of distributions. *The Annals of Statistics* 49(4), 2271–2291.
- Horváth, L. and G. Rice (2024). *Change Point Analysis for Time Series*. Springer.
- Horváth, L. and L. Trapani (2025). Real-time monitoring with RCA models. *Econometric Theory* (forthcoming).
- Horváth, L., L. Trapani, and S. Wang (2025). Sequential monitoring for changes in GARCH(1, 1) models without assuming stationarity. *Journal of Time Series Analysis*.
- Hušková, M. and S. G. Meintanis (2006). Change point analysis based on empirical characteristic functions: empirical characteristic functions. *Metrika* 63, 145–168.
- Inoue, A. (2001). Testing for distributional change in time series. *Econometric Theory* 17(1), 156–187.
- Kallenberg, O. (2002). *Foundations of Modern Probability* (2nd ed.). Springer.
- Kirch, C. and C. Stoehr (2022a). Asymptotic delay times of sequential tests based on U-statistics for early and late change points. *Journal of Statistical Planning and Inference* 221, 114–135.
- Kirch, C. and C. Stoehr (2022b). Sequential change point tests based on U-statistics. *Scandinavian Journal of Statistics* 49(3), 1184–1214.
- Koltchinskii, V. and E. Giné (2000). Random matrix approximation of spectra of integral operators. *Bernoulli* 6(1), 113–167.
- Lyons, R. (2013). Distance covariance in metric spaces. *The Annals of Probability* 41(5), 3284–3305.
- Malekloo, A., E. Ozer, M. AlHamaydeh, and M. Girolami (2022). Machine learning and structural health monitoring overview with emerging technology and high-dimensional data source highlights. *Structural Health Monitoring* 21(4), 1906–1955.
- Matteson, D. S. and N. A. James (2014). A nonparametric approach for multiple change point analysis of multivariate data. *Journal of the American Statistical Association* 109(505), 334–345.
- Meckes, M. W. (2013). Positive definite metric spaces. *Positivity* 17(3), 733–757.
- Nason, G. P., R. Von Sachs, and G. Kroisandt (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *Journal of the Royal Statistical Society: Series B* 62(2), 271–292.
- Page, E. S. (1954). Continuous inspection schemes. *Biometrika* 41(1/2), 100–115.
- Riesz, F. and B. Sz.-Nagy (1990). *Functional Analysis*. Dover.
- Sejdinovic, D., B. Sriperumbudur, A. Gretton, and K. Fukumizu (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *The Annals of Statistics* 41(5), 2263–2291.

- Serfling, R. J. (2009). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons.
- Shorack, G. R. and J. A. Wellner (1986). *Empirical Processes with Applications to Statistics*. Wiley.
- Smith, S. C. and A. Timmermann (2021). Break risk. *Review of Financial Studies* 34(4), 2045–2100.
- Sriperumbudur, B. K., K. Fukumizu, and G. R. Lanckriet (2011). Universality, characteristic kernels and RKHS embedding of measures. *Journal of Machine Learning Research* 12(7), 2389–2410.
- Sriperumbudur, B. K., A. Gretton, K. Fukumizu, B. Schölkopf, and G. R. Lanckriet (2010). Hilbert space embeddings and metrics on probability measures. *Journal of Machine Learning Research* 11, 1517–1561.
- Sun, L., Z. Shang, Y. Xia, S. Bhowmick, and S. Nagarajaiah (2020). Review of bridge structural health monitoring aided by big data and artificial intelligence: From condition assessment to damage detection. *Journal of Structural Engineering* 146(5), 04020073.
- Székely, G. J. and M. L. Rizzo (2005). Hierarchical clustering via joint between-within distances: extending Ward’s minimum variance method. *Journal of Classification* 22(2), 151–183.
- Székely, G. J. and M. L. Rizzo (2013). Energy statistics: A class of statistics based on distances. *Journal of Statistical Planning and Inference* 143(8), 1249–1272.
- Trapani, L. (2016). Testing for (in) finite moments. *Journal of Econometrics* 191(1), 57–68.
- Van der Vaart, A. W. (2000). *Asymptotic statistics*, Volume 3. Cambridge university press.

A. FURTHER MONTE CARLO EVIDENCE

Tables [A.1](#) and [A.2](#) - complementing Tables [5.4](#) and [5.5](#) in the main paper, are reported hereafter.

TABLE A.1. Empirical power - strong changes (randomised alternative $H_{A,i}$)

Scheme	β	$k_* = 10$			$k_* = m$			$k_* = 5m$		
		$h^{(1)}$	$h^{(2)}$	$h^{(3)}$	$h^{(1)}$	$h^{(2)}$	$h^{(3)}$	$h^{(1)}$	$h^{(2)}$	$h^{(3)}$
$\mathcal{D}_m^{(1)}$	0	1.000	1.000	1.000	0.995	0.997	0.998	0.608	0.526	0.608
	0.5	1.000	0.999	0.999	0.996	0.991	0.997	0.544	0.491	0.538
	0.9	0.999	0.999	0.999	0.993	0.985	0.992	0.410	0.350	0.385
$\mathcal{D}_m^{(2)}$	0	1.000	1.000	1.000	0.997	0.999	1.000	0.703	0.647	0.699
	0.5	1.000	1.000	0.999	0.998	0.997	0.999	0.618	0.609	0.619
	0.9	0.999	0.999	0.999	0.997	0.992	0.996	0.475	0.378	0.462
$\mathcal{D}_m^{(3)}$	0	0.991	0.973	0.975	0.994	0.988	0.994	0.997	0.963	0.985
	0.5	0.981	0.952	0.965	0.986	0.929	0.947	0.901	0.706	0.739
	0.9	0.950	0.913	0.932	0.807	0.754	0.787	0.404	0.468	0.517

TABLE A.2. Empirical power - weak changes (randomised alternative $H_{A,i}$)

Scheme	β	$k_* = 10$			$k_* = m$			$k_* = 5m$		
		$h^{(1)}$	$h^{(2)}$	$h^{(3)}$	$h^{(1)}$	$h^{(2)}$	$h^{(3)}$	$h^{(1)}$	$h^{(2)}$	$h^{(3)}$
$\mathcal{D}_m^{(1)}$	0	0.960	0.874	0.873	0.874	0.825	0.818	0.238	0.223	0.211
	0.5	0.945	0.854	0.864	0.844	0.785	0.811	0.207	0.201	0.202
	0.9	0.849	0.768	0.762	0.681	0.645	0.697	0.144	0.115	0.116
$\mathcal{D}_m^{(2)}$	0	0.958	0.872	0.876	0.891	0.838	0.827	0.256	0.256	0.231
	0.5	0.945	0.854	0.858	0.868	0.802	0.818	0.223	0.222	0.211
	0.9	0.854	0.762	0.764	0.702	0.666	0.718	0.147	0.114	0.124
$\mathcal{D}_m^{(3)}$	0	0.707	0.669	0.664	0.659	0.666	0.656	0.508	0.435	0.477
	0.5	0.687	0.607	0.614	0.570	0.416	0.485	0.184	0.199	0.168
	0.9	0.585	0.429	0.492	0.258	0.214	0.234	0.062	0.036	0.043

We report the plot of the density for the delay times in the cases of “strong” and “weak” changes in Figures A.1 and A.2 respectively - in all cases, we consider the unweighted versions of our detection schemes (i.e., $\beta = 0$), and use the kernel $h^{(1)}$ defined in the main paper.¹⁷

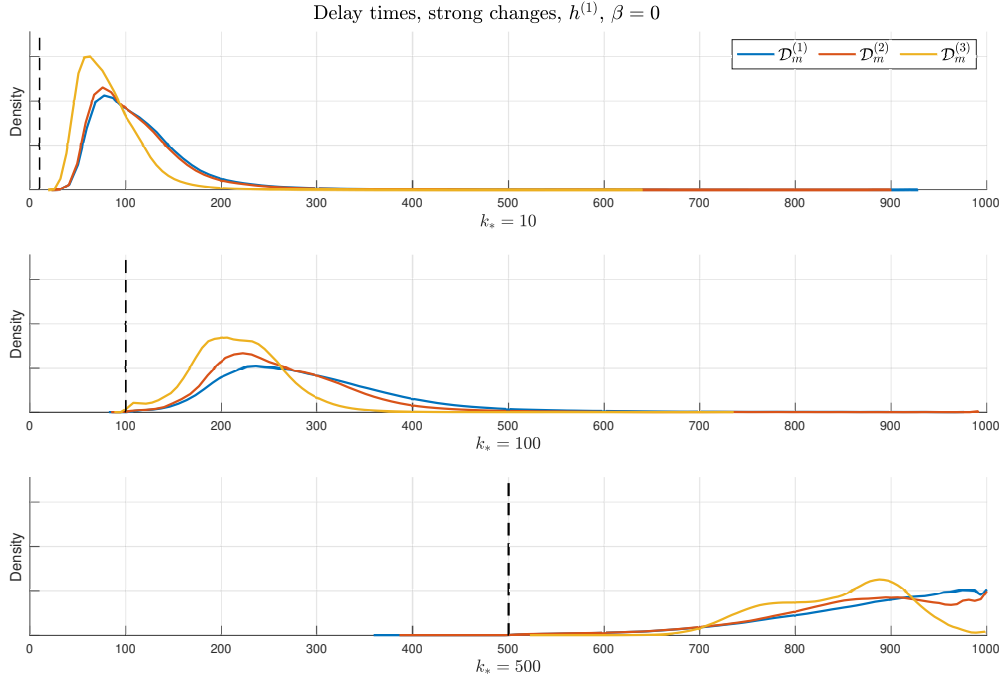


Figure A.1. Detection delays - density under *strong* breaks

¹⁷Results for different kernels and values of β are very similar, and are available under request.

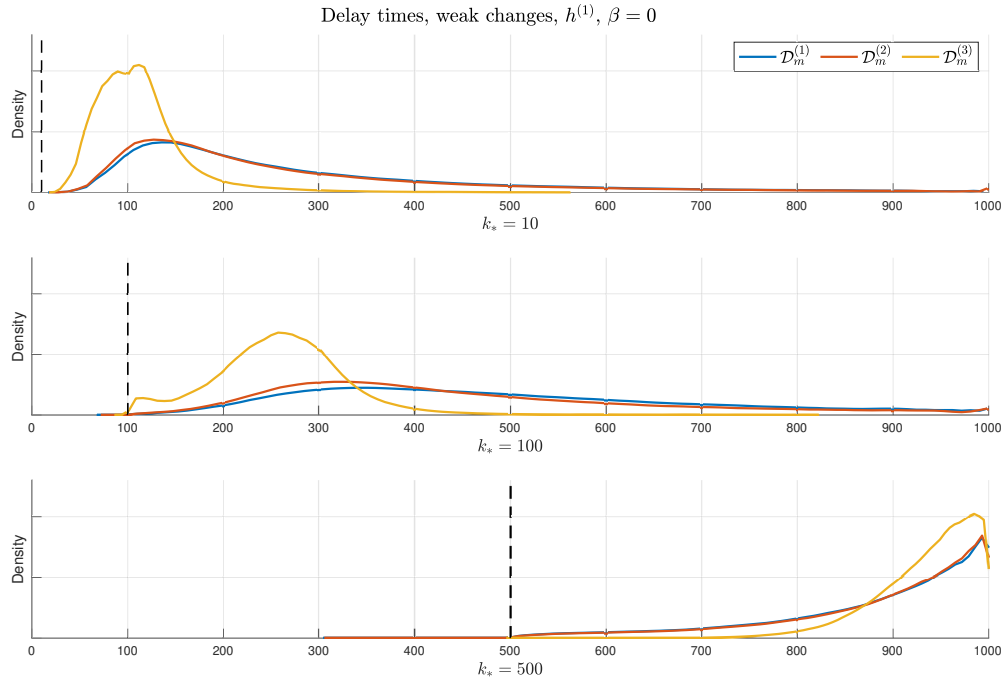


Figure A.2. Detection delays - density under *weak* breaks

In Tables A.3 and A.4, we consider a comparison between our proposed detection schemes $\mathcal{D}_m^{(1)}$, $\mathcal{D}_m^{(2)}$ and $\mathcal{D}_m^{(3)}$, and a CUSUM based detector, i.e.,

$$(A.1) \quad Z_m(k) = \left\| \sum_{i=m+1}^{m+k} \mathbf{X}_i - \frac{k}{m} \sum_{i=1}^m \mathbf{X}_i \right\|$$

We also use a different statistic, called “CUSUM-cov”, based on the detector

$$(A.2) \quad Z_m(k) = \left\| \sum_{i=m+1}^{m+k} \mathbf{Y}_i - \frac{k}{m} \sum_{i=1}^m \mathbf{Y}_i \right\|$$

where $\mathbf{Y}_i = \text{Vech}(\mathbf{X}_i \mathbf{X}_i^\top)$. When using $\mathcal{D}_m^{(1)}$, $\mathcal{D}_m^{(2)}$ and $\mathcal{D}_m^{(3)}$, we use the kernel $h^{(1)}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_1^{1/2}$, and only consider their *unweighted* version, setting $\beta = 0$.¹⁸

TABLE A.3. Power under alternatives - comparison with CUSUM schemes under strong changes

Scheme	$H_{A,1}$	$H_{A,2}$	$H_{A,3}$
$\mathcal{D}_m^{(1)}$	0.836	0.801	0.971
$\mathcal{D}_m^{(2)}$	0.892	0.827	1.000
$\mathcal{D}_m^{(3)}$	0.988	1.000	1.000
CUSUM	0.864	0.057	0.051
CUSUM-vec	0.264	1.000	0.721

TABLE A.4. Power under alternatives - comparison with CUSUM schemes under weak changes

Scheme	$H_{A,1}$	$H_{A,2}$	$H_{A,3}$
$\mathcal{D}_m^{(1)}$	0.751	0.586	0.764
$\mathcal{D}_m^{(2)}$	0.774	0.590	0.776
$\mathcal{D}_m^{(3)}$	0.856	0.240	0.884
CUSUM	0.780	0.062	0.058
CUSUM-vec	0.128	1.000	0.484

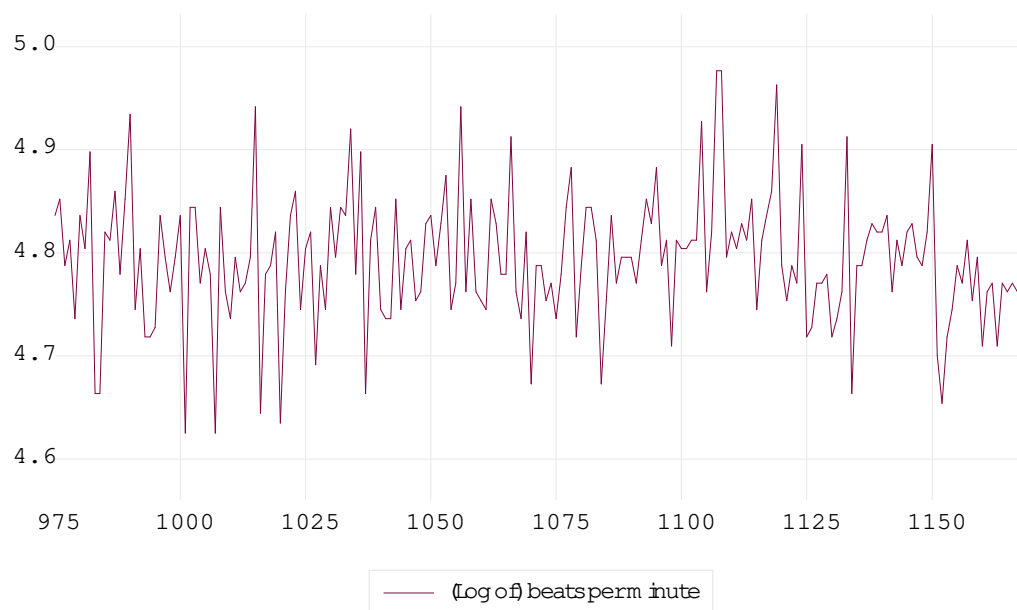
¹⁸Results for different kernels and values of β are very similar, and are available under request.

Results show that CUSUM-based schemes have limited power versus certain alternatives - for example, the scheme (A.1) is designed to detect changes in location, and it performs comparably with the detectors $\mathcal{D}_m^{(1)}$, $\mathcal{D}_m^{(2)}$ and $\mathcal{D}_m^{(3)}$ under $H_{A,1}$; however, it displays virtually no power versus $H_{A,2}$ and $H_{A,3}$. Similarly, the scheme considered in (A.2) is designed to pick up changes in the second moment, and whilst it does have good power versus $H_{A,2}$ and (albeit to a lesser extent) versus $H_{A,3}$, its power is considerably lower versus $H_{A,1}$.

B. COMPLEMENTS TO SECTION 5.2

We begin by reporting a graphical representation of the training sample - spanning from observation 975 till observation 1,169 (i.e., from 01 : 37 : 59 till 02 : 29 : 43) - in Figure B.1.

Figure B.1. Logs of beats per minute - training sample, from 01 : 37 : 59 till 02 : 29 : 43



In Table B.1, we test for the non-contamination condition of Assumption 2.1 based on the methodology developed in Section 4.2 in the main paper. The assumption is virtually always satisfied, and this result is robust to the weight β and the choice of kernel; we note that there are some instances of rejection, when $\beta = 0.9$ and the nominal level is set to a “liberal” value, but otherwise the non-contamination assumption is never rejected.

TABLE B.1. Test for in-sample changepoint - non-contamination assumption

Kernel	β	nominal level	0.01	0.05	0.10
$\ \mathbf{x} - \mathbf{y}\ _2$	0		“not reject”	“not reject”	“not reject”
	0.5		“not reject”	“not reject”	“not reject”
	0.9		“not reject”	“not reject”	“reject”
$[1 - \exp(-\ \mathbf{x} - \mathbf{y}\ _2^2/(2a^2))]^{1/2}$	0		“not reject”	“not reject”	“not reject”
	0.5		“not reject”	“not reject”	“not reject”
	0.9		“not reject”	“reject”	“reject”

In Table B.2, we report a test for Assumption 2.2 (independence), based on Broock et al. (1996); the null of independence is not rejected, indicating that the assumption is satisfied.

TABLE B.2. BDS test for the null of no time dependence

Dimension	BDS statistic	P-value
2	0.0143	0.053
3	0.0132	0.224
4	0.0088	0.446
5	0.0056	0.595
6	0.0027	0.719

BDS test for serial independence. We have used the “fraction of pairs” statistic, calculated so as to ensure that 70% of the total number of pairs of points in the sample lie within a distance of each other equal to 0.7. Bootstrapped p-values are used for the test statistics.

In Table B.3, we report a set of descriptive statistics, including several tests to determine how many moments the data admit. We have used the tests developed in [Trapani \(2016\)](#) and [Degiannakis et al. \(2023\)](#), which we summarize here for completeness. The test for the null hypothesis that the moment of order k of a random variable X does not exist, viz.

$$H_0 : E |X|^k = \infty,$$

is implemented by constructing the statistic

$$\mu_k = \frac{m^{-1} \sum_{i=1}^m |X_i|^k}{(m^{-1} \sum_{i=1}^m |X_i|^2)^{k/2}},$$

for the training sample $\{X_i, 1 \leq i \leq m\}$, and subsequently

$$\psi_k = \exp(\mu_k) - 1.$$

The statistic ψ_k is then randomised according to the following algorithm:

- Step 1:** Generate an artificial sample $\{\xi_n^{(k)}, 1 \leq n \leq N\}$, *i.i.d.* across n and independently across k , with $\xi_1^{(k)} \sim N(0, 1)$, and define $\left\{\psi_k^{1/2} \times \xi_n^{(k)}\right\}_{n=1}^N$.
- Step 2:** For $u \in \{\pm\sqrt{2}\}$, generate $\zeta_{n,m}^{(k)}(u) = I\left(\psi_k^{1/2} \times \xi_n^{(k)} \leq u\right)$, $1 \leq n \leq N$.
- Step 3:** For each u , define

$$\vartheta_{m,N}^{(k)}(u) = \frac{2}{N^{1/2}} \sum_{n=1}^N \left[\zeta_{n,N}^{(k)}(u) - \frac{1}{2} \right],$$

and then the test statistic

$$\Theta_{m,N}^{(k)} = \frac{1}{2} \left[\left(\vartheta_{m,N}^{(k)}(\sqrt{2}) \right)^2 + \left(\vartheta_{m,N}^{(k)}(-\sqrt{2}) \right)^2 \right].$$

Trapani (2016) shows that, as $\min(m, N) \rightarrow \infty$ with $N = O(m)$

$$\begin{aligned} \Theta_{m,N}^{(k)} &\xrightarrow{D^*} \chi_1^2 && \text{under } H_0, \\ N^{-1}\Theta_{m,N}^{(k)} &\xrightarrow{P^*} c_0 > 0 && \text{under } H_A, \end{aligned}$$

where P^* denotes the conditional probability with respect of the sample, and “ $\xrightarrow{D^*}$ ” and “ $\xrightarrow{P^*}$ ” denote conditional convergence in distribution and in probability according to P^* . In order to wash out dependence on the randomness, FILIS proposes running the test for $1 \leq b \leq B$ iterations, each time defining a test statistic ${}_{(b)}\Theta_{m,N}^{(k)}$, and computing the *randomised confidence function*

$$Q_{m,N,B}(\alpha) = \frac{1}{B} \sum_{b=1}^B I \left[{}_{(b)}\Theta_{m,N}^{(k)} \leq c_\alpha \right],$$

where c_α is defined as $P\{\chi_1^2 \geq c_\alpha\} = \alpha$, for a given nominal level $\alpha \in (0, 1)$. Hence, the decision rule in favour of H_0 is

$$(B.1) \quad Q_{M,N,B}(\alpha) \geq (1 - \alpha) - \frac{\sqrt{\alpha(1 - \alpha)}}{f(B)},$$

where the function $f(B)$ is user-defined such that

$$(B.2) \quad \liminf_{B \rightarrow \infty} \frac{B^{1/2}}{f(B)} \geq c_\alpha.$$

Following the indications in Trapani (2016) and Degiannakis et al. (2023), we have used: $M = N = B$, and $f(B) = B^{1/4}$. As can be seen from the table, the test shows that the data admit at least 32 moments. Indeed, the Jarque-Bera test reported in the table barely rejects the null of normality at 5% nominal level (not rejecting it at 1% nominal level). This indicates that the kernels $h(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_2$ and $h(\mathbf{x}, \mathbf{y}) = [1 - \exp(-\|\mathbf{x} - \mathbf{y}\|_2^2/(2a^2))]^{1/2}$ employed in our application satisfy Assumption 2.3.

TABLE B.3. Descriptive statistics and moment existence

Descriptive statistics		Tests for moment existence	
Mean	4.793	Degiannakis et al. (2023)	$H_0 : E X ^8 = \infty$ [reject H_0]
St. Dev.	0.063		$H_0 : E X ^{16} = \infty$ [reject H_0]
Skewness	0.098		$H_0 : E X ^{32} = \infty$ [reject H_0]
Kurtosis	3.864	Jarque-Bera	$H_0 : \text{Gaussian data}$ [p-value=0.041]

The table contains the outcomes for the test by [Degiannakis et al. \(2023\)](#) described above for the null that the moments of order 8, 16 and 32 are non-existent. The tests always reject the null, leading to the conclusion that the data admit at least 32 moments.

We finally report a graphical representation of the training and monitoring sample, including the interquartile range of the identified break dates. This has been (roughly) calculated as the interquartile range of all the break dates found by all our procedures (e.g., for the case $M = 81$, all the dates in the upper panel of Table 5.6, found with both kernels, across the three different values of β , and the three detectors), setting the break date equal to the end of the monitoring period when no break is found.

