2002

On nonparametric methods for strongly and weakly dependent lattice data

Daniel John Nordman
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Statistics and Probability Commons

Recommended Citation
https://lib.dr.iastate.edu/rtd/1018

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand comer and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®
On nonparametric methods for strongly and weakly dependent lattice data

by

Daniel John Nordman

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY

Major: Statistics

Program of Study Committee:
Soumendra N. Lahiri, Major Professor
Mark S. Kaiser
William Q. Meeker, Jr.
Sunder Sethuraman
Yuhong Yang

Iowa State University
Ames, Iowa
2002

Copyright © Daniel John Nordman, 2002. All rights reserved.
This is to certify that the doctoral dissertation of

Daniel John Nordman

has met the dissertation requirements of Iowa State University

Signature was redacted for privacy.

Major Professor

Signature was redacted for privacy.

For the Major Program
## TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>On optimal spatial subsample size for variance estimation</td>
<td>3</td>
</tr>
<tr>
<td>Abstract</td>
<td>3</td>
</tr>
<tr>
<td>1  Introduction</td>
<td>3</td>
</tr>
<tr>
<td>2  Variance estimators via subsampling</td>
<td>7</td>
</tr>
<tr>
<td>2.1 The sampling structure</td>
<td>7</td>
</tr>
<tr>
<td>2.2 Subsampling designs and variance estimators</td>
<td>9</td>
</tr>
<tr>
<td>2.3 Assumptions</td>
<td>11</td>
</tr>
<tr>
<td>3  Variance expansions</td>
<td>13</td>
</tr>
<tr>
<td>4  Bias expansions</td>
<td>15</td>
</tr>
<tr>
<td>5  Asymptotically optimal subsample sizes</td>
<td>19</td>
</tr>
<tr>
<td>5.1 Theoretical optimal subsample sizes</td>
<td>19</td>
</tr>
<tr>
<td>5.2 Empirical subsample size determination</td>
<td>21</td>
</tr>
<tr>
<td>6  Examples</td>
<td>22</td>
</tr>
<tr>
<td>6.1 Examples in $\mathbb{R}^2$</td>
<td>22</td>
</tr>
<tr>
<td>6.2 Examples in $\mathbb{R}^d, d \geq 3$</td>
<td>23</td>
</tr>
<tr>
<td>7  Subsample design extensions</td>
<td>25</td>
</tr>
<tr>
<td>8  Preliminaries for the proofs</td>
<td>26</td>
</tr>
<tr>
<td>9  Proof of Theorem 3.1</td>
<td>29</td>
</tr>
<tr>
<td>9.1 Proof of Theorem 3.1(a)</td>
<td>30</td>
</tr>
<tr>
<td>9.2 Proof of Theorem 3.1(b)</td>
<td>34</td>
</tr>
<tr>
<td>10 Proofs for Sections 4, 5 and 6</td>
<td>42</td>
</tr>
<tr>
<td>10.1 Preliminaries</td>
<td>42</td>
</tr>
<tr>
<td>10.2 Proofs: Lemma 4.1; Corollaries 4.1, 6.1; Theorems 4.1, 4.2, 5.1, 6.1</td>
<td>49</td>
</tr>
<tr>
<td>11 Appendix 1: Proof of Theorem 9.1</td>
<td>54</td>
</tr>
</tbody>
</table>
On the approximation of differenced lattice point counts with application to statistical bias expansions

Abstract

1 Introduction

2 A counting problem in statistics

3 Problem statement

4 Main results

4.1 Approximations in \( \mathbb{R}^2 \)

4.2 An \( \mathbb{R}^2 \) extension

4.3 Results for \( \mathbb{R}^3 \)

5 Applications to statistical bias expansions

6 Proofs

6.1 Proof of Theorem 1

6.2 Proof of Theorem 2

6.3 Proof of Corollary 1

7 Proof of Theorem 3

8 Appendix: Supplementary proofs

Empirical likelihood confidence intervals for the mean of a long-range dependent process

Abstract

1 Introduction

2 Preliminaries

2.1 Process assumptions

2.2 Asymptotic distribution of the sample mean

3 Asymptotic distribution of blockwise empirical likelihood

3.1 Construction of blockwise empirical likelihood for the mean

3.2 Limit distribution of empirical likelihood ratio
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Empirical calibration of empirical likelihood confidence</td>
<td>125</td>
</tr>
<tr>
<td>4.1</td>
<td>Sampling window method</td>
<td>125</td>
</tr>
<tr>
<td>4.2</td>
<td>Confidence intervals for the mean</td>
<td>127</td>
</tr>
<tr>
<td>4.3</td>
<td>Empirical block selection</td>
<td>128</td>
</tr>
<tr>
<td>5</td>
<td>Proofs</td>
<td>129</td>
</tr>
<tr>
<td>5.1</td>
<td>Proof of Theorem 3</td>
<td>129</td>
</tr>
<tr>
<td>5.2</td>
<td>Proofs of Theorems 2 and 4</td>
<td>134</td>
</tr>
<tr>
<td>1</td>
<td>Introduction</td>
<td>140</td>
</tr>
<tr>
<td>1.1</td>
<td>Related Literature</td>
<td>143</td>
</tr>
<tr>
<td>2</td>
<td>Definition of spectral empirical likelihood</td>
<td>143</td>
</tr>
<tr>
<td>2.1</td>
<td>Estimating functions</td>
<td>143</td>
</tr>
<tr>
<td>2.2</td>
<td>Construction of spectral empirical likelihood</td>
<td>146</td>
</tr>
<tr>
<td>3</td>
<td>Preliminaries</td>
<td>148</td>
</tr>
<tr>
<td>3.1</td>
<td>Assumptions</td>
<td>148</td>
</tr>
<tr>
<td>3.2</td>
<td>Discussion: bootstrap and empirical likelihood mechanics</td>
<td>151</td>
</tr>
<tr>
<td>4</td>
<td>Main Results</td>
<td>152</td>
</tr>
<tr>
<td>4.1</td>
<td>A Wilks's theorem</td>
<td>152</td>
</tr>
<tr>
<td>4.2</td>
<td>Maximum empirical likelihood estimation</td>
<td>154</td>
</tr>
<tr>
<td>5</td>
<td>Applications</td>
<td>157</td>
</tr>
<tr>
<td>5.1</td>
<td>Whittle estimation</td>
<td>157</td>
</tr>
<tr>
<td>5.2</td>
<td>Goodness-of-fit tests</td>
<td>158</td>
</tr>
<tr>
<td>6</td>
<td>Extensions</td>
<td>160</td>
</tr>
<tr>
<td>6.1</td>
<td>Inference with parameter restrictions</td>
<td>160</td>
</tr>
<tr>
<td>6.2</td>
<td>Results with tapering</td>
<td>161</td>
</tr>
<tr>
<td>7</td>
<td>Proofs</td>
<td>162</td>
</tr>
<tr>
<td>8</td>
<td>Appendix</td>
<td>191</td>
</tr>
</tbody>
</table>
Introduction

This dissertation addresses the development of nonparametric inference procedures for data exhibiting dependencies of varied form and structure. Current statistical methodology for dependent data relies heavily on parametric models (such as ARMA, Spatial autoregressive models, etc.) to faithfully represent the underlying time or spatial process. In practical applications, such models often prove difficult to choose and at times even inadequate. Parametric modelling presupposes a fair amount of knowledge on the part of the statistician about the nature of dependence among samples, which is not available in many situations and can lead to model misspecification. As a consequence, inference drawn from a misspecified model is not reliable. It is important to develop alternative inference tools based on nonparametric methods (eg. resampling) that are inherently less sensitive to model misspecification and require little knowledge of the exact dependence structure of the data-generating process.

There are four completed research papers which constitute the dissertation:


Each paper deals with some aspect of inference on dependent samples observed (or located) on some lattice (eg. indexed by the integers $\mathbb{Z}^d$). Observational stretches in time series are a special case of such lattice data.

Let $R_n$ be a sampling region in $\mathbb{R}^d$, $t \in \mathbb{R}^d$, and suppose data are collected at $t + Z^d$ lattice points (sampling sites) within $R_n$. Paper [1] considers the problem of optimally implementing a spatial subsampling method for nonparametrically estimating the variance of statistics on $R_n$. The results are applicable to a wide variety of stationary random fields exhibiting weak dependence, characterized by fairly mild mixing conditions, and a broad range of sampling regions $R_n$. 
Paper [2] frames and addresses a mathematical problem in lattice point theory. Theoretical considerations for statistics of spatial lattice data (e.g., bias or expectation expansions) often require calculating, or adequately approximating, subtracted lattice point counts for large sets based on \( R_n \) and certain intersections of the form \( R_n \cap (t + R_n) \). Paper [2] provides new lattice point count estimation results which can facilitate asymptotic bias expansions for spatial statistics for both rectangular and *non-rectangular* sampling regions \( R_n \). The count approximation tools are valid for a large class of sampling regions in \( \mathbb{R}^2 \) and \( \mathbb{R}^3 \), including all convex sets in these dimensions, and help determine the influence of \( R_n \)'s geometry on the properties of a spatial statistic.

Papers [3] and [4] consider nonparametric inference on dependent time series data, allowing for many possible process dependence strengths. Let \( X_1, X_2, \ldots \) be a sequence of stationary random variables with autocovariance function \( r(k) = \text{Cov}(X_i, X_{i+k}) \) of the form \( r(k) \sim k^{-\alpha} L(k) \) for some \( \alpha \in (0, 1) \) and slowly varying function \( L \). Since \( \sum_{k=1}^{\infty} |r(k)| = \infty \), the process \( \{X_i\} \) exhibits strong or long-range dependence (LRD). Limit behaviors of estimators and test statistics under LRD are typically very different from their behaviors with weakly dependent or iid processes. Consequently, inference methods proposed for iid or weakly dependent data may not work for strongly dependent data, or at least require suitable modifications.

Paper [3] develops new empirical likelihood confidence intervals for the process mean \( \mathbb{E}(X_1) = \mu \) under LRD, using (time-domain) "blocking" techniques like those involved in subsampling and the block bootstrap. The paper shows that empirical likelihood provides valid nonparametric estimation of \( \mu \), even in those instances of LRD where the block bootstrap is known to fail. The confidence interval procedure remains valid for weakly dependent processes.

Paper [4] develops a spectral version of empirical likelihood which allows nonparametric, likelihood-based confidence region estimation and hypothesis testing for processes which could exhibit *either* LRD or weak dependence (including mixing processes). Inference is possible for spectral parameters like autocorrelations and Whittle parameters and spectral goodness-of-fit tests can also be made.
On optimal spatial subsample size for variance estimation

A paper to be submitted to the Annals of Statistics

Daniel J. Nordman and Soumendra N. Lahiri

Abstract

In this paper, we consider the problem of determining the optimal block size for a spatial subsampling method for spatial processes observed on regular grids. We derive expansions for the mean square error of the subsampling variance estimator, which yields an expression for the theoretical optimal block size. The theoretical optimal block size is shown to depend in an intricate way on the geometry of the spatial sampling region as well as on the characteristics of the underlying random field. Final expressions for the optimal block size make use of some nontrivial estimates of lattice point counts in shifts of convex sets. The expressions for the optimal block size are computed for sampling regions of a number of commonly-encountered shapes.

Key Words: Block bootstrap, block size, lattice point count, mixing, nonparametric variance estimation, random fields, spatial statistics, subsamples

1 Introduction

In this paper, we examine the problem of choosing subsample sizes to maximize the performance of subsampling methods for variance estimation. The data at hand are viewed as realizations of a stationary, weakly dependent, spatial lattice process. We consider the common scenario of sampling from sites of regular distance (e.g., indexed by the integer lattice $\mathbb{Z}^d$), lying within some region $R_n$ embedded in $\mathbb{R}^d$. Such lattice data appear often in time series, agricultural field trials, and remote sensing/image analysis (medical and satellite image processing).
For variance estimation via subsampling, the basic idea is to construct several "scaled-down" copies of the sampling region \( R_n \) (subsamples) that fit inside \( R_n \), evaluate the analog of \( \hat{\theta}_n \) on each of these subregions, and then compute a properly normalized sample variance from the resulting values. The \( R_n \)-sampling scheme is essentially recreated at the level of the subregions. Two subsampling designs are most typical: subregions can be maximally overlapping (OL) or devised to be non-overlapping (NOL). The accuracy (e.g. variance and bias) of subsample-based estimators depends crucially on the choice of subsample size.

To place our work into perspective, we briefly outline previous research in variance estimation with subsamples and theoretical size considerations. The concept of variance estimation through subsampling originated from analysis of weakly dependent time processes. To obtain subsamples from a stationary stretch \( Z(1),...,Z(n) \) in time, Carlstein (1986) first proposed the use of NOL blocks of length \( m < n \):

\[
\{Z(1 + (i-1)m),...,Z(im)\}, \text{ for } i = 1,...,\lfloor n/m \rfloor,
\]

while the sequence of subseries:

\[
\{Z(i),...,Z(i + m - 1)\} \text{ for } i = 1,...,n - m + 1
\]

provides OL subsamples of length \( m \) (cf. Künsch, 1989; Politis and Romano, 1993b). In each respective collection, evaluations of an analog statistic \( \hat{S}_i \) are made for each subseries and a normalized sample variance is calculated to estimate the parameter \( n \text{Var}(\hat{\theta}_n) \):

\[
S_n^2 = \frac{1}{J} \sum_{i=1}^{J} \frac{m(\hat{\theta}_i - \bar{\theta})^2}{J}, \quad \hat{\theta} = \frac{1}{J} \sum_{i=1}^{J} \hat{\theta}_i,
\]

where \( J = \lfloor n/m \rfloor \) \((J = n - m + 1)\) for the NOL (OL) subsample-based estimator. Carlstein (1986) and Fukuchi (1999) established the L2 consistency of the NOL and OL estimators, respectively, for the variance of a general (not necessarily linear) statistic. Politis and Romano (1993b) determined asymptotic orders of the variance \( O(m/n) \) and bias \( O(1/m) \) of the subsample variance estimators for linear statistics. For mixing time series, they found that a subsample size \( m \) proportional to \( n^{1/3} \) is optimal in the sense of minimizing the Mean Square Error (MSE) of variance estimation, concurring also with optimal block order for the moving block bootstrap variance estimator [Hall et al. (1995), Lahiri (1996)].

Cressie (1991, p. 492) conjectured the recipe for extending Carlstein's variance estimator to the general spatial setting, obtaining subsamples by tiling the sample region \( R_n \) with disjoint "congruent" subregions. Politis and Romano (1993a, 1994) have shown the consistency of subsample-based variance estimators for rectangular sampling/subsampling regions in \( \mathbb{R}^d \) (e.g. sampling sites observed on
Z^d \cap \prod_{i=1}^d [1, n_i]; integer translates of \prod_{i=1}^d [1, m_i] for subsamples. Garcia-Soidan and Hall (1997) and Possolo (1991) proposed similar estimators under an identical sampling scenario. For linear statistics, Politis and Romano (1993a) determined that a subsampling scaling choice \prod_{i=1}^d m_i = C[\prod_{i=1}^d n_i]^{d/(d+2)}, for some unknown C, minimizes the order of a variance estimator’s asymptotic MSE. Sherman and Carlstein (1994) and Sherman (1996) proved the MSE-consistency of NOL and OL subsample estimators, respectively, for the variance of general statistics in \mathbb{R}^2. Their work allowed for a more flexible sampling scheme: the “inside” of a simple closed curve defines a set \( D \subset [-1, 1]^2 \), \( Z^2 \cap nD \) (using a scaled-up copy of \( D \)) constitutes the set of sampling sites, and translates of \( mD \) within \( nD \) form subsamples. Sherman (1996) minimized a bound on the asymptotic order of the OL estimator’s MSE to argue that the best size choice for OL subsamples involves \( m = O(n^{1/2}) \) (coinciding with the above findings of Politis and Romano (1993a) for rectangular regions in \( \mathbb{R}^2 \)). Politis and Sherman (1998) have developed consistent subsampling methods for variance estimation with marked point process data [cf. Politis et al. (1999), Chapter 6].

Few theoretical and numerical recommendations for choosing subsamples have been offered in the spatial setting, especially with the intent of variance estimation. As suggested in the literature, an explicit theoretical determination of optimal subsample “scaling” (size) requires calculation of an order and associated proportionality constant for a given sampling region \( R_n \). Even for the few sampling situations where the order of optimal subsample size has been established, the exact adjustments to these orders are unknown and, quoting Politis and Romano (1993a), “important (and difficult) in practice.” Beyond the time series case with the univariate sample mean, the influence of the geometry and dimension of \( R_n \), as well as the structure of \( \hat{\theta}_n \), on precise subsample selection has not been explored. We attempt here to advance some ideas on the best size choice, both theoretically and empirically, for subsamples.

We work under the “smooth function” model of Hall (1992), where the statistic of interest \( \hat{\theta}_n \) can be represented as a function of sample means. We formulate a framework for sampling in \( \mathbb{R}^d \) where the sampling region \( R_n \) (say) is obtained by “inflating” a prototype set in the unit cube in \( \mathbb{R}^d \) and the subsampling regions are given by suitable translates of a scaled down copy of the sampling region \( R_n \). We consider both a non-overlapping version and a (maximal) overlapping version of the subsampling method. For each method, we derive expansions for the variance and the bias of the corresponding subsample estimator of \( \text{Var}(\hat{\theta}_n) \). The asymptotic variance of the spatial subsample estimator for the OL version turns out to be smaller than that of the NOL version by a constant factor \( K_1 \) (say) which depends solely on the geometry of the sampling region \( R_n \). In the time series case, Meketon and...
Schmeiser (1984), Künsch (1989), Hall et al. (1995) and Lahiri (1996) have shown in different degrees of generality that the asymptotic variance under the OL subsampling scheme, compared to the NOL one, is $K_1 = \frac{2}{3}$ times smaller. Results of this paper show that for rectangular sampling regions $R_n$ in $d$-dimensional space, the factor $K_1$ is given by $(\frac{2}{3})^d$. We list the factor $K_1$ for sampling regions of some common shapes in Table 1.

<table>
<thead>
<tr>
<th>Shape of $R_n$</th>
<th>Rectangle in $\mathbb{R}^d$</th>
<th>Sphere in $\mathbb{R}^3$</th>
<th>Circle in $\mathbb{R}^2$</th>
<th>Right triangle in $\mathbb{R}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_1$</td>
<td>$(2/3)^d$</td>
<td>$17\pi/315$</td>
<td>$\pi/4 - 4/(3\pi)$</td>
<td>$1/5$</td>
</tr>
</tbody>
</table>

In contrast, the bias parts of both the OL and NOL subsample variance estimators are (usually) asymptotically equivalent and depend on the covariance structure of the random field as well as on the geometry of the sampling region $R_n$. Since the bias term is typically of the same order as the number of lattice points lying near a subsample's boundary, determination of the leading bias term involves some nontrivial estimates of the lattice point counts over translated subregions. Counting lattice points in scaled-up sets is a hard problem and has received a lot of attention in Analytic Number Theory and in Combinatorics. Even for the case of the plane (i.e., $d = 2$), the counting results available in the literature are directly applicable to our problem only for a very restricted class of subregions (that have the so-called “smoothly winding border” [cf. van der Corput (1920), Huxley (1993, 1996)]. Here explicit expressions for the bias terms are derived for a more general class of sampling regions using some new estimates on the discrepancy between the number of lattice points and the volume of the shifted subregions in the plane and in three dimensional Euclidean space. In particular, our results are applicable to sampling regions that do not necessarily have “smoothly winding borders”.