Figure B.2. Logs of beats per minute - training and monitoring sample, from 01 : 37 : 59 till 02 : 51 : 19

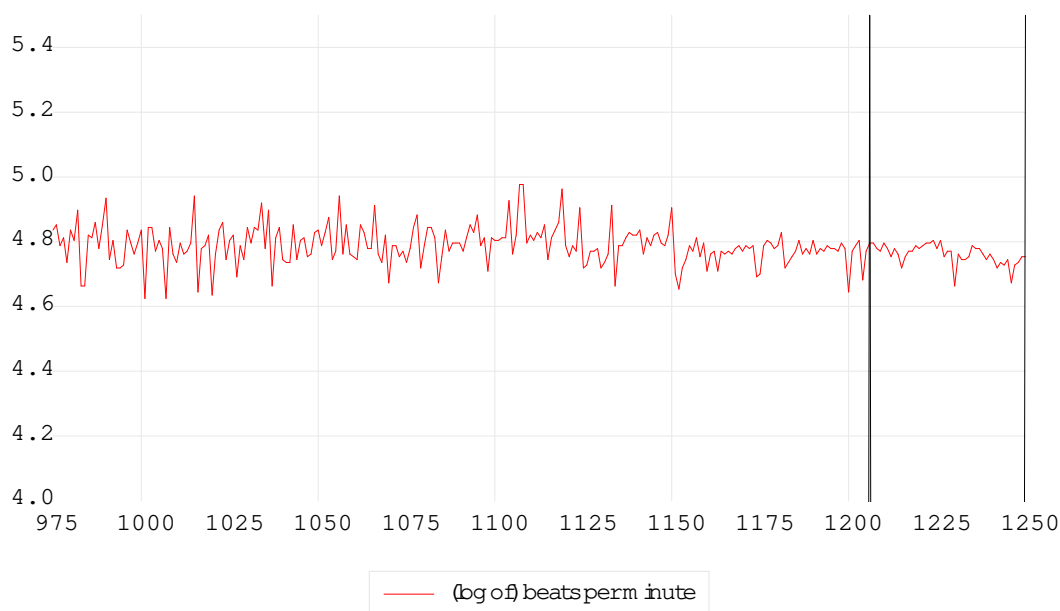
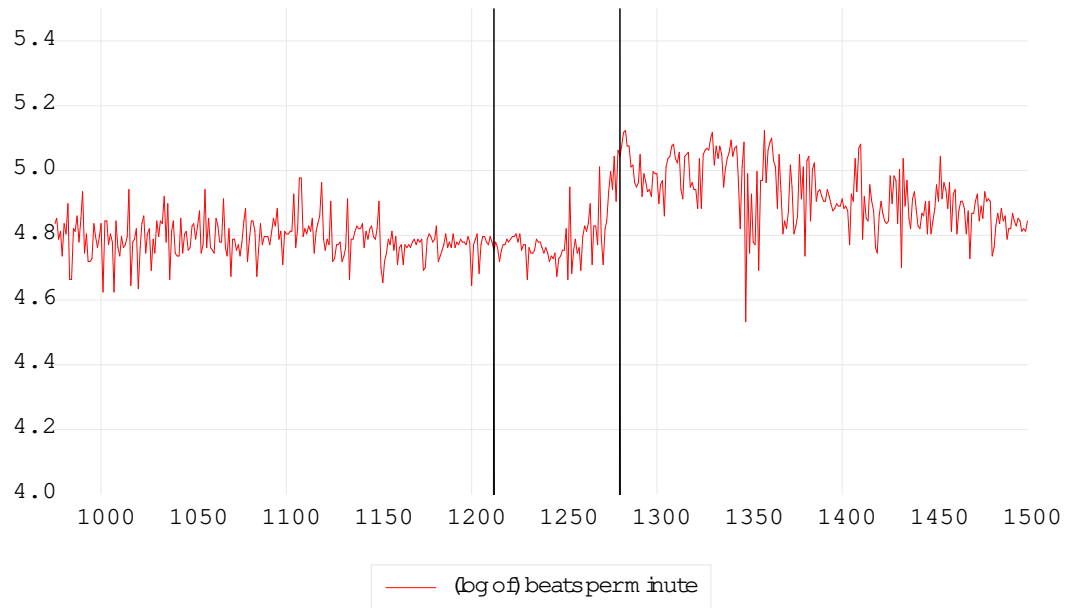


Figure B.3. Logs of beats per minute - training and monitoring sample, from 01 : 37 : 59 to 03 : 55 : 59



C. PRELIMINARY LEMMAS

We begin by collecting a series of lemmas which will be used to prove the main results under H_0 (Lemmas C.1-C.5), under H_A (Lemmas C.6-C.9), and the ones reported in Section 4 (Lemmas C.10-C.12). Throughout this section, Assumptions 2.1, 2.2, and 2.3 are in force, and hence we omit them from statements. Prior to reporting the lemmas, we spell out some notation and several facts which will be used throughout this section and the next one.

In all proofs, $C > 0$ denotes a generic, finite constant independent of m whose value may change line-to-line. For any interval $I \subseteq [0, \infty)$, we write $\mathbf{C}(I)$ to denote the space of continuous real-valued functions on I with the uniform topology, and $\mathbf{D}(I)$ the space of càdlàg functions endowed with the Skorokhod topology, and $\mathbf{C}^r(I)$ and $\mathbf{D}^r(I)$ for their $\mathbb{R}^r$ -valued counterparts, with $r \geq 2$. We use $\Rightarrow$ to denote weak convergence. When convenient for any $a, b \in \mathbb{R}$ we write $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$. Throughout, $\mathcal{F} = (\mathcal{F}_k)_{k \geq 1}$ denotes the natural filtration generated by the sequence $\{\mathbf{X}_k, k \geq 1\}$, i.e., $\mathcal{F}_k = \sigma(\mathbf{X}_1, \dots, \mathbf{X}_k)$.

It can be readily checked that for any function $f : \mathfrak{X} \times \mathfrak{X} \rightarrow \mathbb{R}$ of the form

$$f(\mathbf{x}, \mathbf{y}) = f_0(\mathbf{x}) + f_0(\mathbf{y}),$$

with some function $f_0 : \mathfrak{X} \rightarrow \mathbb{R}$, then for $U_m(\cdot; r, k)$ as in (2.7),

$$(C.1) \quad U_m(f; r, k) = 0, \quad m, k \geq 2 \quad 0 \leq r < k.$$

In particular for $\bar{h}$ as in (3.1), we have $U_m(\bar{h}; r, k) = U_m(h; r, k)$, and hereinafter we may assume without loss of generality that $h = \bar{h}$. We also note that, under Assumption (2.3),

we may write

$$(C.2) \quad h(\mathbf{x}, \mathbf{y}) = \sum_{\ell=1}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{x}) \phi_{\ell}(\mathbf{y}),$$

where the equality holds in the $\mathcal{L}^2(F \times F)$ sense, and for $\mathbf{X} \sim F$,

$$(C.3) \quad \mathbb{E} \phi_{\ell}(\mathbf{X}) \phi_{\ell'}(\mathbf{X}) = \begin{cases} 1, & \text{if } \ell = \ell', \\ 0, & \text{if } \ell \neq \ell'. \end{cases}$$

Moreover, by (3.1), $\mathbb{E} h(\mathbf{X}, \mathbf{y}) = 0$ F -a.s., i.e., the operator A has $\phi(\mathbf{x}) \equiv 1$ as eigenvector (with corresponding eigenvalue 0), so by orthogonality, for all ℓ such that $\lambda_{\ell} \neq 0$, we have

$$(C.4) \quad \mathbb{E} \phi_{\ell}(\mathbf{X}) = 0.$$

Define, for each integer $m, k \geq 1$,

$$(C.5) \quad S_{\ell}(m) = \sum_{i=1}^m \phi_{\ell}(\mathbf{X}_i), \quad S_{\ell}(k, m) = \sum_{j=m+1}^{m+k} \phi_{\ell}(\mathbf{X}_j).$$

Define

$$(C.6) \quad f_{\ell}(\mathbf{x}, \mathbf{y}) = \phi_{\ell}(\mathbf{x}) \phi_{\ell}(\mathbf{y}),$$

$$(C.7) \quad U_m(h; k) = \sum_{\ell=1}^{\infty} \lambda_{\ell} U_m(f_{\ell}; 0, k),$$

and the truncated version

$$(C.8) \quad U_{m,L}(h; k) = \sum_{\ell=1}^L \lambda_{\ell} U_m(f_{\ell}; 0, k).$$

A straightforward calculation shows that, letting $w = k - r$, we have

$$(C.9) \quad \begin{aligned} & m^{-1}w^2U_m(f_\ell; k - w, k) \\ &= -m^{-1} \left(S_\ell(k, m) - S_\ell(k - w, m) - \frac{w}{m}S_\ell(m) \right)^2 + \frac{w(m + w)}{m^2} + R_\ell(k, w, m), \end{aligned}$$

where

$$(C.10) \quad \begin{aligned} & R_\ell(k, w, m) \\ &= -\frac{w^2S_\ell^2(m)}{m^3(m - 1)} + \frac{w^2}{m^2(m - 1)} \sum_{i=1}^m (\phi_\ell^2(\mathbf{X}_i) - 1) + \frac{w^2}{m^2(m - 1)} \\ & - \frac{[S_\ell(k, m) - S_\ell(k - w, m)]^2}{m(w - 1)} + \frac{w}{m(w - 1)} \sum_{j=m+(k-w)+1}^{m+k} (\phi_\ell^2(\mathbf{X}_j) - 1) + \frac{w}{m(w - 1)}. \end{aligned}$$

Lastly, to simplify some expressions in the proofs, for any kernel $f(\mathbf{x}, \mathbf{y})$ we set

$$U_m(f; r, k) = U_m(f; r \wedge (k - 2), k \vee 2), \quad \text{if } k \leq 2 \text{ or } r \geq k - 1.$$

C.1. Lemmas under H_0 . We are now in a position to present our lemmas. We begin with a series of lemmas which are valid under H_0 .

Lemmas C.1 and C.2, below, are used to provide uniform control over the the difference between the process $U_m(h; r, k)$ and its finite-expansion counterpart $U_{m,L}(h; r, k)$ defined in (C.22).

Lemma C.1. *Under H_0 , then for all $L, m, n \geq 1$, we have*

$$(C.11) \quad \mathbb{E} \max_{1 \leq k \leq n} \left(\sum_{1 \leq i \neq j \leq k} \sum_{\ell=L+1}^{\infty} \lambda_\ell \phi_\ell(\mathbf{X}_i) \phi_\ell(\mathbf{X}_j) \right)^2 \leq Cn^2 \sum_{\ell=L+1}^{\infty} \lambda_\ell^2,$$

and

$$(C.12) \quad \mathbb{E} \max_{1 \leq k \leq n} \left(\sum_{i=1}^m \sum_{j=m+1}^{m+k} \sum_{\ell=L+1}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right)^2 \leq Cnm \sum_{\ell=L+1}^{\infty} \lambda_{\ell}^2.$$

Proof. We first establish (C.11). For each integer $q \geq 2$, write

$$Y_q = 2 \sum_{i=1}^{q-1} \sum_{\ell=L+1}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_q),$$

so that

$$\sum_{1 \leq i \neq j \leq k} \sum_{\ell=L+1}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) = \sum_{q=1}^k Y_q.$$

By (C.4), $\mathbb{E}Y_q = 0$, and clearly Y_q is $\mathcal{F}_q$ -measurable, with $\mathbb{E}(Y_q | \mathcal{F}_{q-1}) = 0$, implying $\sum_{q=1}^k Y_q$ is an $(\mathcal{F}_k)_{k \geq 1}$ -martingale. Moreover,

$$\begin{aligned} \mathbb{E}Y_q^2 &= 4 \sum_{i,i'=1}^{q-1} \sum_{\ell,\ell'=L+1}^{\infty} \lambda_{\ell} \lambda_{\ell'} \mathbb{E}[\phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_q) \phi_{\ell'}(\mathbf{X}_{i'}) \phi_{\ell'}(\mathbf{X}_q)] \\ &= 4 \sum_{i,i'=1}^{q-1} \sum_{\ell,\ell'=L+1}^{\infty} \lambda_{\ell} \lambda_{\ell'} \mathbb{E}[\phi_{\ell}(\mathbf{X}_q) \phi_{\ell'}(\mathbf{X}_q)] \mathbb{E}[\phi_{\ell}(\mathbf{X}_i) \phi_{\ell'}(\mathbf{X}_{i'})] \\ &= 4(q-1) \sum_{\ell=L+1}^{\infty} \lambda_{\ell}^2. \end{aligned}$$

Hence, Doob's maximal inequality gives

$$\mathbb{E} \max_{1 \leq k \leq n} \left(\sum_{q=1}^k Y_q \right)^2 \leq 4 \sum_{q=1}^n \mathbb{E}Y_q^2 = 16 \sum_{q=1}^n (q-1) \sum_{\ell=L+1}^{\infty} \lambda_{\ell}^2 \leq Cn^2 \sum_{\ell=L+1}^{\infty} \lambda_{\ell}^2.$$

For (C.12), let $Y_{m,j} = \sum_{i=1}^m \sum_{\ell=L+1}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_{j+m})$. Then, for each fixed m , $\sum_{j=1}^k Y_{j,m}$ is an $(\mathcal{F}_{k+m})_{k \geq 1}$ martingale, and, arguing as before, we have

$$\mathbb{E}Y_{m,j}^2 = m \sum_{\ell=L+1}^{\infty} \lambda_{\ell}^2.$$

(C.12) now follows immediately by applying Doob's inequality. $\square$

Lemma C.2. *Let $x > 0$. Under H_0 , for any integer $L \geq 0$,*

$$(C.13) \quad P \left\{ \sup_{0 \leq r < k < \infty} \frac{(k-r)^2 m^{-1} |U_m(h; r, k) - U_{m,L}(h; r, k)|}{g_m(k)} > x \right\} \leq Cx^{-2} \sum_{\ell=L+1}^{\infty} \lambda_{\ell}^2,$$

where $U_{m,L}$ is defined as in (C.22). Moreover,

$$(C.14) \quad \limsup_{m \rightarrow \infty} P \left\{ \max_{0 \leq r < k \leq m\delta} \frac{|(k-r)^2 m^{-1} U_m(h; r, k)|}{g_m(k)} > x \right\} = O(\delta^{1-\beta}), \quad \delta \rightarrow 0,$$

and

$$(C.15) \quad \limsup_{m \rightarrow \infty} P \left\{ \sup_{k \geq mT} \max_{0 < r < k} \frac{|(k-r)^2 m^{-1} U_m(h; r, k)|}{g_m(k)} > x \right\} = O(1/T), \quad T \rightarrow \infty.$$

Proof. Note, to begin with, that

$$\begin{aligned} & |U_m(h; r, k) - U_{m,L}(h; r, k)| \\ & \leq \frac{2}{(k-r)m} \left| \sum_{i=1}^m \sum_{j=m+r+1}^{m+k} \sum_{\ell=L+1}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right| \\ & \quad + \binom{m}{2}^{-1} \left| \sum_{1 \leq i < j \leq m} \sum_{\ell=L+1}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right| + \binom{k-r}{2}^{-1} \left| \sum_{m+r < i < j \leq m+k} \sum_{\ell=L+1}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right|. \end{aligned}$$

Let now $0 < \delta \leq 1$. Since $g_m(k) \geq C(k/m)^{\beta}$ for all $1 \leq k \leq m\delta$, any integer $L \geq 0$ we have

$$\begin{aligned} & P \left\{ \max_{0 < r < k \leq m\delta} \frac{(k-r)^2}{mg_m(k)} \frac{1}{(k-r)m} \left| \sum_{i=1}^m \sum_{j=m+r+1}^{m+k} \sum_{\ell=L+1}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right| > x \right\} \\ & \leq P \left\{ \max_{0 \leq r < k \leq m\delta} \frac{k^{1-\beta}}{m^{2-\beta}} \left| \sum_{i=1}^m \sum_{j=m+r+1}^{m+k} \sum_{\ell=L+1}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right| > Cx \right\} \\ & \leq P \left\{ \max_{1 \leq q \leq \lfloor \log(m\delta) \rfloor} \max_{e^{q-1} \leq k < e^q} \max_{0 \leq r < k} \frac{k^{1-\beta}}{m^{2-\beta}} \left| \sum_{i=1}^m \sum_{j=m+r+1}^{m+k} \sum_{\ell=L+1}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right| > Cx \right\} \end{aligned}$$

(C.16)

$$\leq \sum_{q=1}^{\lceil \log(m\delta) \rceil} P \left\{ \frac{e^{q(1-\beta)}}{m^{2-\beta}} \max_{e^{q-1} \leq k < e^q} \max_{0 \leq r < k} \left| \sum_{i=1}^m \sum_{j=m+r+1}^{m+k} \sum_{\ell=L+1}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right| > Cx \right\}$$

Using the bound

$$\begin{aligned} & \max_{0 \leq r < k} \left| \sum_{i=1}^m \sum_{j=m+r+1}^{m+k} \sum_{\ell=L+1}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right| \\ (C.17) \quad & \leq \left| \sum_{i=1}^m \sum_{j=m+1}^{m+k} \sum_{\ell=L+1}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right| + \max_{1 \leq r < k} \left| \sum_{i=1}^m \sum_{j=m+1}^{m+r} \sum_{\ell=L+1}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right|, \end{aligned}$$

which holds for each fixed k , (C.16) is bounded by

$$\begin{aligned} & \sum_{q=1}^{\lceil \log(m\delta) \rceil} P \left\{ \frac{e^{q(1-\beta)}}{m^{2-\beta}} \max_{e^{q-1} \leq k < e^q} \left| \sum_{i=1}^m \sum_{j=m+1}^{m+k} \sum_{\ell=L+1}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right| > Cx/2 \right\} \\ & + \sum_{q=1}^{\lceil \log(m\delta) \rceil} P \left\{ \frac{e^{q(1-\beta)}}{m^{2-\beta}} \max_{1 \leq r < e^q} \left| \sum_{i=1}^m \sum_{j=m+1}^{m+r} \sum_{\ell=L+1}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right| > Cx/2 \right\} \\ (C.18) \quad & \leq \frac{C}{x^2 m^{4-2\beta}} \sum_{q=1}^{\lceil \log(m\delta) \rceil} m e^{2q(1-\beta)} e^q \sum_{\ell=L+1}^{\infty} \lambda_{\ell}^2 \leq \frac{C \delta^{3-2\beta}}{x^2} \sum_{\ell=L+1}^{\infty} \lambda_{\ell}^2. \end{aligned}$$

For any $T \geq 1$, we have $g_m(k) \geq C(k/m)^2$ for all $k \geq Tm$, and applying (C.17) again we obtain

$$\begin{aligned} & P \left\{ \sup_{k \geq mT} \max_{0 \leq r < k} \frac{k}{m^2 g_m(k)} \left| \sum_{i=1}^m \sum_{j=m+r+1}^{m+k} \sum_{\ell=L+1}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right| > x \right\} \\ & \leq P \left\{ \sup_{k \geq mT} \max_{0 \leq r < k} \frac{1}{k} \left| \sum_{i=1}^m \sum_{j=m+r+1}^{m+k} \sum_{\ell=L+1}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right| > Cx \right\} \\ & \leq \sum_{q=\lceil \log(mT) \rceil}^{\infty} P \left\{ \max_{e^{q-1} \leq k < e^q} \frac{1}{e^{q-1}} \left| \sum_{i=1}^m \sum_{j=m+1}^{m+k} \sum_{\ell=L+1}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right| > Cx/2 \right\} \\ & + \sum_{q=\lceil \log(mT) \rceil}^{\infty} P \left\{ \max_{1 \leq r < e^q} \frac{1}{e^{q-1}} \left| \sum_{i=1}^m \sum_{j=m+1}^{m+r} \sum_{\ell=L+1}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right| > Cx/2 \right\} \end{aligned}$$

$$(C.19) \quad \leq \frac{C}{x^2} \sum_{\ell=L+1}^{\infty} \lambda_{\ell}^2 \sum_{q=\lfloor \log(mT) \rfloor}^{\infty} m e^{-q} \leq \frac{C}{T x^2} \sum_{\ell=L+1}^{\infty} \lambda_{\ell}^2.$$

In particular, if we take $\delta = T = 1$, we obtain

$$P \left\{ \sup_{2 \leq w \leq k < \infty} \frac{(k-r)^2}{m g_m(k)} \left| \frac{1}{(k-r)m} \sum_{i=1}^m \sum_{j=m+1}^{m+k} \sum_{\ell=L+1}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right| > x \right\} \leq \frac{C}{x^2} \sum_{\ell=L+1}^{\infty} \lambda_{\ell}^2.$$

Analogous arguments leading to (C.18) and (C.19) give

$$\begin{aligned} \limsup_{m \rightarrow \infty} P \left\{ \max_{2 \leq k \leq \delta m} \frac{k^2}{m g_m(k)} \left| \frac{1}{m^2} \sum_{1 \leq i < j \leq m} \sum_{\ell=L+1}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right| > x \right\} &= O(\delta^{2-\beta}), \\ \limsup_{m \rightarrow \infty} P \left\{ \max_{2 \leq k \leq \delta m} \max_{0 \leq r < k} \frac{(k-r)^2}{m g_m(k)} \left| \frac{1}{(k-r)^2} \sum_{m+r < i < j \leq m+k} \sum_{\ell=L+1}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right| > x \right\} &= O(\delta^{1-\beta}), \end{aligned}$$

as $\delta \rightarrow 0$, and

$$\begin{aligned} \limsup_{m \rightarrow \infty} P \left\{ \sup_{k \geq mT} \frac{k^2}{m g_m(k)} \left| \frac{1}{m^2} \sum_{1 \leq i < j \leq m} \sum_{\ell=L+1}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right| > x \right\} &= O(1/T), \\ \limsup_{m \rightarrow \infty} P \left\{ \sup_{k \geq mT} \max_{0 \leq r < k} \frac{(k-r)^2}{m g_m(k)} \left| \frac{1}{(k-r)^2} \sum_{m+r < i < j \leq m+k} \sum_{\ell=L+1}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right| > x \right\} &= O(1/T), \end{aligned}$$

as $T \rightarrow \infty$, which gives (C.14). $\square$

The next lemma shows that the $U_{m,L}(\ell, k)$ can be approximated by a weighted sum of squared CUSUM-type statistics, based on the eigenfunctions of h .

Lemma C.3. *Under H_0 , for any sequence $T > 0$ with and fixed $L \geq 1$,*

$$\sup_{2 \leq w \leq k \leq mT} \sum_{\ell=1}^L \left| \frac{\lambda_{\ell} R_{\ell}(k, w, m)}{g_m(k)} \right| = o_P(1),$$

with $R_{\ell}(k, w, m)$ as defined in (C.10).

Proof. Fix any $1 \leq \ell \leq L$; we proceed to analyze each term in $R_\ell(k, w, m)$ separately. Since $g_m(k) = g(k/m)$, it is easily seen that

$$\sup_{2 \leq w \leq k \leq mT} \frac{1}{g_m(k)} \frac{w^2}{m^2(m-1)} \leq \sup_{0 \leq t \leq T} \left(\frac{t}{1+t} \right)^{-\beta} (1+t)^{-2} \frac{t^2}{(m-1)} = O((T \wedge 1)^{2-\beta} m^{-1})$$

and

$$\sup_{2 \leq k \leq mT} \frac{1}{g_m(k)} \frac{w}{m(w-1)} \leq C \sup_{0 \leq t \leq T} \left(\frac{t}{1+t} \right)^{-\beta} (1+t)^{-2} \frac{1}{m} = O((T \wedge 1)^{2-\beta} m^{-1}).$$

Similarly, from (C.3), we have

$$\begin{aligned} & \sup_{1 \leq w \leq k \leq mT} \frac{1}{g_m(k)} \frac{w^2}{m^2(m-1)} \left| \sum_{i=1}^m (\phi_\ell^2(\mathbf{X}_i) - 1) \right| \\ & \leq C (T \wedge 1)^{2-\beta} \frac{1}{m-1} \left| \sum_{i=1}^m (\phi_\ell^2(\mathbf{X}_i) - 1) \right| = O_P(m^{-1/2} (T \wedge 1)^{2-\beta}), \end{aligned}$$

and from (C.4),

$$\sup_{1 \leq w \leq k \leq mT} \frac{1}{g_m(k)} \frac{w^2 |S_\ell(m)|}{m^3(m-1)} \leq C \frac{(T \wedge 1)^{2-\beta}}{m(m-1)} |S_\ell(m)| = O_P(m^{-3/2} (T \wedge 1)^{2-\beta}).$$

Also,

$$\begin{aligned} & P \left\{ \max_{2 \leq w \leq k \leq mT} \left| \frac{w}{m(w-1)} \sum_{j=m+(k-w)}^{m+k} (\phi_\ell^2(\mathbf{X}_j) - 1) \right| > x \right\} \\ & \leq P \left\{ \max_{2 \leq k \leq mT} \left| \frac{1}{m} \sum_{j=m+1}^{m+k} (\phi_\ell^2(\mathbf{X}_j) - 1) \right| > x/4 \right\} \\ & \quad + P \left\{ \max_{2 \leq w \leq k \leq mT} \left| \frac{1}{m} \sum_{j=m+1}^{m+(k-w)} (\phi_\ell^2(\mathbf{X}_j) - 1) \right| > x/4 \right\} \\ & \leq C m^{-1} \mathbb{E} \left| \sum_{j=1}^{\lceil Tm \rceil} (\phi_\ell^2(\mathbf{X}_j) - 1) \right| = o(1), \end{aligned}$$

where the last inequality follows from Kolmogorov's maximal inequality and the $o(1)$ statement follows from uniform integrability of $m^{-1} \sum_{j=1}^{mT} (\phi_\ell^2(\mathbf{X}_j) - 1)$. Finally, we will show

$$(C.20) \quad P \left\{ \max_{2 \leq w \leq k \leq Tm} \frac{|S_\ell(k, m) - S_\ell(k - w, m)|^2}{m(w - 1)} > x \right\} = o(1),$$

which will complete the statement. Fix $0 < \delta < 1/3$. Then

$$\begin{aligned} & P \left\{ \max_{2 \leq w \leq k \leq Tm} \frac{|S_\ell(k, m) - S_\ell(k - w, m)|^2}{m(w - 1)} \mathbf{1}_{\{w > m^\delta\}} > x \right\} \\ & \leq P \left\{ \max_{2 \leq w \leq k \leq Tm} \frac{|S_\ell(k, m)|^2 + |S_\ell(k - w, m)|^2}{m^{1+\delta}} > x/4 \right\} \\ & \leq P \left\{ \max_{2 \leq k \leq Tm} \frac{|S_\ell(k, m)|^2}{m^{1+\delta}} > x/8 \right\} \\ & \leq C \frac{\mathbf{E} |S_\ell(mT, m)|^2}{xm^{1+\delta}} \\ & \leq Cm^{-\delta}. \end{aligned}$$

Next, we have

$$\begin{aligned} & P \left\{ \max_{2 \leq w \leq k \leq Tm} \frac{|S_\ell(k, m) - S_\ell(k - w, m)|^2}{m(w - 1)} \mathbf{1}_{\{w \leq m^\delta\}} > x \right\} \\ (C.21) \quad & \leq P \left\{ \max_{2 \leq w \leq k \leq m^\delta} \frac{|S_\ell(k, m) - S_\ell(k - w, m)|^2}{m} > x \right\} + P \left\{ m^{-1} \max_{m^\delta < k \leq Tm} Y_{k,m} > x \right\}, \end{aligned}$$

with

$$\begin{aligned} Y_{k,m} &= \max_{2 \leq w \leq m^\delta} \left| \frac{1}{\sqrt{w-1}} \sum_{j=m+(k-w)+1}^k \phi_\ell(\mathbf{X}_j) \right|^2 \\ &= \max_{2 \leq w \leq m^\delta} \frac{1}{w-1} |S_\ell(k, m) - S_\ell(k-w, m)|^2. \end{aligned}$$

Note

$$\mathbb{E}Y_{k,m} \leq \left| \frac{1}{\sqrt{m^\delta - 1}} \sum_{j=m+(k-m^\delta)+1}^k \phi_\ell(\mathbf{X}_j) \right|^2 \leq \frac{m^\delta}{m^\delta - 1} \leq 2.$$

Thus, for all $\varepsilon > 0$,

$$\mathbb{E} \max_{2 \leq k \leq m} Y_{k,m} \leq m\varepsilon + \int_{\varepsilon m}^{\infty} mP(Y_2 > t)dt = \varepsilon m + o(m),$$

i.e. $\mathbb{E} \max_{2 \leq k \leq m} Y_{k,m} = o(m)$, giving $P\{m^{-1} \max_{m^\delta < k \leq Tm} Y_{k,m} > x\} = o(1)$ in (C.21). For the second term in (C.21), note

$$\begin{aligned} \max_{2 \leq w \leq k \leq m^\delta} \frac{|S_\ell(k, m) - S_\ell(k - w, m)|^2}{m} &\leq \max_{2 \leq w \leq k \leq m^\delta} \frac{w^2 \max_{m+(k-w) < j \leq m+k} \phi_\ell(\mathbf{X}_j)}{m} \\ &\leq \frac{\max_{2 \leq j \leq m^\delta} \phi_\ell^2(\mathbf{X}_{j+m})}{m^{1-2\delta}} = \frac{o_P(m^\delta)}{m^{1-2\delta}} = o_P(1), \end{aligned}$$

implying (C.21) tends to zero, which gives the result. $\square$

We now report two approximation lemmas that are central to the main proofs. The first shows the weighted truncated processes $U_{m,L}$ can be approximated by limits driven by a linear combination of squares of Gaussian processes.

Lemma C.4. *Fix $L \geq 1$, and set*

$$(C.22) \quad U_{m,L}(r, k) = \sum_{\ell=1}^L \lambda_\ell U_m(f_\ell; r, k),$$

where f_ℓ is given in (C.6). Let

$$(C.23) \quad \mathbb{U}_{m,L}(s, t) = m^{-1}((\lfloor mt \rfloor - \lfloor ms \rfloor) \vee 2)^2 U_{m,L}(\lfloor ms \rfloor, \lfloor mt \rfloor), \quad 0 \leq s \leq t.$$

Also, for every $s, t \geq 0$, set

$$(C.24) \quad \mathbb{V}_L(s, t) = - \sum_{\ell=1}^L \lambda_\ell [(W_{2,\ell}(t) - W_{2,\ell}(s) - (t-s)W_{1,\ell}(1))^2 - (t-s)(1+t-s)],$$

where $\{W_{1,1}(t), t \geq 0\}$, $\{W_{2,1}(t), t \geq 0\}$, $\{W_{1,2}(t), t \geq 0\}$, $\{W_{2,2}(t), t \geq 0\}$, $\dots$ are independent Wiener processes. Then, we may define a sequence $\{\mathbb{V}_{m,L}, m \geq 1\}$ of processes $\mathbb{V}_{m,L} = \{\mathbb{V}_{m,L}(s, t), s, t \geq 0\}$ such that for each m , $\mathbb{V}_{m,L} \stackrel{\mathcal{D}}{=} \mathbb{V}_L$, and for any $0 < \delta < T$,

$$(C.25) \quad \sup_{s, t \in I_{\delta, T}} \left| \frac{\mathbb{V}_{m,L}(s, t)}{g(t)} - \frac{\mathbb{U}_{m,L}(s, t)}{g_m(\lfloor mt \rfloor)} \right| = o_P(1),$$

with $I_{\delta, T} = \{(s, t) : \delta \leq t \leq T, 0 \leq s \leq t\}$.

Proof. For $0 \leq s \leq t$, write

$$(C.26) \quad \begin{aligned} & \mathbb{U}_{m,L}^\circ(s, t) \\ &= \sum_{\ell=1}^L \lambda_\ell \left(\frac{1}{m} \left(S_\ell(\lfloor mt \rfloor, m) - S_\ell(\lfloor ms \rfloor, m) - \frac{\lfloor mt \rfloor - \lfloor ms \rfloor}{m} S_\ell(m) \right)^2 \right. \\ & \quad \left. - \frac{(\lfloor mt \rfloor - \lfloor ms \rfloor)(\lfloor mt \rfloor - \lfloor ms \rfloor + m)}{m^2} \right). \end{aligned}$$

The Dudley-Wichura-Skorokhod Theorem (see e.g. [Shorack and Wellner, 1986](#), p. 47) entails that, for each m , one can construct independent Wiener processes $\{W_{1,1,m}(t), t \geq 0\}$, $\{W_{2,1,m}(t), t \geq 0\}$, $\dots$, $\{W_{1,L,m}(t), t \geq 0\}$, $\{W_{2,L,m}(t), t \geq 0\}$ such that

$$|m^{-1/2} S_\ell(m) - W_{1,\ell,m}(1)| + \sup_{0 \leq t \leq T} |m^{-1/2} S_\ell(\lfloor mt \rfloor, m) - W_{2,\ell,m}(t)| = o_P(1), \quad 1 \leq \ell \leq L.$$

Hence, for all $1 \leq \ell \leq L$,

$$\sup_{0 \leq s \leq t \leq T} |m^{-1/2} [S_\ell(\lfloor mt \rfloor, m) - S_\ell(\lfloor ms \rfloor, m)] - W_{2,\ell,m}(t) + W_{2,\ell,m}(s)| = o_P(1).$$

and

$$\sup_{0 \leq s \leq t \leq T} \left| \frac{\lfloor mt \rfloor - \lfloor ms \rfloor}{m} m^{-1/2} S_\ell(m) - (t - s) W_{1,\ell,m}(m) \right| = o_P(1).$$

Thus, if we define, for all $0 \leq s \leq t$,

$$(C.27) \quad \begin{aligned} & \mathbb{V}_{m,L}(s, t) \\ &= - \sum_{\ell=1}^L \lambda_{\ell} \left[(W_{2,\ell,m}(t) - W_{2,\ell,m}(s) - (t-s)W_{1,\ell,m}(1))^2 - (t-s)(1+t-s) \right] \end{aligned}$$

then

$$(C.28) \quad \sup_{0 \leq s \leq t \leq T} |\mathbb{V}_{m,L}(s, t) - \mathbb{U}_{m,L}^{\circ}(s, t)| = o_P(1).$$

Since $\sup_{\delta \leq t \leq T} |g_m(\lfloor mt \rfloor) - g(t)| \rightarrow 0$ and $\inf_{\delta \leq t \leq T} |g(t)| > 0$, from (C.28) we obtain

$$(C.29) \quad \sup_{s, t \in I_{\delta, T}} \left| \frac{\mathbb{V}_{m,L}(s, t)}{g(t)} - \frac{\mathbb{U}_{m,L}^{\circ}(s, t)}{g_m(\lfloor mt \rfloor)} \right| = o_P(1).$$

Lastly, Lemma C.3 yields

$$\sup_{s, t \in I_{\delta, T}} \frac{|\mathbb{U}_{m,L}(s, t) - \mathbb{U}_{m,L}^{\circ}(s, t)|}{g_m(\lfloor mt \rfloor)} \leq \sup_{1 \leq w < k \leq mT} \left| \sum_{\ell=1}^L \frac{\lambda_{\ell} R_{\ell}(k, k-w, m)}{g_m(k)} \right| = o_P(1),$$

which combined with (C.29) gives (C.25). $\square$

The next lemma shows the weak limit of the (weighted) $U_{m,L}$ can be itself approximated when L is large.

Lemma C.5. *For each $r, s, t \geq 0$, let*

$$(C.30) \quad \mathbb{V}(s, t) = \sum_{\ell=1}^{\infty} \lambda_{\ell} \left[(W_{2,\ell}(t) - W_{2,\ell}(s) - (t-s)W_{1,\ell}(1))^2 - (t-s)(1+t-s) \right],$$

where $\{W_{1,1}(t), t \geq 0\}, \{W_{2,1}(t), t \geq 0\}, \{W_{1,2}(t), t \geq 0\}, \{W_{2,2}(t), t \geq 0\}, \dots$ are independent Wiener processes and the sums in (C.30) are understood as limits in $\mathcal{L}^2(P)$. Also, set

$$\mathcal{V}(s, t) = \frac{\mathbb{V}(s \wedge t, t)}{g(t)}, \quad \mathcal{V}_L(s, t) = \frac{\mathbb{V}_L(s \wedge t, t)}{g(t)} \quad s \geq 0, t > 0,$$

with $\mathbb{V}_L$ as in (C.24), and set $\mathcal{V}(s, 0) = \mathcal{V}_L(s, 0) = 0$ for all $s \geq 0$. Then, $\{\mathcal{V}(s, t) \mid s, t \geq 0\}$ admits a continuous version and $\mathcal{V}_L \Rightarrow \mathcal{V}$ in $\mathbf{C}([0, \infty) \times [0, \infty))$ as $L \rightarrow \infty$. Moreover, $\sup_{s, t \geq 0} |\mathcal{V}(s, t)| < \infty$ a.s., $i = 1, 2$.

Proof. Set $\mathcal{V}_0 \equiv 0$. For any $t > 0$, $0 \leq s \leq t$, and $L \geq 0$,

$$\begin{aligned} & \mathbb{E} |\mathcal{V}_L(s, t) - \mathcal{V}(s, t)|^2 \\ &= \mathbb{E} \left| \frac{1}{g(t)} \sum_{\ell=L+1}^{\infty} \lambda_{\ell} [(W_{2,\ell}(t) - W_{2,\ell}(s) - (t-s)W_{1,\ell}(1))^2 - (t-s)(1+t-s)] \right|^2 \\ &= \frac{2}{g^2(t)} ((t-s)(1+t-s))^2 \sum_{\ell=L+1}^{\infty} \lambda_{\ell}^2. \end{aligned}$$

This implies for each $n \geq 1$ and any $s_1, t_1, \dots, s_n, t_n \geq 0$, $(\mathcal{V}_L(s_1, t_1), \dots, \mathcal{V}_L(s_n, t_n)) \Rightarrow (\mathcal{V}(s_1, t_1), \dots, \mathcal{V}(s_n, t_n))$ as $L \rightarrow \infty$, so it remains to show tightness and continuity. Write

$$\begin{aligned} Y_{\ell}(s, t) &= t^{-\beta} (W_{2,1}(t) - W_{2,1}(s \wedge t) - (t - (s \wedge t))W_{1,1}(1))^2 \\ &= t^{-\beta} (Z_{\ell}(t) - Z_{\ell}(t \wedge s))^2, \end{aligned}$$

where

$$Z_{\ell}(t) = W_{2,\ell}(t) - tW_{1,\ell}(1).$$

With $m(s, t) = t^{-\beta}(t - (t \wedge s))(1 + t - (t \wedge s))$, we have

$$\begin{aligned} \mathcal{V}_L(s, t) &= (1+t)^{\beta-2} \sum_{\ell=1}^L \lambda_{\ell} [Y_{\ell}(s, t) - m(s, t)] \\ (C.31) \quad &= (1+t)^{\beta-2} \sum_{\ell=1}^L \lambda_{\ell} \tilde{Y}_{\ell}(s, t). \end{aligned}$$

Further, note for $0 \leq s_i \leq t_i \leq T$ $i = 1, 2$ and $r > 0$, Rosenthal's inequality yields

$$\mathbb{E} |(1+t_1)^{2-\beta} V_L(s_1, t_1) - (1+t_2)^{2-\beta} V_L(s_2, t_2)|^{2r}$$

$$(C.32) \quad \leq C_r \left[\sum_{\ell=1}^L |\lambda_\ell|^{2r} \mathbb{E} \left| \tilde{Y}_\ell(s_1, t_1) - \tilde{Y}_\ell(s_2, t_2) \right|^{2r} + \left(\sum_{\ell=1}^L \lambda_\ell^2 \mathbb{E} |\tilde{Y}_\ell(s_1, t_1) - \tilde{Y}_\ell(s_2, t_2)|^2 \right)^r \right].$$

Now,

$$(C.33) \quad |Y_\ell(s_1, t_1) - Y_\ell(s_2, t_2)|^{2r} \leq C \left(|t_1^{-\beta} Z_\ell^2(t_1) - t_2^{-\beta} Z_\ell^2(t_2)|^r + |t_1^{-\beta} Z_\ell^2(s_1) - t_2^{-\beta} Z_\ell^2(s_2)|^{2r} \right. \\ \left. + |t_1^{-\beta} Z_\ell(t_1) Z_\ell(s_1) - t_2^{-\beta} Z_\ell(t_2) Z_\ell(s_2)|^{2r} \right).$$

We proceed to bound the expectation of each term in (C.33). Suppose for the moment that for any $T > 0$, $0 \leq s_i \leq t_i \leq T$, $i = 1, 2$,

$$(C.34) \quad \mathbb{E} \left(t_1^{-\beta/2} Z_1(s_1) - t_2^{-\beta/2} Z_1(s_2) \right)^2 \leq C_T (|t_1 - t_2| + |s_1 - s_2|)^a,$$

for some $0 < a < 1 - \beta$. Then, for any $r > 0$, Gaussianity of Z_ℓ gives

$$\mathbb{E} \left| t_1^{-\beta/2} Z_\ell(s_1) - t_2^{-\beta/2} Z_\ell(s_2) \right|^{2r} \leq C_{r,T} (|t_1 - t_2| + |s_1 - s_2|)^{ar}.$$

from which we obtain

$$\begin{aligned} & \mathbb{E} |t_1^{-\beta} Z_\ell^2(s_1) - t_2^{-\beta} Z_\ell^2(s_2)|^{2r} \\ & \leq C \left(\mathbb{E} |t_1^{-\beta/2} Z_\ell(s_1) - t_2^{-\beta/2} Z_\ell(s_2)|^{4r} \right)^{1/2} \left(\mathbb{E} |t_1^{-\beta/2} Z_\ell(s_1)|^{4r} + \mathbb{E} |t_2^{-\beta/2} Z_\ell(s_2)|^{4r} \right)^{1/2} \\ & \leq C (|t_1 - t_2| + |s_1 - s_2|)^{2ar}. \end{aligned}$$

Similarly,

$$\begin{aligned} & \mathbb{E} |t_1^{-\beta} Z_\ell(s_1) Z_\ell(t_1) - t_2^{-\beta} Z_\ell(t_2) Z_\ell(s_2)|^{2r} \\ & \leq C \left(\left(\mathbb{E} |t_1^{-\beta/2} Z_\ell(s_1) - t_2^{-\beta/2} Z_\ell(s_2)|^{4r} \right)^{1/2} \left(\mathbb{E} |t_1^{-\beta/2} Z_\ell(t_1)|^{4r} \right)^{1/2} \right. \\ & \quad \left. + \left(\mathbb{E} |t_1^{-\beta/2} Z_\ell(t_1) - t_2^{-\beta/2} Z_\ell(t_2)|^{4r} \right)^{1/2} \left(\mathbb{E} |t_2^{-\beta/2} Z_\ell(s_2)|^{4r} \right)^{1/2} \right) \\ & \leq C (|t_1 - t_2| + |s_1 - s_2|)^{2ar}. \end{aligned}$$

Moreover, since $m(s, t) = m(s \wedge t, t)$ it is easily seen $m(s, t)$ is locally a -Hölder continuous for any $0 < a < 1 - \beta$. Hence, with $\tilde{Y}_\ell$ as in (C.31),

$$\mathbb{E}|\tilde{Y}_\ell(s_1, t_1) - \tilde{Y}_\ell(s_2, t_2)|^{2r} \leq C(|t_1 - t_2| + |s_1 - s_2|)^{2ar}.$$

From (C.32), since $\sum_{\ell \geq 1} \lambda_\ell^2 < \infty$ we deduce,

$$\mathbb{E}|(1+t_1)^{2-\beta}V_L(s_1, t_1) - (1+t_2)^{2-\beta}V_L(s_2, t_2)|^{2r} \leq C(|t_1 - t_2| + |s_1 - s_2|)^{2ar},$$

Taking r sufficiently large and applying Corollary 14.9 in [Kallenberg \(2002\)](#) yields a continuous version of $\{(1+t)^{2-\beta}\mathcal{V}(s, t), s, t \geq 0\}$ and tightness of the sequence $\{(1+t)^{2-\beta}V_L(s, t), s, t \geq 0\}$ in $\mathbf{C}([0, T] \times [0, T])$ for each $T > 0$. Thus, we have

$$\{(1+t)^{2-\beta}\mathcal{V}_L(s, t), s, t \geq 0\} \Rightarrow \{(1+t)^{2-\beta}\mathcal{V}(s, t), s, t \geq 0\} \quad \text{in} \quad \mathbf{C}([0, \infty) \times [0, \infty)),$$

which yields the desired weak convergence $\mathcal{V}_L \Rightarrow \mathcal{V}$ in $C([0, \infty) \times [0, \infty))$. To see that $\sup_{s, t \geq 0} |\mathcal{V}(s, t)| < \infty$ a.s., since $\mathcal{V} \in \mathbf{C}([0, \infty) \times [0, \infty))$ it suffices that $\sup_{s, t \geq T} |\mathcal{V}(s, t)| < \infty$ for some $T > 0$. Observe that from (D.12) we have

$$(C.35) \quad \{\mathcal{V}(s, t), s, t > 0\} \stackrel{\mathcal{D}}{=} \left\{ \left(\frac{t}{1+t} \right)^{-\beta} G(s \wedge t, t), s, t > 0 \right\},$$

where

$$G(s, t) = \sum_{\ell=1}^{\infty} \lambda_\ell \left[\left(W_\ell \left(\frac{t}{1+t} \right) - \frac{1+s}{1+t} W_\ell \left(\frac{s}{1+s} \right) \right)^2 - \frac{(t-s)(1+t-s)}{(1+t)^2} \right].$$

Since

$$\begin{aligned} & \sup_{0 < s \leq t < \infty} |G(s, t)| \\ &= \sup_{0 < u \leq v < 1} \left| G \left(\frac{u}{1-u}, \frac{v}{1-v} \right) \right| \end{aligned}$$

$$= \sup_{0 < u \leq v < 1} \sum_{\ell=1}^{\infty} \lambda_{\ell} \left[\left(W_{\ell}(v) - \frac{1-v}{1-u} W_{\ell}(u) \right)^2 - \left(v - u \left(\frac{1-v}{1-u} \right) \right) \left(1 - u \left(\frac{1-v}{1-u} \right) \right) \right],$$

and $\{G(\frac{u}{1-u}, \frac{v}{1-v}), 0 \leq u \leq v < 1\}$ is easily seen to extend to a continuous version on the region $\{0 \leq u \leq v \leq 1\}$, which by (C.35) shows $\sup_{s, t \geq T} |\mathcal{V}(s, t)| < \infty$ for any $T > 0$, thereby giving the statement.

We now conclude the proof by showing (C.34). Note for any $0 < s_i \leq t_i$, $i = 1, 2$,

$$\begin{aligned} & \mathbb{E} \left(t_1^{-\beta/2} Z_{\ell}(s_1) - t_2^{-\beta/2} Z_{\ell}(s_2) \right)^2 \\ & \leq C \left(\mathbb{E} \left(t_1^{-\beta/2} W_{2,\ell}(s_1) - t_2^{-\beta/2} W_{2,\ell}(s_2) \right)^2 + \left(t_1^{-\beta/2} s_1 - t_2^{-\beta/2} s_2 \right)^2 \right). \end{aligned}$$

Without loss of generality suppose $s_1 \geq s_2$. We have

$$\begin{aligned} \mathbb{E} \left(t_1^{-\beta/2} W_{2,\ell}(s_1) - t_2^{-\beta/2} W_{2,\ell}(s_2) \right)^2 &= t_1^{-\beta} s_1 + t_2^{-\beta} s_2 - 2(t_1 t_2)^{-\beta/2} s_2 \\ &= y^{-\beta/a} x^{1/a} + t_2^{-\beta} s_2 - 2y^{-\beta/(2a)} t_2^{-\beta/2} s_2 \\ &= f(x, y), \end{aligned}$$

where $x = s_1^a$, $y = t_1^a$. Note $x \leq y$. Since $0 < a < 1 - \beta$, the mean value theorem applied to $f(x, y)$ at $x_0 = s_2^a$, $y_0 = t_2^a$ gives an x_*, y_* with $y_* \geq x_* \geq s_2^a$ and

$$\begin{aligned} |f(x, y)| &\leq C \left((x_*^{-\beta/a-1} y_*^{1/a} + x_*^{-\beta/(2a)-1} t_2^{-\beta/2} s_2) |x - x_0| + x_*^{-\beta/a} y_*^{1/a-1} |y - y_0| \right) \\ &\leq C \left((y_*^{1/a-\beta/a-1} + s_2^{1-\beta-a}) |x - x_0| + y_*^{1/a-1-\beta/a} |y - y_0| \right) \\ (C.36) \quad &\leq C(|x - x_0| + |y - y_0|) \leq C(|t_1 - t_2|^a + |s_1 - s_2|^a). \end{aligned}$$

Similarly if $s_2 = 0$, we have $\mathbb{E}|t_1^{-\beta/2} W_{2,\ell}(s_1) - t_2^{-\beta/2} W_{2,\ell}(s_2)|^2 = t_1^{-\beta} s_1 \leq s_1^{1-\beta} \leq s^a$, and thus (C.36) holds for all $0 \leq s_i \leq t_i \leq T$. Analogous arguments for (C.36) show $\left(t_1^{-\beta/2} s_1 - t_2^{-\beta/2} s_2 \right)^2 \leq C(|t_1 - t_2|^a + |s_1 - s_2|^a)$, which gives (C.34). □

C.2. **Lemmas under H_A .** The next few lemmas are used under H_A . We first set up some notation. Let

$$\mu_1 = \iint h(\mathbf{x}, \mathbf{y}) dF(\mathbf{x}) dF(\mathbf{y}), \quad \mu_2 = \iint h(\mathbf{x}, \mathbf{y}) dF_*(\mathbf{x}) dF_*(\mathbf{y}),$$

$$\mu_{12} = \iint h(\mathbf{x}, \mathbf{y}) dF(\mathbf{x}) dF_*(\mathbf{y}).$$

$$h_1(\mathbf{x}) = \int h(\mathbf{x}, \mathbf{y}) dF(\mathbf{y}), \quad h_2(\mathbf{x}) = \int h(\mathbf{x}, \mathbf{y}) dF_*(\mathbf{y})$$

Also, with ν_1, ν_2 as in (3.11), we note

$$\nu_1 = \theta^{-1}(\mu_1 - \mu_{12}), \quad \nu_2 = \theta^{-1}(\mu_{12} - \mu_2), \quad \nu_1 - \nu_2 = \theta^{-1}(\mu_1 + \mu_2 - 2\mu_{12}).$$

Whenever convenient we write $\mathbf{X}_i^*$ in place of $\mathbf{X}_i$ for $i > m + k_*$. We also set

$$(C.37) \quad z_i = v(\mathbf{X}_i) - \nu_1, \quad z_i^* = v(\mathbf{X}_i^*) - \nu_2.$$

Below, we set any sum $\sum_{j=a}^b (\dots) = 0$ whenever $b < a$. We proceed to decompose the summations appearing in (2.7) for $k \geq k_* + 1$ into drift, degenerate, and nondegenerate terms. For any $k \geq k_* + 1$, $0 \leq r < k_*$, write

$$\begin{aligned} & \sum_{i=1}^m \sum_{j=m+r+1}^{m+k} h(\mathbf{X}_i, \mathbf{X}_j) - m(k_* - r)\mu_1 - m(k - k_*)\mu_{12} \\ &= R_{m,1}(r, k) + (k_* - r) \sum_{i=1}^m [h_1(\mathbf{X}_i) - \mu_1] + m \sum_{i=m+r+1}^{m+k_*} [h_1(\mathbf{X}_i) - \mu_1] \\ & \quad + (k - k_*) \sum_{i=1}^m [h_2(\mathbf{X}_i) - \mu_{12}] + m \sum_{j=m+k_*+1}^{m+k} [h_1(\mathbf{X}_j^*) - \mu_{12}] \\ &= R_{m,1}(r, k) + (k - r) \sum_{i=1}^m [h_1(\mathbf{X}_i) - \mu_1] + m \sum_{i=m+r+1}^{m+k_*} [h_1(\mathbf{X}_i) - \mu_1] \end{aligned}$$

$$\begin{aligned}
& -\theta(k - k_*) \sum_{i=1}^m z_i + m \sum_{j=m+k_*+1}^{m+k} [h_1(\mathbf{X}_j^*) - \mu_{12}], \\
& = R_{m,1}(r, k) + T_{m,1}(r, k),
\end{aligned}$$

with

$$\begin{aligned}
R_{m,1}(r, k) &= \sum_{i=1}^m \sum_{j=m+r+1}^{m+k_*} [h(\mathbf{X}_i, \mathbf{X}_j) - h_1(\mathbf{X}_i) - h_1(\mathbf{X}_j) + \mu_1] \\
&+ \sum_{i=1}^m \sum_{j=m+k_*+1}^{m+k} [h(\mathbf{X}_i, \mathbf{X}_j^*) - h_2(\mathbf{X}_i) - h_1(\mathbf{X}_j^*) + \mu_{12}] \\
&= R_{m,1,1}(r) + R_{m,1,2}(k_*, k).
\end{aligned} \tag{C.38}$$