Minimizing the combined expansions for the bias and the variance parts, we derive explicit expressions for the theoretical optimal block size for sampling regions of different shapes. To briefly describe the result for a few common shapes, suppose the sampling region $R_n$ is obtained by inflating a given set $R_0 \in (-\frac{1}{2}, \frac{1}{2})^d$ by a scaling constant $\lambda_n$ as $R_n = \lambda_n R_0$ and that the subsamples are formed by considering the translates of $\lambda_n R_0$. Then, the theoretically optimal choice of the block size $\lambda_n$ for the OL version is of the form

$$\lambda_{n}^{opt} = \left(\frac{\lambda_n^d B_0^2}{dK_0 r^d}\right) \left(1 + o(1)\right) \quad \text{as} \quad n \to \infty$$

for some constants $B_0$ and $K_0$ (coming from the bias and the variance terms, respectively), where $r^2$ is
a population parameter that does not depend on the shape of the sampling region \( R_n \) (see Theorem 5.1 for details). The following table lists the constants \( B_0 \) and \( K_0 \) for some shapes of \( R_n \). It follows from

<table>
<thead>
<tr>
<th>( R_n )</th>
<th>Sphere in ( \mathbb{R}^2 )</th>
<th>Cross in ( \mathbb{R}^2 )</th>
<th>Right triangle in ( \mathbb{R}^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( B_0 )</td>
<td>( \frac{3}{2} \sum_{k \in \mathbb{Z}^2}</td>
<td></td>
<td>k</td>
</tr>
<tr>
<td>( K_0 )</td>
<td>( \frac{34}{105} )</td>
<td>( \frac{4}{9} \cdot \frac{191}{192} )</td>
<td>( \frac{2}{5} )</td>
</tr>
</tbody>
</table>

Table 2 that, unlike the time series case, in higher dimensions the optimal block size critically depends on the shape of the spatial sampling region \( R_n \). It simplifies only slightly for the NOL subsampling scheme as the constant \( K_0 \) is unnecessary for computing optimal NOL subsamples, but the bias constant \( B_0 \) is often the same for both estimators from each version of subsampling. These expressions may be readily used to obtain 'plug-in' estimates of the theoretical optimal block lengths for use in practice.

The rest of the paper is organized as follows. In Section 2, we describe the spatial subsampling method and state the Assumptions used in the paper. In Sections 3 and 4, we respectively derive expansions for the variance and the bias parts of the subsampling estimators. Theoretical optimal block lengths are derived in Section 5. The results are illustrated with some common examples in Section 6. In Section 7, we describe the properties of variance estimators based on an alternative subsampling scheme. Proofs of all results are partitioned among Sections 8-12.

2 Variance estimators via subsampling

In Section 2.1, we frame the sampling design and the structure of the sampling region. Two methods of subsampling are presented in Section 2.2 along with corresponding nonparametric variance estimators. Assumptions and Conditions used in the paper are given in Section 2.3.

2.1 The sampling structure

To establish the sampling scheme used, we first assume all potential sampling sites are located on a translate of the rectangular integer lattice in \( \mathbb{R}^d \). For a fixed (chosen) vector \( t \in [-1/2, 1/2]^d \), we identify the \( t \)-translated integer lattice as \( Z^d \equiv t + Z^d \). Let \( \{Z(s) \mid s \in Z^d\} \) be a stationary weakly
dependent random field (hereafter r.f.) taking values in $\mathbb{R}^d$. (We use bold font as a standard to denote vectors in the space of sampling $\mathbb{R}^d$ and normal font for vectors in $\mathbb{R}^d$, including $Z(\cdot)$.) We suppose that the process $Z(\cdot)$ is observed at sampling sites lying within the sampling region $R_n \subset \mathbb{R}^d$. Namely, the collection of available sampling sites is

$$\{Z(s) : s \in R_n \cap Z^d\}.$$ 

To obtain the results in the paper, we assume that the sampling region $R_n$ becomes unbounded as the sample size increases. This will provide a commonly used "increasing domain" framework for studying asymptotics with spatial lattice data [cf. Cressie (1991)]. We next specify the structure of the regions $R_n$ and employ a formulation similar to that of Lahiri (1999ab).

Let $R_0$ be a Borel subset of $(-1/2, 1/2)^d$ containing an open neighborhood of the origin such that for any sequence of positive real numbers $\alpha_n \to 0$, the number of cubes of the scaled lattice $\alpha_n Z^d$ which intersect the closures $\overline{R_0}$ and $\overline{R_0}$ is $O((\alpha_n^{-1})^{d-1})$ as $n \to \infty$. Let $\Delta_n$ be a sequence of $d \times d$ diagonal matrices, with positive diagonal elements $\lambda_1^{(n)}, \ldots, \lambda_d^{(n)}$, such that each $\lambda_i^{(n)} \to \infty$ as $n \to \infty$. We assume that the sampling region $R_n$ is obtained by "inflating" the template set $R_0$ by the directional scaling factors $\Delta_n$; that is,

$$R_n = \Delta_n R_0.$$ 

Because the origin is assumed to lie in $R_0$, the sampling region $R_n$ grows outward in all directions as $n$ increases. Furthermore, if the scaling factors are all equal ($\lambda_1^{(n)} = \cdots = \lambda_d^{(n)}$), the shape of $R_n$ remains the same for different values of $n$. Note that the location of sampling sites on $Z^d$, rather than $Z^d$, implies that the region $R_n$ does not necessarily expand around a potential sampling site at the origin.

The formulation given above allows the sampling region $R_n$ to have a large variety of fairly irregular shapes with the boundary condition on $R_0$ imposed to avoid pathological cases. Some common examples of such regions are convex subsets of $\mathbb{R}^d$, such as spheres, ellipsoids, polyhedrons, as well as certain non-convex subsets with irregular boundaries, such as star-shaped regions. Sherman and Carlstein (1994) and Sherman (1996) consider a similar class of such regions in the plane (i.e. $d = 2$) where the boundaries of the sets $R_0$ are delineated by simple rectifiable curves with finite lengths. The border requirements on $R_0$ ensure that the number of observations near the boundary of $R_n$ is negligible compared to the totality of data values. This assumption is crucial to the derivation of results to follow. In addition, Perera (1997) demonstrates that the border geometry of $R_0$ can significantly influence the asymptotic distribution of sample means taken with observations from "expanding" sets, like $R_n$. 
2.2 Subsampling designs and variance estimators

We suppose that the relevant statistic, whose variance we wish to estimate, can be represented as a function of sample means. Let \( \hat{\theta}_n = H(\bar{Z}_{N_n}) \) be an estimator of the population parameter of interest \( \theta = H(\mu) \), where \( H : \mathbb{R}^p \to \mathbb{R} \) is a smooth function, \( \mu \in \mathbb{R}^p \) is the mean of the stationary r.f. \( Z(\cdot) \), and \( \bar{Z}_{N_n} \) is the sample mean of the \( N_n \) sites within \( R_n \). We can write

\[
\bar{Z}_{N_n} = N_n^{-1} \sum_{s \in \mathbb{Z}^{d \times R_n}} Z(s).
\]

This parameter and estimator formulation is what Hall (1992) calls the "smooth function" model and it has been used in other scenarios, such as with the moving block bootstrap (MBB) and empirical likelihood, for studying approximately linear functions of a sample mean [Lahiri (1996), DiCiccio et al. (1991)]. By considering suitable functions of the \( Z(s) \)'s, one can represent a wide range of estimators under the present framework. In particular, these include means, products and ratios of means, autocorrelation estimators, sample lag cross-correlation estimators [Politis and Romano (1993b)], and Yule-Walker estimates for autoregressive processes [cf. Guyon (1996)].

The quantity which we seek to estimate nonparametrically is the variance of the normalized statistic \( \sqrt{N_n} \hat{\theta}_n \), say, \( \tau_n^2 = N_n E(\hat{\theta}_n - \hat{\mu}_n)^2 \). In our problem, this goal is equivalent to consistently estimating the limiting variance \( \tau^2 = \lim_{n \to \infty} \tau_n^2 \).

2.2.1 Overlapping subsamples

Variance estimation with OL subsampling regions has been presented previously in the literature, though in more narrow sampling situations. Sherman (1996) considered an OL subsample-based estimator for sampling regions in \( \mathbb{R}^2 \); Politis and Romano (1994) extended a similar estimator for rectangular regions in \( \mathbb{R}^d \) with faces parallel to the coordinate axes (i.e. \( R_0 = (-1/2, 1/2]^d \)); and a host of authors, in a variety of contexts, have examined OL subsample estimators applied to time series data [cf. Song and Schmeiser (1988), Politis and Romano (1993a), Fukuchi (1999)].

We first consider creating a smaller version of \( R_n \), which will serve as a template for the OL subsampling regions. To this end, let \( \Delta_n \) be a \( d \times d \) diagonal matrix with positive diagonal elements, \( \{\lambda_1^n, \ldots, \lambda_d^n\} \), such that \( \lambda_i^n / \lambda_i^n \to 0 \) and \( \lambda_i^n \to \infty \), as \( n \to \infty \), for each \( i = 1, \ldots, d \). (The matrix \( \Delta_n \) represents the determining scaling factors for \( R_n \) and \( \Delta_n \) shall be factors used to define the subsamples.) We make the "prototype" subsampling region,

\[
\hat{R}_n = \hat{\Delta}_n R_0,
\]
and identify a subset of \( \mathbb{Z}^d \), say \( J_{OL} \), corresponding to all integer translates of \( sR_n \) lying within \( R_n \). That is,

\[
J_{OL} = \{ i \in \mathbb{Z}^d : i + sR_n \subset R_n \}.
\]

The desired OL subsampling regions are precisely the translates of \( sR_n \) given by: \( R_{1,n} \equiv i + sR_n \), \( i \in J_{OL} \). Note that the origin belongs to \( J_{OL} \) and, clearly, some of these subregions overlap.

Let \( sN_n = |Z^d \cap sR_n| \) be the number of sampling sites in \( sR_n \) and let \( |J_{OL}| \) denote the number of available subsampling regions. (The number of sampling sites within each OL subsampling region is the same, namely for any \( i \in J_{OL} \), \( sN_n = |Z^d \cap R_{i,n}| \).) For each \( i \in J_{OL} \), compute \( \hat{\theta}_i = H(Z_{i,n}) \), where

\[
Z_{i,n} = sN_n^{-1} \sum_{z \in Z^d \cap R_{i,n}} z(z)
\]

denotes the sample mean of observations within the subregion. We then have the OL subsample variance estimator of \( \tau_n^2 \) as

\[
\hat{\tau}_{n,OL}^2 = |J_{OL}|^{-1} \sum_{i \in J_{OL}} sN_n \left( \hat{\theta}_{i,n} - \hat{\theta}_n \right)^2,
\]

\[
\hat{\theta}_n = |J_{OL}|^{-1} \sum_{i \in J_{OL}} \hat{\theta}_{i,n}.
\]

2.2.2 Non-overlapping subsamples

Carlstein (1986) first proposed a variance estimator involving NOL subsamples for time processes. Politis and Romano (1993a) and Sherman and Carlstein (1994) demonstrated the consistency of variance estimation, via NOL subsampling, for certain rectangular regions in \( \mathbb{R}^d \) and some sampling regions in \( \mathbb{R}^3 \), respectively. Here we adopt a formulation similar to those of Sherman and Carlstein (1994) and Lahiri (1999a).

The sampling region \( R_n \) is first divided into disjoint "cubes". Let \( s\Delta_n \) be the previously described \( d \times d \) diagonal matrix from (2), which will determine the "window width" of the partitioning cubes. Let

\[
J_{NOL} = \{ i \in \mathbb{Z}^d : s\Delta_n (i + (-1/2, 1/2]^d) \subset R_n \}
\]

represent the set of all "inflated" subcubes that lie inside \( R_n \). Denote its cardinality as \( |J_{NOL}| \). For each \( i \in J_{NOL} \), define the subsampling region \( R_{i,n} = s\Delta_n(i + R_0) \) by inscribing the translate of \( s\Delta_n R_0 \) such that the origin is mapped onto the midpoint of the subcube \( s\Delta_n(1 + (-1/2, 1/2]^d) \). This provides a collection of NOL subsampling regions, which are "smaller" versions of the original sampling region \( R_n \), that lie inside \( R_n \).

For each \( i \in J_{NOL} \), the function \( H(\cdot) \) is evaluated at the sample mean, say \( \tilde{Z}_{k,n} \), for a corresponding subsampling region \( \tilde{R}_{k,n} \) to obtain \( \tilde{\theta}_{k,n} = H(\tilde{Z}_{k,n}) \). The NOL subsample estimator of \( \tau_n^2 \) is again an
appropriately scaled sample variance:

\[ \hat{\sigma}_{n,\text{NOL}}^2 = |J_{\text{NOL}}|^{-1} \sum_{i \in J_{\text{NOL}}} \hat{N}_{i,n} (\hat{\theta}_{i,n} - \hat{\theta}_n)^2, \quad \hat{\theta}_n = |J_{\text{NOL}}|^{-1} \sum_{i \in J_{\text{NOL}}} \hat{\theta}_{i,n} \]

where \( \hat{N}_{i,n} = |Z^d \cap \tilde{A}_{i,n}| \) denotes the number of sampling sites within a given NOL subsample.

We note that \( \hat{N}_{i,n} \) may differ between NOL subsamples, but all such subsamples will have exactly \( \hat{N}_{i,n} = \hat{N}_n \) sites available if the diagonal elements of \( \hat{\Delta}_n \) are integers.

### 2.3 Assumptions

For stating the assumptions, we need to introduce some notation. For a vector \( x = (x_1, ..., x_d)' \in \mathbb{R}^d \), let \( ||x|| \) and \( ||x||_1 = \sum_{i=1}^d |x_i| \) denote the usual Euclidean and \( l^1 \) norms of \( x \), respectively. Denote the \( l^\infty \) norm as \( ||x||_\infty = \max_{1 \leq k \leq d} |x_k| \). Define \( \text{dis}(E_1, E_2) = \inf\{||x - y||_\infty : x \in E_1, y \in E_2\} \) for two sets \( E_1, E_2 \subset \mathbb{R}^d \). We shall use the notation \( |\cdot| \) also in two other cases: for a countable set \( B \), \( |B| \) would denote the cardinality of the set \( B \); for an uncountable set \( A \subset \mathbb{R}^d \), \( |A| \) would refer to the volume (i.e., the \( \mathbb{R}^d \) Lebesgue measure) of \( A \).

Let \( \mathcal{F}_Z(T) = \sigma\{Z(s) : s \in T\} \) be the \( \sigma \)-field generated by the variables \( \{Z(s) : s \in T\}, T \subset \mathbb{Z}^d \). For \( T_1, T_2 \subset \mathbb{Z}^d \), write

\[ \tilde{\alpha}(T_1, T_2) = \sup\{|P(A \cap B) - P(A)P(B)| : A \in \mathcal{F}_Z(T_1), B \in \mathcal{F}_Z(T_2)\} \]

Then, the strong mixing coefficient for the r.f. \( Z(\cdot) \) is defined as

\[ \alpha(k, l) = \sup\{\tilde{\alpha}(T_1, T_2) : T_1, T_2 \subset \mathbb{Z}^d, |T_i| \leq l, i = 1, 2; \text{dis}(T_1, T_2) \geq k\} \tag{3} \]

Note that the supremum in the definition of \( \alpha(k, l) \) is taken over sets \( T_1, T_2 \) which are bounded. For \( d > 1 \), this is important. A r.f. on the (rectangular) lattice \( \mathbb{Z}^d \) with \( d \geq 2 \) that satisfies a strong mixing condition of the form

\[ \lim_{k \to \infty} \sup\{\tilde{\alpha}(T_1, T_2) : T_1, T_2 \subset \mathbb{Z}^d, \text{dis}(T_1, T_2) \geq k\} = 0 \tag{4} \]

with supremum taken over possibly unbounded sets necessarily belongs to the more restricted class of \( \rho \)-mixing r.f.'s [cf. Bradley (1989)]. Politis and Romano (1993a) used moment inequalities based on the mixing condition in (4) to determine the orders of the bias and variance of \( \hat{\sigma}_{n,\text{OL}}^2, \hat{\sigma}_{n,\text{NOL}}^2 \) for rectangular sampling regions.

For proving the subsequent theorems, the assumptions below are needed along with two conditions stated as functions of a positive argument \( r \in \mathbb{Z}_+ = \{0, 1, 2, \ldots\} \). In the following, \( \det(\Delta) \) represents
the determinant of a square matrix Δ. For α = (α₁, ..., αₚ)' ∈ (Z_+)^p, let D^α denote the αth order partial differential operator ∂^{α₁+⋯+αₚ}/∂x₁^{α₁}...∂xₚ^{αₚ} and ∇ = (∂H(μ)/∂x₁, ..., ∂H(μ)/∂xₚ)' be the vector of first order partial derivatives of H at μ. Limits in order symbols are taken letting n tend to infinity.

Assumptions:

(A.1) There exists a d × d diagonal matrix Δ₀, det(Δ₀) > 0, such that
\[ \frac{1}{\sqrt{\lambda_1^{(n)}}} Δₙ \rightarrow Δ₀. \]

(A.2) For the scaling factors of the sampling and subsampling regions:
\[ \sum_{i=1}^{d} \frac{1}{\sqrt{\lambda_1^{(n)}}} + \sum_{i=1}^{d} \lambda_1^{(n)} + \frac{\det(Δₙ)}{\det(Δ₀)^{(d+1)/d}} = o(1), \quad \max_{1 ≤ i ≤ d} \lambda_i^{(n)} = O\left( \min_{1 ≤ i ≤ d} \lambda_i^{(n)} \right). \]

(A.3) There exist nonnegative functions α₁(·) and g(·) such that \( \lim_{l \to \infty} α₁(k) = 0 \), \( \lim_{l \to \infty} g(l) = \infty \) and the strong-mixing coefficient α(k, l) from (3) satisfies the inequality
\[ α(k, l) ≤ α₁(k)g(l) \quad k > 0, \quad l > 0. \]

(A.4) \( \sup \{ α(T₁, T₂) : T₁, T₂ ⊂ Z^d, |T₁| = 1, \text{dis}(T₁, T₂) ≥ k \} = o(k⁻ᵈ). \)

(A.5) \( r^2 > 0 \), where \( r^2 = \sum_{k \in Z^d} σ(k), \quad σ(k) = \text{Cov}(∇'Z(t), ∇'Z(t + k)). \)

Conditions:

\( D_r : H : R^p \rightarrow R \) is r-times continuously differentiable and, for some \( a \in Z_+ \) and real \( C > 0 \),
\[ \max \{ ||D^rH(x)|| : ||x||_1 = r \} ≤ C(1 + ||x||^a), \quad x \in R^p. \]

\( M_r \): For some \( 0 < δ ≤ 1, \quad 0 < κ < (2r - 1 - 1/d)(2r + δ)/δ \), and \( C > 0 \),
\[ E||Z(t)||^{2r+δ} < \infty, \quad \sum_{x=1}^{∞} x^{(2r-1)d-1} α₁(x)^{δ/(2r+δ)} < \infty, \quad g(x) ≤ Cx^κ. \]

Some comments about the assumptions and the conditions are in order.