When $k_* < r < k$,

$$\begin{aligned}
& \sum_{i=1}^m \sum_{j=m+r+1}^{m+k} h(\mathbf{X}_i, \mathbf{X}_j) - m(k - r)\mu_{12} \\
&= R_{m,1}(r, k) + (k - r) \sum_{i=1}^m (z_i - [h_1(\mathbf{X}_i) - \mu_1]) + m \sum_{j=m+r+1}^{m+k} [h_1(\mathbf{X}_j^*) - \mu_{12}] \\
&= R_{m,1}(r, k) + T_{m,1}(r, k)
\end{aligned}$$

with

$$\begin{aligned}
R_{m,1}(r, k) &= \sum_{i=1}^m \sum_{j=m+r+1}^{m+k} [h(\mathbf{X}_i, \mathbf{X}_j^*) - h_2(\mathbf{X}_i) - h_1(\mathbf{X}_j^*) + \mu_{12}] \\
&= R_{m,1,2}(r, k),
\end{aligned} \tag{C.39}$$

Similarly,

$$\sum_{1 \leq i < j \leq m} h(\mathbf{X}_i, \mathbf{X}_j) - \binom{m}{2} \mu_1 = R_{m,2} + (m - 1) \sum_{i=1}^m [h_1(\mathbf{X}_i) - \mu_1],$$

$$= R_{m,2} + T_{m,2},$$

with

$$(C.40) \quad R_{m,2} = \sum_{1 \leq i < j \leq m} [h(\mathbf{X}_i, \mathbf{X}_j) - h_1(\mathbf{X}_i) - h_1(\mathbf{X}_j) + \mu_1].$$

For the third summation in (2.7), when $0 \leq r \leq k_*$,

$$\begin{aligned} & \sum_{m+r < i < j \leq m+k} h(\mathbf{X}_i, \mathbf{X}_j) - \left[\binom{k_* - r}{2} \mu_1 + \binom{k - k_*}{2} \mu_2 + (k_* - r)(k - k_*) \mu_{12} \right] \\ &= \sum_{m < i < j \leq m+k_*} [h(\mathbf{X}_i, \mathbf{X}_j) - \mu_1] + \sum_{m+k_* < i < j \leq m+k} [h(\mathbf{X}_i^*, \mathbf{X}_j^*) - \mu_2] \\ & \quad + \sum_{i=m+1}^{m+k_*} \sum_{j=m+k_*+1}^{m+k} [h(\mathbf{X}_i, \mathbf{X}_j^*) - \mu_{12}] \\ &= R_{m,3}(r, k) + (k - r - 1) \sum_{i=m+r+1}^{m+k_*} [h_1(\mathbf{X}_i) - \mu_1] - \theta(k - k_* - 1) \sum_{j=m+k_*+1}^{m+k} z_j^* \\ & \quad - \theta(k - k_*) \sum_{i=m+r+1}^{m+k_*} z_i + (k - r - 1) \sum_{j=m+k_*+1}^{m+k} [h_1(\mathbf{X}_j^*) - \mu_{12}] \\ &= R_{m,3}(r, k) + T_{m,3}(r, k), \end{aligned}$$

with

$$\begin{aligned} R_{m,3}(r, k) &= \sum_{m+r < i < j \leq m+k_*} [h(\mathbf{X}_i, \mathbf{X}_j) - h_1(\mathbf{X}_i) - h_1(\mathbf{X}_j) + \mu_1] \\ & \quad + \sum_{m+k_* < i < j \leq m+k} [h(\mathbf{X}_i^*, \mathbf{X}_j^*) - h_2(\mathbf{X}_i^*) - h_2(\mathbf{X}_j^*) + \mu_2] \\ & \quad + \sum_{i=m+r+1}^{m+k_*} \sum_{j=m+k_*+1}^{m+k} [h(\mathbf{X}_i, \mathbf{X}_j^*) - h_2(\mathbf{X}_i) - h_1(\mathbf{X}_j^*) + \mu_{12}] \\ (C.41) \quad &=: R_{m,3,1}(r) + R_{m,3,2}(k_*, k) + R_{m,3,3}(r, k), \end{aligned}$$

and when $k_* < r < k$,

$$\begin{aligned} \sum_{m+r < i < j \leq m+k} h(\mathbf{X}_i, \mathbf{X}_j) - \binom{k-r}{2} \mu_2 &= R_{m,3}(r, k) + (k-r-1) \sum_{j=m+r}^{m+k} [h_2(\mathbf{X}_j^*) - \mu_2] \\ &= R_{m,3}(r, k) + T_{m,3}(r, k), \end{aligned}$$

with

$$\begin{aligned} R_{m,3}(r, k) &= \sum_{m+r < i < j \leq m+k} [h(\mathbf{X}_i^*, \mathbf{X}_j^*) - h_2(\mathbf{X}_i^*) - h_2(\mathbf{X}_j^*) + \mu_2] \\ &= R_{m,3,2}(r, k). \end{aligned}$$

This gives, for $k \geq k_* + 1$,

$$(C.42) \quad (k-r)^2 U_m(h; r, k) = q_1(r, k) + q_2(r, k) + q_3(r, k),$$

with

$$(C.43) \quad q_1(r, k) = p_1(r, k) \mu_1 + p_{12}(r, k) \mu_{12} + p_2(r, k) \mu_2,$$

where

$$\begin{aligned} p_1(r, k) &= \begin{cases} \left(2(k-r)(k_*-r) - (k-r)^2 - \frac{(k-r)(k_*-r)(k_*-r-1)}{k-r-1} \right) & 0 \leq r \leq k_* \\ -(k-r)^2 & k_* < r < k \end{cases} \\ p_{12}(r, k) &= \begin{cases} 2 \left((k-r)(k-k_*) - \frac{(k-r)(k_*-r)(k-k_*)}{k-r-1} \right) & 0 \leq r \leq k_* \\ 2(k-r)^2 & k_* < r < k \end{cases} \\ p_2(r, k) &= \begin{cases} \frac{(k-r)(k-k_*)(k-k_*-1)}{k-r-1} & 0 \leq r \leq k_* \\ -(k-r)^2 & k_* < r < k \end{cases}, \end{aligned}$$

and after some cancellation,

$$\begin{aligned}
q_2(r, k) &= (k-r)^2 \left[\frac{2T_{m,1}(r, k)}{(k-r)m} - \frac{T_{m,2}}{\binom{m}{2}} - \frac{T_{m,3}(r, k)}{\binom{k-r}{2}} \right] \\
\text{(C.44)} \quad &= \begin{cases} 2\theta(k - k_*) \left[-\frac{k-r}{m} \sum_{i=1}^m z_i - \frac{k-r}{k-r-1} \sum_{i=m+r+1}^{m+k_*} z_i \right. \\ \quad \left. - \frac{k-r}{k-r-1} \left(\frac{k-k_*-1}{k-k_*} \right) \sum_{i=m+k_*+1}^{m+k} z_i^* \right], & 0 \leq b \leq k_* \\ 2\theta(k-r) \left[-\frac{k-r}{m} \sum_{i=1}^m z_i + \sum_{i=m+r+1}^{m+k} z_i^* \right] & k_* < b < k. \end{cases}
\end{aligned}$$

Lastly,

$$\text{(C.45)} \quad q_3(r, k) = (k-r)^2 \left[\frac{2R_{m,1}(r, k)}{(k-r)m} - \frac{R_{m,2}}{\binom{m}{2}} - \frac{R_{m,3}(r, k)}{\binom{k-r}{2}} \right].$$

The next lemma provides an approximation of the drift term q_1 and nondegenerate term q_2 by asymptotically equivalent but simpler terms.

Lemma C.6. *Let $y_m > k_*$ be any sequence with $y_m \rightarrow \infty$, and for $1 \leq r \leq k$, set*

$$\begin{aligned}
q_1(r, k) &= -(k - (k_* \vee r))^2 \theta(\nu_1 - \nu_2), \\
\text{(C.46)} \quad q_2(r, k) &= 2\theta(k - (k_* \vee r)) \left[-\frac{k-r}{m} \sum_{i=1}^m z_i + \mathbf{1}_{\{r < k_*\}} \sum_{i=m+r+1}^{m+k_*} z_i + \sum_{i=m+(k_* \vee r)+1}^{m+k} z_i^* \right].
\end{aligned}$$

Then, for $q_1(r, k)$ and $q_2(r, k)$ as in (C.42),

$$\text{(C.47)} \quad \max_{k_* < k \leq y_m} \max_{0 \leq r < k} \frac{|q_1(r, k) - \bar{q}_1(r, k)|}{mg_m(k)} \leq C\theta|\nu_1 - \nu_2| \left(\frac{y - k_*}{k_*} \wedge 1 \right) ((k_*/m)^{1-\beta} \wedge (k_*/m)^{-1}),$$

and for any $\delta > 0$

$$\begin{aligned}
& \max_{k_* < k \leq y_m} \max_{0 \leq r < k} \frac{|q_2(r, k) - \mathfrak{q}_2(r, k)|}{mg_m(k)} \\
\text{(C.48)} \quad & \leq C\theta k_*^{-1} \left((k_*/m)^{1-\beta} \wedge (k_*/m)^{-1} \right) \left((1 - k_*/y) O_P(\sigma k_*^{1/2}) + O_P(\sigma_*(y - k_*)^{1/2+\delta}) \right)
\end{aligned}$$

Proof. The bounds are immediate when $k_* \leq r < k$, so we only consider $0 \leq r < k_*$. Note with $p_i(r, k)$ as in (C.43),

$$\begin{aligned}
p_1(r, k) &= 2(k - r)(k - k_*) - (k - r)^2 - (k_* - r)^2 + \varepsilon_1(k) \\
&= -(k - k_*)^2 + \varepsilon_1(k),
\end{aligned}$$

with

$$\varepsilon_1(r, k) = -\frac{(k_* - r)(k_* - k)}{k - r - 1}.$$

Similarly,

$$\begin{aligned}
p_{12}(r, k) &= 2(k - k_*)^2 - 2\varepsilon_1(r, k), \\
p_2(r, k) &= -(k - k_*)^2 + \varepsilon_1(r, k),
\end{aligned}$$

Hence,

$$\begin{aligned}
\max_{k_* < k \leq y_m} \frac{\max_{0 \leq r < k_*} |\varepsilon_1(r, k)|}{mg_m(k)} &\leq C \frac{1}{mg_m(k_*)} \max_{k_* < k \leq y_m} \varepsilon_1(0, k) \\
&\leq C \left(\frac{y - k_*}{k_*} \wedge 1 \right) \left((k_*/m)^{1-\beta} \wedge (k_*/m)^{-1} \right),
\end{aligned}$$

which gives (C.47). Likewise,

$$\begin{aligned}
q_2(r, k) &= \mathfrak{q}_2(r, k) - 2\theta(k - k_*) \left[\frac{1}{k - r - 1} \sum_{i=m+1}^{m+k_*} z_i + \frac{k_* - r}{(k - r - 1)(k - k_*)} \sum_{i=m+k_*+1}^{m+k} z_i^* \right] \\
&= \mathfrak{q}_2(r, k) + 2\theta\varepsilon_2(r, k),
\end{aligned}$$

and

$$\begin{aligned} \max_{k_* < k \leq y_m} \frac{\max_{0 \leq r < k} |\varepsilon_2(k)|}{mg_m(k)} &\leq \max_{k_* < k \leq y_m} \frac{C}{mg_m(k)} \left(\frac{k - k_*}{k} \left| \sum_{i=m+1}^{m+k_*} z_i \right| + \frac{k_*}{k} \left| \sum_{i=m+k_*+1}^{m+k} z_i^* \right| \right) \\ &\leq \frac{C\theta}{mg_m(k_*)} \left((1 - k_*/y) O_P(\sigma k_*^{1/2}) + O_P(\sigma_*(y - k_*)^{1/2+\delta}) \right), \end{aligned}$$

which gives (C.48). □

The next few lemmas concern bounds and approximations for q_3 , under H_A .

Lemma C.7. *With $q_3(r, k)$ as in (C.45), for any sequence $y_m \geq k_*$ with $y_m \rightarrow \infty$,*

$$\max_{k_* < k \leq y_m} \max_{0 \leq r < k} \frac{|q_3(r, k)|}{mg_m(k)} = O_P \left(((y/m)^{2(1-\beta)} \log^4(y)) \wedge 1 \right),$$

and

$$\max_{k \geq k_*} \max_{0 \leq r < k} \frac{|q_3(r, k)|}{mg_m(k)} = O_P(1).$$

Proof. Write

$$\begin{aligned} \frac{q_3(r, k)}{mg_m(k)} &= \frac{(k - r)^2}{mg_m(k)} \left[\frac{2R_{m,1}(r, k)}{(k - r)m} - \binom{m}{2}^{-1} R_{m,2} - \binom{k - r}{2}^{-1} R_{m,3}(r, k) \right] \\ &= A_1(r, k) - A_2(r, k) - A_3(r, k). \end{aligned}$$

It suffices to establish

$$(C.49) \quad \max_{k_* < k \leq y_m} \max_{0 \leq r < k} |A_i(r, k)| = O_P \left(((y/m)^{2(1-\beta)} \log^4(y)) \wedge 1 \right),$$

$$(C.50) \quad \max_{k \geq k_*} \max_{0 \leq r < k} |A_i(r, k)| = O_P(1),$$

for $i = 1, 2, 3$. For brevity we consider only $i = 3$ since $i = 1, 2$ are essentially the same but simpler. Write

$$\begin{aligned}
\bar{h}_{11}(\mathbf{x}, \mathbf{y}) &= h(\mathbf{x}, \mathbf{y}) - h_1(\mathbf{x}) - h_1(\mathbf{y}) + \mu_1, \\
\bar{h}_{22}(\mathbf{x}, \mathbf{y}) &= h(\mathbf{x}, \mathbf{y}) - h_2(\mathbf{x}) - h_2(\mathbf{y}) + \mu_2, \\
\bar{h}_{12}(\mathbf{x}, \mathbf{y}) &= h(\mathbf{x}, \mathbf{y}) - h_1(\mathbf{x}) - h_2(\mathbf{y}) + \mu_{12},
\end{aligned}
\tag{C.51}$$

So that

$$R_{m,3}(r, k) = R_{m,3,1}(r)\mathbf{1}_{\{r \leq k_*\}} + R_{m,3,2}(r \vee k_*, k) + R_{m,3,3}(r, k)\mathbf{1}_{\{r \leq k_*\}},$$

with

$$\begin{aligned}
R_{m,3,1}(r) &= \sum_{m+r < i < j \leq m+k_*} \bar{h}_{11}(\mathbf{X}_i, \mathbf{X}_j), & R_{m,3,2}(r, k) &= \sum_{m+r < i < j \leq m+k} \bar{h}_{22}(\mathbf{X}_i^*, \mathbf{X}_j^*), \\
R_{m,3,3}(r, k) &= \sum_{i=m+r+1}^{m+k_*} \sum_{j=m+k_*+1}^{m+k} \bar{h}_{12}(\mathbf{X}_i, \mathbf{X}_j^*).
\end{aligned}$$

For $R_{m,3,1}(r)$, note

$$\mathbb{E}[R_{m,3,1}(r-1) - R_{m,3,1}(r) | \sigma(\mathbf{X}_{m+r+1}, \dots, \mathbf{X}_{m+k_*})] = 0.$$

Hence, for each fixed m , $M(r) = R_{m,3,1}(-r)$ for $-k_* < r \leq 0$ is a martingale with respect to the filtration $\mathcal{G}_r = \sigma(\mathbf{X}_{m+k_*}, \dots, \mathbf{X}_{m-r+1})$; Doob's maximal inequality gives

$$\mathbb{E} \left(\max_{0 \leq r \leq k_*} R_{m,3,1}(r) \right)^2 \leq 4 \mathbb{E} (R_{m,3,1}(0))^2 \leq k_*^2 \mathbb{E} \bar{h}_{11}^2(\mathbf{X}, \mathbf{Y}).$$

Now, since $g_m(k) \geq C(k/m)^\beta$, we have

$$\begin{aligned}
(C.52) \quad P \left\{ \max_{k_* < k \leq y_m} \max_{0 \leq r < k_*} \frac{1}{mg_m(k)} |R_{m,3,1}(r)| > x \right\} &\leq P \left\{ \max_{0 \leq r < k_*} \frac{y^{-\beta}}{m^{1-\beta}} |R_{m,3,1}(r)| > Cx \right\} \\
&\leq Cx^{-2} \frac{k_*^2 y^{-2\beta}}{m^{2(1-\beta)}}.
\end{aligned}$$

On the other hand, since $g_m(k) \geq C((k/m)^\beta \mathbf{1}_{\{k \leq m\}} + (k/m)^2 \mathbf{1}_{\{k > m\}})$, it follows that

$$\begin{aligned}
(C.53) \quad P \left\{ \sup_{k \geq k_*} \max_{0 \leq r < k_*} \frac{1}{mg_m(k)} |R_{m,3,1}(r)| > x \right\} \\
\leq P \left\{ \max_{0 \leq r < k_*} \left(\frac{m}{k_*^2} \mathbf{1}_{\{k_* > m\}} + \mathbf{1}_{\{k_* \leq m\}} \frac{m^{\beta-1}}{k_*^\beta} \right) |R_{m,3,1}(r)| > Cx \right\} \\
(C.54) \quad \leq Cx^{-2}.
\end{aligned}$$

Now, for $R_{m,3,2}(r, k)$, suppose first $k_* \leq y \leq Cm$. Using Lemma C.8, we have

$$\begin{aligned}
(C.55) \quad P \left\{ \max_{k_* < k \leq y_m} \max_{0 \leq r < k} \frac{1}{mg_m(k)} |R_{m,3,2}(r \vee k_*, k)| > x \right\} \\
\leq P \left\{ \max_{k_* < k \leq y_m} \frac{k^{-\beta}}{m^{1-\beta}} \max_{k_* \leq r < k} |R_{m,3,2}(r, k)| > Cx \right\} \\
\leq P \left\{ \max_{\lfloor \log(k_*) \rfloor \leq q \leq \lceil \log y \rceil} \max_{e^{q-1} \leq k < e^q} \max_{k_* \leq r < k} \frac{e^{-\beta(q-1)}}{m^{1-\beta}} |R_{m,3,2}(r, k)| > Cx \right\} \\
\leq Cx^{-2} \sum_{q=\lfloor \log k_* \rfloor + 1}^{\lceil \log y \rceil} \frac{e^{-2\beta q}}{m^{2(1-\beta)}} (e^q - k_*)^2 \log^4(e^q - k_*) \\
(C.56) \quad \leq Cx^{-2} \frac{y^{2(1-\beta)}}{m^{2(1-\beta)}} \log^4(y - k_*).
\end{aligned}$$

On the other hand, if $y > Cm$, since $g_m(k) \geq C(k/m)^2$ for $k \geq m$, we have

$$\begin{aligned}
P \left\{ \max_{m \leq k \leq y} \frac{1}{mg_m(k)} \max_{k_* \leq r < k} |R_{m,3,2}(r, k)| > x \right\} \\
\leq P \left\{ \max_{m \leq k \leq y} mk^{-2} \max_{k_* \leq r < k} |R_{m,3,2}(r, k)| > Cx \right\}
\end{aligned}$$

$$\begin{aligned}
&\leq P \left\{ \max_{\lfloor \log(m) \rfloor < q \leq \lceil \log y \rceil} \max_{e^{q-1} \leq k < e^q} m e^{-2(q-1)} \max_{k_* \leq r < k} |R_{m,3,2}(r, k)| > Cx \right\} \\
&\leq Cx^{-2} \sum_{q=\lfloor \log m \rfloor + 1}^{\lceil \log(y) \rceil} m^2 e^{-2q} \log^4(e^q) \\
(C.57) \quad &\leq Cx^{-2},
\end{aligned}$$

which, combined with (C.56), gives (C.49). Likewise, analogous steps leading to (C.57) show

$$\max_{k \geq k_*} \max_{0 \leq r < k} \frac{1}{mg_m(k)} |R_{m,3,2}(r \vee k_*, k)| = O_P(1).$$

Repeating the above arguments *mutatis mutandis* for $R_{m,3,3}(r, k)$ then gives the claim. $\square$

Lemma C.8. *Let $\bar{h}_{12}$ and $\bar{h}_{22}$ be as in (C.51). Then for any $x > 0$, $y \geq k_* + 2$,*

(C.58)

$$P \left\{ \max_{k_* < k \leq y} \max_{k_* \leq r < k} \left| \sum_{i=1}^m \sum_{j=m+r}^{m+k} \bar{h}_{12}(\mathbf{X}_i, \mathbf{X}_j^*) \right| > x \right\} \leq Cx^{-2}m(y - k_*),$$

$$(C.59) \quad \mathbb{E} \max_{k_* < k \leq y} \max_{k_* \leq r < k} \left| \sum_{m+r < i < j \leq m+k} \bar{h}_{22}(\mathbf{X}_i^*, \mathbf{X}_j^*) \right|^2 \leq C(y - k_*)^2 \log^4(y - k_*).$$

(C.60)

$$\mathbb{E} \max_{0 \leq r < k_*} \max_{k_* < k \leq y} \left| \sum_{i=m+r+1}^{m+k_*} \sum_{j=m+k_*+1}^{m+k} \bar{h}_{12}(\mathbf{X}_i, \mathbf{X}_j^*) \right|^2 \leq Cx^{-2}k_*(y - k_*) \log^2(k_*) \log^2(y - k_*)$$

Proof. We have

$$\begin{aligned}
\max_{k_* < k \leq y} \max_{k_* \leq r < k} \left| \sum_{i=1}^m \sum_{j=m+r}^{m+k} \tilde{h}(\mathbf{X}_i, \mathbf{X}_j^*) \right| &= \max_{k_* < k \leq y} \max_{k_* \leq r < k} \left| \sum_{i=1}^m \left(\sum_{j=m+k_*+1}^{m+k} - \sum_{j=m+k_*+1}^{m+r-1} \right) \tilde{h}(\mathbf{X}_i, \mathbf{X}_j^*) \right| \\
&\leq 2 \max_{k_* < k \leq y} \left| \sum_{i=1}^m \sum_{j=m+k_*+1}^{m+k} \tilde{h}(\mathbf{X}_i, \mathbf{X}_j^*) \right|.
\end{aligned}$$

Conditioning on $\mathbf{X}_1, \dots, \mathbf{X}_m$ and applying Kolmogorov's maximal inequality yields

$$P \left\{ \max_{k_* < k \leq y} \left| \sum_{i=1}^m \sum_{j=m+k_*+1}^{m+k} \tilde{h}(\mathbf{X}_i, \mathbf{X}_j^*) \right| > x \right\} \leq C(y - k_*)^2,$$

from which we deduce (C.58). For (C.59), note

$$\begin{aligned} & \max_{k_* < k \leq y} \max_{k_* \leq r < k} \left| \sum_{m+r < i < j \leq m+k} \bar{h}_*(\mathbf{X}_i^*, \mathbf{X}_j^*) \right| \\ &= \frac{1}{2} \max_{k_* < k \leq y} \max_{k_* \leq r < k} \left| \left(\sum_{i,j=m+k_*+1}^{m+k} + \sum_{i,j=m+k_*+1}^{m+r} - 2 \sum_{i=m+k_*+1}^{m+r} \sum_{j=m+k_*+1}^{m+k} \right) \bar{h}_*(\mathbf{X}_i^*, \mathbf{X}_j^*) \mathbf{1}_{\{i \neq j\}} \right| \\ &\leq \max_{k_* < k \leq y} \left| \sum_{m+k_* < i < j \leq m+k} \bar{h}_*(\mathbf{X}_i^*, \mathbf{X}_j^*) \right| + \max_{k_* < k \leq y} \max_{k_* < r \leq k} \left| \sum_{i=m+k_*+1}^{m+r} \sum_{j=m+k_*+1}^{m+k} \bar{h}_*(\mathbf{X}_i^*, \mathbf{X}_j^*) \mathbf{1}_{\{i \neq j\}} \right|. \end{aligned}$$

The bound for the first term on the last line above can be argued as in Lemma C.1 so we proceed to bound the second term. For each r, k we may write

$$\sum_{i=m+k_*+1}^{m+r} \sum_{j=m+k_*+1}^{m+k} \bar{h}_*(\mathbf{X}_i^*, \mathbf{X}_j^*) \mathbf{1}_{\{i \neq j\}} = \sum_{i=1}^a \sum_{j=1}^b \xi_{i,j},$$

where $\xi_{i,j} = \bar{h}_*(\mathbf{X}_{m+k_*+i}^*, \mathbf{X}_{m+k_*+j}^*)$ for $i \neq j$, $\xi_{j,j} = 0$, and $1 \leq a \leq b \leq y - k_*$. Note $\mathbb{E} \xi_{i,j} \xi_{i',j'} = 0$ whenever $(i, j) \neq (i', j')$.