Assumption A.5 implies a positive, finite asymptotic variance \( r^2 \) for the standardized estimator, \( \sqrt{N_n} \hat{θ}_n \). We would like \( r^2 \in (0, \infty) \) for a purposeful variance estimation procedure.

In Assumption A.3, we formulate a conventional bound on the mixing coefficient α(k, l) from (3) that is applicable to many r.f.s and resembles a mixing assumption of Lahiri (1999a,b). For r.f.s satisfying
Assumption A.3, the “distance” component of the bound, \( \alpha_1(\cdot) \), often decreases at an exponential rate while the function of “set size,” \( g(\cdot) \), increases at a polynomial rate [cf. Guyon (1996)]. Examples of r.f.s that meet the requirements of A.3 and \( M_r \) include Gaussian fields with analytic spectral densities, certain linear fields with a moving average or autoregressive (AR) representation (like \( m \)-dependent fields), separable AR(1) x AR(1) lattice processes suggested by Martin (1990) for modelling in \( \mathbb{R}^2 \), many Gibbs and Markov fields, and important time series models [cf. Doukhan (1994)]. Condition \( M_r \) combined with A.3 also provides useful moment bounds for normed sums of observations (see Lemma 8.2).

In conjunction with the boundary condition on \( R_0 \), Assumption A.3, and Condition \( M_r \), Assumption A.4 permits the CLT in Bolthausen (1982) to be applied to sums of \( Z(\cdot) \) on sets of increasing domain. This version of the CLT (Stein’s method) is derived from \( \alpha \)-mixing conditions which ensure asymptotic independence between a single point and observations in arbitrary sets of increasing distance [Perera (1997)].

Assumptions A.1 and A.2 set additional guidelines for how sampling and subsampling design parameters, \( \Delta_n \) and \( \Delta_n \), may be chosen. The assumptions provide a flexible framework for handling “increasing domains” of many shapes. For \( d = 1 \), A.1-A.2 are equivalent to the requirements of Lahiri (1999) who provides variance and bias expansions for the MBB variance estimator with weakly dependent time processes.

### 3 Variance expansions

We now give expansions for the asymptotic variance of the OL/NOL subsample variance estimators \( r_{n,OL}^2 \) and \( r_{n,NOL}^2 \) of \( \hat{\sigma}_n = \sqrt{n \text{Var}(\hat{\theta}_n)} \).

**Theorem 3.1** Suppose that Assumptions A.1 - A.5 and Conditions D2 and \( M_{5+2\alpha} \) hold, then

\[
\begin{align*}
(a) \quad \text{Var}(r_{n,OL}^2) &= K_0 \cdot \frac{\det(\Delta_n)}{\det(\Delta_n)} [2\pi^d] \left( 1 + o(1) \right), \\
(b) \quad \text{Var}(r_{n,NOL}^2) &= \frac{1}{|R_0|} \cdot \frac{\det(\Delta_n)}{\det(\Delta_n)} [2\pi^d] \left( 1 + o(1) \right),
\end{align*}
\]

where \( K_0 = \frac{1}{|R_0|} \int_{\mathbb{R}^d} \frac{|(x + R_0) \cap R_0|^2}{|R_0|^2} dx \) is an integral with the Lebesgue measure on \( \mathbb{R}^d \).

The constant \( K_0 \) appearing in the variance expansion of the estimator \( r_{n,OL}^2 \) is a property of the shape of the sampling template \( R_0 \) but not of its exact embedding in space \( \mathbb{R}^d \) or the volume of the set, \( |R_0| \). \( K_0 \) can be computed from either \( R_0 \) or \( R_n = \Delta_n R_0 \) because the constant is invariant to invertible affine transformations applied to \( R_0 \). Values of \( K_0 \) for some template shapes are given in Table 3 and Section 6.
Table 3  Examples of $K_0$ from Theorem 3.1 for several shapes of $R_0 \subset \mathbb{R}^d$.

*The trapezoid has a 90°Z and parallel sides $b_2 \geq b_1$; $c = (b_2/b_1 + 1)^{-2}[1 + 2(b_2/b_1 - 1)/(b_2/b_1 + 1)]$.

<table>
<thead>
<tr>
<th>$R_0$ Shape</th>
<th>$\mathbb{R}^d$ Rectangle</th>
<th>$\mathbb{R}^3$ Ellipsoid</th>
<th>$\mathbb{R}^3$ Cylinder</th>
<th>$\mathbb{R}^2$ Ellipse</th>
<th>$\mathbb{R}^2$ Trapezoid*</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_0$</td>
<td>(2/3)$^d$</td>
<td>34/105</td>
<td>2/3 $(1 - 16/(3\pi^2))$</td>
<td>1 $- 16/(3\pi^2)$</td>
<td>2/5 $(1 + 4c/9)$</td>
</tr>
</tbody>
</table>

A stationary time sequence $Z_1, \ldots, Z_n$ can be obtained within our sampling formulation by choosing $R_0 = (-1/2,1/2]^d$ and $\lambda_1^{(n)} = n$ on the untranslated integer lattice $\mathbb{Z} = \mathbb{Z}$. In this special sampling case, an application of Theorem 3.1 yields

$$\text{Var}(\hat{\tau}^2_{n,OL}) = 2/3 \cdot \text{Var}(\hat{\tau}^2_{n,NOL}), \quad \text{Var}(\hat{\tau}^2_{n,NOL}) = \lambda_1^{(n)}/\lambda_1^{(n)} \cdot [2r^2](1 + o(1)), $$

a result which is well-known for “nearly” linear functions $\hat{\theta}_n$ of a time series sample mean [cf. Künsch (1989)].

Theorem 3.1 implies that, under the “smooth” function model, the asymptotic variance of the OL subsample-based variance estimator is always strictly less than the NOL version because

$$K_1 = \lim_{n \to \infty} \frac{\text{Var}(\hat{\tau}^2_{n,OL})}{\text{Var}(\hat{\tau}^2_{n,NOL})} = K_0 |R_0| < 1.$$ (5)

The percentage $100(1 - K_1)%$ represents the relative reduction in the variance of the estimator based on the OL subsampling scheme over the NOL one. If both estimators have biases of the same order, (5) implies that the variance estimation with OL subsamples is more efficient than the NOL subsample alternative owing to a smaller asymptotic MSE.

Unlike $K_0$, $K_1$ does depend on the volume $|R_0|$, which in turn is constrained by the $R_0$-template’s geometry. That is, $K_1$ is ultimately bounded, through $|R_0|$ in (5), by the amount of space that an object of $R_0$’s shape can possibly occupy within $(-1/2,1,2]^d$; or by how much volume can be filled by a given geometrical body (eg. circle) compared to a cube. The constants $K_1$ in Table 1 are computed with templates of prescribed shape and largest possible volume in $(-1/2,1,2]^d$. These values most accurately reflect the influence of $R_0$’s (or $R_n$’s) geometry on the relative performance (ie. variance, efficiency) of the estimators $\hat{\tau}^2_{n,OL}$ and $\hat{\tau}^2_{n,NOL}$.

To conclude this section, we remark that both subsample-based variance estimators can be shown to be (MSE) consistent under Theorem 3.1 conditions, allowing for more general spatial sampling regions, in both shape and dimension, than previously considered. Inference on the parameter $\theta$ can be made through the limiting standard normal distribution of $\sqrt{N_n}(\hat{\theta}_n - \theta)/\hat{\tau}_n$ for $\hat{\tau}_n = \hat{\tau}_{n,OL}$ or $\hat{\tau}_{n,NOL}$. 
4 Bias expansions

We now try to capture and precisely describe the leading order terms in the asymptotic bias of each subsample-based variance estimator, similar to the variance determinations from the previous section. We first establish and note the order of the dominant component in the bias expansions of $\hat{\sigma}_{n,OL}^2$ and $\hat{\sigma}_{n,NOL}^2$, which is the subject of the following lemma.

Lemma 4.1. With Assumptions A.1 - A.5, suppose that Conditions D$_2$ and M$_{2+a}$ hold for $d \geq 2$; and D$_d$ and M$_{d+a}$ for $d = 1$. Then, the subsample estimators of $r_n^2 = \text{Var}(\hat{\theta}_n)$ have expectations:

$$E(\hat{r}_{n,OL}^2) = r_n^2 + O(1/\alpha^{(i)})$$ and $$E(\hat{r}_{n,NOL}^2) = r_n^2 + O(1/\alpha^{(i)}).$$

The lemma shows that, under the smooth function model, the asymptotic bias of each estimator is $O(1/\alpha^{(i)})$ for all dimensions of sampling. Politis and Romano (1993a, p.323) and Sherman (1996) showed this same size for the bias of $\hat{\sigma}_{n,OL}^2$ with sampling regions based on rectangles $R_0 = (-1/2, 1/2]^d$ or simple closed curves in $\mathbb{R}^2$, respectively. Lemma 4.1 extends these results to a broader class of sampling regions. However, we would like to precisely identify the $O(1/\alpha^{(i)})$ bias component for $\hat{\sigma}_{n,OL}^2$ or $\hat{\sigma}_{n,NOL}^2$ to later obtain theoretically correct proportionality constants associated with optimal subsample scaling.

To achieve some measure of success in determining the exact bias of the subsampling estimators, we reformulate the subsampling design slightly so that $\alpha^{(i)} \equiv \alpha^{(i)} = \cdots = \alpha^{(i)}$. That is, a common scaling factor in all directions is now used to define the subsampling regions (as in Sherman and Carlstein (1994); Sherman (1996)). This constraint will allow us to deal with the counting issues at the heart of the bias expansion.

Adopting a common scaling factor $\alpha^{(i)}$ for the subsamples also is sensible for a few other reasons at this stage:

- "Unconstrained" optimum values of $\alpha^{(i)}$ cannot always be found by minimizing the asymptotic MSE of $\hat{\sigma}_{n,OL}^2$ or $\hat{\sigma}_{n,NOL}^2$, even for variance estimation of some desirable statistics on geometrically "simple" sampling and subsampling regions. Consider estimating the variance of a real-valued sample mean over a rectangular sampling region in $\mathbb{R}^d$ based on $R_0 = (-1/2, 1/2]^d$, with observations on $Z^d = Z^d$. If Assumptions A.1-A.5 and Condition M$_1$ hold, the leading term in the bias expansion can be shown to be:

$$\text{Bias of } \hat{\sigma}_{n,OL}^2 = \left( - \sum_{i=1}^{d} \frac{L_i}{\alpha^{(i)}} \right) (1 + o(1)); \quad L_i = \sum_{k \in \mathbb{Z}^d} |k_i| \text{Cov}(Z(\mathbf{0}), Z(\mathbf{k})).$$
In using the parenthetical sum above to expand the MSE of $\hat{\tau}^2_{n,OL}$, one finds that the resulting MSE cannot be minimized over the permissible, positive range of $\Delta_n$ if the signs of the $L_i$ values are unequal. That is, for $d > 1$, the subsample estimator MSE cannot always be globally minimized to obtain optimal subsample factors $\Delta_n$ by considering just the leading order bias terms. An effort to determine and incorporate (into the asymptotic MSE) second or third order bias components quickly becomes intractable, even with rectangular regions.

- The diagonal components of $\Delta_n$ are asymptotically scalar multiples of each other by Assumption A.1. If so desired, a template choice for $R_0$ could be used to scale the expansion of the subsampling regions in each direction.

In the continued discussion, we assume

$$R_n = \lambda_n R_0. \quad (6)$$

We frame the components necessary for determining the biases of the spatial subsample variance estimators in the next theorem. Let $C_n(k) \equiv |\mathbb{Z}^d \cap \lambda_n R_n \cap (k + \lambda_n)|$ denote the number of pairs of observations in the subsampling region $\lambda_n R_n$ separated by a translate $k \in \mathbb{Z}^d$.

Theorem 4.1 Suppose that $d \geq 2$, $\lambda_n R_n = \lambda_n R_0$ and Assumptions A.1 - A.5, Conditions D3 and M3+8 hold. If, in addition, $\lambda_n \in \mathbb{Z}^+$ for NOL subsamples and

$$\lim_{n \to \infty} \frac{N_n - C_n(k)}{(\lambda_n)^d} = C(k) \quad (7)$$

exists for all $k \in \mathbb{Z}^d$, then

$$\mathbb{E} \hat{\tau}^2_n - \tau^2_n = \frac{1}{\lambda_n |R_0|} \left( \sum_{k \in \mathbb{Z}^d} C(k) \sigma(k) \right) (1 + o(1))$$

where $\sigma(k) = \text{Cov}(\nabla Z(t), \nabla Z(t+k))$ and where $\hat{\tau}^2_n$ is either $\hat{\tau}^2_{n,OL}$ or $\hat{\tau}^2_{n,OL}$.

Note that the numerator on the left side of (7) is the number of integer grid points that lie in the subregion $\lambda_n R_n$, but not in the translate $\lambda_n R_n + k$. Hence, computing the bias above actually requires counting the number of lattice points inside intersections like $\lambda_n R_n \cap (k + \lambda_n)$, which is difficult in general. To handle the problem, one may attempt to approximate $C_n(k)$ with the corresponding Lebesgue volume, $|\mathbb{Z}^d \cap \lambda_n R_n \cap (k + \lambda_n)|$, and then examine the discrepancy between the two values. The determination of volumes or areas may not be easy either but hopefully more manageable. For example, if $R_0$ is a circle, the area of $\lambda_n R_0$ can be readily computed, but the number of $\mathbb{Z}^2$ integers inside $\lambda_n R_0$ is not so simple and was in fact a famous consideration of Gauss [cf. Krätzel (1988, p.141)].
The boundary condition on \( R_0 \) implies that the discrepancy between the volume \(|sR_n \cap k + sR_n|\) and the count \( C_n(k)\) is of order \( O((s\lambda_n)^{d-1})\), uniformly for \( k \in \mathbb{Z}^d\). Applying this bound to majorize the approximation error incurred by exchanging the volumes \(|sR_n|, |sR_n \cap k + sR_n|\) for the counts \( N_n, C_n(k)\) in the numerator of (7) will ultimately introduce a new \( O(1/s\lambda_n) \) sum of “estimation errors” into the bias expansion. This bound from the \( R_0\)-boundary condition, on the discrepancy between \(|sR_n \cap k + sR_n|\) and \( C_n(k)\), thus is too large to help in pinpointing the leading terms of the bias expansions of \( \tau_{n,OL}^2 \) and \( \tau_{n,NO}^2 \), which are also \( O(1/s\lambda_n) \) from Lemma 4.1.

Bounds on the difference between lattice point counts and volumes have received much attention in analytic number theory, which we briefly mention. Research has classically focused on sets outlined by “smooth” simple closed curves in the plane \( \mathbb{R}^2 \) and on one question in particular [Huxley (1996)]: When a curve with interior area \( A \) is ‘blown up’ by a factor \( b \), how large is the difference between the number of \( \mathbb{Z}^2 \) integer points inside the new curve and the area \( b^2 A \)? For convex sets with a smoothly winding border, van der Corput’s (1920) answer to the posed question above is \( O(b^{45/69+\epsilon}) \), while the best answer is \( O(b^{45/73+\epsilon}) \) for curves with sufficiently differentiable radius of curvature [Huxley (1993, 1996)]. These types of bounds, however, are invalid for many convex polygonal templates \( R_0 \) in \( \mathbb{R}^2 \) such as triangles, trapezoids, etc., where often the difference between number of \( \mathbb{Z}^2 \) integer points in \( sR_n = s\lambda_n R_0 \) and its area is of exact order \( O(s\lambda_n) \) (set also by the boundary condition on \( R_0 \) or the perimeter length of \( sR_n \)). The formulated problem above, as considered by number theorists, does not directly address counts for intersections between an expanding region and its vector translates, e.g. \( sR_n \cap k + sR_n \).

To eventually compute “closed-form” bias expansions for \( \tau_{n,OL}^2 \), we use the following two-step approach for counting the number of lattice points inside the intersection of \( sR_n \) and its translates. For each \( k \in \mathbb{Z}^d \),

1. Approximate the numerator of the coefficient in (7) with the difference of the corresponding Lebesgue volumes: \( (s\lambda_n)^d|R_0| - |sR_n \cap k + sR_n| \)

2. Show the discrepancy in approximation is of sufficiently small order, \( o((s\lambda_n)^{d-1}) \).

We do approximate the number of lattice points in \( sR_n \) and \( sR_n \cap k + sR_n \) by set volumes, though the Lebesgue volume may not adequately capture the lattice point count in either set. However, when subtracted, the errors incurred in the approximation of both sets can cancel each other to a great extent. The difference between the two discrepancies \( N_n - (s\lambda_n)^d|R_0| \) and \( C_n(k) - |sR_n \cap k + sR_n| \) can be shown to be asymptotically small enough, for some templates \( R_0 \), to justify replacing counts with volumes. The
above approach becomes slightly more complicated for NOL subsamples, \( \tilde{R}_{i,n} = s \Delta_n(i + R_0), i \in J_{\text{NOL}} \), which may vary in the number of sampling sites. In this case, one tries to demonstrate additionally that, for each \( k \in \mathbb{Z}^d \), approximation errors incurred by the substitution of volumes (e.g., replacing the count \( |Z^d \cap \tilde{R}_{i,n} \cap (k + \tilde{R}_{i,n})| \) with the volume \( |\tilde{R}_{i,n} \cap (k + \tilde{R}_{i,n})| \) are asymptotically negligible, uniformly in \( i \in J_{\text{NOL}} \).

In the following theorem, we use this technique to give bias expansions for a large class of sampling regions in \( \mathbb{R}^d, d \leq 3 \), which are "nearly" convex. That is, the sampling region \( R_n \) differs from a convex set possibly only at its boundary, but sampling sites on the border may be arbitrarily included or excluded from \( R_n \).

Some notation is additionally required. For \( \alpha = (\alpha_1, ..., \alpha_p)' \in (\mathbb{Z}_+)^p, x \in \mathbb{R}^p \), write \( x^\alpha = \prod_{i=1}^p x_i^{\alpha_i} \), \( \alpha! = \prod_{i=1}^p \alpha_i! \), and \( c_{\alpha} = D^\alpha H(\mu)/\alpha! \). Let \( Z_{\infty} \) denote a random vector with a normal \( \mathcal{N}(0, \Sigma_\infty) \) distribution on \( \mathbb{R}^p \), where \( \Sigma_\infty \) is the limiting covariance matrix of the scaled sample mean \( \sqrt{N_n} (\bar{Z}_{n} - \mu) \) from (1). Let \( B^*, \overline{B} \) denote the interior and closure of \( B \in \mathbb{R}^d \), respectively.