We adapt the argument in (Doob, 1990, p.156). Let r be an integer such that $2^r < (y - k_*) \leq 2^{r+1}$; for convenience set $\xi_{i,j} = 0$ if $i \wedge j \geq (y - k_*)$. Let $\mathcal{S}$ be the sum of all (squared) partial sums of the form

$$\left(\sum_{i=\alpha_1}^{\beta_1} \sum_{j=\alpha_2}^{\beta_2} \xi_{i,j} \right)^2,$$

where for $i = 1, 2$,

$$\alpha_i = 2^{v_i}k, \quad \beta_i = 2^{v_i}(k+1), \quad k = 0, \dots, 2^{r-v_i+1} - 1,$$

$$v_i = 0, 1, \dots, r+1.$$

Then for each fixed pair (v_i, v_j) , the sum of all terms entering into $\mathcal{S}$ corresponding to v_i, v_j have expectation bounded by $\sum_{i,j=1}^{y-k_*} \mathbb{E} y_{i,j}^2 \leq C(y-k_*)^2$, so that

$$\mathbb{E}\mathcal{S} \leq C(r+2)^2(y-k_*)^2.$$

Then by considering the binary expansions of a, b we can write

$$\sum_{i=1}^a \sum_{j=1}^b \xi_{i,j} = \sum_{i=1}^{a'} \sum_{j=1}^{b'} \eta_{i,j}$$

where $a', b' \leq r+2$, and for each i, j ,

$$\eta_{i,j} = \sum_{\ell=\alpha_{1,i}}^{\beta_{1,i}} \sum_{m=\alpha_{2,i}}^{\beta_{2,i}} \xi_{\ell,m},$$

with $\beta_{1,i} - \alpha_{1,i} = 2^{r_{1,i}}$, $\beta_{2,j} - \alpha_{2,j} = 2^{r_{2,j}}$, $(r+1) \geq r_{1,1} > r_{1,2} > \dots > r_{1,a'} \geq 0$ and $(r+1) \geq r_{2,1} > \dots > r_{2,b'} \geq 0$. Then Cauchy-Schwarz gives

$$\left(\sum_{i=1}^{a'} \sum_{j=1}^{b'} \eta_{i,j} \right)^2 \leq a'b' \sum_{i=1}^{a'} \sum_{j=1}^{b'} \eta_{i,j}^2 \leq (r+2)^2 \sum_{i=1}^{a'} \sum_{j=1}^{b'} \eta_{i,j}^2 \leq (r+2)^2 \mathcal{S}.$$

Finally, we obtain

$$\mathbb{E} \max_{1 \leq a, b \leq (y-k_*)} \left(\sum_{i=1}^a \sum_{j=1}^b \xi_{i,j} \mathbf{1}_{\{i \neq j\}} \right)^2 \leq (r+2)^2 \mathbb{E}\mathcal{S}$$

$$\leq C(r+2)^4(y-k_*)^2 \leq C \log^4(y-k_*)(y-k_*)^2.$$

This gives (C.59). The argument for (C.60) is essentially the same so it is omitted. $\square$

Lemma C.9. Suppose $k_* = c_*m$. With $R_{m,1,i}, R_{m,2,i}$ and $R_{m,3,i}$ as in (C.38), (C.40), and (C.41), respectively, let

$$(C.61) \quad q_3(r, k) = (k - r)^2 \left(\frac{2\mathbf{1}_{\{r \leq k_*\}} R_{m,1,1}(r)}{(k - r)m} - \frac{2R_{m,2}}{m(m - 1)} - \frac{2\mathbf{1}_{\{r \leq k_*\}} R_{m,3,1}(r)}{(k - r)(k - r - 1)} \right).$$

Then, for any $T > 0$, and $0 < \delta < 1$,

$$\begin{aligned} & \max_{k_* < k \leq k_* + Tm^{1-\delta}} \max_{0 \leq r < k} \left| \frac{q_3(r, k)}{mg_m(k)} - \frac{q_3(r, k)}{mg_m(k)} \right| \\ &= 2 \max_{k_* < k \leq k_* + Tm^{1-\delta}} \max_{0 \leq r < k} \frac{(k - r)^2}{mg_m(k)} \left| \frac{R_{m,1,2}(r \vee k_*, k)}{(k - r)m} - \frac{R_{m,3,2}(r \vee k_*, k) + \mathbf{1}_{\{r \leq k_*\}} R_{m,3,3}(r, k)}{(k - r)(k - r - 1)} \right| \\ &= o_P(1). \end{aligned}$$

Proof. We treat each of the terms $R_{m,1,2}(r, k)$, $R_{m,3,2}(r, k)$, and $R_{m,3,3}(r, k)$ separately. Since $g_m(k) \geq C(k/m)^2$ for all $k > k_*$, using Lemma C.8 we get

$$\begin{aligned} & \mathbb{E} \max_{k_* < k \leq k_* + Tm^{1-\delta}} \max_{0 \leq r < k} \left| \frac{(k - r)}{m^2 g_m(k)} R_{m,1,2}(k_* \vee r, k) \right|^2 \\ & \leq Ck_*^{-2} \mathbb{E} \max_{k_* < k \leq k_* + Tm^{1-\delta}} \max_{k_* \leq r < k} |R_{m,1,2}(r, k)|^2 \\ & = Cm^{-\delta}. \end{aligned}$$

Similarly, again using Lemma C.8,

$$\begin{aligned} & \mathbb{E} \max_{k_* < k \leq k_* + Tm^{1-\delta}} \max_{0 \leq r < k} \left| \frac{1}{mg_m(k)} R_{m,3,2}(r \vee k_*, k) \right|^2 \\ & \leq Cm^2 k_*^{-4} \mathbb{E} \max_{k_* \leq k \leq k_* + Tm^{1-\delta}} \max_{k_* \leq r < k} |R_{m,3,2}(r, k)|^2 \\ & \leq Cm^{-2\delta} \log^4(m). \end{aligned}$$

Again applying Lemma C.8 we obtain

$$\mathbb{E} \max_{k_* < k \leq k_* + Tm^{1-\delta}} \max_{0 \leq r < k} \left| \frac{1}{mg_m(k)} R_{m,3,3}(r, k) \right|^2 \leq Cm^{-\delta} \log^4(m).$$

□

C.3. Lemmas for Section 4. We conclude this section with a set of lemmas which will be used for the proofs of the results in Section 4.

Lemma C.10. *If Assumptions 2.1–2.3 hold, then we have*

$$\max_{1 \leq k \leq n} \left(\sum_{i=1}^k \sum_{j=k+1}^m \sum_{\ell=K}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right)^2 \leq cn(m-n) \sum_{\ell=K}^{\infty} \lambda_{\ell}^2,$$

$$\max_{1 \leq k \leq n} \left(\sum_{1 \leq i \neq j \leq k} \sum_{\ell=K}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right)^2 \leq cn^2 \sum_{\ell=K}^{\infty} \lambda_{\ell}^2,$$

and

$$\max_{1 \leq k \leq n} \left(\sum_{k+1 \leq i \neq j \leq m} \sum_{\ell=K}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right)^2 \leq c(m-n)^2 \sum_{\ell=K}^{\infty} \lambda_{\ell}^2,$$

for all $K \geq 1$ and $2 \leq n \leq m-2$.

Proof. The argument is essentially the same as in Lemma C.1, so it is omitted. □

Let

$$\begin{aligned} \mathfrak{R}_K(k) = & \frac{2}{k(m-k)} \sum_{i=1}^k \sum_{j=k+1}^m \sum_{\ell=K}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) - \frac{1}{k(k-1)} \sum_{1 \leq i \neq j \leq k} \sum_{\ell=K}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \\ & - \frac{1}{(m-k)(m-k-1)} \sum_{k+1 \leq i \neq j \leq m} \sum_{\ell=K}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j). \end{aligned}$$

Lemma C.11. *If Assumptions 2.1–2.3 hold, then we have*

$$P \left\{ \max_{2 \leq k \leq m-2} \frac{1}{\mathfrak{q}(k/m)} \frac{k^2(m-k)^2}{m^3} |\mathfrak{R}_K(k)| > x \right\} \leq \frac{c}{x^2} \sum_{\ell=K}^{\infty} \lambda_{\ell}^2,$$

for all $x > 0$ and $K \geq 1$.

Proof. We note

$$\begin{aligned} \frac{k^2(m-k)^2}{m^3} |\Re_K(k)| &\leq \frac{2k(m-k)}{m^3} \left| \sum_{i=1}^k \sum_{j=k+1}^m \sum_{\ell=K}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right| \\ &\quad + \frac{2(m-k)^2}{m^3} \left| \sum_{1 \leq i \neq j \leq m} \sum_{\ell=K}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right| \\ &\quad + \frac{2k^2}{m^3} \left| \sum_{k+1 \leq i \neq j \leq m} \sum_{\ell=K}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right|. \end{aligned}$$

Lemma C.1 yields via Markov's inequality that

$$\begin{aligned} &P \left\{ \max_{2 \leq k \leq m-2} \left(\frac{m}{k} \right)^{\zeta} \frac{k(m-k)}{m^3} \left| \sum_{i=1}^k \sum_{j=k+1}^m \sum_{\ell=K}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right| > x \right\} \\ &\leq P \left\{ \max_{2 \leq k \leq m-2} k^{1-\zeta} \left| \sum_{i=1}^k \sum_{j=k+1}^m \sum_{\ell=K}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right| > xm^{2-\zeta} \right\} \\ &\leq \sum_{z=1}^{\log m+1} P \left\{ \max_{e^{z-1} \leq k \leq e^z} k^{1-\zeta} \left| \sum_{i=1}^k \sum_{j=k+1}^m \sum_{\ell=K}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right| > xm^{2-\zeta} \right\} \\ &\leq \sum_{z=1}^{\log m+1} P \left\{ \max_{e^{z-1} \leq k \leq e^z} \left| \sum_{i=1}^k \sum_{j=k+1}^m \sum_{\ell=K}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right| > xm^{2-\zeta} e^{(z-1)(\zeta-1)} \right\} \\ &\leq \frac{c}{x^2} m^{2\zeta-4} \sum_{z=1}^{\log(m-2)} e^{2z(1-\zeta)} e^{z\zeta} (m - e^{z\zeta}) \sum_{\ell=k}^{\infty} \lambda_{\ell}^2 \\ &\leq \frac{c}{x^2} \sum_{\ell=K}^{\infty} \lambda_{\ell}^2. \end{aligned}$$

Similar arguments yield

$$P \left\{ \max_{2 \leq k \leq m-2} \left(\frac{m}{k} \right)^{\zeta} \frac{(m-k)^2}{m^3} \left| \sum_{1 \leq i \neq j \leq k} \sum_{\ell=K}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right| > x \right\}$$

$$\begin{aligned}
&\leq P \left\{ \max_{2 \leq k \leq m-2} k^{-\zeta} \left| \sum_{1 \leq i \neq j \leq m} \sum_{\ell=K}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right| > x m^{1-\zeta} \right\} \\
&\leq \sum_{z=1}^{\log m+1} P \left\{ \max_{e^{z-1} \leq k \leq e^z} k^{-\zeta} \left| \sum_{1 \leq i \neq j \leq m} \sum_{\ell=K}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right| > x m^{1-\zeta} \right\} \\
&\leq \sum_{z=1}^{\log m+1} P \left\{ \max_{e^{z-1} \leq k \leq e^z} \left| \sum_{1 \leq i \neq j \leq m} \sum_{\ell=K}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right| > x m^{1-\zeta} e^{(z-1)\zeta} \right\} \\
&\leq \frac{C}{x^2} m^{2\zeta-2} \sum_{z=1}^{\log m} e^{2z} e^{-2z\zeta} \sum_{\ell=K}^{\infty} \lambda_{\ell}^2 \\
&\leq \frac{C}{x^2} \sum_{\ell=K}^{\infty} \lambda_{\ell}^2,
\end{aligned}$$

and

$$\begin{aligned}
&P \left\{ \max_{2 \leq k \leq m-2} \left(\frac{m}{k} \right)^{\zeta} \frac{k^2}{m^3} \left| \sum_{k+1 \leq i \neq j \leq m} \sum_{\ell=K}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right| > x \right\} \\
&= P \left\{ \max_{2 \leq k \leq m-2} k^{2-\zeta} \left| \sum_{k+1 \leq i \neq j \leq m} \sum_{\ell=K}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right| > x m^{3-\zeta} \right\} \\
&\leq \sum_{z=1}^{\log m+1} P \left\{ \max_{e^{z-1} \leq k \leq e^z} \left| \sum_{k+1 \leq i \neq j \leq m} \sum_{\ell=K}^{\infty} \lambda_{\ell} \phi_{\ell}(\mathbf{X}_i) \phi_{\ell}(\mathbf{X}_j) \right| > x m^{3-\zeta} e^{(z-1)(\zeta-2)} \right\} \\
&\leq \frac{C}{x^2} m^{2\zeta-6} \sum_{z=1}^{\log m+1} (m - e^{\zeta-1})^2 e^{-2z(\zeta-2)} \sum_{\ell=K}^{\infty} \lambda_{\ell}^2 \\
&\leq \frac{C}{x^2} \sum_{\ell=K}^{\infty} \lambda_{\ell}^2.
\end{aligned}$$

By symmetry, we have the same inequalities when $(k/m)^{\zeta}$ is replaced with $(1 - k/m)^{\zeta}$ above.

Hence the proof of Lemma C.11 is proven. $\square$

According to Lemma C.11 it is enough to study

$$\begin{aligned}\mathfrak{R}_K(k) &= \frac{2}{k(m-k)} \sum_{i=1}^k \sum_{j=k+1}^m \sum_{\ell=1}^K \lambda_\ell \phi_\ell(\mathbf{X}_i) \phi_\ell(\mathbf{X}_j) - \frac{1}{k(k-1)} \sum_{1 \leq i \neq j \leq k} \sum_{\ell=1}^K \lambda_\ell \phi_\ell(\mathbf{X}_i) \phi_\ell(\mathbf{X}_j) \\ &\quad - \frac{1}{(m-k)(m-k-1)} \sum_{k+1 \leq i \neq j \leq m} \sum_{\ell=1}^K \lambda_\ell \phi_\ell(\mathbf{X}_i) \phi_\ell(\mathbf{X}_j)\end{aligned}$$

for all $K \geq 1$. Let

$$S_{\ell,3}(k) = \sum_{i=1}^k \phi_\ell(\mathbf{X}_i).$$

Elementary algebra yields

$$\begin{aligned}\mathfrak{R}_K(k) &= - \sum_{\ell=1}^K \lambda_\ell \left(\frac{S_\ell(k)}{k} - \frac{S_\ell(m) - S_\ell(k)}{m-k} \right)^2 + \frac{1}{k^2} \sum_{i=1}^k \sum_{\ell=1}^K \lambda_\ell \phi_\ell^2(\mathbf{X}_i) + \frac{1}{(m-k)^2} \sum_{j=k+1}^m \sum_{\ell=1}^K \lambda_\ell \phi_\ell^2(\mathbf{X}_j) \\ &\quad + \frac{1}{k^2(k-1)} \sum_{1 \leq i \neq j \leq k} \sum_{\ell=1}^K \lambda_\ell \phi_\ell(\mathbf{X}_i) \phi_\ell(\mathbf{X}_j) \\ &\quad + \frac{1}{(m-k)^2(m-k-1)} \sum_{k+1 \leq i \neq j \leq m} \sum_{\ell=1}^K \lambda_\ell \phi_\ell(\mathbf{X}_i) \phi_\ell(\mathbf{X}_j) \\ &= - \frac{m^2}{k^2(m-k)^2} \sum_{\ell=1}^K \lambda_\ell \left[\left(S_\ell(k) - \frac{k}{m} S_\ell(m) \right)^2 - \frac{k(m-k)}{m} \right] + \frac{1}{k^2} \sum_{\ell=1}^K \lambda_\ell \sum_{i=1}^k [\phi_\ell^2(\mathbf{X}_i) - 1] \\ &\quad + \frac{1}{(m-k)^2} \sum_{\ell=1}^K \lambda_\ell \sum_{i=k+1}^m [\phi_\ell^2(\mathbf{X}_i) - 1] + \frac{1}{k^2(k-1)} \sum_{1 \leq i \neq j \leq k} \sum_{\ell=1}^K \lambda_\ell \phi_\ell(\mathbf{X}_i) \phi_\ell(\mathbf{X}_j) \\ &\quad + \frac{1}{(m-k)^2(m-k-1)} \sum_{k+1 \leq i \neq j \leq m} \sum_{\ell=1}^K \lambda_\ell \phi_\ell(\mathbf{X}_i) \phi_\ell(\mathbf{X}_j).\end{aligned}$$

For every fixed integer $2 < a < m$ and $1 \leq \ell \leq K$ and

$$\begin{aligned}&\max_{2 \leq k \leq m-2} \left(\frac{k}{m} \frac{m-k}{m} \right)^{-\zeta} \frac{k^2(m-k)^2}{m^3} \frac{1}{k^2} \left| \sum_{i=1}^k [\phi_\ell^2(\mathbf{X}_i) - 1] \right| \\ &\leq \max_{2 \leq k \leq a} \left(\frac{k}{m} \frac{m-k}{m} \right)^{-\zeta} \frac{k^2(m-k)^2}{m^3} \frac{1}{k^2} \left| \sum_{i=1}^k [\phi_\ell^2(\mathbf{X}_i) - 1] \right|\end{aligned}$$

$$+ \max_{a \leq k \leq m-2} \left(\frac{k}{m} \frac{m-k}{m} \right)^{-\zeta} \frac{k^2(m-k)^2}{m^3} \frac{1}{k^2} \left| \sum_{i=1}^k [\phi_\ell^2(\mathbf{X}_i) - 1] \right|.$$

Lemma C.12. *If Assumptions 2.1–2.3 hold, then we have*

$$(C.62) \quad \max_{2 \leq k \leq m-2} \frac{1}{\mathbf{q}(k/m)} \frac{k^2(m-k)^2}{m^3} \frac{1}{k^2} \left| \sum_{i=1}^k [\phi_\ell^2(\mathbf{X}_i) - 1] \right| = o_P(1),$$

$$(C.63) \quad \max_{2 \leq k \leq m-2} \frac{1}{\mathbf{q}(k/m)} \frac{k^2(m-k)^2}{m^3} \frac{1}{(m-k)^2} \left| \sum_{i=k+1}^m [\phi_\ell^2(\mathbf{X}_i) - 1] \right| = o_P(1),$$

$$(C.64) \quad \max_{2 \leq k \leq m-2} \frac{1}{\mathbf{q}(k/m)} \frac{k^2(m-k)^2}{m^3} \frac{1}{k^3} \left| \sum_{1 \leq i \neq j \leq k} \phi_\ell(\mathbf{X}_i) \phi_\ell(\mathbf{X}_j) \right| = o_P(1),$$

$$(C.65) \quad \max_{2 \leq k \leq m-2} \frac{1}{\mathbf{q}(k/m)} \frac{k^2(m-k)^2}{m^3} \frac{1}{(m-k)^3} \left| \sum_{k+1 \leq i \neq j \leq m} \phi_\ell(\mathbf{X}_i) \phi_\ell(\mathbf{X}_j) \right| = o_P(1).$$

Proof. For every fixed $2 < a < m$

$$\max_{2 \leq k \leq a} \left(\frac{k}{m} \frac{m-k}{m} \right)^{-\zeta} \frac{k^2(m-k)^2}{m^3} \frac{1}{k^2} \left| \sum_{i=1}^k [\phi_\ell^2(\mathbf{X}_i) - 1] \right| = O_P(m^{-1+\zeta}) = o_P(1), \quad \text{as } m \rightarrow \infty.$$

Also,

$$\begin{aligned} & \max_{a \leq k \leq m-2} \left(\frac{k}{m} \frac{m-k}{m} \right)^{-\zeta} \frac{k^2(m-k)^2}{m^3} \frac{1}{k^2} \left| \sum_{i=1}^k [\phi_\ell^2(\mathbf{X}_i) - 1] \right| \\ & \leq \max_{a \leq k \leq m-2} \left(\frac{k}{m} \frac{m-k}{m} \right)^{-\zeta} \frac{k^2(m-k)^2}{m^3} \frac{1}{k} \max_{a \leq k \leq m-2} \frac{1}{k} \left| \sum_{i=1}^k [\phi_\ell^2(\mathbf{X}_i) - 1] \right| \end{aligned}$$

and by the law of large numbers

$$\max_{a \leq k < \infty} \frac{1}{k} \left| \sum_{i=1}^k [\phi_\ell^2(\mathbf{X}_i) - 1] \right| \xrightarrow{P} 0.$$

Hence the proof of (C.62) is complete. By symmetry, (C.62) implies (C.63). Finally, arguing as before, but using Lemma C.10 instead of the law of large numbers, one could verify (C.64) and (C.65). $\square$

Lemma C.13. *If Assumptions 2.1–2.3 hold, then*

$$\begin{aligned} & \left\{ \frac{1}{\mathbf{q}^{1/2}(t)} \frac{1}{m^{1/2}} \left(S_\ell(mt) - \frac{t}{m} S_\ell(m) \right), 0 \leq t \leq 1, 1 \leq \ell \leq K \right\} \\ & \Rightarrow \left\{ \frac{1}{\mathbf{q}^{1/2}(t)} B_\ell(t), 0 \leq t \leq 1, 1 \leq \ell \leq K \right\}, \quad \text{in } \mathbf{D}^K[0, 1] \end{aligned}$$

where $\{B_1(t), 0 \leq t \leq 1\}, \dots, \{B_K(t), 0 \leq t \leq 1\}$ are independent Brownian bridges.

Proof. The result is taken from Chapter 1 of Horváth and Rice [Horváth and Rice \(2024\)](#). $\square$

D. MAIN PROOFS

Proof of Theorem 3.1. Recall $\mathcal{D}_m^{(1)}(k) = m^{-1}k^2|U_m(h; k)|$, (C.7), and (C.8). From (C.9), we have

$$\begin{aligned}
 & m^{-1}k^2U_{m,L}(h; k) \\
 &= \sum_{\ell=1}^L m^{-1}k^2\lambda_\ell U_m(f_\ell; 0, k) \\
 (D.1) \quad &= -\sum_{\ell=1}^L \lambda_\ell \left(\frac{1}{m} \left(S_\ell(k, m) - \frac{k}{m} S_\ell(m) \right)^2 - \frac{k(k+m)}{m^2} \right) + \sum_{\ell=1}^L \lambda_\ell R_\ell(k, k, m).
 \end{aligned}$$

For each real number $t \geq 2/m$, let

$$(D.2) \quad \mathcal{U}_m(t) = \frac{m^{-1}\lfloor mt \rfloor^2 U_m(h; \lfloor mt \rfloor)}{g_m(\lfloor mt \rfloor)}, \quad \mathcal{U}_{m,L}(t) = \frac{m^{-1}\lfloor mt \rfloor^2 U_{m,L}(h; \lfloor mt \rfloor)}{g_m(\lfloor mt \rfloor)} = \frac{\mathbb{U}_{m,L}(0, t)}{g_m(\lfloor mt \rfloor)},$$

where $\mathbb{U}_{m,L}$ is given in (C.23), and set $\mathcal{U}_m(t) = \mathcal{U}_m(2/m)$, $\mathcal{U}_{m,L}(t) = \mathcal{U}_{m,L}(2/m)$ for $0 \leq t < 2/m$. We have

$$\sup_{t \geq 0} |\mathcal{U}_m(t)| = \sup_{k \geq 2} \frac{\mathcal{D}_m(k)}{g_m(k)}.$$

With $\mathcal{V}_{m,L}(t) = \mathbb{V}_{m,L}(0, t)/g(t)$, where $\mathbb{V}_{m,L}$ is defined in Lemma C.4, applying Lemma C.4 we have, for any $0 < \delta < T$,

$$(D.3) \quad \sup_{\delta \leq t \leq T} |\mathcal{U}_{m,L}(t) - \mathcal{V}_{m,L}(t)| = \sup_{\delta \leq t \leq T} \left| \frac{\mathbb{V}_{m,L}(0, t)}{g(t)} - \frac{\mathbb{U}_{m,L}(0, t)}{g_m(\lfloor mt \rfloor)} \right| = o_P(1).$$

On the other hand, setting

$$(D.4) \quad \mathcal{V}(t) = -\frac{1}{g(t)} \sum_{\ell=1}^{\infty} \lambda_\ell \left[(W_{2,\ell}(t) - tW_{1,\ell}(1))^2 - t(1+t) \right], \quad t > 0,$$

and $\mathcal{V}(0) = 0$, Lemma C.5 implies *that* $\mathcal{V}$ is well-defined, and for any fixed m , as $L \rightarrow \infty$,

$$(D.5) \quad \mathcal{V}_{m,L} \Rightarrow \mathcal{V} \quad \text{in} \quad \mathbf{D}[\delta, T].$$

Additionally, from Lemma C.2, we have

$$(D.6) \quad \lim_{L \rightarrow \infty} \sup_{m \geq 1} P \left\{ \sup_{t \geq 0} |\mathcal{U}_m(t) - \mathcal{U}_{m,L}(t)| > x \right\} = 0,$$

which combined with (D.3) and (D.5) implies (see e.g. Theorem 3.2 in Billingsley, 1968)

$$(D.7) \quad \mathcal{U}_m \Rightarrow \mathcal{V} \quad \text{in} \quad \mathbf{D}[\delta, T].$$

On the other hand, Lemma C.5 implies

$$(D.8) \quad \sup_{0 \leq t \leq \delta} |\mathcal{V}(t)| \rightarrow 0, \quad \text{a.s.} \quad \delta \rightarrow 0.$$

Further, by Lemma C.2, expression (C.14),

$$(D.9) \quad \lim_{\delta \rightarrow 0} \limsup_{m \rightarrow \infty} P \left\{ \sup_{0 \leq t \leq \delta} |\mathcal{U}_m(t)| > x \right\} = 0.$$

Combining (D.8), and (D.9) gives $\mathcal{U}_m \Rightarrow \mathcal{V}$ in $D[0, T]$ for every $T > 0$, and since $\mathcal{V}$ is continuous, we therefore have (see e.g. Theorem 16.7 in Billingsley, 1968)

$$(D.10) \quad \mathcal{U}_m \Rightarrow \mathcal{V} \quad \text{in} \quad D[0, T],$$

for any $T > 0$. Further, a.s. boundedness of $\mathcal{V}$ implies $\sup_{t \geq 0}(\cdot)$ is continuous at $\mathcal{V}$, giving

$$(D.11) \quad \sup_{t \geq 0} |\mathcal{U}_m(t)| \Rightarrow \sup_{t \geq 0} |\mathcal{V}(t)|.$$

Now, checking covariance functions, one can easily verify that

$$(D.12) \quad \left\{ \frac{W_{2,\ell}(t) - tW_{1,\ell}(1)}{1+t}, \quad t \geq 0, \ell \geq 1 \right\} \stackrel{\mathcal{D}}{=} \left\{ W_\ell \left(\frac{t}{1+t} \right), \quad t \geq 0, \ell \geq 1 \right\},$$

where $\{W_1(t), t \geq 0\}, \{W_2(t), t \geq 0\}, \dots$ are independent Wiener processes. Thus, recalling (2.8), we have

$$\begin{aligned} \sup_{t \geq 0} |\mathcal{V}(t)| &\stackrel{\mathcal{D}}{=} \sup_{t \geq 0} \left(\frac{t}{1+t} \right)^{-\beta} \left| \sum_{\ell=1}^{\infty} \lambda_{\ell} \left[W_{\ell}^2 \left(\frac{t}{1+t} \right) - \frac{t}{1+t} \right] \right| \\ (D.13) \quad &\stackrel{\mathcal{D}}{=} \sup_{0 < u \leq 1} u^{-\beta} \left| \sum_{\ell=1}^{\infty} \lambda_{\ell} [W_{\ell}^2(u) - u] \right|, \end{aligned}$$

yielding part (i) of the theorem. Turning to part (ii), for simplicity write $M_m = M$. Since $M/m \rightarrow a_0$, and

$$(D.14) \quad \sup_{2 \leq k \leq M} \frac{\mathcal{D}_m(k)}{g_m(k)} = \sup_{0 \leq t \leq M/m} |\mathcal{U}_m(t)|,$$

the same arguments above yield $\sup_{0 \leq t \leq M/m} |\mathcal{U}_m(t)| \Rightarrow \sup_{0 \leq t \leq a_0} |\mathcal{V}(t)|$, and the result follows from the change of variables in (D.13).

Turning now to part (iii) of the theorem, for any $t \geq 2/M$, define

$$\tilde{\mathcal{U}}_m(t) = \frac{\lfloor Mt \rfloor^2 U_m(h; \lfloor Mt \rfloor)}{M(\lfloor Mt \rfloor/M)^{\beta}}, \quad \tilde{\mathcal{U}}_{m,L}(t) = \frac{\lfloor Mt \rfloor^2 U_m(h; \lfloor Mt \rfloor)}{M(\lfloor Mt \rfloor/M)^{\beta}},$$

so that

$$\max_{2 \leq k \leq M} \frac{\mathcal{D}_m^{(1)}(k)}{g_m(k)} = \sup_{2/m \leq t \leq 1} |\tilde{\mathcal{U}}_m(t)|.$$

Also, for each $t > 0$ let

$$\begin{aligned} &\tilde{\mathcal{U}}_{m,L}^{\circ}(t) \\ &= - \left(\frac{\lfloor Mt \rfloor}{M} \right)^{\beta} \sum_{\ell=1}^L \lambda_{\ell} \left(\frac{1}{M} \left(S_{\ell}(\lfloor Mt \rfloor, m) - \frac{\lfloor Mt \rfloor}{m} S_{\ell}(m) \right)^2 - \frac{\lfloor Mt \rfloor (\lfloor Mt \rfloor + m)}{m^2} \right), \end{aligned}$$

and

$$\tilde{\mathcal{V}}_L(t) = -t^{-\beta} \sum_{\ell=1}^L \lambda_{\ell} [W_{\ell}^2(t) - t], \quad \tilde{\mathcal{V}}(t) = -t^{-\beta} \sum_{\ell=1}^L \lambda_{\ell} [W_{\ell}^2(t) - t].$$

Arguing as in the case of part (i), we need only establish the weak convergence of

$$(D.15) \quad \tilde{\mathcal{U}}_{m,L}^\circ \Rightarrow \tilde{\mathcal{V}}_L, \quad \text{in } \mathbf{D}[\delta, 1],$$

for every fixed $L \geq 1$ and $0 < \delta < 1$. However, since $m^{-1}S_\ell(m) = O_P(m^{-1/2})$, we have

$$\begin{aligned} & \max_{2 \leq k \leq M} \left(\frac{k}{M} \right)^\beta \left| \sum_{\ell=1}^L \lambda_\ell \left(\frac{1}{M} \left(\frac{k}{m} S_\ell(m) \right)^2 - \frac{k^2}{m^2} \right) \right| \\ & \leq \left| \sum_{\ell=1}^L \lambda_\ell \left(M \left(\frac{1}{m} S_\ell(m) \right)^2 + \frac{M^2}{m^2} \right) \right| = O_P \left(\frac{M}{m} \right) + O \left(\left(\frac{M}{m} \right)^2 \right) = o_P(1). \end{aligned}$$

Finally

$$\begin{aligned} M^{-1/2} (S_1(\lfloor Mt \rfloor, m), \dots, S_L(\lfloor Mt \rfloor, m)) &\stackrel{\mathcal{D}}{=} M^{-1/2} (S_1(\lfloor Mt \rfloor, 0), \dots, S_L(\lfloor Mt \rfloor, 0)) \\ &\Rightarrow (W_1(t), \dots, W_L(t)) \quad \text{in } \mathbf{D}[0, 1], \end{aligned}$$

and the continuous mapping theorem yields

$$\left(\frac{\lfloor Mt \rfloor}{M} \right)^\beta \sum_{\ell=1}^L \lambda_\ell \left(\frac{1}{M} (S_\ell(\lfloor Mt \rfloor, m))^2 - \frac{\lfloor Mt \rfloor}{m} \right) \Rightarrow \tilde{\mathcal{V}}_L(t) \quad \text{in } \mathbf{D}[\delta, 1],$$

giving (D.15). The remainder of the proof is the same as in case (i). $\square$

Proof of Theorem 3.2. The proof is largely the same as Theorem 3.1, though we provide details where there are important differences. Let

$$\mathbb{U}_m(s, t) = m^{-1} ((\lfloor mt \rfloor - \lfloor ms \rfloor) \vee 2)^2 U_m(h; \lfloor ms \rfloor, \lfloor mt \rfloor), \quad 0 \leq s \leq t,$$

and let $\mathbb{U}_{m,L}(s, t)$ be as in (C.23). For any $0 \leq s \leq t$, let

$$(D.16) \quad \mathcal{U}_m(s, t) = \frac{\mathbb{U}_m(s, t)}{g_m(\lfloor mt \rfloor \vee 2)}, \quad \mathcal{U}_{m,L}(s, t) = \frac{\mathbb{U}_{m,L}(s, t)}{g_m(\lfloor mt \rfloor \vee 2)},$$

and for any real-valued function $\{u(s, t), s, t \geq 0\}$, write

$$(D.17) \quad \Psi u(t) = \sup_{0 \leq s \leq t} |u(s, t)|, \quad t > 0.$$

With $\mathbb{V}_{m,L}$ as defined in Lemma C.4, set $\mathcal{V}_{m,L}(s, t) = \mathbb{V}_{m,L}(s, t)/g(t)$. Lemma C.4 gives, for any $0 < \delta < T$,

$$\sup_{\delta \leq t \leq T} |\Psi \mathcal{U}_m(t) - \Psi \mathcal{U}_{m,L}(t)| \leq \sup_{s, t \in I_{\delta, T}} |\mathcal{U}_{m,L}(s, t) - \mathcal{V}_{m,L}(s, t)| = o_P(1).$$

We again have from Lemma C.2

$$(D.18) \quad \limsup_{L \rightarrow \infty} \sup_{m \geq 1} P \left\{ \sup_{\delta \leq t < \infty} |\Psi \mathcal{U}_m(t) - \Psi \mathcal{U}_{m,L}(t)| > x \right\}$$

$$(D.19) \quad \leq \limsup_{L \rightarrow \infty} \sup_{m \geq 1} P \left\{ \sup_{0 \leq s \leq t < \infty} |\mathcal{U}_m(s, t) - \mathcal{U}_{m,L}(s, t)| > x \right\}.$$

With $\mathcal{V}(s, t) = \mathbb{V}(s, t)/g(t)$, Lemma C.5 shows $\mathcal{V}$ admits a version $\mathcal{V} \in \mathbf{C}[0, \infty)$ and Ψ is continuous at $\mathcal{V}$; hence for any fixed m , and any $T > 0$,

$$\{\Psi \mathcal{V}_{m,L}(t), t \geq 0\} \Rightarrow \{\Psi \mathcal{V}(t), t \geq 0\} \quad \text{in } \mathbf{C}[0, T], \quad L \rightarrow \infty,$$

which combined with (D.3) and (D.5) implies

$$(D.20) \quad \{\Psi \mathcal{U}_m(t), t \geq 0\} \Rightarrow \{\Psi \mathcal{V}(t), t \geq 0\} \quad \text{in } \mathbf{D}[\delta, T].$$

On the other hand, Lemma C.5 implies

$$(D.21) \quad \sup_{0 \leq t \leq \delta} \Psi \mathcal{V}(t) \rightarrow 0, \quad \text{a.s.} \quad \delta \rightarrow 0.$$

Further, by Lemma C.2, expression (C.14),

$$(D.22) \quad \lim_{\delta \rightarrow 0} \limsup_{m \rightarrow \infty} P \left\{ \sup_{0 \leq t \leq \delta} \Psi \mathcal{U}_m(t) > x \right\} = 0.$$

Combining (D.20), (D.21), and (D.22) gives, for every $T > 0$

$$(D.23) \quad \{\Psi \mathcal{U}_m(t), t \geq 0\} \Rightarrow \{\Psi \mathcal{V}(t), t \geq 0\} \quad \text{in} \quad \mathbf{D}[0, T].$$

Now, using (D.23) and that $\{\Psi \mathcal{V}(t), t \geq 0\} \in \mathbf{C}[0, \infty)$, we readily deduce convergence $\{\sup_{t \geq 0} \Psi \mathcal{U}_m(t)\} \Rightarrow \{\sup_{t \geq 0} \Psi \mathcal{V}(t)\}$.

From (D.12), writing

$$H(s, t) = \sum_{\ell=1}^{\infty} \lambda_{\ell} \left[\left(W_{\ell} \left(\frac{t}{1+t} \right) - \frac{1+s}{1+t} W_{\ell} \left(\frac{s}{1+s} \right) \right)^2 - \frac{(t-s)(1+t-s)}{(1+t)^2} \right],$$

we have

$$(D.24) \quad \begin{aligned} \sup_{t \geq 0} |\Psi \mathcal{V}(t)| &\stackrel{\mathcal{D}}{=} \sup_{0 \leq s \leq t < \infty} \left(\frac{t}{1+t} \right)^{-\beta} |H(s, t)| \\ &= \sup_{0 < u \leq v < 1} v^{-\beta} \left| H \left(\frac{u}{1-u}, \frac{v}{1-v} \right) \right| \\ &= \sup_{0 < u \leq v < 1} v^{-\beta} \left| \sum_{\ell=1}^{\infty} \lambda_{\ell} \left[\left(W_{\ell}(v) - \frac{1-v}{1-u} W_{\ell}(u) \right)^2 \right. \right. \\ &\quad \left. \left. - \left(v - u \left(\frac{1-v}{1-u} \right) \right) \left(1 - u \left(\frac{1-v}{1-u} \right) \right) \right] \right|. \end{aligned}$$

The proof of parts (ii) and (iii) are similar to the proofs of Theorem 3.1(ii)-(iii) and thus omitted. $\square$

Proof of Theorem 3.3. First suppose $k_* = O(m)$. Then, with $\mathbf{q}_1(r, k)$ and $\mathbf{q}_2(r, k)$ as in (C.46), Lemma C.6 gives, along any sequence $y = y_m \rightarrow \infty$,

$$(D.25) \quad \max_{k_* \leq k \leq y} \left| \frac{k^2 U_m(h; k)}{mg_m(k)} - \frac{\mathbf{q}_1(0, k) + \mathbf{q}_2(0, k) + q_{m,3}(0, k)}{mg_m(k)} \right| = O_P(1).$$

In particular, (D.25) holds when $y = 2k_* \vee m$. Moreover, it is easily seen that

$$(D.26) \quad \frac{|\mathbf{q}_1(0, y)|}{mg_m(y)} \geq Cm\theta|\nu_1 - \nu_2|,$$

whereas

$$(D.27) \quad \frac{|q_2(0, y)|}{mg_m(y)} \leq C\theta \left(\left| \sum_{i=1}^m z_i \right| + \left| \sum_{i=m+1}^{m+k_*} z_i \right| + \left| \sum_{i=m+k_*+1}^{m+y} z_i^* \right| \right) = O_P(\theta m^{1/2}),$$

and from Lemma C.7,

$$(D.28) \quad \frac{|q_{m,3}(0, y)|}{mg_m(y)} = O_P(1).$$

Putting together (D.25)–(D.28), we then obtain

$$\mathcal{D}_m^{(2)}(m) \geq \mathcal{D}_m^{(1)}(m) = \frac{k^2 |U_m(h; k)|}{mg_m(m)} \geq Cm\theta |\nu_1 - \nu_2| (1 + o_P(1)) \xrightarrow{P} \infty,$$

which implies $P(\tau_m < \infty)$ when $k_* = O(m)$ under either monitoring scheme. If instead we have $m = o(k_*)$, taking $y = 2k_*$, Lemma C.6 again yields (D.25), and it is easily verified (D.26) still holds. Moreover, since $g_m(2k_*) \geq C(k_*/m)^2$,

$$(D.29) \quad \frac{|q_2(0, 2k_*)|}{mg_m(2k_*)} \leq C \frac{\theta}{(k_*/m)} \left(\frac{k_*}{m} \left| \sum_{i=1}^m z_i \right| + \left| \sum_{i=m+1}^{m+k_*} z_i \right| + \left| \sum_{i=m+k_*+1}^{m+k} z_i^* \right| \right) = O_P(\theta m^{1/2}).$$

Further, Lemma (C.7) again gives (D.28), and the statement follows. $\square$

Proof of Theorem 3.4. We begin with part (i). We first proceed to find a sequence $y_m \rightarrow \infty$ for which $P(\kappa_m \leq y_m)$ has a nontrivial limit. Set

$$(D.30) \quad q_{2,1}(r, k) = 2\theta(k - (k_* \vee r)) \sum_{i=m+k_*+1}^{m+k} z_i^*,$$

If we choose y_m in such a way that

$$(D.31) \quad y_m \rightarrow \infty, \quad m^{-1}y_m \rightarrow 0$$

then for $\mathbf{q}_1(r, k)$ as in (C.46),

$$(D.32) \quad a_m = \frac{|\mathbf{q}_1(0, y_m)|}{mg_m(y_m)} = \theta|\nu_1 - \nu_2|m \left(\frac{y_m/m}{1 + y_m/m} \right)^{2-\beta} \sim \theta|\nu_1 - \nu_2|m^{\beta-1}y_m^{2-\beta}.$$

Under (D.31), we also have

$$(D.33) \quad \begin{aligned} \frac{\mathbf{q}_{2,1}(0, y)}{mg_m(y)} &= \mathbf{q}_{2,1}(0, y)m^{\beta-1}y_m^{-\beta}(1 + y_m/m)^{\beta-2} \\ &= 2\theta m^{\beta-1}y_m^{1-\beta} \left(\sum_{i=m+k_*+1}^{m+y_m} z_i^* \right) \frac{(1 - k_*/y_m)}{(1 + y_m/m)^{2-\beta}} \\ &= b_m \left(\frac{1}{\sigma_* y_m^{1/2}} \sum_{i=m+k_*+1}^{m+y_m} z_i^* \right) \frac{(1 - k_*/y_m)}{(1 + y_m/m)^{2-\beta}}, \end{aligned}$$

with

$$(D.34) \quad b_m = b_m(y_m) = 2\sigma_*\theta m^{\beta-1}y_m^{3/2-\beta}.$$

With $\rho = (1 - \beta)/(2 - \beta)$, we may pick y_m satisfying (D.31) as a solution to

$$y_m = k_* + w_1 m^\rho (1 + w_2 b_m(y_m)),$$

where $w_1^{2-\beta} = \mathfrak{c}(\theta|\nu_1 - \nu_2|)^{-1}$, and w_2 is a constant to be later specified such that

$$(D.35) \quad a_m \rightarrow \mathfrak{c}, \quad b_m^{-1}(\mathfrak{c} - a_m) \rightarrow -x.$$

Indeed, since $\theta|\nu_1 - \nu_2|m \rightarrow \infty$ under Assumption 3.1 and $(\mathfrak{c}\theta|\nu_1 - \nu_2|)^{1/2} \sim C\theta\sigma_*$ under Assumption 3.2,

$$\begin{aligned} b_m(2w_1 m^\rho) &= C(\theta|\nu_1 - \nu_2|)^{-(\rho+1)/2} m^{-\rho/2} \sigma_* \theta \\ &= C(m\theta|\nu_1 - \nu_2|)^{-\rho/2} \sigma_* \theta / (\theta|\nu_1 - \nu_2|)^{1/2} = o(1). \end{aligned}$$

Thus, the function $\varphi(y) = y - k_* - w_1 m^\rho (1 + w_2 b_m(y))$ is easily seen to have a root in the region $(0, 2w_1 m^\rho)$ for all large m and any fixed w_2 , which satisfies $y_m \sim w_1 m^\rho$ and in particular satisfies (D.31). From (D.35) we obtain

$$\begin{aligned}
 P\{\kappa_m < y_m\} &= P\left\{\max_{k_* < k \leq y_m} \frac{k^2 |U_m(h; k)|}{mg_m(k)} > c\right\} \\
 &= P\left\{b_m^{-1} \left(\max_{k_* < k \leq y_m} \frac{k^2 |U_m(h; k)|}{mg_m(k)} - a_m\right) > -x + o(1)\right\}.
 \end{aligned}
 \tag{D.36}$$

Recall the decomposition (C.42). Applying Lemma C.6, we have

$$\begin{aligned}
 b_m^{-1} \max_{k_* < k \leq y_m} \frac{\max_{0 \leq b < k} |q_1(0, k) - \mathfrak{q}_1(0, k)|}{mg_m(k)} &\leq C b_m^{-1} \theta |\nu_1 - \nu_2| (k_*/m)^{1-\beta} \\
 &= C \sigma_*^{-1} |\nu_1 - \nu_2| (k_*/y_m)^{(1-\beta)} y_m^{-1/2} \\
 &= o(1),
 \end{aligned}
 \tag{D.37}$$

and for all small $\delta > 0$,

$$\begin{aligned}
 b_m^{-1} \max_{k_* < k \leq y_m} \frac{\max_{0 \leq b < k} |q_2(0, k) - \mathfrak{q}_2(0, k)|}{mg_m(k)} &\leq C b_m^{-1} \theta k_*^{-1} (k_*/m)^{1-\beta} O_P(y_m^{1/2+\delta}) \\
 &= O_P(\theta k_*^{-1} (k_*/y_m)^{1-\beta} y_m^\delta) = o_P(1).
 \end{aligned}
 \tag{D.38}$$

Also,

$$\begin{aligned}
 b_m^{-1} \max_{k_* < k \leq y_m} \frac{\max_{0 \leq b < k} |\mathfrak{q}_{2,1}(0, k) - \mathfrak{q}_2(0, k)|}{mg_m(k)} &\leq C b_m^{-1} \max_{k_* < k \leq y_m} \theta (k/m)^{1-\beta} \left| -\frac{k}{m} \sum_{i=1}^m z_i + \sum_{i=m+1}^{m+k_*} z_i \right| \\
 &\leq C \theta b_m^{-1} y_m^{1-\beta} m^{\beta-1} \sigma \left(O_P(y_m m^{-1/2}) + O_P(m^{1/2}) \right) \\
 &\leq C y_m^{-1/2} \left(O_P(y_m m^{-1/2}) + O_P(1) \right) = o_P(1),
 \end{aligned}
 \tag{D.39}$$

and from Lemma C.7,

$$(D.40) \quad b_m^{-1} \max_{k_* < k \leq y_m} \frac{|q_3(0, k)|}{mg_m(k)} \leq C b_m^{-1} (y_m/m)^{2(1-\beta)} = O_P \left(\frac{(y/m)^{1-\beta}}{\sigma_* y^{1/2}} \right) = o_P(1),$$

where we used Assumption 3.2 to conclude $\sigma_* y^{1/2} = O(m^{\rho/2} \sigma_*/w_1) \rightarrow \infty$. From the bounds (D.37)–(D.40), in view of (D.36), it suffices to show

$$(D.41) \quad b_m^{-1} \left(\max_{k_* < k \leq y_m} \frac{|q_1(0, k) + q_{2,1}(0, k)|}{mg_m(k)} - a_m \right) \Rightarrow \mathcal{N}(0, 1).$$

So, note (c.f. (D.32)) $|q_1(0, k)|/mg_m(k)$ is increasing in k . Hence, for any $0 \leq \delta < 1$,

$$(D.42) \quad \max_{k_* \leq k \leq y_m(1-\delta)} \frac{|q_1(0, k)|}{mg_m(k)} = \theta |\nu_1 - \nu_2| m^{\beta-1} [y_m(1-\delta)]^{2-\beta} (1 + o(1)),$$

$$(D.43) \quad \min_{(1-\delta)y_m \leq k \leq y_m} \frac{|q_1(0, k)|}{mg_m(k)} = \theta |\nu_1 - \nu_2| m^{\beta-1} [y_m(1-\delta)]^{2-\beta} (1 + o(1)).$$

Also, from (D.33), for all $k_*/y_m < s < 1$,

$$(D.44) \quad b_m^{-1} \frac{q_{2,1}(0, \lfloor y_m s \rfloor)}{mg_m(\lfloor y_m s \rfloor)} = \left(\frac{\lfloor y_m s \rfloor}{y_m} \right)^{1-\beta} \left(\frac{1}{y_m^{1/2}} \sum_{i=m+k_*+1}^{m+\lfloor y_m s \rfloor} z_i^* \right) \frac{(1 - k_*/\lfloor y_m s \rfloor)}{(1 + \lfloor y_m s \rfloor/m)^{2-\beta}},$$

the functional central limit theorem gives

$$b_m^{-1} \frac{q_{2,1}(0, \lfloor y_m s \rfloor \vee (k_* + 1))}{mg_m(\lfloor y_m s \rfloor \vee (k_* + 1))} \Rightarrow s^{1-\beta} W(s) \quad \text{in } \mathbf{D}[0, 1],$$

where $\{W(s), s \geq 0\}$ is a Wiener process. Now, from (D.42) and (D.43),

$$(D.45) \quad \begin{aligned} & b_m^{-1} \left(\max_{k_* < k \leq y_m(1-\delta)} \frac{|q_1(0, k) + q_{2,1}(0, k)|}{mg_m(k)} - a_m \right) \\ & \leq O_P(1) + (2\sigma_*)^{-1} ([1 - \delta]^{2-\beta} - 1) |\nu_1 - \nu_2| y_m^{1/2} (1 + o(1)) \xrightarrow{P} -\infty. \end{aligned}$$

On the other hand, if we let

$$A_m = \left\{ \omega : b_m^{-1} \max_{(1-\delta)y \leq k \leq y} \frac{|q_{2,1}(0, k)|}{mg_m(k)} < b_m^{-1} \min_{(1-\delta)y \leq k \leq y} \frac{|q_1(0, k)|}{mg_m(k)} \right\},$$

then (D.42), (D.43) and (D.44) give $P(A_m) \rightarrow 1$, and for each $\omega \in A_m$,

$$(D.46) \quad \frac{|q_1(0, k) + q_{2,1}(0, k)|}{mg_m(k)} = \frac{|q_1(0, k)|}{mg_m(k)} + \text{sgn}(q_1(0, k)) \frac{q_{2,1}(0, k)}{mg_m(k)}, \quad (1-\delta)y_m \leq k \leq y_m.$$

Note $\text{sgn}(q_1(r, k)) = \text{sgn}(\nu_1 - \nu_2)$. Thus, if $(\nu_1 - \nu_2) > 0$,

$$\begin{aligned} & P \left(A_m \cap \left\{ b_m^{-1} \left(\max_{y_m(1-\delta) \leq k \leq y_m} \frac{|q_1(0, k) + q_{2,1}(k)|}{mg_m(k)} - \frac{|q_1(y_m) + q_{2,1}(0, y_m)|}{mg_m(y_m)} \right) > x \right\} \right) \\ &= P \left(A_m \cap \left\{ b_m^{-1} \left(\max_{y_m(1-\delta) \leq k \leq y_m} \frac{q_1(0, k) + q_{2,1}(0, k)}{mg_m(k)} - \frac{q_1(y_m) + q_{2,1}(0, y_m)}{mg_m(y_m)} \right) > x \right\} \right) \\ &\leq P \left(b_m^{-1} \left(\max_{y_m(1-\delta) \leq k \leq y_m} \left[\frac{q_1(0, k)}{mg_m(k)} - \frac{q_1(0, y_m)}{mg_m(y_m)} \right] + \max_{y_m(1-\delta) \leq k \leq y_m} \left[\frac{q_{2,1}(0, k)}{mg_m(k)} - \frac{q_{2,1}(0, y_m)}{mg_m(y_m)} \right] \right) > x \right) \\ &= P \left\{ b_m^{-1} \max_{y_m(1-\delta) \leq k \leq y_m} \left| \frac{q_{2,1}(0, k)}{mg_m(k)} - \frac{q_{2,1}(0, y_m)}{mg_m(y_m)} \right| > x/2 \right\} \\ &\rightarrow P \left\{ \sup_{(1-\delta) \leq s \leq 1} |s^{1-\beta} W(s) - W(1)| > x/2 \right\}, \end{aligned}$$

where on the third line we used increasingness of $q_1(0, k)/mg_m(k)$ and on the last line we used (D.44). Analogous reasoning holds in the case $\nu_1 - \nu_2 < 0$. Thus, by continuity of W ,

$$(D.47) \quad \lim_{\delta \rightarrow 0} \limsup_{m \rightarrow \infty} P \left\{ b_m^{-1} \left(\max_{y_m(1-\delta) \leq k \leq y_m} \frac{|q_1(0, k) + q_{2,1}(0, k)|}{mg_m(k)} - \frac{|q_1(0, y) + q_{2,1}(0, y)|}{mg_m(y)} \right) > x \right\} = 0.$$

Now, from (D.46),

$$(D.48) \quad b_m^{-1} \left(\frac{|q_1(0, y_m) + q_{2,1}(0, y_m)|}{mg_m(y_m)} - a_m \right) = b_m^{-1} \text{sgn}(\nu_1 - \nu_2) \frac{q_{2,1}(0, y_m)}{mg_m(y_m)} + o_P(1) \Rightarrow W(1),$$

which, together with (D.69) and (D.47) yields the limit (D.41). From (D.36), we then obtain

$$(D.49) \quad P\{\kappa_m < y_m\} \rightarrow 1 - \Phi(-x) = \Phi(x), \quad m \rightarrow \infty.$$