**Theorem 4.2** Suppose \( R_n = s \lambda_n R_0 \) and there exists a convex set \( B \) such that \( B^* \subset R_0 \subset \overline{B} \). With Assumptions A.2 - A.5, assume Conditions Ds-d and M5−d+α hold for \( d \in \{1, 2, 3\} \). Then,

\[
C(k) = V(k) \equiv \lim_{n \to \infty} \frac{|s R_n| - |s R_n \cap (k + s R_n)|}{(s \lambda_n)^{d-1}} \quad k \in \mathbb{Z}^d
\]

whenever \( V(k) \) exists and the biases \( E(\tau_{n,\text{OLS}}^2 - \tau_n^2), E(\tau_{n,\text{NOL}}^2 - \tau_n^2) \) are equal to:

for \( d = 1 \),

\[
\frac{-1}{s \lambda_n |R_0|} \left( \sum_{k \in \mathbb{Z}} |k| \sigma(k) + C_\infty \right) \left( 1 + o(1) \right);
\]

for \( d = 2 \) or 3,

\[
\left( -\sum_{k \in \mathbb{Z}^d} \frac{|s R_n| - |s R_n \cap (k + s R_n)|}{|s R_n|} \sigma(k) \right) \left( 1 + o(1) \right)
\]

or \( \frac{-1}{s \lambda_n |R_0|} \left( \sum_{k \in \mathbb{Z}^d} V(k) \sigma(k) \right) \left( 1 + o(1) \right) \), provided each \( V(k) \) exists;

where \( \sigma(k) = \text{Cov}(\nabla' Z(t), \nabla' Z(t+k)) \) and \( C_\infty = \text{Var} \left( \sum_{\|\alpha\|=2} \frac{c_\alpha}{\alpha!} Z_\infty \right) + 2 \sum_{\|\alpha\|=2} \frac{c_\alpha c_{\beta}}{\beta!} E \left( Z_\infty^2 Z_{\infty,2} \right) + 2 \sum_{k_1, k_2 \in \mathbb{Z}} \sum_{\|\gamma\|=1} \frac{c_\alpha c_{\beta+\gamma}}{(\beta+\gamma)!} E \left( [Z(t) - \mu]^\alpha [Z(t+k_1) - \mu]^\beta [Z(t+k_2) - \mu]^\gamma \right).
Remark 1: If Condition $D_2$ holds with $C = 0$ for $x \in \{2,3,4\}$, then Condition $M_{2-1}$ is sufficient in Theorem 4.2.

Remark 2: For each $k \in \mathbb{Z}^d$, the numerator in $V(k)$ is of order $O((\lambda_n)^{d-1})$ by the $R_0$-boundary condition which holds for convex templates. We may then expand the bias of the estimators through the limiting, scaled volume differences $V(k)$. For $d = 1$, with samples and subsamples based on intervals, it can be easily seen that $V(k) = |k|$ which appears in Theorem 4.2.

The function $H(\cdot)$ needs to be increasingly “smoother” to determine the bias component of $\tilde{\sigma}^2_{n,OL}$, $\tilde{\sigma}^2_{n,NOL}$ in lower dimensional spaces $d = 1$ or $2$. For a real-valued time series sample mean $\hat{\theta}_n = \bar{Z}_n$, the well-known bias of the subsample variance estimators follows from Theorem 4.2 under our sampling framework $R_0 = (-1/2,1/2]$, $Z = Z$:

$$-\sum_{k \in \mathbb{Z}} |k| \text{Cov}(\nabla'Z(0), \nabla'Z(k))$$

with $\nabla = 1$. In general though, terms in the Taylor expansion of $\tilde{\theta}_{1,n}$ (around $\mu$) up to fourth order can contribute to the bias of $\tilde{\sigma}^2_{n,OL}$ and $\tilde{\sigma}^2_{n,NOL}$ when $d = 1$. The asymptotic bias of the time series MBB variance estimator with “smooth” model statistics is very different from its subsample-based counterpart, appearing in (8) [cf. Lahiri (1996)]. That is, only the linear component from the Taylor’s expansion of $\tilde{\theta}_{1,n}$ determines the bias of the MBB variance estimator.

5 Asymptotically optimal subsample sizes

In the following, we consider “size” selection for the subsampling regions to maximize the large-sample accuracy of the subsample variance estimators. For reasons discussed in Section 4, we examine a theoretically optimal scaling choice $\lambda_n$ for subregions in (6).

5.1 Theoretical optimal subsample sizes

Generally speaking, there is a trade off in the effect of subsample size on the bias and variance of $\tilde{\sigma}^2_{n,OL}$ or $\tilde{\sigma}^2_{n,NOL}$. For example, increasing $\lambda_n$ reduces the bias but increases the variance of the estimators. The best value of $\lambda_n$ optimizes the over-all performance of a subsample variance estimator by balancing the contributions from both the estimator’s variance and bias. An optimal $\lambda_n$ choice can be found by minimizing the asymptotic order of a variance estimator’s MSE under a given OL or NOL sampling scheme.

Theorem 4.1 implies that the bias of the estimators $\tilde{\sigma}^2_{n,OL}$ and $\tilde{\sigma}^2_{n,NOL}$ is of exact order $O(1/\lambda_n)$. For
a broad class of sampling regions \( R_n \), the leading order bias component can be determined explicitly with
Theorem 4.2. We bring these variance and bias expansions together to obtain an optimal subsample
scaling factor \( \lambda_n \).

**Theorem 5.1** Let \( R_n = \lambda_n R_0 \). With Assumptions A.2 - A.5, assume Conditions D_2 and \( M_{5+2\alpha} \) hold if \( d > 2 \); and Conditions D_3 and \( D_{7+2\alpha} \) if \( d = 1 \). If \( B_0 \equiv |R_0|^{-1} \sum_{k \in \mathbb{Z}^d} C(k) \sigma(k) \neq 0 \), then

\[
\lambda_{n, \text{OL}}^{\text{opt}} = \left( \frac{\det(\Delta_n)(B_0)^2}{dK_0 r^4} \right)^{1/(d+2)} (1 + o(1))
\]

\( s_{n, \text{NOL}}^{\text{opt}} = \left( \frac{\det(\Delta_n)|R_0|(B_0)^2}{d^4 r^4} \right)^{1/(d+2)} (1 + o(1))

**Remark 3:** If Condition D_2 holds with \( C = 0 \) for \( x \in \{2, 3\} \), then Condition \( M_{2x-1} \) is sufficient.

Theorem 5.1 suggests that optimally scaled OL subsamples should be larger than the NOL ones
by a scalar: \( (K_1)^{-1/(d+2)} > 1 \) where \( K_1 = K_0 |R_0| \) the limiting ratio of variances from (5). It is
well-known in the time series case that the OL subsampling scheme produces an asymptotically more
efficient variance estimator than its NOL counterpart. We can quantify the relative efficiency of the two
subsampling procedures in \( d \)-dimensional sampling space. With optimally selected subsamples from
(6), the asymptotic relative efficiency of \( \hat{\tau}_{n, \text{NOL}}^2 \) to \( \hat{\tau}_{n, \text{OL}}^2 \) depends solely of the geometry of \( R_0 \):

\[
\text{ARE}_d = \lim_{n \to \infty} \frac{\mathbb{E}(\hat{\tau}_{n, \text{NOL}}^2 - \tau_n^2)^2}{\mathbb{E}(\hat{\tau}_{n, \text{OL}}^2 - \tau_n^2)^2} = (K_1)^{2/(d+2)} < 1.
\]

Possolo (1991), Politis and Romano (1993a, 1994), Hall and Jing (1996), and Garcia-Soidan and
Hall (1997) examined subsampling with rectangular sampling regions based essentially on \( R_0 = (-1/2, 1/2]^d \).
While \( \hat{\tau}_{n, \text{OL}}^2 \) is always more efficient than \( \hat{\tau}_{n, \text{NOL}}^2 \), the geometry of rectangular regions interestingly im­
poses a bound on the relative improvement of \( \hat{\tau}_{n, \text{OL}}^2 \) over \( \hat{\tau}_{n, \text{NOL}}^2 \) as the sampling dimension increases.
In this case, we have that \( K_1 = (\frac{d}{2})^d \) and \( \lim_{d \to \infty} \text{ARE}_d = \frac{1}{2} \), a dimensional limit on the efficacy of the
optimized OL subsampling scheme over the NOL one.

We note also that “unconstrained” optimal subsample parameters \( \Delta_n \) for \( \hat{\tau}_{n, \text{OL}}^2 \) can, in principle,
be found by minimizing the estimator’s asymptotic MSE to achieve an appropriate scalar adjustment
in each component direction. This is equivalent to choosing an optimally scaled sample template \( R_0 \).
To further explain, let \( \Lambda_t \) be a \( d \times d \) diagonal matrix with positive diagonal entries \( \{l_1, \ldots, l_d\} \). The
sampling regions \( \Delta_n R_0 \) and \( \Delta_n \tilde{R}_0 \) are the same for \( \Delta_n = \Delta_n \Lambda_t^{-1}, \tilde{R}_0 = \Lambda_t R_0 \). Using Theorem 3.1
with template \( \tilde{R}_0 \) and \( \Delta_n \), the variance of \( \hat{\tau}_{n, \text{OL}}^2 \) becomes a multiple of \( [\det(\Lambda_t) \cdot (\lambda n)^d] / \det(\Delta_n) \).
The factors \( \Lambda_t \) can often enter the bias of \( \hat{\tau}_{n, \text{OL}}^2 \) through

\[
\frac{-1}{\lambda_n |R_0|} \left( \sum_{k \in \mathbb{Z}^d} V(\Lambda_t^{-1} k) \sigma(k) \right) (1 + o(1)).
\]
with the limiting values $V(\cdot)$ from Theorem 4.2. One could then attempt to minimize the asymptotic 
MSE of $\tilde{\sigma}_n^2$ for $\lambda_n$ and $l_1, \ldots, l_d$. However, the success of this endeavor depends usually on the 
covariances $\sigma(k)$, $k \in \mathbb{Z}^d$ and a global minimum of the MSE may not exist over the positive range of 
values for $\lambda_n, l_1, \ldots, l_d$ (as described in Section 4).

By construction, a template choice $\Lambda_i R_0$ may change the bias of the NOL variance estimator but 
the variance of $\tilde{\sigma}_{n,\text{NOL}}^2$ remains the same (as in Theorem 3.1).

5.2 Empirical subsample size determination

In this section we consider data-based estimation of the theoretical optimal subsample sizes. In 
particular, we propose methods for estimating the scaling factor $\lambda^{\text{opt}}_n$ for subsamples in (6), which are 
applicable to both OL and NOL subsampling schemes. Inference on “best” subsample scaling closely 
resembles the problem of empirically gauging the theoretically optimal block size with the MBB variance 
estimator, which has been much considered for time series. We first frame the tactics used in this 
special setting, which can be modified for data-based subsample selection.

Two distinct techniques have emerged for estimating the (time series) MBB block size. One approach involves “plug-in” type estimators of optimal block length [cf. Bühlman and Künsch (1999)].

For subsample inference on $n \text{Var}(\tilde{Z}_n)$, Carlstein (1986) (with AR(1) models) and Politis et al. (1999) 
have described “plug-in” estimators for subsample length. The second MBB block selection method, 
suggested by Hall et al. (1995) and Hall and Jing (1996), uses subsampling to create a data-based 
version of the MSE of the MBB estimator (as a function of block size) which then is minimized to 
obtain an estimate of optimal block length. With spatial data, Garcia-Soidan and Hall (1997) extended 
this empirical MSE selection procedure with subsample-based distribution estimators. For variance 
estimation of a univariate time series sample mean, Lèger et al. (1992) proposed a subsample length 
estimate based on asymptotic MSE considerations; namely, selecting a size by simultaneously solv­
ing $(\lambda_n^{\text{opt}})^3 = (3/2)(\hat{B}/\tilde{\sigma}_{n,\text{OL}}^2)^2 n$ for $\lambda_n^{\text{opt}}$ and the associated estimator $\tilde{\sigma}_{n,\text{OL}}^2$, where $\hat{B}$ estimates the 
quantity in (8).

A "plug-in" procedure for inference on the theoretical, optimal subsample size involves constructing, 
and subsequently substituting, (consistent) estimates of unknown population parameters in $\lambda^{\text{opt}}_n$ from 
Theorem 5.1. The limiting variance $\tau^2$ appears in the formulation of $\lambda_n^{\text{opt}}$ and could be approximated 
with a subsample variance estimate based on some preliminary size choice. The value $K_0$ can be 
determined from the available sampling region $R_n$, but selection of a template $R_0$ is also required in 
the “plug-in” approach. The best solution may be to pick a sampling template $R_0$ to be the largest set
of the form \(\Lambda^{-1} R_n\) for a positive diagonal matrix \(\Lambda\) that fits within \((-1/2, 1/2)^d\); this choice appears to reduce the magnitude of the bias of \(\hat{\omega}_{n,\hat{A}}^2\) and \(\hat{\omega}_{n,\hat{\theta}}^2\), which contributes heavily to the MSE of an estimator.

In a spirit similar to the "plug-in" approach, one could empirically select \(\omega_n^*\) as in Léger et al. (1992) and Politis and Romano (1993b). After evaluating \(K_0\) and an estimate \(\hat{B}_0\) of the leading bias component \(B_0\), compute \(\hat{\omega}_{n,\hat{A}}^2\) (or analogously \(\hat{\omega}_{n,\hat{\theta}}^2\)) for a series of \(\omega_n\) values and simultaneously solve the asymptotical MSE-based formulation \((\omega_n)^{d+2} = \text{det}(\Delta_n) \cdot (\hat{B}_0)^2 / \{d \cdot K_0 \cdot (\hat{\omega}_{n,\hat{A}}^2)^2\}\) for \(\omega_n\) and an associated \(\hat{\omega}_{n,\hat{A}}^2\).

6 Examples

We now provide some examples of the important quantities \(K_0, K_1, B_0\) associated with optimal scaling \(\omega_n^*\) with some common sampling region templates, determined from Theorems 3.1 and 4.2. For subsamples from (5), the theoretically best \(\omega_n^*\) can also be formulated in terms of \(|R_n| = \text{det}(\Delta_n)|R_0|\) (sampling region volume), \(K_1\), and \(B_0\).

6.1 Examples in \(\mathbb{R}^2\)

**example 1.** Rectangular regions in \(\mathbb{R}^2\) (potentially rotated):

If \(R_0 = \{(l_1 \cos \theta, l_2 \sin \theta)x, (-l_1 \sin \theta, l_2 \cos \theta)x) : x \in (-1/2, 1/2)^2\}\) for \(\theta \in [0, \pi], 0 < l_1, l_2,\) then

\[
K_0 = \frac{4}{9}, \quad B_0 = \sum_{k \in \mathbb{Z}^2} \left(\frac{|k_1 \cos \theta - k_2 \sin \theta|}{l_1} + \frac{|k_1 \sin \theta + k_2 \cos \theta|}{l_2}\right) \sigma(k).
\]

The characteristics \(K_1, B_0\) for determining optimal subsamples based on two rectangular templates are further described in Table 4.

**example 2.** If \(R_0\) is a circle of radius \(r \leq 1/2\) centered at the origin, then \(K_0\) appears in Table 3 and \(B_0 = 2/(\pi r) \sum_{k \in \mathbb{Z}^2} ||k|| \sigma(k)\).

**example 3.** For any triangle, \(K_0 = 2/5\). Two examples are provided in Tables 2 and 4.

**example 4.** For any parallelogram in \(\mathbb{R}^2\) with interior angle \(\gamma\) and adjacent sides of ratio \(b \geq 1, K_0 = 4/9 + 2/15 \cdot b^{-2} \cdot |\cos \gamma|/(1 - |\cos \gamma|)\). In particular, if a parallelogram \(R_0\) is formed by two vectors \((0, l_1), (l_2 \cos \gamma, l_2 \sin \gamma)\) extended from a point \(x \in (-1/2, 1/2)^2\), then

\[
B_0 = \frac{1}{|\sin \theta|} \sum_{k \in \mathbb{Z}^2} \left(\frac{|k_1 \cdot \cos \theta - k_2 \cdot \sin \theta|}{\max\{l_1, l_2\}} + \frac{|k_2|}{\min\{l_1, l_2\}}\right) \sigma(k), \quad \gamma \in (0, \pi), l_1, l_2 > 0.
\]
Figure 1. Examples of templates $R_0 \subset (-1/2,1/2]^2$ are outlined by solid lines. Cross-shaped sampling regions $R_n$ described in Table 2 are based on $R_0$ in (v).

Table 4  Examples of several shapes of $R_0 \subset \mathbb{R}^2$ and associated $K_1$, $B_0$ for $\lambda_n^{opt}$

<table>
<thead>
<tr>
<th>$R_0$</th>
<th>$K_1$</th>
<th>$B_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(-1/2,1/2]^2$</td>
<td>$4/9$</td>
<td>$\sum_{k \in \mathbb{Z}^2}</td>
</tr>
<tr>
<td>&quot;Diamond&quot; in Figure 1(i)</td>
<td>$2/9$</td>
<td>$2\sum_{k \in \mathbb{Z}^2}</td>
</tr>
<tr>
<td>Circle of radius 1/2 at origin</td>
<td>$\pi/4 - 4/(3 \pi)$</td>
<td>$4/\pi \sum_{k \in \mathbb{Z}^2}</td>
</tr>
<tr>
<td>Right triangle in Figure 1(ii)</td>
<td>$1/5$ (Table 1)</td>
<td>Table 2</td>
</tr>
<tr>
<td>Triangle in Figure 1(iii)</td>
<td>$1/5$</td>
<td>$\sum_{k \in \mathbb{Z}^2} 2</td>
</tr>
<tr>
<td>Parallelogram in Figure 1(iv)</td>
<td>$2/9 + (\sqrt{3} - 1)/375$</td>
<td>$4/\sqrt{3} \sum_{k \in \mathbb{Z}^2} (</td>
</tr>
</tbody>
</table>

6.2 Examples in $\mathbb{R}^d$, $d \geq 3$

Example 5. For any sphere, $K_0$ is given in Table 3. The properties $B_0$, $K_1$ of the sphere described in Table 1 and 2 correspond to the template sphere $R_0$ of radius 1/2 (and maximal volume in $(-1/2,1/2]^3$).

Example 6. The $K_0$ value for any $\mathbb{R}^3$ cylinder appears in Table 3. If $R_0$ is a cylinder with circular base (parallel to x-y plane) of radius $r$ and height $h$, then

$$B_0 = \sum_{k \in \mathbb{Z}^3} \left( \frac{|k_3|}{h} + \frac{2\sqrt{k_1^2 + k_2^2}}{\pi r} \right) \sigma(k).$$

With the counting techniques used in the proof of Theorem 4.2, bias expansions for $\tilde{\tau}_{n,\text{OL}}^2$, $\tilde{\tau}_{n,\text{OL}}^2$ seem possible for sampling templates $R_0$ created from finite unions of convex sets whose borders do not intersect more than finitely often. More irregular sampling regions, including polygonal and non-convex regions, could then be constructed through unions of convex sets. We do not explore this possibility at
great length in this paper, but we provide a small step toward an extension in the following corollary concerning sampling regions which are a union of two “nearly” convex sets.

**Corollary 6.1** Let \( R_0 = R_1 \cup R_2 \subset (-1/2, 1/2)^2 \). Suppose there exists convex sets \( A_1, A_2 \) such that \( A_i^0 \neq \emptyset, A_i^0 \subset R_i \subset \overline{A_i} \) for \( i = 1, 2 \). Assume that \( \partial A_i \cap \partial A_2 = (\overline{A_i} \setminus A_i^0) \cap (\overline{A_2} \setminus A_2^0) \) is empty or finite. Then, the results of Theorem 4.2 still hold under the same conditions. Furthermore, for \( k \in \mathbb{Z}^d \),

\[
V(k) = V^*(k) = \lim_{n \to \infty} \left( \sum_{i=1}^n V_i^*(k) - V_n^*(k) \right),
\]

whenever \( V^*(k) \) exists, where:

- \( V_i^*(k) = |\lambda_i n R_i| - |\lambda_i n (R_i \cap (k + \lambda_i n R_i))| \) for \( i = 1, 2 \)
- \( V_n^*(k) = |\lambda_1 n (R_1 \cap R_2)| - |\lambda_1 n (R_1 \cap R_2) \cap (k + \lambda_2 n (R_1 \cap R_2))| \).