Now we choose w_2 so that (D.35) holds. Note $w_1^{2-\beta} = \mathfrak{c}(\theta|\nu_1 - \nu_2|)^{-1}$ clearly gives $a_m \rightarrow \mathfrak{c}$. Moreover,

$$\begin{aligned}\mathfrak{c} - a_m &= \mathfrak{c} - \theta|\nu_1 - \nu_2|m^{\beta-1}y_m^{2-\beta}(1 + y_m/m)^{\beta-2}(1 - k_*/y_m)^2 \\ &= \mathfrak{c} - \mathfrak{c}(1 + w_2b_m)^{2-\beta}(1 + y_m/m)^{\beta-2}(1 - k_*/y_m)^2 \\ &= -\mathfrak{c}(2 - \beta)w_2b_m + o(b_m).\end{aligned}$$

where we used that $b_m \gg y_m m^{-1}$ since (noting that $\theta\sigma_* \sim C(\theta|\nu_1 - \nu_2|)^{1/2} \sim Cw_1^{-(2-\beta)/2}$ under Assumption 3.2),

$$\begin{aligned}b_m \gg y_m m^{-1} &\iff \sigma_*\theta m^{-\beta}(w_1 m^\rho)^{1/2-\beta} \rightarrow \infty \\ &\iff m^{\rho/2+\beta(1-\rho)}w_1^{-(1+\beta)/2} \rightarrow \infty, \\ &\iff (m\theta|\nu_1 - \nu_2|)^{(1+\beta)/(4-2\beta)} \rightarrow \infty,\end{aligned}$$

which holds under Assumption 3.1, and also we used that $b_m \gg k_*y_m^{-1}$, which holds since

$$\begin{aligned}b_m \gg k_*y_m^{-1} &\iff \sigma_*\theta m^{-(2-\beta)\rho}(w_1 m^\rho)^{5/2-\beta} \rightarrow \infty \\ &\iff m^{\rho/2}w_1^{(3-\beta)/2} \rightarrow \infty,\end{aligned}$$

which always holds. So, choosing $w_2 = (\mathfrak{c}(2 - \beta))^{-1}x$, we obtain the second statement in (D.35), implying (D.70) holds for the sequence y_m . Now, since $y_m \sim w_1 m^\rho$, we have

$$\begin{aligned}y_m - k_* - w_1 m^\rho &= w_1 w_2 m^\rho b_m \\ &\sim w_1 w_2 m^\rho m^{-\rho(2-\beta)}(w_1 m^\rho)^{3/2-\beta} \\ &= (2\sigma_*\theta)w_2 w_1^{2-\beta}(w_1 m^\rho)^{1/2} \\ &= \frac{2\sigma_*x}{(2 - \beta)|\nu_1 - \nu_2|}(w_1 m^\rho)^{1/2},\end{aligned}$$

from which we obtain

$$P\{\kappa_m < y_m\} \sim P\left\{\frac{(2-\beta)|\nu_1 - \nu_2|}{2\sigma_*} \frac{\kappa_m - w_1 m^\rho}{(w_1 m^\rho)^{1/2}} > -x\right\} \rightarrow \Phi(x).$$

We now turn to part (ii). First we set up some notation used in the proof and define the limit variable appearing (3.16). Let $\{W_1(t), t \geq 0\}$, $\{W_2(t), t \geq 0\}, \dots$ be independent Wiener processes, and let $\{V_1(t), t \geq 0\}$, $\{V_2(t), t \geq 0\}$, each be Wiener processes with

$$(D.50) \quad \begin{aligned} \mathbb{E}V_1(t)V_2(t) &= 0, & \mathbb{E}V_1(t)W_{1,\ell}(t) &= \eta_\ell t, & \mathbb{E}V_1(t)W_{2,\ell}(t) &= 0, \\ \mathbb{E}V_2(t)W_{1,\ell}(t) &= 0, & \mathbb{E}V_2(t)W_{2,\ell}(t) &= \eta_\ell t, \end{aligned}$$

where, with $v(\mathbf{x})$ as in (3.10), and $\phi_\ell(\mathbf{x})$ as in (C.2),

$$\sigma^{-1}\mathbb{E}v(\mathbf{X}_1)\phi_\ell(\mathbf{X}_1) = \eta_\ell.$$

Also, let

$$Z(t, c_*) = \frac{\zeta t^2 + 2t(V_2(c_*) - c_* V_1(1)) + \mathbb{V}(0, c_*)}{g(c_*)},$$

with $\mathbb{V}(s, t)$ as in (C.30). Finally, we define

$$(D.51) \quad \mathcal{H}_{c_*}(u) = \inf \left\{ x \geq 0 : \sup_{0 \leq s \leq x} |Z(t, c_*)| \geq u \right\},$$

i.e., $\mathcal{H}_{c_*}(u)$ is the left-continuous inverse of $x \mapsto \sup_{0 \leq s \leq x} |Z(t, c_*)|$. We are now ready to proceed with the proof.

For simplicity write $\Delta = \theta|\nu_1 - \nu_2| = \theta^2|\mathfrak{D}_h(F, G)|$. We first show, for any $T > 0$, (c.f. (C.42))

$$(D.52) \quad \max_{k_* < k \leq k_*(m/\Delta)^{1/2}T} \left| \frac{U_m(h; k)}{g_m(k)} - \frac{q_1(0, k) + q_{2,2}(0, k) + q_3(0, k_*)}{mg_m(k)} \right| = o_P(1),$$

where $q_1(r, k)$ is given in (C.46), and

$$\begin{aligned} q_{2,2}(r, k) &= 2\theta(k - (k_* \vee r)) \left[-\frac{k-r}{m} \sum_{i=1}^m z_i + \mathbf{1}_{\{r < k_*\}} \sum_{i=m+r+1}^{m+k_*} z_i \right] \\ (D.53) \quad &=: 2\theta\sigma(k - (k_* \vee r)) \left[-\frac{k-r}{m} V_{1,m} + V_{2,m}(r) \right], \end{aligned}$$

$$(D.54) \quad q_3(k, r) = (k-r)^2 \left(\frac{2\mathbf{1}_{\{r \leq k_*\}} R_{m,1,1}(r)}{(k-r)m} - \frac{2R_{m,2}}{m(m-1)} - \frac{2\mathbf{1}_{\{r \leq k_*\}} R_{m,3,1}(r)}{(k-r)(k-r-1)} \right),$$

with $R_{m,1,1}(r)$, $R_{m,1,1}$, and $R_{m,3,1}$ as in (C.38), (C.40) and (C.41).

Lemma C.6 immediately gives

$$(D.55) \quad \max_{k_* < k \leq k_* + (m/\Delta)^{1/2}T} \max_{0 \leq r < k} \left(\frac{|q_1(r, k) - q_1(r, k)| + |q_2(r, k) + q_2(r, k)|}{mg_m(k)} \right) = o_P(1).$$

With $q_2(r, k)$ in (C.46), we have

$$(D.56) \quad \max_{k_* < k \leq k_* + (m/\Delta)^{1/2}T} \max_{0 \leq r < k} \frac{|q_2(r, k) + q_{2,2}(r, k)|}{mg_m(k)} = o_P(1).$$

Indeed, for any $T > 0$, the law of the iterated logarithm gives

$$\begin{aligned} &\max_{k_* < k \leq k_* + (m/\Delta)^{1/2}T} \max_{0 \leq r < k} \frac{|q_2(r, k) - q_{2,2}(r, k)|}{mg_m(k)} \\ &\leq \max_{k_* \leq k \leq k_* + T(m/\Delta)^{1/2}} \frac{\theta(k - k_*)}{mg_m(k)} \max_{k_* \leq r < k} \left| \sum_{i=m+r+1}^{m+k} z_i^* \right| \\ &\leq C\theta m^{-1/2} \max_{k_* \leq k \leq k_* + (m/\Delta)^{1/2}T} \left| \sum_{i=m+k_*+1}^{m+k} z_i^* \right| \\ &= O_P \left(\theta \sigma_* m^{-1/2} ((m/\Delta)^{1/2} \log \log(m/\Delta))^{1/2} \right) \\ &= O_P \left(((m/\Delta)^{-1/2} \log \log(m/\Delta))^{1/2} \right), \end{aligned}$$

where we used that $\sigma_*^2 \sim C|\mathfrak{D}_h(F, G)| \sim C|\Delta|/\theta^2$ due to Assumption 3.2, giving (D.56).

Applying Lemma C.9,

$$(D.57) \quad \max_{k_* < k \leq k_* + (m/\Delta)^{1/2}T} \max_{0 \leq r \leq k} \frac{|q_3(r, k) + \mathfrak{q}_3(r, k)|}{mg_m(k)} = o_P(1).$$

Next, we claim that, for any $\delta > 0$

$$(D.58) \quad \max_{k_* < k \leq k_* + (m/\Delta)^{1/2}T} \max_{0 \leq r \leq k_*} \frac{|\mathfrak{q}_3(r, k) - (k_* - r)^2 U_m(\bar{h}; r, k_*)|}{mg_m(k)} = o_P(1).$$

First note (C.1) implies

$$\frac{\mathfrak{q}_3(k_*, r)}{(k_* - r)^2} = \frac{2R_{m,1,1}(r)}{(k_* - r)m} - \frac{2R_{m,2}}{m(m-1)} - \frac{2R_{m,3,1}(r)}{(k_* - r)(k_* - r - 1)} = U_m(\bar{h}; r, k_*).$$

With $R_{m,1,1}(r)$ as in (C.38), we have

$$\begin{aligned} & \max_{k_* < k \leq k_* + (m/\Delta)^{1/2}T} \max_{0 \leq r \leq k_*} \left| \frac{(k_* - r)^2}{m} \frac{R_{m,1,1}(r)}{(k_* - r)m} - \frac{(k - r)^2}{m} \frac{R_{m,1,1}(r)}{(k - r)m} \right| \\ &= \max_{k_* < k \leq k_* + (m/\Delta)^{1/2}T} \frac{k - k_*}{m} \max_{0 \leq r \leq k_*} \left| \frac{R_{m,1,1}(r)}{m} \right| \\ &= O_P((\Delta m)^{-1/2}), \end{aligned}$$

where we used that $\max_{0 \leq r \leq k_*} |R_{m,1,1}(r)| = O_P(m)$ due to Lemma C.8. Similarly, with

$R_{m,3,1}(r)$ as in (C.41), using the mean value theorem applied to $f(k) = (k - r)/(k - r - 1)$,

$$\begin{aligned} & \max_{k_* < k \leq k_* + (m/\Delta)^{1/2}T} \max_{0 \leq r \leq k_*} \left| \frac{(k - r)^2}{m} \frac{R_{m,3,1}(r)}{(k_* - r)(k_* - r - 1)} - \frac{(k_* - r)^2}{m} \frac{R_{m,3,1}(r)}{(k_* - r)(k_* - r - 1)} \right| \\ &\leq 2 \max_{k_* < k \leq k_* + (m/\Delta)^{1/2}T} \frac{C(k - k_*)}{m} \max_{0 \leq r \leq k_*} \frac{|R_{m,3,1}(r)|}{(k_* - r - 1)^2} \\ &= O_P((\Delta m)^{-1/2}), \end{aligned}$$

since $\max_{0 \leq r \leq k_*-1} |R_{m,3,1}(r)|/(k_* - r)^2 = o_P(1)$ again due to Lemma C.8. Lastly,

$$(D.59) \quad \begin{aligned} & \max_{k_* < k \leq k_*(m/\Delta)^{1/2}T} \max_{0 \leq r \leq k_*} \frac{|(k-r)^2 - (k_*-r)^2|}{m} \frac{|R_{m,2}|}{m(m-1)} \\ & \leq C \max_{k_* < k \leq k_*(m/\Delta)^{1/2}T} \frac{k-k_*}{m} \frac{k_*|R_{m,2}|}{m(m-1)} = O_P((\Delta m)^{-1/2}). \end{aligned}$$

Since $g_m(k_*) \geq C > 0$, we therefore have (D.58), which combined with (D.55), (D.56), and (D.57) gives (D.52). Noting that $\mathbf{q}_3(0, k_*) = k_*^2 U_m(\bar{h}; k_*)$, we now show

$$\max_{k_* < k \leq k_*(m/\Delta)^{1/2}T} \frac{|\mathbf{q}_1(0, k) + \mathbf{q}_{2,2}(0, k) + k_*^2 U_m(\bar{h}; k_*)|}{m g_m(k)} \Rightarrow \sup_{0 \leq t \leq 1} |Z(t, c_*)|.$$

For each $0 \leq t \leq T$, let

$$(D.60) \quad Z_m(t) = \frac{\mathbf{q}_1(0, k_* + \lfloor (m/\Delta)^{1/2}t \rfloor) + \mathbf{q}_{2,2}(0, k_* + \lfloor (m/\Delta)^{1/2}t \rfloor) + k_*^2 U_m(\bar{h}; k_*)}{m}$$

and

$$Z_{m,L}(t) = \frac{\mathbf{q}_1(0, k_* + \lfloor (m/\Delta)^{1/2}t \rfloor) + \mathbf{q}_{2,2}(0, k_* + \lfloor (m/\Delta)^{1/2}t \rfloor) + k_*^2 U_{m,L}(\bar{h}; k_*)}{m},$$

where $U_{m,L}$ is given by (C.22). Clearly,

$$\mathbf{E} z_i \phi_\ell(\mathbf{X}_i) = \sigma^{-1} \mathbf{E} v(\mathbf{X}_i) \phi_\ell(\mathbf{X}_i) = \eta_\ell.$$

Hence, we deduce the joint weak convergence

$$(D.61) \quad \begin{aligned} & m^{-1/2} (S_1(m), \dots, S_L(m), S_1(\lfloor mt \rfloor, m), \dots, S_L(\lfloor mt \rfloor, m), V_{1,m}, V_{2,m}) \\ & \Rightarrow (W_{1,1}(1), \dots, W_{1,L}(1), W_{2,1}(t), \dots, W_{2,L}(t), V_1(1), V_2(c_*)), \quad \text{in } \mathbf{D}[0, T]. \end{aligned}$$

Lemma C.3 implies (c.f. (D.1))

$$\left| \frac{k_*^2 U_{m,L}(\bar{h}; k_*)}{m} - \sum_{\ell=1}^L \lambda_\ell \left(\frac{1}{m} \left(S_\ell(\lfloor mt \rfloor, m) - \frac{\lfloor mt \rfloor}{m} S_\ell(m) \right)^2 - \frac{\lfloor mt \rfloor (\lfloor mt \rfloor + m)}{m^2} \right) \right| = o_P(1).$$

Hence, we deduce that as $m \rightarrow \infty$,

$$\begin{aligned}
& Z_{m,L}(t) \\
&= \frac{\lfloor (m/\Delta)^{1/2} t \rfloor^2 \theta (\nu_1 - \nu_2)}{m} + \frac{2\theta \sigma_* \lfloor (m/\Delta)^{1/2} t \rfloor}{m^{1/2}} \left[\frac{V_{2,m}(0)}{m^{1/2}} - \frac{c_* m + \lfloor (m/\Delta)^{1/2} t \rfloor}{m} \frac{V_{1,m}}{m^{1/2}} \right] \\
&\quad + \frac{k_*^2 U_{m,L}(\bar{h}, k_*)}{m} \\
& \Rightarrow Z_L(t), \quad \text{in } \mathbf{D}[0, T],
\end{aligned}
\tag{D.62}$$

where (recalling $\sigma_* \theta / \Delta^{1/2} = \sigma_* / \mathfrak{D}_h(F, G)^{1/2}$ and ζ_* in Assumption 3.2),

$$Z_L(t) = t^2 + 2\zeta_* t (V_2(c_*) - c_* V_1(1)) - \sum_{\ell=1}^L \lambda_\ell [(W_{2,\ell}(c_*) - c_* W_{1,\ell}(1))^2 - c_*(1 + c_*)].$$

Moreover, since $\sum_{\ell \geq 1} \lambda_\ell^2 < \infty$, an application of Cauchy-Schwarz gives

$$\lim_{L \rightarrow \infty} \limsup_{m \rightarrow \infty} P \left\{ \sup_{0 \leq t \leq T} |Z_{m,L}(t) - Z_L(t)| > x \right\} = 0.
\tag{D.63}$$

So, if we now let

$$\begin{aligned}
& Z(t) = t^2 + 2\zeta_* t (V_2(c_*) - c_* V_1(1)) - \sum_{\ell=1}^{\infty} \lambda_\ell [(W_{2,\ell}(c_*) - c_* W_{1,\ell}(1))^2 - c_*(1 + c_*)] \\
& \Rightarrow \zeta t^2 + 2\zeta_* t (V_2(c_*) - c_* V_1(1)) + \mathbb{V}(0, c_*),
\end{aligned}
\tag{D.64}$$

it is easily seen that $\sup_{0 \leq t \leq T} |Z_L(t) - Z(t)| = o_P(1)$, implying $Z_L \Rightarrow Z$ in $\mathbf{D}[0, T]$, which together with (D.62) and (D.63) gives

$$Z_m \Rightarrow Z \text{ in } \mathbf{D}[0, T].$$

Then, the continuous mapping theorem gives

$$(D.65) \quad \max_{k_* < k \leq k_* + (m/\Delta)^{1/2}T} \frac{\mathcal{D}_m(k)}{g_m(k)} = \sup_{0 \leq t \leq T} \frac{|Z_m(t)|}{g(c_* + \lfloor (m/\Delta)^{1/2}t \rfloor / m)} + o_P(1) \Rightarrow \sup_{0 \leq t \leq T} \frac{|Z(t)|}{g(c_*)}.$$

In other words,

$$\max_{k_* < k \leq k_* + (m/\Delta)^{1/2}T} \frac{\mathcal{D}_m(k)}{g_m(k)} \Rightarrow \sup_{0 \leq t \leq T} |Z(t, c_*)|.$$

Thus,

$$\begin{aligned} P \left\{ \frac{\kappa_m - k_*}{(m/\Delta)^{1/2}} < x \right\} &= P \{ \kappa_m < k_* + x(m/\Delta)^{1/2} \} \\ &= P \left\{ \max_{k_* \leq k \leq k_* + x(m/\Delta)^{1/2}} \frac{\mathcal{D}_m^{(1)}(k)}{g_m(k)} > c \right\} \\ &\rightarrow P \left\{ \sup_{0 \leq t \leq x} |Z(t, c_*)| > c \right\} \\ &= P \{ \mathcal{H}_{c_*}(c) < x \}, \end{aligned}$$

as was to be shown. □

Proof of Theorem 3.5. For any $y > k_*$,

$$(D.66) \quad P \{ \kappa_m < y \} = P \left\{ \max_{k_* < k \leq y} \frac{\mathcal{D}_m^{(2)}(k)}{mg_m(k)} > c \right\}$$

$$(D.67) \quad = P \left\{ \max_{k_* < k \leq y} \frac{\max_{0 \leq r < k} w^2 |U_m(h; r, k)|}{mg_m(k)} > c \right\}.$$

The argument for part (i) is essentially the same as in the proof of Theorem 3.4(i), so we highlight only the main differences. With $y = y_m > k_*$ as in (D.31), from the bounds (D.37)–(D.40), it suffices to show

$$(D.68) \quad b_m^{-1} \left(\max_{k_* < k \leq y_m} \frac{\max_{0 \leq r < k} |\mathfrak{q}_1(r, k) + \mathfrak{q}_{2,1}(r, k)|}{mg_m(k)} - a_m \right) \Rightarrow \mathcal{N}(0, 1),$$

where $\mathbf{q}_1(r, k)$ is given in (C.46) and $\mathbf{q}_{2,1}(r, k)$ is in (D.30). Now, since $\max_{0 \leq r < k} |\mathbf{q}_1(r, k)| = |\mathbf{q}_1(0, k)|$, from (D.41), we have

$$(D.69) \quad b_m^{-1} \left(\max_{k_* < k \leq y_m(1-\delta)} \frac{\max_{0 \leq r < k} |\mathbf{q}_1(r, k) + \mathbf{q}_{2,1}(r, k)|}{mg_m(k)} - a_m \right) \xrightarrow{P} -\infty.$$

On the other hand, an analogous argument leading to (D.47) shows

$$\lim_{\delta \rightarrow 0} \limsup_{m \rightarrow \infty} P \left\{ b_m^{-1} \left(\max_{y_m(1-\delta) \leq k \leq y_m} \frac{\max_{0 \leq r < k} |\mathbf{q}_1(r, k) + \mathbf{q}_{2,1}(r, k)|}{mg_m(k)} - \frac{|\mathbf{q}_1(y) + \mathbf{q}_{2,1}(r, y)|}{mg_m(r, y)} \right) > x \right\} = 0.$$

Hence from (D.48), we obtain

$$(D.70) \quad P\{\kappa_m < y_m\} \rightarrow 1 - \Phi(-x) = \Phi(x), \quad m \rightarrow \infty,$$

and the rest of the proof is identical to that of Theorem 3.4(i).

Now we turn to part (ii). Recall $k_* = c_* m$. Write

$$\overline{\mathbf{q}}(r, k) = \begin{cases} \mathbf{q}_1(r, k) + \mathbf{q}_{2,2}(r, k) + (k_* - r)^2 U_m(\bar{h}; r, k_*) & 0 \leq r \leq k_*, \\ \mathbf{q}_1(r, k) + \mathbf{q}_{2,2}(r, k) + \mathbf{q}_3(r, k) & k > k_*. \end{cases}$$

Using the bounds above, and Lemma C.6, it can be shown that

$$(D.71) \quad \max_{k_* < k \leq k_* + (m/\Delta)^{1/2} T} \max_{0 \leq r \leq k} \left| \frac{U_m(h; r, k)}{g_m(k)} - \frac{\overline{\mathbf{q}}(r, k)}{mg_m(k)} \right| = o_P(1).$$

Define

$$Y_{m,1}(s, t) = \overline{\mathbf{q}}(\lfloor ms \rfloor, k_* + \lfloor (m/\Delta)^{1/2} t \rfloor), \quad 0 \leq s \leq c_* \quad 0 \leq t \leq T,$$

$$Y_{m,2}(s, t) = \overline{\mathbf{q}}(k_* + \lfloor (m/\Delta)^{1/2} (s \wedge t) \rfloor, k_* + \lfloor (m/\Delta)^{1/2} t \rfloor), \quad 0 \leq s, t \leq T,$$

so that

$$(D.72) \quad \begin{aligned} \sup_{0 \leq s \leq c_*} |Y_{m,1}(s, t)| &= \max_{0 \leq r \leq k_*} |\overline{\mathcal{Q}}(r, k_* + \lfloor (m/\Delta)^{1/2} t \rfloor)|, \\ \sup_{0 \leq s \leq t} |Y_{m,2}(s, t)| &= \max_{k_* < r < k_* + \lfloor (m/\Delta)^{1/2} t \rfloor} |\overline{\mathcal{Q}}(r, k_* + \lfloor (m/\Delta)^{1/2} t \rfloor)|. \end{aligned}$$

With $V_{m,1}$ and $V_{m,2}(r)$ as in (D.53), we have

$$(D.73) \quad \begin{aligned} Y_{m,1}(s, t) &= \frac{\lfloor (m/\Delta)^{1/2} t \rfloor^2 \theta(\nu_1 - \nu_2)}{m} \\ &\quad + \frac{2\theta\sigma_* \lfloor (m/\Delta)^{1/2} t \rfloor}{m^{1/2}} \left[\frac{V_{2,m}(\lfloor ms \rfloor)}{m^{1/2}} - \frac{\lfloor (m/\Delta)^{1/2} t \rfloor + k_* - \lfloor ms \rfloor}{m} \frac{V_{1,m}}{m^{1/2}} \right] \\ &\quad + \frac{(k_* - \lfloor ms \rfloor)^2 U_m(\bar{h}, \lfloor ms \rfloor, k_*)}{m}, \end{aligned}$$

and

$$(D.74) \quad \begin{aligned} Y_{m,2}(t) &= \frac{(\lfloor (m/\Delta)^{1/2} t \rfloor - \lfloor (m/\Delta)^{1/2} s \rfloor)^2 \theta(\nu_1 - \nu_2)}{m} \\ &\quad - \frac{2\theta\sigma_* (\lfloor (m/\Delta)^{1/2} t \rfloor - \lfloor (m/\Delta)^{1/2} s \rfloor)}{m^{1/2}} \left[\frac{\lfloor (m/\Delta)^{1/2} t \rfloor - \lfloor (m/\Delta)^{1/2} s \rfloor}{m} \frac{V_{1,m}}{m^{1/2}} \right] \\ &\quad - (\lfloor (m/\Delta)^{1/2} t \rfloor - \lfloor (m/\Delta)^{1/2} s \rfloor)^2 \frac{2R_{m,2}}{m(m-1)} \end{aligned}$$

Note (e.g., [Shorack and Wellner \(1986\)](#))

$$m \frac{2R_{m,2}}{m(m-1)} \Rightarrow \sum_{\ell \geq 1} \lambda_\ell (W_{1,\ell}(1)^2 - 1) = \chi,$$

where $W_{1,\ell}$ are as in (D.50). Arguing as in (D.61), we deduce the joint weak convergence

$$\left(V_{1,m}, V_{2,m}(\lfloor ms \rfloor), (k_* - \lfloor ms \rfloor)^2 U_m(\bar{h}, \lfloor ms \rfloor, k_*), \frac{2R_{m,2}}{(m-1)} \right) \Rightarrow (V_1(1), V_2(s), \mathbb{V}(s, c_*), \chi),$$

in $\mathbf{D}^4[0, c_*]$. Then, the Dudley-Wichura-Skorokhod Theorem gives for each $m \geq 1$, Wiener processes $V_1^{(m)}, V_2^{(m)}$, a process $\{\mathbb{V}^{(m)}(s, c_*), 0 \leq s \leq c_*\} \stackrel{\mathcal{D}}{=} \{\mathbb{V}(s, c_*), 0 \leq s \leq c_*\}$ and a variable $\chi^{(m)}$ such that $(V_1^{(m)}(1), V_2^{(m)}(s), \mathbb{V}^{(m)}(s, c_*), \chi^{(m)})^\top \stackrel{\mathcal{D}}{=} (V_1(1), V_2(s), \mathbb{V}(s, c_*), \chi)^\top$ in