The formulation of the asymptotic bias involving \( V^*(k) \) can be determined by separately working with the areas of the (nearly) convex regions \( \lambda_i n R_i \), \( \lambda_i n R_2 \) and their convex intersection \( \lambda_i n (R_1 \cap R_2) \). We provide a small application of Corollary 6.1 for a non-convex, star-shaped sampling template in Table 2, which is illustrated in Figure 1(v). The template is a union of two perpendicular rectangles with boundaries that intersect at exactly four points.

The results of Theorem 4.2 for determining the bias \( B_0 \) also seem plausible for convex sampling regions in \( \mathbb{R}^d, d \geq 4 \), but require further study and extension of lattice point counting techniques in higher dimensions. However, bias expansions of the OL and NOL subsample variance estimators are relatively straightforward for an important class of rectangular sampling regions based on the prototype \( R_0 = (-1/2, 1/2)^d \). These hypercubes have “faces” parallel to the coordinate axes which simplifies the task of counting sampling sites, or lattice points, within such regions. We give precise bias expansions in the following theorem, while allowing for potentially missing sampling sites at the borders, or faces, of the sampling region \( R_n \).

**Theorem 6.1** Let \( (-1/2, 1/2)^d \subset A^{-1}_\ell R_0 \subset [-1/2, 1/2]^d \), \( d \geq 3 \) and \( A_\ell \) is a \( d \times d \) diagonal matrix with entries \( 0 < \ell_i \leq 1, i = 1, ..., d \). Suppose \( \eta R_n = \lambda R_0 \) and Assumptions A.2 - A.5, Conditions D2 and M2+a hold. Then, the biases \( E(\hat{\tau}^2_{n,OL} - \tau_n), E(\hat{\tau}^2_{n,NOL} - \tau_n) \) are equal to:

\[
-\frac{1}{\lambda n} \left( \sum_{k \in \mathbb{Z}^d} \left( \sum_{i=1}^d |k_i| \ell_i \right) \sigma(k) \right) (1 + o(1)); \quad \sigma(k) = \text{Cov}(\nabla' Z(t), \nabla' Z(t + k)).
\]

Exact first-order expansions of the bias and variance of \( \hat{\tau}^2_{n,OL}, \hat{\tau}^2_{n,NOL} \) require mild differentiability conditions on \( H(\cdot) \) for higher sampling dimensions \( d \geq 3 \).
7 Subsample design extensions

We mention here a small reformulation of the sampling and subsampling scheme from Section 2. Instead of making scaled down copies of the original observation region \( R_n = \Delta_n R_0 \) to serve as subsamples, one may choose an entirely different template shape, say \( R_0' \), for defining the subregions. For example, one might wish to use rectangular subsampling regions based on \( R_n' = (-1/2, 1/2]^d \) regardless of the shape of \( R_n \). This variation on subsampling has received little consideration in the literature, but it can also yield consistent variance estimators.

There may, as well, be practical justification for this alternative subsampling scheme. If \( R_n \) is quite irregular in shape, subsamples based on scaled copies of \( R_n \) could become computationally undesirable. Also, asymptotic bias expansions for some subsample shape choices can also be particularly tractable, which then simplifies the construction of empirical "plug-in" estimates of optimal subsample scaling.

In the following, we focus on the behavior of modified \( OL \) and \( NOL \) subsample variance estimators \( \hat{\tau}^2_{OL} \) and \( \hat{\tau}^2_{NOL} \), assuming sampling sites are observed in \( R_n = \Delta_n R_0 \) while a potential \( OL \) subsample is given by \( i + \Delta_n R_0' \subset R_n \), and \( NOL \) subsample consists of \( i \Delta_n (i + R_0') \subset R_n \), for \( i \in \mathbb{Z}^d \). (Both templates \( R_0, R_0' \) are assumed to contain the origin as an interior point and satisfy the same previously mentioned boundary condition.)

Under the conditions and assumptions of Theorem 3.1, the variance of \( \hat{\tau}^2_{OL} \) can be expanded as

\[
\text{Var}(\hat{\tau}^2_{OL}) = K'_0 |R_0'| \cdot \frac{\det(\Delta_n)}{|R_n|} \cdot [2r^4](1 + o(1)).
\]

The variance constant \( K'_0 \) is a function of the shape of the subsample prototype \( R_0' \), found by substituting \( R_0' \) for \( R_0 \) in the computation of \( K_0 \) from Theorem 3.1. For example, if rectangular subregions are chosen based on \((-1/2, 1/2]^d \), the above variance of \( \hat{\tau}^2_{OL} \) becomes

\[
\text{Var}(\hat{\tau}^2_{OL}) = \left( \frac{2}{3} \right)^d \cdot \frac{\det(\Delta_n)}{|R_n|} \cdot [2r^4](1 + o(1)).
\]

The variance of the non-overlapping version of \( \hat{\tau}^2_{NOL} \) is the same for all subsample template selections \( R_0' \) (or \( R_0 \)) and appears in Theorem 3.1.

The bias of each variance estimator \( \hat{\tau}^2_{OL} \) or \( \hat{\tau}^2_{NOL} \) can be deduced from the expressions given in Theorems 4.2-4.3 after replacing \( R_0 \) with the alternative subsample template \( R_0' \) in the result of each theorem.
8 Preliminaries for the proofs

For proving the theorems, we will use \( C, C(-) \) to denote generic positive constants that depend on their arguments (if any) but not on \( n \) or any \( \mathbb{Z}^d \) integers (or \( \mathbb{Z}^d \) lattice points). The real number \( r \), appearing in some proofs, always assume the value stated under Condition \( M_r \) with respect to the lemma or theorem under consideration. Unless otherwise specified, limits in order symbols are taken letting \( n \) tend to infinity.

In the following, we denote the indicator function as \( I_{\{\cdot\}} \) (ie. \( I_{\{\cdot\}} \in \{0,1\} \) and \( I_{\{A\}} = 1 \) if and only if an event \( A \) holds). For two sequences \( \{r_n\} \) and \( \{t_n\} \) of positive real numbers, we write \( r_n \sim t_n \) if \( r_n/t_n \to 1 \) as \( n \to \infty \). We adopt the standard that for a set \( A \): \( A^\circ, \overline{A} \text{ and } A^c \) represent the interior, closure, and complement of the set, respectively. We write \( \lambda_n^{max} \) and \( \lambda_n^{min} \) for the largest diagonal entries of \( \Delta_n \) and \( \Delta_n \), respectively, while \( \lambda_n^{(a)} \) will denote the smallest diagonal entry of \( \Delta_n \).

We require a few lemmas for the proofs.

Lemma 8.1 Suppose \( T_1, T_2 \subset \mathbb{Z}^d \equiv t + \mathbb{Z}^d \) are bounded. Let \( p, q > 0 \) where \( 1/p + 1/q < 1 \). If \( X_1, X_2 \) are random variables, with \( X_i \) measurable with respect to \( \mathcal{F}_{Z(T_i)}, i = 1, 2 \), then,

\[
|\text{Cov}(X_1, X_2)| \leq 8 (\text{E}|X_1|^p)^{1/p} (\text{E}|X_2|^q)^{1/q} \alpha(\text{dis}(T_1, T_2); \max_{i=1,2} |T_i|)^{1-1/p-1/q},
\]

provided expectations are finite and \( \text{dis}(T_1, T_2) > 0 \).

Proof: Follows from Chapter 1, Doukhan (1994). □

Lemma 8.2 Let \( r \in \mathbb{Z}_+ \). Under Assumption A.3 and Condition \( M_r \), for \( 1 \leq m \leq 2r \) and any \( T \subset \mathbb{Z}^d \equiv t + \mathbb{Z}^d \),

\[
\text{E}(\| \sum_{s \in T} Y(s) \|_r^r) \leq C(\alpha)|T|^{m/2}, \quad Y(s) = Z(s) - \mu;
\]

\( C(\alpha) \) is a constant that depends only on the coefficients \( \alpha(k, l), l \leq 2r \), and \( \text{E}(\|Z(t)\|_r^{2r+4}) \).

Proof: This follows from Doukhan (1994, Theorem 1, p. 26-31) and Jensen's inequality. □

We next determine the asymptotic sizes of important sets relevant to the sampling or subsampling designs. For \( i \in \mathbb{Z}^d \), let

\[
\lambda N_{i,n} = |\lambda \Delta_n (i + R_0) \cap \mathbb{Z}^d|
\]
denote number of sampling sites in a NOL subregion. (Note \( \lambda N_{0,n} = \lambda N_n \) by definition.)
Lemma 8.3 Under Assumptions A.1-A.2, the number of sampling sites within
(a) the sampling region $R_n$: $N_n = |R_n \cap \mathbb{Z}^d| \sim |R_0| \cdot \det(\Delta_n)$;
(b) an OL subsample, $i + s\Delta_n R_0$, $i \in J_{OL}$: $sN_n = |sR_n \cap \mathbb{Z}^d| \sim |sR_0| \cdot \det(s\Delta_n)$;
(c) a NOL subsample, $s\Delta_n (i + R_0)$, $i \in J_{NOL}$: $sN_{1,n} \sim |R_0| \cdot \det(s\Delta_n)$.

The number of
(d) OL subsamples within $R_n$: $|J_{OL}| \sim |R_0| \cdot \det(\Delta_n)$;
(e) NOL subsamples within $R_n$: $|J_{NOL}| \sim |R_0| \cdot \det(\Delta_n) \cdot \det(s\Delta_n)^{-1}$.

Proof: To establish Lemma 8.3(a), it is enough to show

$$|N_n - \det(\Delta_n)|R_0| \leq C \cdot (\lambda_n^{m*})^{d-1}. \tag{11}$$

We first bound $N_n$ from above as follows:

$$N_n \leq \left| \{i \in \mathbb{Z}^d : (i + (-1/2, 1/2]^d) \cap \Delta_n R_0^c = \emptyset \} \right| + \left| \{i \in \mathbb{Z}^d : T^d \cap \Delta_n R_0^c = \emptyset, T^d \cap \Delta_n R_0 = \emptyset, T^d = i + (-1/2, 1/2]^d \} \right|$$

$$\leq \det(\Delta_n)|R_0| + \left| \{i \in \Delta_n^{-1} \mathbb{Z}^d : T^d \cap \Delta_n R_0^c = \emptyset, T^d \cap \Delta_n R_0^c = \emptyset, T^d = i + (\lambda_n^{m*})^{-1}[-1, 1]^d \} \right|$$

$$\leq \det(\Delta_n)|R_0| + 2^d \left| \{i \in (\lambda_n^{m*})^{-1} \mathbb{Z}^d : T^d \cap \Delta_n R_0^c = \emptyset, T^d \cap \Delta_n R_0^c = \emptyset, T^d = i + (\lambda_n^{m*})^{-1}[0, 1]^d \} \right|$$

$$\leq \det(\Delta_n)|R_0| + C \cdot (\lambda_n^{m*})^{d-1},$$

where the last inequality follows from the boundary condition on $R_0$. We use the previous argument to find a lower bound on $N_n$:

$$N_n \geq \left| \{i \in \mathbb{Z}^d : (i + (-1/2, 1/2]^d) \cap \Delta_n R_0^c = \emptyset \} \right|$$

$$\geq \det(\Delta_n)|R_0| - \left| \{i \in \mathbb{Z}^d : T^d \cap \Delta_n R_0^c = \emptyset, T^d \cap \Delta_n R_0 = \emptyset, T^d = i + (-1/2, 1/2]^d \} \right|$$

$$\geq \det(\Delta_n)|R_0| - C \cdot (\lambda_n^{m*})^{d-1}.$$

We now have (11).

Lemma 8.3(b) follows by

$$\left| sN_n - \det(s\Delta_n)|R_0| \right| \leq C \cdot (s\lambda_n^{m*})^{d-1}$$

which can be justified simply by replacing $(N_n, \Delta_n, \lambda_n^{m*})$ in the proof of (11) with the subsample versions $(sN_n, s\Delta_n, s\lambda_n^{m*})$. 
To show Lemma 8.3(c), we fix $i \in J_{\text{poll}}$ and note $|\Delta_n(i + R_0)| = \det(\Delta_n) |R_0|$. With arguments similar to those above, we find

$$
|\Delta_{n,1,n} - |\Delta_n(i + R_0)|
\leq |\{ k \in Z^d : T^k \cap \Delta_n(i + R_0) = \emptyset, T^k \cap \Delta_n(i + \overline{R_0}) \neq \emptyset ; T^k = k + [-1/2, 1/2]^d \}|
\leq |\{ k \in Z^d : T^k \cap \Delta_n(i + \overline{R_0}) = \emptyset, T^k \cap \Delta_n(i + R_0) \neq \emptyset ; T^k = k + [-1, 1]^d \}|
\leq |\{ k \in Z^d : T^k \cap \Delta_n \overline{R_0} = \emptyset, T^k \cap \Delta_n R_0 \neq \emptyset ; T^k = k + [-2, 2]^d \}|
\leq C \cdot (\lambda_n^{\max})^{d-1},
$$

by the $R_0$ boundary condition. Lemma 8.3(c) then follows.

To establish Lemma 8.3(d), we show

$$
|J_{\text{oll}} - \det(\Delta_n) |R_0| \leq C \cdot \rho_n^{\max} (\lambda_n^{\max})^{d-1}.
$$

Define the set of subsamples near the boundary of $R_n$, say $K_n$, which we then bound.

$$
|\{ i \in Z^d \cap \Delta_n R_0 : (i + \rho_n R_0) \cap \Delta_n \overline{R_0} \neq \emptyset \}|
\leq |\{ i \in Z^d : T^i \cap \Delta_n \overline{R_0} = \emptyset, T^i \cap \Delta_n R_0 \neq \emptyset ; T^i = i + \rho_n^{\max} [-1/2, 1/2]^d \}|
\leq |\{ i \in \rho_n^{\max} (2\lambda_n^{\max})^{-1} Z^d : T^i \cap \overline{R_0} = \emptyset, T^i \cap R_0 \neq \emptyset ; T^i = 2(\rho_n^{\max})^{-1} i + \rho_n^{\max} (2\lambda_n^{\max})^{-1} [-1, 1]^d \}|
\leq |\{ i \in \rho_n^{\max} (2\lambda_n^{\max})^{-1} Z^d : T^i \cap \overline{R_0} = \emptyset, T^i \cap R_0 \neq \emptyset ; T^i = i + \rho_n^{\max} (2\lambda_n^{\max})^{-1} [-1, 1]^d \}|
\cdot |\{ j \in Z^d : 2(\rho_n^{\max})^{-1} j \in [-2, 2]^d \}|
\leq C \cdot (\rho_n^{\max})^{d} (2\rho_n^{\max})^{-1} \lambda_n^{\max})^{d-1}
\leq C \cdot \rho_n^{\max} \cdot (\lambda_n^{\max})^{d-1}
$$

using the boundary condition on $R_0$ in the second-to-last inequality. The result follows by applying the bound on $K_n$, $|J_{\text{oll}}| = N_n - K_n$, and (11).

Finally, for proving Lemma 8.3(e), it suffices to show

$$
|J_{\text{oll}}| - \{ \det(\Delta_n) \}^{-1} \det(\Delta_n) |R_0| \leq C \cdot (\lambda_n^{\max})^{d-1}
$$

where $\lambda_n^{\max}$ denotes the largest entry of $\Delta_n^{-1} \Delta_n$. Because $|J_{\text{oll}}| = \{ i \in Z^d \cap \Delta_n R_0 : i + (-1/2, 1/2]^d \subset (\Delta_n)^{-1} \Delta_n R_0 \}$, we may use the boundary condition on $R_0$ and proceed as in the proof of (11) to obtain the desired result. The proof of Lemma 8.3 is now complete. $\square$
We require the next lemma for counting the number of subsampling regions which are separated by an appropriately "small" integer translate; we shall apply this lemma in the proof of Theorem 3.1. For \( k = (k_1, \ldots, k_d)' \in \mathbb{Z}^d \), define the following sets,

\[
J_n(k) = |\{i \in J_{OL} : i + k + \varepsilon \Delta_n R_0 \subset \Delta_n R_0\}|, \quad E_n = \{k \in \mathbb{Z}^d : |k_j| \leq \lambda_j^{-1}, \ j = 1, \ldots, d\}.
\]

**Lemma 8.4** Under Assumption A.2,

\[
\max_{k \in E_n} \left| \frac{1 - J_n(k)}{|J_{OL}|} \right| = o(1).
\]

**Proof:** For \( k \in E_n \), define the set \( J_n^*(k) \) and bound its cardinality:

\[
J_n^*(k) = |\{i \in J_{OL} : (i + k + \varepsilon \Delta_n R_0) \cap \Delta_n R_0^c \neq \emptyset\}|
\]

\[
\leq |\{i \in J_{OL} : (i + \varepsilon \Delta_n(-2,2)^d) \cap \Delta_n R_0^c \neq \emptyset\}|
\]

\[
\leq |\{i \in \mathbb{Z}^d : T^4 \cap \Delta_n R_0^c \neq \emptyset, T^4 \cap \Delta_n R_0 \neq \emptyset \text{ for } T^4 = i + \varepsilon \lambda_{max}^{-1}[-2,2]^d\}|
\]

\[
\leq C \cdot \varepsilon \lambda_{max} \cdot (\lambda_{max})^{d-1},
\]

by the boundary condition on \( R_0 \). We have then that for all \( k \in E_n \),

\[
|J_{OL}| \geq J_n(k) = |J_{OL}| - J_n^*(k) \geq |J_{OL}| - C \cdot \varepsilon \lambda_{max} \cdot (\lambda_{max})^{d-1},
\]

where \( C > 0 \) does not depend on \( k \in E_n \). By Assumption A.2 and the growth rate of \( |J_{OL}| \) from Lemma 8.3, the proof is complete. \( \Box \)

**9 Proof of Theorem 3.1**

We now provide a theorem which captures the main contribution to the asymptotic variance expansion of the OL subsample variance estimator \( \hat{\sigma}^2_{OL} \) (presented in Theorem 3.1).

**Theorem 9.1** For \( i \in \mathbb{Z}^d \), let \( R_{in} = i + \varepsilon R_n; Y_{in} = \nabla'(Z_{i,n} - \mu) \), where \( Z_{i,n} = \varepsilon N_n^{-1} \sum_{s \in Z_{i,n} R_n} Z(s) \) and \( \varepsilon R_n = \varepsilon \Delta_n R_0 \). Under the Assumptions and Conditions of Theorem 3.1,

\[
\varepsilon N_n \sum_{k \in E_n} \text{Cov}(Y_{0,n,k}^2, Y_{k,n}^2) = K_0 \cdot 2\tau^2(1 + o(1)),
\]

where the constant \( K_0 \) is defined in Theorem 3.1.