$\mathbf{C}^4[0, c_*]$ satisfying

$$\sup_{0 \leq s \leq c_*} |\mathbb{V}^{(m)}(s, c_*) - (k_* - \lfloor ms \rfloor)^2 U_m(\bar{h}, \lfloor ms \rfloor, k_*)| = o_P(1), \quad \left| m^{-1/2} V_{m,1} - V_1^{(m)}(1) \right| = o_P(1),$$

$$\sup_{0 \leq s \leq c_*} \left| V_2^{(m)}(s) - \frac{V_{2,m}(0) - V_{2,m}(\lfloor ms \rfloor)}{m^{1/2}} \right| = o_P(1), \quad \left| \chi^{(m)} - \frac{2R_{m,2}}{(m-1)} \right| = o_P(1).$$

This gives

$$\sup_{0 \leq t \leq T} \sup_{0 \leq s \leq c_*} \left| Y_{m,1}(s, t) - \zeta t^2 - 2\zeta_* t \left(V_2^{(m)}(c_*) - V_2^{(m)}(s) - c_*(1-s)V_1^{(m)}(1) \right) - \mathbb{V}^{(m)}(s, c_*) \right| = o_P(1),$$

and

$$\sup_{0 \leq s, t \leq T} |Y_{m,2}(s, t) - \zeta(t-s)^2 + \chi^{(m)}| = o_P(1).$$

In particular, in view of (D.72), and the convergence $\max_{k_* < k \leq k_* + (m/\Delta)^{1/2}T} g_m(k) \rightarrow g(c_*)$, we obtain

$$\begin{aligned} \max_{k_* < k \leq k_* + (m/\Delta)^{1/2}T} \max_{0 \leq r \leq k} \left| \frac{\overline{q}(r, k)}{mg_m(k)} \right| &\Rightarrow \frac{1}{g(c_*)} \sup_{0 \leq t \leq T} \max \left\{ \sup_{0 \leq s \leq c_*} |Y_1(s, t)|, \sup_{0 \leq s \leq t} |Y_2(s, t)| \right\} \\ &= \sup_{0 \leq t \leq T} Y(t, c_*). \end{aligned}$$

with

$$(D.75) \quad Y_1(s, t) = \zeta t^2 + 2\zeta_* t (V_2(c_*) - V_2(s) - c_*(1-s)V_1(1)) + \mathbb{V}(s, c_*),$$

$$(D.76) \quad Y_2(s, t) = \zeta(t-s)^2 + \chi.$$

Since Y_1 and Y_2 are continuous, $Y(t)$ is nondecreasing and continuous, hence we may define

$$(D.77) \quad \tilde{\mathcal{H}}_{c_*}(u) = \inf \left\{ x \geq 0 : \sup_{0 \leq s \leq x} |Y(t, c_*)| \geq u \right\}.$$

Recalling (D.71), we finally have

$$\begin{aligned}
P \left\{ \frac{\kappa_m - k_*}{(m/\Delta)^{1/2}} < x \right\} &= P \{ \kappa_m < k_* + x(m/\Delta)^{1/2} \} \\
&= P \left\{ \max_{k_* \leq k \leq k_* + x(m/\Delta)^{1/2}} \frac{\mathcal{D}_m^{(2)}(k)}{g_m(k)} > \mathfrak{c} \right\} \\
&\rightarrow P \left\{ \sup_{0 \leq t \leq x} |Y(t, c_*)| > \mathfrak{c} \right\} \\
&= P \left\{ \tilde{\mathcal{H}}_{c_*}(\mathfrak{c}) < x \right\},
\end{aligned}$$

as was to be shown. □

Proof of Theorem 4.1. The proof is largely the same as Theorem 3.1; we provide a sketch and highlight main differences. Note $k > w$ if and only if $k > c_w m$. With f_ℓ as in (C.6), we clearly have $U_m(f_\ell, w; k) = U_m(f_\ell; k)$ for $2 \leq k \leq w$. For $k > w$,

$$\begin{aligned}
&m^{-1} w^2 U_m(f_\ell, w; k) \\
&= -m^{-1} \left(S_\ell(k, m+r) - \frac{w}{m+r} S_\ell(m+r) \right)^2 + \frac{w(m+k)}{m(m+r)} + \tilde{R}_\ell(k, m),
\end{aligned}$$

where $r = (k \vee w) - w$. Above, $\tilde{R}_\ell(k, m) = R_\ell(k, m)$ as in (C.10) when $k \leq w$, and when $k > w$,

(D.78)

$$\begin{aligned}
&\tilde{R}_\ell(k, m) \\
&= \frac{1}{m} \left(-\frac{w^2 S_\ell(m+r)}{(m+r)^2(m+r-1)} + \frac{w^2}{(m+r)(m+r-1)} \sum_{i=1}^{m+r} (\phi_\ell^2(\mathbf{X}_i) - 1) \right. \\
&\quad \left. - \frac{S_\ell(k, m+r)}{(w-1)} + \frac{w}{w-1} \sum_{j=m+r+1}^{m+k} (\phi_\ell^2(\mathbf{X}_j) - 1) + \frac{w}{(w-1)} + \frac{w^2}{(m+r)(m+r-1)} \right),
\end{aligned}$$

Then, from (4.2), with $U_{m,L}(h, w; k) = \sum_{\ell=1}^L \lambda_\ell U_m(f_\ell, w; k)$, we have for $k > w$,

$$\begin{aligned} & m^{-1} w^2 U_{m,L}(h, w; k) \\ &= - \sum_{\ell=1}^L \lambda_\ell \left(\frac{1}{m} \left(S_\ell(k, m+r) - \frac{w}{m+r} S_\ell(m+r) \right)^2 - \frac{w(m+k)}{m(m+k-w)} \right) + \sum_{\ell=1}^L \lambda_\ell \tilde{R}_\ell(k, m). \end{aligned}$$

Note for any $2 \leq k_1 < k_2$,

$$S_\ell(k_2, k_1) = S_\ell(k_2 + k_1) - S_\ell(k_1).$$

So, letting

$$\begin{aligned} \bar{S}_{\ell,m}(w, t) &= \begin{cases} S_\ell(\lfloor mt \rfloor, m) - \frac{\lfloor mt \rfloor}{m} S_\ell(m) & 0 \leq t \leq c_w \\ S_\ell(\lfloor mt \rfloor, \lfloor mt \rfloor - w + m) - \frac{w}{\lfloor mt \rfloor - w + m} S_\ell(\lfloor mt \rfloor - w + m) & t > c_w, \end{cases} \\ &= \begin{cases} S_\ell(\lfloor mt \rfloor + m) - S_\ell(m) - \frac{\lfloor mt \rfloor}{m} S_\ell(m) & 0 \leq t \leq c_w \\ S_\ell(\lfloor mt \rfloor + m) - S_\ell(\lfloor mt \rfloor - w + m) - \frac{w}{\lfloor mt \rfloor - w + m} S_\ell(\lfloor mt \rfloor - w + m) & t > c_w, \end{cases} \end{aligned}$$

we readily deduce, for any $T > 0$,

$$\{m^{-1/2} (\bar{S}_{1,m}(w, t), \dots, S_{L,m}(w, t)), 0 \leq t \leq T\} \Rightarrow \{\bar{W}_\ell(t), 0 \leq t \leq T\},$$

in $\mathbf{D}[0, T]$, where, writing $w(t) = c_w + b(t - c_w)$,

$$\begin{aligned} \bar{W}_\ell(t) &= \begin{cases} W_\ell(t+1) - W_\ell(1) - tW_\ell(1), & 0 \leq t \leq c_w, \\ W_\ell(t+1) - W_\ell(t+1-w(t)) - \frac{w(t)}{t-w(t)+1} W_\ell(t-w(t)+1) & t > c_w. \end{cases} \\ &= \begin{cases} W_{2,\ell}(t) - tW_{1,\ell}(1), & 0 \leq t \leq c_w, \\ W_{2,\ell}(t) - \frac{t+1}{t-w(t)+1} W_{2,\ell}(t-w(t)) - \frac{w(t)}{t-w(t)+1} W_{1,\ell}(1) & t > c_w. \end{cases} \end{aligned}$$

Above, $W_{2,\ell}(t) = W_\ell(t+1) - W_\ell(1)$, and $W_{1,\ell}(1) = W_\ell(1)$, which are clearly independent.

Thus, with $\mathbf{v}(t) = (t - \mathbf{w}(t)) \vee 0$, we have

$$\mathbf{v}(t) = \begin{cases} 0 & 0 \leq t \leq c_w \\ t - \mathbf{w}(t) & t > c_w \end{cases}, \quad t - \mathbf{v}(t) = \begin{cases} t & 0 \leq t \leq c_w \\ \mathbf{w}(t) & t > c_w. \end{cases}$$

Thus, writing $\mathbf{v} = \mathbf{v}(t)$ for simplicity, we have

$$\begin{aligned} (\mathbf{v} + 1)\overline{W}_\ell(t) &= (\mathbf{v} + 1)W_{2,\ell}(t) - (t + 1)W_{2,\ell}(\mathbf{v}) + (t - \mathbf{v})W_{1,\ell}(1) \\ &= (\mathbf{v} + 1)(W_{2,\ell}(t) - tW_{1,\ell}(1)) - (t + 1)(W_{2,\ell}(\mathbf{v}) - \mathbf{v}W_{1,\ell}(1)). \end{aligned}$$

From (D.12), we have

$$\begin{aligned} &\{\overline{W}_\ell(t), t \geq 0, \ell \geq 1\} \\ &\stackrel{\mathcal{D}}{=} \left\{ \frac{1}{\mathbf{v} + 1} \left[(\mathbf{v} + 1)(t + 1)W_\ell\left(\frac{t}{1 + t}\right) - (1 + t)(1 + \mathbf{v})W_\ell\left(\frac{\mathbf{v}}{1 + \mathbf{v}}\right) \right], t \geq 0, \ell \geq 1 \right\} \\ (D.79) \quad &\stackrel{\mathcal{D}}{=} \left\{ (t + 1)W_\ell\left(\frac{t}{1 + t}\right) - (1 + t)W_\ell\left(\frac{\mathbf{v}(t)}{1 + \mathbf{v}(t)}\right), t \geq 0, \ell \geq 1 \right\}. \end{aligned}$$

If we write

$$\mathcal{U}_{m,L}(w, t) = \frac{m^{-1}(\lfloor mt \rfloor \wedge w)^2 U_{m,L}(h, w; \lfloor mt \rfloor)}{g_m(\lfloor mt \rfloor)},$$

we may argue analogously as in (D.3) and (D.5) to obtain the weak convergence

$$\mathcal{U}_{m,L}(w, \cdot) \xrightarrow{m \rightarrow \infty} \mathcal{V}_L(b_w, c_w, \cdot), \quad \text{and} \quad \mathcal{V}_L(b_w, c_w, \cdot) \xrightarrow{L \rightarrow \infty} \mathcal{V}(c_w, \cdot) \quad \text{in } \mathbf{D}[\delta, T],$$

for any $0 < \delta < T$, where, for $L \in \{1, 2, \dots, \infty\}$,

$$\mathcal{V}_L(b_w, c_w, t) = -\frac{1}{g(t)} \sum_{\ell=1}^L \lambda_\ell \left[\overline{W}_\ell^2(t) - \frac{(t - \mathbf{v}(t))(1 + t)}{(1 + \mathbf{v}(t))} \right].$$

Arguing analogously to (D.10), we may obtain $\mathcal{U}_m(w, t) \Rightarrow \mathcal{V}_\infty(c_w, t)$. Finally, (D.79) gives

$$\begin{aligned}
& \sup_{t \geq 0} |\mathcal{V}_\infty(t, c_w)| \\
& \stackrel{\mathcal{D}}{=} \sup_{t \geq 0} \left(\frac{t}{1+t} \right)^{-\beta} \left| \sum_{\ell=1}^{\infty} \lambda_\ell \left[\left(W_\ell \left(\frac{t}{1+t} \right) - W_\ell \left(\frac{\mathbf{v}(t)}{1+\mathbf{v}(t)} \right) \right)^2 - \frac{t - \mathbf{v}(t)}{1+\mathbf{v}(t)} \right] \right| \\
& = \sup_{t \geq 0} \left(\frac{t}{1+t} \right)^{-\beta} \left| \sum_{\ell=1}^{\infty} \lambda_\ell \left[\left(W_\ell \left(\frac{t}{1+t} \right) - W_\ell \left(\frac{\mathbf{v}(t)}{1+\mathbf{v}(t)} \right) \right)^2 - \left(\frac{t}{1+t} - \frac{\mathbf{v}(t)}{1+\mathbf{v}(t)} \right) \right] \right| \\
& = \sup_{0 \leq u \leq 1} u^{-\beta} \left| \sum_{\ell=1}^{\infty} \lambda_\ell \left[(W_\ell(u) - W_\ell(\mathbf{y}(u)))^2 - (u - \mathbf{y}(u)) \right] \right|.
\end{aligned}$$

On the last line, we used the relation $\mathbf{v}(t) = t - \mathbf{w}(t) = (t - c_w)(1 - b_w)$ for $t > c_w$, giving

$$\frac{\mathbf{v} \left(\frac{u}{1-u} \right)}{1 + \mathbf{v} \left(\frac{u}{1-u} \right)} = \frac{\left(\frac{u}{1-u} - c_w \right) (1-b)}{1 + \left(\frac{u}{1-u} - c_w \right) (1-b)}, \quad \frac{c_w}{1+c_w} \leq u \leq 1$$

The remainder of the proof is the same as that of Theorem 3.1(ii), mutatis mutandis. $\square$

Proof of Theorem 4.2. We begin by noting that, following the proofs of Theorems 3.1–3.2, one can show that $\mathfrak{R}(k)$ can be written as

$$\begin{aligned}
\mathfrak{R}(k) &= \frac{2}{k(m-k)} \sum_{i=1}^k \sum_{j=k+1}^m \sum_{\ell=1}^{\infty} \lambda_\ell \phi_\ell(\mathbf{X}_i) \phi_\ell(\mathbf{X}_j) - \frac{1}{k(k-1)} \sum_{1 \leq i \neq j \leq k} \sum_{\ell=1}^{\infty} \lambda_\ell \phi_\ell(\mathbf{X}_i) \phi_\ell(\mathbf{X}_j) \\
&\quad - \frac{1}{(m-k)(m-k-1)} \sum_{k+1 \leq i \neq j \leq m} \sum_{\ell=1}^{\infty} \lambda_\ell \phi_\ell(\mathbf{X}_i) \phi_\ell(\mathbf{X}_j),
\end{aligned}$$

up to negligible terms. The theorem now follows immediately from Lemma C.10–C.13. $\square$

Proof of Theorem 4.3. Recall that, by the Moore-Aronszajn theorem, the positive (semi)definite kernel $K(x, y)$ yields a unique RKHS $\mathcal{H}_K$ of real-valued functions on $\mathfrak{X}$ with reproducing kernel $K(\cdot, \cdot)$. Consider the map $x \rightarrow K(\cdot, x) = \varphi(x)$. By assumption, this is injective; moreover, seeing as $K(\cdot, \cdot)$ is continuous, so is $\varphi(x)$ by Lemma 4.29 in Christmann and Steinwart

(2008), and $\mathcal{H}_K$ is separable (Lemma 4.33 in Christmann and Steinwart, 2008). By the reproducing property, it follows that

$$K(x, y) = \langle \varphi(x), \varphi(y) \rangle_{\mathcal{H}_K},$$

and therefore

$$\delta^{1/2}(x, y) = [K(x, x) + K(y, y) - 2K(x, y)]^{1/2} = \|\varphi(x) - \varphi(y)\|_{\mathcal{H}_K}.$$

Hence, by Theorem 3.16 in Lyons (2013), the space $(\mathcal{H}_K, \rho)$ is of strong negative type, having defined $\rho(x, y) = \|x - y\|_{\mathcal{H}_K}$. In other words, if $\mathbb{P}_1$ and $\mathbb{P}_2$ are two Borel measures defined on $\mathcal{H}_K$, given $Z, Z' \stackrel{i.i.d.}{\sim} \mathbb{P}_1$ and $W, W' \stackrel{i.i.d.}{\sim} \mathbb{P}_2$, the quantity

$$\mathcal{D}_\rho(\mathbb{P}_1, \mathbb{P}_2) = 2\mathbb{E} \|W - Z\|_{\mathcal{H}_K} - \mathbb{E} \|W - W'\|_{\mathcal{H}_K} - \mathbb{E} \|Z - Z'\|_{\mathcal{H}_K},$$

is zero if and only if $\mathbb{P}_1 = \mathbb{P}_2$. Consider now any two Borel probability measures ν_1 and ν_2 on $\mathcal{X}$, and let $\mathbb{P}_i = \nu_i \circ \varphi^{-1}$. Then, if $X, X' \stackrel{i.i.d.}{\sim} \nu_1$ and $Y, Y' \stackrel{i.i.d.}{\sim} \nu_2$, it holds that $\varphi(X) \sim \mathbb{P}_1$ and $\varphi(Y) \sim \mathbb{P}_2$, and

$$\begin{aligned} \mathcal{D}_\delta(\nu_1, \nu_2) &= 2\mathbb{E} \|\varphi(X) - \varphi(Y)\|_{\mathcal{H}_K} - \mathbb{E} \|\varphi(X) - \varphi(X')\|_{\mathcal{H}_K} - \mathbb{E} \|\varphi(Y) - \varphi(Y')\|_{\mathcal{H}_K} \\ &= \mathcal{D}_\rho(\mathbb{P}_1, \mathbb{P}_2). \end{aligned}$$

Hence, if $\nu_1 = \nu_2$, then for any Borel set $B \subseteq \mathcal{H}_K$, $\mathbb{P}_1(B) = \mathbb{P}_2(B)$. Consider now a compact set $A \subseteq \mathcal{H}_K$; then, $\varphi(A)$ also is compact - and therefore it is a Borel set in $\mathcal{H}_K$ - and therefore

$$\nu_1(A) = \mathbb{P}_1(\varphi(A)) = \mathbb{P}_2(\varphi(A)) = \nu_2(A).$$

Given that $\mathfrak{X}$ is a complete and separate metric space, every Borel measure is Radon (Theorem 7.1.7, [Bogachev and Ruas, 2007](#)); hence, for any Borel set $C \subseteq \mathfrak{X}$

$$\nu_1(C) = \{\sup \nu_1(D) : D \subseteq C, D \text{ compact}\} = \{\sup \nu_2(D) : D \subseteq C, D \text{ compact}\} = \nu_2(C).$$

□

Proof of Theorem 4.4. Fix a collection of independent Wiener processes $\{\{W_{i,\ell}(t), t \geq 0\}, \ell \geq 1, i = 1, 2\}$ independent of $\mathcal{F} = \sigma(X_1, X_2, \dots)$. Defining $\tilde{Y}(s, t)$ as in (C.31) based on $\{W_{i,\ell}, \ell \geq 1, i = 1, 2\}$, let

$$\widehat{\mathcal{V}}_m(s, t) = -(1+t)^{\beta-2} \sum_{\ell=1}^{\infty} \widehat{\lambda}_{\ell,m} \tilde{Y}_{\ell}(s, t) \stackrel{d}{=} \sum_{\ell=1}^{\infty} \widehat{\lambda}_{\pi(\ell),m} \tilde{Y}_{\ell}(s, t) =: \widehat{\mathcal{V}}_m(s, t; \pi)$$

where $\pi : \{1, 2, \dots\} \rightarrow \{1, 2, \dots\}$ is any permutation. Similarly, we may construct $\mathcal{V}(s, t)$ as in Lemma (C.5) based on this same sequence of Wiener processes, so that

$$\mathcal{V}(s, t) - \widehat{\mathcal{V}}_m(s, t; \pi) = -(1+t)^{\beta-2} \sum_{\ell=1}^{\infty} (\lambda_{\ell} - \widehat{\lambda}_{\pi(\ell),m}) \tilde{Y}_{\ell}(s, t).$$

Pick a sequence of permutations π_m such that, for each m ,

$$\sum_{\ell=1}^{\infty} (\lambda_{\ell} - \widehat{\lambda}_{\pi_m(\ell),m})^2 \leq \inf_{\pi} \sum_{\ell=1}^{\infty} (\lambda_{\ell} - \widehat{\lambda}_{\pi(\ell),m})^2 + 1/m.$$

Then,

$$\begin{aligned} \mathbb{E}[(\mathcal{V}(s, t) - \widehat{\mathcal{V}}_m(s, t; \pi_m))^2 | \mathcal{F}] &= (1+t)^{2(\beta-2)} \mathbb{E}[\tilde{Y}_1^2(s, t) | \mathcal{F}] \sum_{\ell=1}^{\infty} (\lambda_{\ell} - \widehat{\lambda}_{\pi_m(\ell),m})^2 \\ &= C(s, t) \sum_{\ell=1}^{\infty} (\lambda_{\ell} - \widehat{\lambda}_{\pi_m(\ell),m})^2 \rightarrow 0, \quad \text{a.s.}, \end{aligned}$$

where we used that $\inf_{\pi} \sum_{\ell=1}^{\infty} (\lambda_{\ell} - \widehat{\lambda}_{\pi(\ell),m})^2 \rightarrow 0$ a.s. as $m \rightarrow \infty$ as a consequence of ([Koltchinskii and Giné, 2000](#), Theorem 3.1)

In particular, this implies for each $n \geq 1$ and any $s_1, t_1, \dots, s_n, t_n \geq 0$,

$$(\mathcal{V}_m(s_1, t_1), \dots, \mathcal{V}_L(s_n, t_n)) \Rightarrow_{\mathcal{F}} (\mathcal{V}(s_1, t_1), \dots, \mathcal{V}(s_n, t_n)).$$

For tightness, recall in the proof of Lemma C.5 we have, for each $r > 1$, $\mathbb{E}|\tilde{Y}_\ell(s_1, t_1) - \tilde{Y}_\ell(s_2, t_2)|^{2r} \leq C(|t_1 - t_2| + |s_1 - s_2|)^{2ar}$ for some $0 < a < 1 - \beta$ (C may depend on r). Rosenthal's inequality yields, for $r > 1$, (c.f. (C.32))

$$\begin{aligned} & \mathbb{E} \left[|(1+t_1)^{2-\beta} \widehat{\mathcal{V}}_m(s_1, t_1) - (1+t_2)^{2-\beta} \widehat{\mathcal{V}}_m(s_2, t_2)|^{2r} | \mathcal{F} \right] \\ & \leq \liminf_{R \rightarrow \infty} \mathbb{E} \left[\left| \sum_{\ell=1}^R \widehat{\lambda}_{\ell,m} (\tilde{Y}_\ell(s_1, t_1) - Y_\ell(s_2, t_2)) \right|^{2r} | \mathcal{F} \right] \\ & \leq C_r \left[\sum_{\ell=1}^{\infty} |\lambda_\ell - \widehat{\lambda}_{\ell,m}|^{2r} \mathbb{E} |\tilde{Y}_\ell(s_1, t_1) - \tilde{Y}_\ell(s_2, t_2)|^{2r} + \left(\sum_{\ell=1}^{\infty} (\lambda_\ell - \widehat{\lambda}_{\ell,m})^2 \mathbb{E} |\tilde{Y}_\ell(s_1, t_1) - \tilde{Y}_\ell(s_2, t_2)|^2 \right)^r \right] \\ & \leq C \left(\sum_{\ell=1}^{\infty} (\lambda_\ell - \widehat{\lambda}_{\ell,m})^2 \right)^r (|t_1 - t_2| + |s_1 - s_2|)^{2ar} \\ & \leq C(\omega) (|t_1 - t_2| + |s_1 - s_2|)^{2ar}, \end{aligned}$$

for some $0 < a < 1 - \beta$ almost surely, where $C(\omega) > 0$ is a constant that depends on ω . By taking r large enough, we obtain that for a.s. ω , the law of $\{(1+t)^{2-\beta} \widehat{\mathcal{V}}_m(s, t), \ s, t \geq 0\}$ in $C([0, \infty) \times [0, \infty))$ is tight under $P(\cdot | \mathcal{F})(\omega)$, and hence the same is true of $\{\widehat{\mathcal{V}}_m(s, t), \ s, t \geq 0\}$. We obtain

$$\{\widehat{\mathcal{V}}_m(s, t), \ s, t \geq 0\} \Rightarrow_{\mathcal{F}} \{\mathcal{V}(s, t), \ s, t \geq 0\} \quad \text{in } C([0, \infty) \times [0, \infty)).$$

Since under $P(\cdot | \mathcal{F})$, for any $a_0 \in (0, \infty]$ it holds a.s. that

$$\sup_{0 < t < a_0} |\widehat{\mathcal{V}}_m(0, t)| \stackrel{\mathcal{D}}{=} \sup_{0 < u < \frac{a_0}{1+a_0}} u^{-\beta} |\widehat{\Gamma}_m(u)|, \quad \sup_{0 < t < a_0} \sup_{0 \leq s \leq t} |\widehat{\mathcal{V}}_m(s, t)| \stackrel{\mathcal{D}}{=} \sup_{0 < u < \frac{a_0}{1+a_0}} u^{-\beta} |\widehat{\Gamma}_m(u)|,$$

(c.f. (D.13) and (D.24)), the first two convergence statements in (4.11) hold. Similar arguments give the last convergence statement. $\square$