**Proof:** Treated in Section 11. \( \Box \)

For clarity of exposition, we will prove Theorem 3.1, parts (a) and (b), separately for the OL and NOL subsample variance estimators. The arguments involved are quite different in both cases.
9.1 Proof of Theorem 3.1(a)

For \( i \in \mathcal{J}_{OL} \), we use a Taylor's expansion of \( H(\cdot) \) (around \( \mu \)) to rewrite the statistic \( \hat{\theta}_{i,n} = H(Z_{i,n}) \) evaluated on the subregion \( R_{1,n} = i + \Delta, R_{n} = \Delta n R_0 \). Write, for \( i \in \mathbb{Z}^d \),

\[
\hat{\theta}_{i,n} = H(\mu) + \sum_{\|a\|_1=1} c_a (Z_{i,n} - \mu)^a + 2 \sum_{\|a\|_1=2} \frac{(Z_{i,n} - \mu)^a}{a!} \int_0^1 (1 - \omega) D^\omega H(\mu + \omega (Z_{i,n} - \mu)) d\omega \\
= H(\mu) + Y_{i,n} + Q_{i,n} \text{ (say)}. \tag{13}
\]

We also have

\[
\hat{\theta} = |J_{OL}|^{-1} \sum_{i \in \mathcal{J}_{OL}} \hat{\theta}_{i,n} = H(\mu) + |J_{OL}|^{-1} \sum_{i \in \mathcal{J}_{OL}} Y_{i,n} + |J_{OL}|^{-1} \sum_{i \in \mathcal{J}_{OL}} Q_{i,n} \tag{14}
\]

\[
\hat{\theta} = H(\mu) + \hat{Y}_n + \hat{Q}_n.
\]

The overlapping subsample variance estimator \( \hat{s}_{n,OL}^2 \) can be expanded as:

\[
s_{n,OL}^2 = \left[ \frac{1}{|J_{OL}|} \sum_{i \in \mathcal{J}_{OL}} Y_{i,n}^2 + \frac{1}{|J_{OL}|} \sum_{i \in \mathcal{J}_{OL}} Q_{i,n}^2 + \frac{2}{|J_{OL}|} \sum_{i \in \mathcal{J}_{OL}} Y_{i,n} Q_{i,n} - \hat{Y}_n^2 - \hat{Q}_n^2 - 2(\hat{Y}_n)(\hat{Q}_n) \right]. \tag{15}
\]

We will accomplish the proof in two parts by showing

\[
(a) \quad \text{Var} \left( \frac{s_{n,OL}^2}{|J_{OL}|} \sum_{i \in \mathcal{J}_{OL}} Y_{i,n}^2 \right) = K_0 \cdot \frac{\text{det}(\Delta_n)}{\text{det}(\Delta_n)} \cdot [2r^4](1 + o(1)),
\]

\[
(b) \quad \left| \text{Var}(s_{n,OL}^2) - \text{Var} \left( \frac{s_{n,OL}^2}{|J_{OL}|} \sum_{i \in \mathcal{J}_{OL}} Y_{i,n}^2 \right) \right| = o \left( \frac{\text{det}(s\Delta_n)}{\text{det}(\Delta_n)} \right). \tag{16}
\]

We will begin with proving (a) above. WLOG assume \( \mu = 0 \) and that each diagonal component of \( s\Delta_n \) is greater than 1. For \( k \in \mathbb{Z}^d \), let \( \tilde{\sigma}_n(k) = \text{Cov}(Y_{0,n}^2, Y_{k,n}^2) \). We first decompose

\[
(s_{n,OL}^2/|J_{OL}|)^2 \text{Var} \left( \sum_{i \in \mathcal{J}_{OL}} Y_{i,n}^2 \right) = (s_{n,OL}^2/|J_{OL}|)^2 \left( \sum_{k \in \mathbb{Z}^d} \text{Var}(\tilde{\sigma}_n(k)) + \sum_{k \in \mathbb{Z}^d \backslash \mathcal{E}_n} \text{Var}(\tilde{\sigma}_n(k)) \right) \equiv I_{1,n} + I_{2,n} \text{ (say)},
\]

and we will individually handle the quantities \( I_{1,n}, I_{2,n} \).

By stationarity and Lemma 8.2, we bound the covariances between subsampling regions: \( |\tilde{\sigma}_n(k)| \leq E(Y_{0,n}^4) \leq C(s_{n})^{-2}, k \in \mathbb{Z}^d \). Using this covariance bound, Lemmas 8.3 and 8.4, and \( |E_n| \leq 3^d \text{det}(s\Delta_n) \):

\[
\left| (s_{n,OL}^2/|J_{OL}|) \sum_{k \in \mathbb{E}_n} \tilde{\sigma}_n(k) - I_{1,n} \right| \leq C \cdot \frac{|E_n|}{|J_{OL}|} \max_{k \in \mathbb{E}_n} \left| 1 - \frac{J_n(k)}{|J_{OL}|} \right| = o \left( \frac{\text{det}(s\Delta_n)}{\text{det}(\Delta_n)} \right). \tag{17}
\]

Then applying Theorem 9.1 and Lemma 8.3,

\[
\frac{(s_{n,OL}^2)^2}{|J_{OL}|} \sum_{k \in \mathbb{E}_n} \tilde{\sigma}_n(k) = K_0 \cdot \frac{\text{det}(s\Delta_n)}{\text{det}(\Delta_n)} \cdot [2r^4](1 + o(1)). \tag{18}
\]
By (17) and (18), we need only show that $I_{2,n} = o(\det(\Delta_n)/\det(\Delta_n))$ to complete the proof of (16)(a).

For $i \in \mathbb{Z}^d$, denote a set of lattice points within a translated rectangular region:

$$T_i^1 = \left( i + \prod_{j=1}^{d} \left( \left\lceil \frac{-s\lambda_j^{n/1}}{2} \right\rceil, \left\lceil \frac{s\lambda_j^{n/1}}{2} \right\rceil \right) \right) \cap \mathbb{Z}^d,$$

where $\lceil \cdot \rceil$ represents the "ceiling" function above. Note that for $k = (k_1, \ldots, k_d) \in \mathbb{Z}^d \setminus E_n$, there exists $j \in \{1, \ldots, d\}$ such that $|k_j| > s\lambda_j^{n/1}$, implying $\text{dis}(R_{0,n} \cap \mathbb{Z}^d, R_{k,n} \cap \mathbb{Z}^d) \geq \text{dis}(T_{0,n}^1, T_{k,n}^1) \geq 1$. Hence, sequentially using Lemmas 8.1 and 8.2, we may bound the covariances $\delta_n(k), k \in \mathbb{Z}^d \setminus E_n$, in terms of the mixing coefficient $\alpha(\cdot, \cdot)$, as follows:

$$|\delta_n(k)| \leq \delta \left( \frac{E(Y_{0,n}^2(2r+4)/r)}{2r/(2r+4)} \right)^{2r/(2r+4)} \alpha(\text{dis}(R_{0,n} \cap \mathbb{Z}^d, R_{k,n} \cap \mathbb{Z}^d), sN_n)^{\delta/(2r+4)} \leq C \cdot (sN_n)^{-2} \alpha(\text{dis}(T_{0,n}^1, T_{k,n}^1), sN_n)^{\delta/(2r+4)}.$$

We define a scalar term: $\ell = (s\lambda_m^{n/1})^e$ and $\epsilon = 1/2 \cdot (\kappa \cdot \delta/[(2r+\delta)(2rd - 1 - 1/d)] + 1) < 1$ by Condition $M_r$. By the bound on $\delta$ in (19) and the fact that $J_n(k)/|J_{0,1}| \leq 1$ for all $k \in \mathbb{Z}^d$, we have

$$|I_{2,n}| \leq C \cdot |J_{0,1}|^{-1} \sum_{z=1}^{\infty} \left( \sum_{j=1}^{d} C_{(z,j,n)} \right) \alpha(x, sN_n)^{\delta/(2r+4)},$$

$C_{(x,j,n)} = \left| \left\{ l \in \mathbb{Z}^d : \text{dis}(T_{0,n}^1, T_l^1) = z = \inf\{|v_j - w_j| : v \in T_{0,n}^0, w \in T_l^1 \} \right\} \right|$. The function $C_{(x,j,n)}$ counts the number of translated rectangles $T_l^1$ that lie a distance of $x$ (measured by $\text{dis}(\cdot, \cdot)$) from the rectangle $T_{0,n}^0$ and, simultaneously, $x$ is the distance between these rectangles realized in the $j$th coordinate direction for $j = 1, \ldots, d$. If, for $i \in \mathbb{Z}^d$ and $x \geq 1 \in \mathbb{Z}_{+}$ and some $j \in \{1, \ldots, d\}$, $\text{dis}(T_{0,n}^0, T_l^1) = z = \inf\{|v_j - w_j| : v \in T_{0,n}^0, w \in T_l^1 \}$, then $|i_j| = \lfloor s\lambda_m^{n/1} \rfloor + x - 1$ with the remaining components of $i$, namely $i_m$ for $m \in \{1 \ldots d\} \setminus \{j\}$, being consequently constrained by $|i_m| \leq s\lambda_m^{n/1} + x$.

We use this observation to further bound the right hand side of (20) by:

$$C \cdot |J_{0,1}|^{-1} \sum_{z=1}^{\infty} \left( \sum_{j=1}^{d} \prod_{m=1,j \neq d}^{d} 3(s\lambda_m^{n/1} + x) \right) \alpha(x, sN_n)^{\delta/(2r+4)} \leq C \cdot \frac{\det(s\Delta_n)}{|J_{0,1}|} \sum_{z=1}^{\infty} \left( \sum_{j=1}^{d} \frac{x_j^{j-1}}{\left( s\lambda_m^{n/1} \right)^j} \alpha(x, sN_n)^{\delta/(2r+4)} \right) \leq C \cdot \frac{\det(s\Delta_n)}{|J_{0,1}|} \sum_{j=1}^{d} (s\lambda_m^{n/1})^{-j} \left[ \left( \sum_{x=1}^{\ell} + \sum_{x=\ell+1}^{\infty} \right) x^{j-1} \alpha_1(x) \right] \leq C \cdot \frac{\det(s\Delta_n)}{|J_{0,1}|} \sum_{j=1}^{d} (s\lambda_m^{n/1})^{-j} \left[ \ell^j + \frac{(s\lambda_m^{n/1})^{d-j-\delta/(2r+4)}}{\ell_{2rd-d-j}} \cdot \sum_{x=\ell+1}^{\infty} x^{2rd-d-1-\delta} \alpha_1(x)^{\delta/(2r+4)} \right] \leq C \cdot \frac{\det(s\Delta_n)}{|J_{0,1}|} \left[ \ell \cdot (\ell/\ell_{2rd}) + \frac{\ell^{1/\ell}}{\ell_{2rd-d}} \cdot \sum_{x=\ell+1}^{\infty} x^{2rd-d-1} \alpha_1(x)^{\delta/(2r+4)} \right] = o(\det(s\Delta_n)/\det(s\Delta_n)),$$
with Assumptions A.1 and A.3 and where the last line follows from \( \lambda_{\text{min}} \geq 1, \ell = o(\lambda_{\text{min}}) \) (because \( e^{-1} \cdot d \cdot (2r + \delta) < 2rd - d \)), and Condition \( M_r \). The resulting bound can be contrasted to the bound of \(|A|\) in Politis and Romano (1994, p.2024). This completes the proof of (16)(a).

To establish (16)(b), first note that

\[
|\text{Var}(\tau^2_{n,0}) - A_{1,n}| \leq \left[A_{1,n}^{1/2} + A_{2,n}^{1/2} + A_{3,n}^{1/2} + A_{4,n}^{1/2} + 2A_{5,n}^{1/2} + 2A_{6,n}^{1/2}\right]^2 - A_{1,n},
\]

\[
A_{1,n} = (sN_n)^2 \text{Var}\left(|J_{0L}|^{-1} \sum_{i \in J_{0L}} Y_{i,n}^2\right), \\
A_{2,n} = (sN_n)^2 \text{Var}\left(|\bar{J}_{0L}|^{-1} \sum_{i \in J_{0L}} \bar{Q}_{i,n}^2\right), \\
A_{3,n} = (sN_n)^2 \text{Var}\left(|J_{0L}|^{-1} \sum_{i \in J_{0L}} Q_{i,n}^2\right), \\
A_{4,n} = (sN_n)^2 \text{Var}\left(|\bar{J}_{0L}|^{-1} \sum_{i \in J_{0L}} \bar{Q}_{i,n}^2\right), \\
A_{5,n} = (sN_n)^3 \text{Var}\left(|J_{0L}|^{-1} \sum_{i \in J_{0L}} Y_{i,n} Q_{i,n}\right), \\
A_{6,n} = (sN_n)^3 \text{Var}\left(|\bar{J}_{0L}|^{-1} \sum_{i \in J_{0L}} \bar{Q}_{i,n} \bar{Q}_{i,n}\right).
\]

By (16)(a), \( A_{1,n} = O(\det(\Delta_n)/\det(\Delta_n)) \), it suffices to show that \( A_{j,n} = o(\det(\Delta_n)/\det(\Delta_n)) \) for each \( j = 2, \ldots, 6. \)

Consider \( A_{2,n} \) first. We observe that

\[
A_{2,n} \leq (sN_n)^2 \text{E}(\bar{Q}_{n}^2) = \frac{(sN_n)^2(2^d \text{det}(\Delta_n))^4}{|J_{0L}|^4(sN_n)^4} \text{E}\left[\left|\sum_{s \in R_n \cap Z^d} \omega(s) \mathbf{V}(s)\right|^4\right]; \\
0 \leq \omega(s) = \frac{|\{i \in J_{0L} : s \in i + \Delta_n \mathbf{R}_0\}|}{2^d \text{det}(\Delta_n)} \leq 1, \quad s \in R_n \cap Z^d.
\]

By mixing Condition \( M_r \) and Doukhan (1994, Theorem 3, p.31) (similar to Lemma 8.2),

\[
\text{E}\left[\left|\sum_{s \in R_n \cap Z^d} \omega(s) \mathbf{V}(s)\right|^4\right] \leq C \cdot (N_n)^2.
\]

Because \( N_n = O(|J_{0L}|) \), \( \text{det}(\Delta_n) = O(sN_n) \), and \( \text{det}(\Delta_n) = o(|J_{0L}|) \) (by Lemma 8.3 and Assumption A.2), it follows from the above moment bound and (21) that indeed \( A_{2,n} = o(\text{det}(\Delta_n)/\det(\Delta_n)) \).

Now we will consider \( A_{5,n} \). Write \( \delta_{1,n}(k) = \text{Cov}(Y_{0,n}, Q_{0,n}, Y_{k,n}, Q_{k,n}) \), \( k \in Z^d \). Then,

\[
A_{5,n} = \frac{(sN_n)^2}{|J_{0L}|^2} \sum_{k \in Z^d} J_n(k) \overline{\delta}_{1,n}(k) \\
\leq \frac{(sN_n)^2}{|J_{0L}|^2} \left(\sum_{k \in E_n} |\delta_{1,n}(k)| + \sum_{k \in Z^d \setminus E_n} |\delta_{1,n}(k)|\right).
\]

For \( k \in E_n \), note

\[
|\delta_{1,n}(k)| \leq \text{Var}(Y_{0,n} Q_{0,n}) \leq \text{E}(Y_{0,n}^2 Q_{0,n}^2) \leq C \cdot \text{E}(|Z_{0,n}|^6(1 + |Z_{0,n}|^{2a})) \leq C \cdot (sN_n)^{-3}
\]

by Lemmas 8.1 and 8.2 using also that \( |Y_{0,n} Q_{0,n}| \leq C \cdot |Z_{0,n}|^2(1 + |Z_{0,n}|^a) \). By this bound and \( |E_n| \leq 3^d \text{det}(\Delta_n) \), we establish:

\[
\frac{(sN_n)^2}{|J_{0L}|^2} \sum_{k \in E_n} |\delta_{1,n}(k)| = o\left(\frac{\text{det}(\Delta_n)}{\det(\Delta_n)}\right).
\]
after applying the set growth conditions from Lemma 8.3. Likewise, we obtain a bound on the covariances \( \hat{\sigma}_{1,n}(k) \), \( k \in \mathbb{Z}^d \setminus E_n \),

\[
|\hat{\sigma}_{1,n}(k)| \leq 8\{E([|Y_{0,n}Q_{0,n}|^{(2r+s)/r}]^{2r/(2r+s)}) \cdot \alpha(\text{dis}(R_{0,n} \cap \mathbb{Z}^d, R_{k,n} \cap \mathbb{Z}^d), sN_n)^{4/(2r+s)}
\]

\[
\leq C \cdot (sN_n)^{-3} \alpha(\text{dis}(T_n^0, T_n^k), sN_n)^{4/(2r+s)}
\]

by the stationarity of the random field \( Z(\cdot) \), the moment bound from Lemma 8.1, and Lemma 8.2. Using this covariance inequality and repeating the same steps used to majorize \( I_{2,n} \) from the proof of (16)(a) (see (19),(20)), we have:

\[
\frac{(sN_n)^2}{|J_{0n}|} \sum_{k \in \mathbb{Z}^d \setminus E_n} |\hat{\sigma}_{1,n}(k)| = O(|J_{0n}|^{-1}) = o\left(\frac{\det(s\Delta_n)}{\det(\Delta_n)}\right),
\]

and hence \( A_{s,n} = o\left(\frac{\det(s\Delta_n)}{\det(\Delta_n)}\right) \).

By symmetry to \( A_{s,n} \) above, \( A_{3,n} = o(\det(s\Delta_n)/\det(\Delta_n)) \) because for \( k \in E_n \),

\[
|\text{Cov}(Q_{0,n}^2, Q_{k,n}^2)| \leq \text{Var}(Q_{0,n}^2) \leq E(Q_{0,n}^4) \leq C \cdot E([\|Z_{0,n}\|^4(1 + \|Z_{0,n}\|^{4a})]) \leq C \cdot (sN_n)^{-4}
\]

with Lemma 8.2 and, for \( k \in \mathbb{Z}^d \setminus E_n \),

\[
|\text{Cov}(Q_{0,n}^2, Q_{k,n}^2)| \leq C \cdot \left(\frac{E([\|Z_{0,n}\|^{4(2r+s)/r} + \|Z_{0,n}\|^{(4+2a)(2r+s)/r})]^{2r}}{\alpha(\text{dis}(R_{0,n} \cap \mathbb{Z}^d, R_{k,n} \cap \mathbb{Z}^d), sN_n)^{4/(2r+s)}}\right)^{1/(2r+s)}
\]

\[
\leq C \cdot (sN_n)^{-4} \alpha(\text{dis}(T_n^0, T_n^k), sN_n)^{4/(2r+s)}
\]

We will now handle \( A_{4,n} \). Define the function \( \xi_n : \mathbb{Z}_+ \rightarrow \mathbb{Z}_+ \) such that: for \( m \in \mathbb{Z}_+ \),

\[
\xi_n(m) = \left| \{ (v, w, x, y) \in (\mathbb{Z}^d)^4 : |v_j|, |w_j|, |x_j|, |y_j| \leq \lambda_j^{n+1}/2, j = 1, \ldots, d; \right. \\
\text{dis}(T_n^v \cup T_n^w \cap \mathbb{Z}^d, T_n^x \cup T_n^y \cap \mathbb{Z}^d) = m \left. \} \right|.
\]

Then, similar to the counting arguments used in (19),

\[
\xi_n(0) \leq C \cdot [\det(\Delta_n)]^3 \cdot [\det(s\Delta_n)]^2
\]

and for \( m \geq 1 \in \mathbb{Z}_+ \),

\[
\xi_n(m) \leq C \cdot [\det(\Delta_n)]^3 \cdot \left( 3^d \prod_{j=1}^{d} (m + s\lambda_j^{n+1}) \cdot \sum_{j=1}^{d} (m + s\lambda_j^{n+1})^{-1} \right).
\]

Note for \( v, w, x, y \in \mathbb{Z}^d \), \( |\text{Cov}(Q_{v,n}Q_{w,n}, Q_{x,n}Q_{y,n})| \leq E(Q_{0,n}^4) \leq C \cdot (sN_n)^{-4} \) by Holder's inequality.
and stationarity, and if dis([T^*_n \cup T^*_m] \cap Z^d, [T^*_n \cup T^*_m] \cap Z^d) = m \geq 1,

\begin{align*}
|\text{Cov}(Q_{v,n}Q_{w,n}, Q_{z,n}Q_{y,n})| \\
\leq 8 \left[ E(|Q_{v,n}Q_{w,n}|^{2r+4}/r) \cdot E(|Q_{z,n}Q_{y,n}|^{2r+4}/r) \right]^{\alpha(m, 2sN_n)/r/(2r+4)} \\
\leq 8 \left[ E(|Q_{v,n}|^{2r+4}/r) \right]^{2r/(2r+4)} \alpha(m, 2sN_n)^{2r/(2r+4)} \\
\leq C \cdot \left[ E(||Z_{n,n}||^{4(2r+4)}r) + ||Z_{n,n}||^{2(2r+4)}r) \right]^{2r/(2r+4)} \alpha(m, 2sN_n)^{2r/(2r+4)} \\
\leq C \cdot (sN_n)^{-4} \alpha(m, 2sN_n)^{2r/(2r+4)}.
\end{align*}

by Lemmas 8.1 and 8.2. We have:

\begin{align*}
A_{4,n} &\leq C \cdot (sN_n)^2 \cdot \frac{1}{|J_{0}|^4} \left( -\sum_{m=1}^{\infty} \xi_{n}(m)\alpha(m, 2sN_n)^{2r/(2r+4)} \right) \\
&= O\left(|\text{det}(\Delta_n)|^{-1}\right) = o\left(\text{det}(\Delta_n) / \text{det}(\Delta_n)\right),
\end{align*}

from Lemma 8.2, the above bounds on \(\xi_n(\cdot)\) and the involved covariances, and the same arguments used in (20).

For the last term \(A_{6,n}\), we the previous moment bounds for \(A_{2,n}\) from (21) and for \(Q_{0,n}\):

\begin{align*}
A_{6,n} &\leq [sN_n^2 E(\hat{Y}_n^4)]^{1/4} \cdot [sN_n^2 E(\hat{Q}_n^4)]^{1/4} \\
&\leq C \cdot sN_n \cdot |J_{0}|^{-1} \left[ (sN_n^2 E(\hat{Q}_n^4)) \right]^{1/2} \\
&\leq C \cdot |J_{0}|^{-1} = o\left(\text{det}(\Delta_n) / \text{det}(\Delta_n)\right),
\end{align*}

using Lemma 8.3. We have now established (16)(b) and thus the proof of Theorem 3.1(a) is now finished.

9.2 Proof of Theorem 3.1(b)

The non-overlapping subsampling regions, by construction, may not always contain the same number of sampling sites, which complicates the proof slightly. To simplify the counting arguments involved, we first identify the set

\[sP_n = \{i \in s\Delta_n R_0 \cap Z^d : i + [-1/2, 1/2]^d \subset s\Delta_n R_0\}.\]

Note if \(k \in J_{0OL}\) then there exists a \(\nu_k \in Z^d\) such that \(\text{dis}(\nu_k, k) \leq 1/2\). Then we have that

\[\nu_k + sP_n \subset s\Delta_n (k + R_0) \cap Z^d.\]

In essence, there are at least \(|sP_n|\) sampling sites in each NOL subregion, \(s\Delta_n (i + R_0), i \in J_{0O}\). Also, applying the counting arguments used to show Lemma 8.3 (based on the boundary condition on \(R_0\)),

\[|sP_n - s\Delta_n R_0| \leq C \cdot (s\lambda_{m})^{d-1}.
\]
We denote, for each \(i \in \mathbb{Z}^d\), the NOL subregions \(R_{i,n} = \Delta_n(i + R_0)\) and their respective sample means

\[
\hat{Z}_{i,n} = \frac{1}{|N_{i,n}|} \sum_{s \in \hat{R}_{i,n} \cap \mathbb{Z}^d} Z(s), \quad \hat{N}_{i,n} = |\hat{R}_{i,n} \cap \mathbb{Z}^d|.
\]

As in the proof of Theorem 3.1(a), the subsample evaluations of the statistic of interest, \(\hat{H}_{i,n} = \hat{H}(\hat{Z}_{i,n})\), \(i \in J_{\text{NOL}}\), can be expressed through a Taylor's expansion of \(H(\cdot)\) around \(\mu\):

\[
\hat{H}_{i,n} = H(\mu) + \sum_{\|a\|_1 = 1} c_{i,n}(\hat{Z}_{i,n} - \mu)^a + 2 \sum_{\|a\|_1 = 2} \frac{(\hat{Z}_{i,n} - \mu)^a}{a!} \int_0^1 (1 - \omega) D^a H(\mu + \omega(\hat{Z}_{i,n} - \mu)) d\omega
\]

\[
\equiv H(\mu) + \hat{Y}_{i,n} + \hat{Q}_{i,n}.
\]

In the following, we assume \(\mu = 0\) (WLOG) and that each diagonal entry of \(\Delta_n\) is greater than 2, with the smallest diagonal component again denoted as \(\lambda_{\text{min}}^n\).

We will accomplish the proof of Theorem 3.1(b) in two parts by showing

(a) \(\text{Var} \left( \frac{1}{|J_{\text{NOL}}|} \sum_{i \in J_{\text{NOL}}} \hat{N}_{i,n} \hat{Y}_{i,n}^2 \right) = \frac{\text{det}(\Delta_n)}{\text{det}(\Delta_n)/|R_0|} \cdot [2\tau^4(1 + o(1))], \)

(b) \(\text{Var}(\hat{H}_{i,n}^2) - \text{Var} \left( \frac{1}{|J_{\text{NOL}}|} \sum_{i \in J_{\text{NOL}}} \hat{N}_{i,n} \hat{Y}_{i,n}^2 \right) = o(\text{det}(\Delta_n)/\text{det}(\Delta_n)). \)

We will begin with showing that (a) above holds. For \(k \in \mathbb{Z}^d\), let

\[
\hat{J}_n(k) = \{i \in J_{\text{NOL}} : i + k \in J_{\text{NOL}}\},
\]

and define

\[
\hat{Y}_{i,n}(k) = \sum_{i \in \hat{J}_n(k)} \hat{N}_{i,n} \cdot \hat{Y}_{i+k,n} \cdot \text{Cov}(\hat{Y}_{i,n}, \hat{Y}_{i+k,n}).
\]

Then we may express the variance:

\[
\text{Var} \left( \frac{1}{|J_{\text{NOL}}|} \sum_{i \in J_{\text{NOL}}} \hat{N}_{i,n} \hat{Y}_{i,n}^2 \right) = |J_{\text{NOL}}|^{-2} \left( \sum_{k \in \mathbb{Z}^d} \hat{\sigma}^2_n(k) + \sum_{k \in \mathbb{Z}^d} \hat{\sigma}^2_n(k) + \hat{\sigma}^2_n(0) \right)
\]

\[
\equiv U_{1,n} + U_{2,n} + |J_{\text{NOL}}|^{-2} \hat{\sigma}^2_n(0) \quad \text{(say)}. \]

We first prove \(U_{2,n} = o(|J_{\text{NOL}}|^{-1})\), noting that \(\text{det}(\Delta_n)/\text{det}(\Delta_n) = O(|J_{\text{NOL}}|^{-1}).\)

If \(k = (k_1, \ldots, k_d)' \in \mathbb{Z}^d\) such that \(\|k\|_\infty > 1\) and \(i \in \hat{J}_n(k)\), then

\[
\text{dis} (\hat{R}_{i,n} \cap \mathbb{Z}^d, \hat{R}_{i+k,n} \cap \mathbb{Z}^d) \geq \text{dis} \left( \Delta_n[-1/2, 1/2]^d, \Delta_n(k + [-1/2, 1/2]^d) \right)
\]

\[
= \max_{1 \leq j \leq d} (|k_j| - 1) \lambda_j^* \geq (|k_m^*| - 1) \lambda_m^* \quad \text{for some} \ m^* \in \{1, \ldots, d\}.
\]
The above follows from the fact that for $x \in \Delta_n[-1/2, 1/2]^d$, $y \in \Delta_n(k + [-1/2, 1/2]^d)$, $\text{dis}(x, y) \geq \max\{0, (|k_j| - 1) \cdot \lambda_{\max}^{(n)}\}$ for each $j = 1, \ldots, d$. And $\text{dis}(x^*, y^*) = (|k_{m^*}| - 1) \cdot \lambda_{\max}^{(n)}$ for $x^* = 1/2 \cdot \Delta_n(\text{sign}(k_1), \ldots, \text{sign}(k_d))$, $y^* = \Delta_n(k - x^*) \in \Delta_n(k + [-1/2, 1/2]^d)$.

Thus, if $j \in \{1, \ldots, d\}, j \neq m^*$, we then have $|k_j| \leq (|k_{m^*}| - 1) \cdot \lambda_{\max}^{(n)} \cdot (\lambda_{\max}^{(n)} - 1)$ + 1. Note also if $k \in \mathbb{Z}^d$, $||k||_\infty > 1$, and $i \in J_n(k)$, then

$$s_{N_{i,n} \cdot s_{N_{i+k,n}}}|\text{Cov}(\tilde{Y}_{i,n}^2, \tilde{Y}_{i+k,n}^2)| \leq C \cdot s_{N_{i,n} \cdot s_{N_{i+k,n}}} \cdot \frac{\text{E}(||\tilde{z}_{i,n}||^2)^{(2r+\delta)/r} \text{E}(||\tilde{z}_{i+k,n}||^2)^{(2r+\delta)/r})}{\alpha((|k_{m^*}| - 1) \cdot \lambda_{\max}^{(n)})^{\delta/(2r+\delta)}}$$

by Lemmas 8.1 and 8.2 and $s_{N_{i,n} \cdot s_{N_{i+k,n}}} \leq \det(\Delta_n)|R_0| + C \cdot (\lambda_{\max}^{(n)})^d < 1 \leq \det(\Delta_n)$ from (24).

Hence, we have

$$|U_{2,n}| \leq \frac{C}{|J_{N_{OL}|}} \sum_{j=1}^{\infty} 3^d(1 + x)^{-d} \sum_{j=1}^{d} (\lambda_{\max}^{(n)})^{-d} \prod_{i=1}^{d} (\text{E}(\text{E}(\tilde{z}_{i,n}^{(n)} \cdot \lambda_{\max}^{(n)}) \cdot \alpha((\text{C} \cdot \det(\Delta_n)))^{\delta/(2r+\delta)})$$

We now handle $U_{1,n}$ and we wish to show, as with $U_{2,n}$ above, that $U_{1,n} = o(|J_{N_{OL}|}^{-1})$. For $k \in \mathbb{Z}^d$, $0 < ||k||_\infty \leq 1$, fix $i \in J_n(k)$. Define the scaling quantity $\ell = (\lambda_{\max}^{(n)})^e$, where $e = 1/2 \cdot (1 + d \cdot \kappa \cdot \delta/((2r + \delta)(2rd - d - 1))) < 1$ (as in the proof of Theorem 3.1(a)).

For each coordinate direction $j = 1, \ldots, d$, let the set

$$T_{j,n}^{k,i} = \begin{cases}
(x \in \mathbb{R}^d : 1/2 \cdot \lambda_{\max}^{(n)} < x_j - i_j \leq 1/2 \cdot \lambda_{\max}^{(n)} + \ell) & \text{if } k_j = 1, \\
(x \in \mathbb{R}^d : -1/2 \cdot \lambda_{\max}^{(n)} - \ell < x_j - i_j \leq -1/2 \cdot \lambda_{\max}^{(n)}) & \text{if } k_j = -1, \\
\emptyset & \text{if } k_j = 0,
\end{cases}
$$

as well as the set $T_{n}^{k,i} = \bigcup_{j=1}^{d} T_{j,n}^{k,i}$. 

With respective to the same fixed \( i \in J_n(k) \), we then decompose the sum:

\[
s_{i+k,n} \tilde{Y}_{i+k,n} = \sum_{s \in \tilde{A}_{i+k,n} \cap \mathbb{Z}^d \setminus \mathbb{T}_n^{k,1}} \nabla' Z(s) + \sum_{s \in \tilde{A}_{i+k,n} \cap \mathbb{Z}^d \setminus \mathbb{T}_n^{k,1}} \nabla' Z(s) = \tilde{S}_{i,n} + \tilde{S}_{i,n}^*.\]

Then, we make the subsequent observations in (1)-(5) below:

(1) It holds that:

\[
\begin{align*}
|E(\tilde{Y}_{i,n}^2 \cdot \tilde{S}_{i,n} \cdot \tilde{S}_{i,n}^*)| &
\leq \left[ E(\tilde{Y}_{i,n}^4)E(|\tilde{S}_{i,n}|^2)E(|\tilde{S}_{i,n}^*|^2) \right]^{1/3} \\
&\leq \frac{C}{s_{N_{i,n}}} \left( (|\tilde{R}_{i+k,n} \cap \mathbb{Z}^d \cap \mathbb{T}_n^{k,1}|)^{1/2} \cdot (|\tilde{R}_{i+k,n} \cap \mathbb{Z}^d \setminus \mathbb{T}_n^{k,1}|)^{1/2} \right) \\
&\leq \frac{C}{s_{P_n}} \left( \ell \det(\Delta_n) \sum_{j=1}^d (s\lambda_j^{*(1)})^{-1} \right)^{1/2} (\det(\Delta_n)|R_0| + C \cdot (s\lambda_j^{**})^{d-1})^{1/2} \\
&= o(1),
\end{align*}
\]

by Lemma 8.2, Assumption A.1, \( |s_{P_n}| \leq s_{N_{i,n}} \leq \det(\Delta_n)|R_0| + C \cdot (s\lambda_j^{**})^{d-1} \) by (24), and the inequalities:

\[
|\tilde{R}_{i+k,n} \cap \mathbb{Z}^d \cap \mathbb{T}_n^{k,1}| \leq \sum_{j=1}^d |\tilde{R}_{i+k,n} \cap \mathbb{Z}^d \cap \mathbb{T}_n^{k,1}| \leq \ell \det(\Delta_n) \sum_{j=1}^d (s\lambda_j^{*(1)})^{-1}.
\]

(2) Likewise,

\[
\begin{align*}
E(\tilde{Y}_{i,n}^2 \cdot \tilde{S}_{i,n}^*) &\leq E(\tilde{Y}_{i,n}^4)E(\tilde{S}_{i,n}^4) \leq C \cdot s_{P_n}^{-1} \left( \ell \det(\Delta_n) \sum_{j=1}^d (s\lambda_j^{*(1)})^{-1} \right) = o(1).
\end{align*}
\]

(3) \( |E(\tilde{Y}_{i,n}^2 \cdot \tilde{S}_{i,n}^2) - E(\tilde{Y}_{i,n}^2)E(\tilde{S}_{i,n}^2) | = |\text{Cov}(\tilde{Y}_{i,n}^2, \tilde{S}_{i,n}^2)| \) and

\[
\begin{align*}
|\text{Cov}(\tilde{Y}_{i,n}^2, \tilde{S}_{i,n}^2)| &\leq E\left[ E(\nabla' Z(s) \cdot \nabla' Z(s)) \right]^{1/(2r+d)} \alpha(\ell, \max\{s_{N_{i,n}}, s_{N_{i+k,n}}\})^{d/(2r+d)} \\
&\leq C \cdot (\det(\Delta_n)|R_0| + C \cdot (s\lambda_j^{**})^{d-1}) |s_{P_n}^{-1} \alpha(\ell, [C \cdot \det(\Delta_n)])^{d/(2r+d)} \\
&= o(1),
\end{align*}
\]

by Lemmas 8.1 and 8.2 and \( \alpha(\ell, [C \cdot \det(\Delta_n)])^{d/(2r+d)} \leq C \cdot \ell^{-2(r+d)} \cdot \det(\Delta_n)^{-d/(2r+d)} = o(1) \) by Assumption A.3 and Condition \( M_r \) (similar arguments appear in the proof of Theorem 3.1(a)), and

\[
\text{dis}(\tilde{R}_{i+k,n} \cap \mathbb{Z}^d \setminus \mathbb{T}_n^{k,1}, \tilde{R}_{i,n} \cap \mathbb{Z}^d) \geq \ell.
\]
(4) Applying Lemma 8.2,
\[ |s_{N_{1+k,n}}E(\tilde{Y}_{i+k,n}^2) - \{s_{N_{1+k,n}}\}^{-1}E(\tilde{S}_{k,1,n}^2) | \]
\[ \leq 4\{s_{N_{1+k,n}}\}^{-1} \max_{x_{1}, \ldots, x_{2}} \left\{ \left[ E(\tilde{S}_{k,1,n}^2) \cdot E(\tilde{S}_{k,1,n}^2) \right]^{1/2}, E(\tilde{S}_{k,1,n}^2) \right\}^{1/2} \]
\[ \leq C \cdot \max_{x_{1}, \ldots, x_{2}} \left( |s_{P_n}|^{-1} \cdot \det(\Delta_n) \sum_{j=1}^{d}(\lambda_{j}^{\ast})^{-1} \right)^{1/2} \leq o(1). \]

(5) By Lemma 8.2, \[ s_{N_{1,n}}E(\tilde{Y}_{i,n}^2) \leq C \] and from (24),
\[ \frac{s_{N_{1,n}}}{s_{N_{1+k,n}}} \leq |s_{P_n}|^{-1} \cdot \det(\Delta_n) + C \cdot (\lambda_{\max}^{\ast})^{d-1} = O(1). \]

Because the bounds in items (1)-(5) above do not depend on \( i \in \tilde{J}(k) \), for \( k \in \mathbb{Z}^{d} \) with \( 0 < ||k||_{\infty} \leq 1 \), we can obtain the desired result for \( U_{1,n} \):
\[ U_{1,n} \leq \frac{1}{|J_{NOL}|} \sum_{i=2}^{2} \left[ \max_{i \in J_{NOL}} s_{N_{1,n}} \cdot s_{N_{1+k,n}} \cdot \text{Cov}(\tilde{Y}_{i,n}^2, \tilde{Y}_{i+k,n}^2) \right] = o \left( \frac{1}{|J_{NOL}|} \right). \]

We now closely examine the remaining quantity in (27), \[ |J_{NOL}|^{-2} \tilde{\sigma}^{\ast}(0) \], which asymptotically determines the variance on the left hand side of (25)(a). First, we note the expansion
\[ \frac{\tilde{\sigma}^{\ast}(0)}{|J_{NOL}|^2} = \frac{1}{|J_{NOL}|^2} \sum_{i \in J_{NOL}} [s_{N_{1,n}}]^2 \text{Var}(\tilde{Y}_{i,n}^2). \]

For \( i \in J_{NOL} \), write
\[ \tilde{S}_{i,n} = \sum_{s \in (\nu_{i}^{\ast}, J_{NOL})} \nabla'Z(s), \]
where the vector \( \nu_{i} \) is defined in (23). Applying the Cauchy-Schwartz inequality (twice),
\[ \left| \left[ s_{N_{1,n}} \right]^2 \text{Var}(\tilde{Y}_{i,n}^2) - \text{Var}(s_{N_{1,n}}^{-1} \tilde{S}_{i,n}^2) \right| \]
\[ \leq 2 \left[ 1 + \text{Var}(s_{N_{1,n}}^{-1} \tilde{S}_{i,n}^2) \right] \cdot \max_{x_{1}=1 \ldots 2} \left\{ \text{Var}(s_{N_{1,n}} \tilde{Y}_{i,n}^2 - s_{N_{1,n}}^{-1} \tilde{S}_{i,n}^2) \right\}^{1/2} \]
\[ \leq 2 \left[ 1 + (s_{N_{1,n}})^{-1} [E(\tilde{S}_{i,n}^2)]^{1/2} \right] \cdot \max_{x_{1}=1 \ldots 2} \left\{ E\left[ (s_{N_{1,n}} \tilde{Y}_{i,n}^2 - s_{N_{1,n}}^{-1} \tilde{S}_{i,n}^2) \right] \right\}^{1/2} \]
\[ \leq C \cdot \max_{x_{1}=1 \ldots 2} \left( \left( s_{N_{1,n}} - |s_{P_n}| \right)^{2}, s_{N_{1,n}} (s_{N_{1,n}} - |s_{P_n}|) \right)^{1/2} \]
\[ \leq C \cdot (\lambda_{\max}^{\ast})^{d-1} = o(1); \]

here we used also Lemma 8.2 for the bound: \[ E(\tilde{S}_{i,n}^2) \leq C \cdot [s_{N_{1,n}}]^2 \], \( i \in J_{NOL} \), as well as inequalities for set sizes: \( |s_{P_n}| \leq s_{N_{1,n}} \) and
\[ |s_{N_{1,n}} - |s_{P_n}| | \leq |s_{N_{1,n}} - \det(\Delta_n)^{1/2} + |s_{P_n} - \det(\Delta_n)^{1/2} | \leq C \cdot (\lambda_{\max}^{\ast})^{d-1}, \]
by the boundary condition for $R_0$ and (24). The above, in turn, yields

$$\frac{1}{|J_{\text{NOL}}|^2} \left| \beta_{\ast}^n(0) - \sum_{i \in J_{\text{NOL}}} \text{Var}(\{sN_{i,n}^{-1}S_{i,n}^2\}) \right| \leq \frac{1}{|J_{\text{NOL}}|} \left( \frac{C}{|sP_n|} \cdot 2(\lambda_{\ast n}^{\max})^{d-1} \right) = o \left( \frac{1}{|J_{\text{NOL}}|} \right)$$

and then

$$\frac{1}{|J_{\text{NOL}}|^2} \left| \sum_{i \in J_{\text{NOL}}} \text{Var}(\{sN_{i,n}^{-1}S_{i,n}^2\}) - |J_{\text{NOL}}| \cdot \text{Var}(|sP_n|^{-1}S_{0,n}^2) \right| \leq \frac{E(\tilde{S}_{0,n}^4)}{|J_{\text{NOL}}|} \max_{i \in J_{\text{NOL}}} \left| \frac{\{sN_{i,n}^2 - |sP_n|^2\}}{\{sN_{i,n}^2|sP_n|^2\}} \right|$$

$$\leq \frac{C}{|J_{\text{NOL}}|} \cdot \frac{|sP_n|^2 \cdot \det(s\Delta_{n})|R_0| + C \cdot (\lambda_{\ast n}^{\max})^{d-1}}{|sP_n|^4} = o \left( \frac{1}{|J_{\text{NOL}}|} \right),$$

using Lemma 8.2 and (28). We now have demonstrated that

$$\left| |J_{\text{NOL}}|^{-2} \tilde{\beta}_{\ast}^n(0) - |J_{\text{NOL}}|^{-1} \text{Var}(|sP_n|^{-1}S_{0,n}^2) \right| = o \left( \frac{1}{|J_{\text{NOL}}|} \right). \quad (29)$$

We then may focus simply on $\tilde{S}_{0,n}$ in evaluating $|J_{\text{NOL}}|^{-2} \tilde{\beta}_{\ast}^n(0)$. We find the limiting distribution of $\tilde{S}_{0,n}$:

$$\frac{1}{|sP_n|^{1/2}} \cdot \tilde{S}_{0,n} \xrightarrow{d} \nabla'Z_{\infty}, \text{ a normal } N(0, \tau^2) \text{ random variable,}$$

by applying the Bolthausen (1982) CLT under Assumptions A.3-A.4 with Condition M, and noting the boundary requirement is satisfied:

for $\partial[sP_n \cap Z^d] = \{i \in sP_n \cap Z^d : \text{there exists some } m \in Z^d \setminus sP_n, \text{dis}(i, m) = 1\},$

$$|\partial[sP_n \cap Z^d]| \leq |\{i \in Z^d : T^1 \cap s\Delta_{n} \bar{R}_0 \neq \emptyset, T^1 \cap s\Delta_{n} \bar{R}_0 \neq \emptyset, T^4 = i + [-2, 2]^d\}|$$

$$= o(|sP_n|).$$

Because, by Lemma 8.2, $E(\tilde{S}_{0,n}^4) \leq C \cdot |sP_n|^3$ for all $n$, we have that the collection of random variables

$\{\{sP_n|^{-1} \tilde{S}_{0,n}^2\}_{n=1}^{\infty}$ for $x = 1$ or 2, is uniformly integrable so that

$$E\left[\{sP_n|^{-1} \tilde{S}_{0,n}^2\} \rightarrow E((\nabla'Z_{\infty})^4) = 3r^4, \quad E\left[\{sP_n|^{-1} \tilde{S}_{0,n}^2\} \rightarrow E((\nabla'Z_{\infty})^2) = r^2, \quad (30)\right]$$

$$\text{Var}(\{sP_n|^{-1/2} \tilde{S}_{0,n}\}) \rightarrow 2r^4.$$ By combining (30) with the limit: $[\det(s\Delta_{n})|R_0|]/[\det(s\Delta_{n})|J_{\text{NOL}}|] \rightarrow 1$ (by Lemma 8.3) and using both (27) and (29) [that is, $U_1,n, U_2,n = o(\det(s\Delta_{n})/\det(\Delta_{n}))$, we establish (25)(a)].

We turn our attention now to showing (25)(b). We first rewrite the following difference, based on an algebraic expansion of $\tilde{S}_{n,\text{NOL}}^2$:

$$\tilde{S}_{n,\text{NOL}}^2 = |J_{\text{NOL}}|^{-1} \sum_{i \in J_{\text{NOL}}} sN_{i,n} \tilde{Y}_{i,n}^2 = \sum_{i=1}^{s} \tilde{A}_{i,n} \quad \text{where,} \quad (31)$$
\[ \dot{A}_{1,n} = 2|J_{\text{NOL}}|^{-1} \sum_{i \in \text{NOL}} s_{N_i,n} \dot{Y}_{i,n} \dot{Q}_{i,n} \quad \dot{A}_{2,n} = |J_{\text{NOL}}|^{-1} \sum_{i \in \text{NOL}} s_{N_i,n} \dot{Q}^2_{i,n} \]
\[ \dot{A}_{3,n} = |J_{\text{NOL}}|^{-1} \sum_{i \in \text{NOL}} s_{N_i,n} \dot{Y}_{i,n}^2 \quad \dot{A}_{4,n} = |J_{\text{NOL}}|^{-1} (\sum_{i \in \text{NOL}} s_{N_i,n}) \dot{Q}^2_{i,n} \]
\[ \dot{A}_{5,n} = -2|J_{\text{NOL}}|^{-1} \dot{Y}_{n} \sum_{i \in \text{NOL}} s_{N_i,n} \dot{Y}_{i,n} \quad \dot{A}_{6,n} = -2|J_{\text{NOL}}|^{-1} \dot{Q}_{n} \sum_{i \in \text{NOL}} s_{N_i,n} \dot{Y}_{i,n} \]
\[ \dot{A}_{7,n} = -2|J_{\text{NOL}}|^{-1} \dot{Y}_{n} \sum_{i \in \text{NOL}} s_{N_i,n} \dot{Q}_{i,n} \quad \dot{A}_{8,n} = -2|J_{\text{NOL}}|^{-1} \dot{Q}_{n} \sum_{i \in \text{NOL}} s_{N_i,n} \dot{Q}_{i,n} ; \]

and here we analogously define

\[ \dot{Y}_{n} = |J_{\text{NOL}}|^{-1} \sum_{i \in \text{NOL}} \dot{Y}_{i,n} ; \quad \dot{Q}_{n} = |J_{\text{NOL}}|^{-1} \sum_{i \in \text{NOL}} \dot{Q}_{i,n} . \]

By the Cauchy-Schwarz inequality, the left hand side of (25)(b) is less than

\[ \left( \sum_{i=1}^{8} \{ \text{Var}(\dot{A}_{i,n}) \}^{1/2} \right)^2 + \left( \sum_{i=1}^{8} \{ \text{Var}(\dot{A}_{i,n}) \}^{1/2} \right)^2 \cdot \text{Var} \left( |J_{\text{NOL}}|^{-1} \sum_{i \in \text{NOL}} s_{N_i,n} \dot{Y}_{i,n}^2 \right) \]

With the established order of the variance in (25)(a), it suffices to show that \( \text{Var}(\dot{A}_{i,n}) = o(|J_{\text{NOL}}|^{-1}), \) \( i = 1, \ldots, 8. \) After noting \( \det(\dot{A}_n)/\det(A_n) \rightarrow |R_0| \) by Lemma 8.3, we will then have the desired proof of (25)(b).

For \( k \in \mathbb{Z}^d \) and \( \hat{J}_n(k) \) from (26), we apply Lemma 8.2 to get the following covariance bound:

\[ \max_{i \in J_n(k)} s_{N_i,n} \cdot s_{N_{i+k,n}} \text{Cov}(\dot{Y}_{i,n} \dot{Q}_{i,n}, \dot{Y}_{i+k,n} \dot{Q}_{i+k,n}) \]
\[ \leq \max_{i \in J_n(k)} s_{N_i,n} \cdot s_{N_{i+k,n}} (\text{E}[(\dot{Y}_{i,n} \dot{Q}_{i,n})^2] \text{E}[(\dot{Y}_{i+k,n} \dot{Q}_{i+k,n})^2])^{1/2} \]
\[ \leq C \cdot \max_{i \in J_n(k)} s_{N_i,n} \cdot s_{N_{i+k,n}} \left[ \text{E}[||\dot{Y}_{i,n}||^6 + ||\dot{Z}_{i,n}||^6] \text{E}[||\dot{Y}_{i+k,n}||^6 + ||\dot{Z}_{i+k,n}||^6] \right]^{1/2} \]
\[ \leq C \cdot \max_{i \in J_n(k)} s_{N_i,n} \cdot s_{N_{i+k,n}} \leq C \cdot |s_{P_n}|^{-1} . \quad (32) \]

And for \( k \in \mathbb{Z}^d \) such that \( ||k|| \leq 1, \) we apply Lemmas 8.1 and 8.2 to bound the covariances

\[ \max_{i \in J_n(k)} s_{N_i,n} \cdot s_{N_{i+k,n}} \text{Cov}(\dot{Y}_{i,n} \dot{Q}_{i,n}, \dot{Y}_{i+k,n} \dot{Q}_{i+k,n}) \]
\[ \leq \max_{i \in J_n(k)} \left( 8 \cdot s_{N_i,n} \cdot s_{N_{i+k,n}} (\text{E}[(\dot{Y}_{i,n} \dot{Q}_{i,n})^{(2r+4)/(r')}] \text{E}[(\dot{Y}_{i+k,n} \dot{Q}_{i+k,n})^{(2r+4)/(r')}] \right)^{r/(2r+4)} \]
\[ \cdot \alpha \left( \max_{1 \leq j \leq d} (||k_j|| - 1) \lambda_j^{(s)}, \max_{s_{N_i,n} \cdot s_{N_{i+k,n}}} \right)^{r/(2r+4)} \]
\[ \leq C \cdot |s_{P_n}|^{-1} \alpha \left( \max_{1 \leq j \leq d} (||k_j|| - 1) \lambda_j^{(s)}, \max_{s_{N_i,n} \cdot s_{N_{i+k,n}}} \right)^{r/(2r+4)} \quad (33) \]

resembling the bounds used for handling \( U_{2,n} \) in (27). Following symmetrical arguments (based on Lemmas 8.1 and 8.2),

\[ \max_{i \in J_n(k)} s_{N_i,n} \cdot s_{N_{i+k,n}} |\text{Cov}(\dot{Q}_{i,n}^2, \dot{Q}_{i+k,n}^2)| \leq C \cdot |s_{P_n}|^{-2} , \]
Using the covariance bounds from (32), (33), and the two in (34), we can repeat the same arguments used to bound $U_{1,n}, U_{2,n}$ from (27) and find:

$$\text{Var}(\hat{A}_{i,n}) = O\left(\frac{1}{|\mathcal{N}_{\text{COL}}| \cdot |sP_n|}\right) = o\left(\frac{1}{|\mathcal{N}_{\text{COL}}|}\right),$$

$$\text{Var}(\hat{A}_{2,n}) = O\left(\frac{1}{|\mathcal{N}_{\text{COL}}| \cdot |sP_n|^2}\right) = o\left(\frac{1}{|\mathcal{N}_{\text{COL}}|}\right).$$

Similar to the treatment of the $'A_{2,n}'$ term in the proof of Theorem 3.1(a) for $\hat{r}_{n,OL}^2$ (see (21)), we rewrite $\tilde{Y}_n = (|sP_n| \cdot |\mathcal{N}_{\text{COL}}|)^{-1} \sum_{l \in \mathcal{N}_{\text{COL}}} sP_n |\tilde{Y}_{l,n}$ as a weighted sum. To this end, let $\omega(s) = |sP_n| \cdot |sN_{l,n}|^{-1} \cdot I_{\{s \in \mathcal{N}_{l,n}\}}, s \in \mathbb{Z}^d$, so that $0 \leq \omega(s) \leq 1$. Now apply the moment bound from Doukhan (1994, Theorem 3, p.31) and (28):

$$\text{Var}(\hat{A}_{3,n}) \leq \frac{C}{|\mathcal{N}_{\text{COL}}|^2} \frac{\text{max}_{l \in \mathcal{N}_{\text{COL}}} |sN_{l,n}|^4}{|sP_n|^4} \leq \frac{C}{|\mathcal{N}_{\text{COL}}|^2} \frac{(\text{det}(s\Delta_n) + C \cdot (s\lambda_{\text{max}}^d)^{-1})^4}{|sP_n|^4} = O(|\mathcal{N}_{\text{COL}}|^{-2}).$$

Likewise, by Lemma 8.2,

$$E\left[\left(\sum_{l \in \mathcal{N}_{\text{COL}}} sN_{l,n} \tilde{Y}_{l,n}\right)^4\right] \leq \frac{C}{|\mathcal{N}_{\text{COL}}|^4} \cdot \left(\sum_{l \in \mathcal{N}_{\text{COL}}} sN_{l,n}\right)^2 \leq \frac{C}{|\mathcal{N}_{\text{COL}}|^2} \cdot \text{(det}(s\Delta_n) + C \cdot (s\lambda_{\text{max}}^d)^{-1})^2$$

so that

$$\text{Var}(\hat{A}_{5,n}) \leq \frac{C}{|\mathcal{N}_{\text{COL}}|^2} \cdot \text{(det}(s\Delta_n) + C \cdot (s\lambda_{\text{max}}^d)^{-1})^2 = O(|\mathcal{N}_{\text{COL}}|^{-2}).$$
With respect to the bounds on $E(Y_n^4)$ and (36), we can use elementary moment inequalities to produce

\[
\text{Var}(\hat{A}_{7,n}) \leq E(\hat{A}_{7,n}^2) \\
\leq \left[ E(Y_n^4) \cdot \max_{i \in J_{NO} \cup} \left[ \left| s_i \right| \cdot \left| \lambda_{n,i}^\ast \right| \right] \right]^{1/2} \\
\leq C \cdot \left[ |s| \cdot \max_{i \in J_{NO} \cup} \left| \lambda_{n,i}^\ast \right| \right]^{1/2} \\
\leq C \cdot |s|^{-2} |J_{NO} \cup| \cdot \left( \max_{i \in J_{NO} \cup} \left| \lambda_{n,i}^\ast \right| \right) \\
\leq O(|J_{NO} \cup|^{-1}) = o(1).
\]

We also have, with counting and moment arguments analogous to those for bounding $|V_{2,n}|$ in (27), that

\[
\text{Var}(\hat{A}_{8,n}) \leq E(\hat{A}_{8,n}^2) \\
\leq \left[ E\left( \left( |J_{NO} \cup|^{-1} \sum_{i \in J_{NO} \cup} \left| s_i \right| \right)^{1/2} \cdot \left[ \left| \max_{i \in J_{NO} \cup} \left| \lambda_{n,i}^\ast \right| \right] \right) \right]^{1/2} \\
\leq C \cdot \left[ |s|^{-2} |J_{NO} \cup| \cdot \left( \sum_{i \in J_{NO} \cup} \left| s_i \right| \right) \right]^{1/2} \\
\leq O(|J_{NO} \cup|^{-1}) = o(1).
\]

This finishes the proof of (25)(b). With (25) established, the proof of Theorem 3.1(b) is now complete.

\[\square\]

10 Proofs for Sections 4, 5 and 6

10.1 Preliminaries

We will use the following lemma concerning $\tau_n^2 = N_n \text{Var}(\tilde{\theta}_n)$ to prove the theorems pertaining to bias expansions of $\hat{\tau}_{n,4}^2$ and $\hat{\tau}_{n,NO}^2$. 
Lemma 10.1 Under the Assumptions and Conditions of Theorem 3.1,

(a) \( \tau_n^2 = \tau^2 + O((\det(\Delta_n))^{-1/2}) \)

(b) \( \sqrt{N_n} (\hat{\theta}_n - \theta) \overset{d}{\to} N(0, \tau^2) \).

Proof. We begin with showing Lemma 10.1(a). By a Taylor's expansion of \( \hat{\theta}_n = H(\hat{Z}_{N_n}) \) around \( \mu \), we have

\[
\hat{\theta}_n = H(\mu) + \sum_{||a||=1} c_a (\hat{Z}_{N_n} - \mu)^a + 2 \sum_{||a||=2} \frac{c_a}{2!} (\hat{Z}_{N_n} - \mu)^a \int_0^1 (1 - \omega) D^a H(\mu + \omega(\hat{Z}_{N_n} - \mu))d\omega
\]

and so \( N_n \text{Var}(\hat{\theta}_n) = N_n \text{Var}(\hat{Z}_{N_n} + \hat{Q}_{N_n}) \). Assume WLOG that \( \mu = 0 \). For \( k \in \mathbb{Z}^d \), let \( N_n(k) = |\{ i \in R_n \cap \mathbb{Z}^d : i + k \in R_n \}| \). Then, for \( k \neq 0 \in \mathbb{Z}^d \), it holds that

\[
N_n(k) \leq N_n,
\]

\[
N_n \leq N_n(k) + |\{ i \in \mathbb{Z}^d : T^1 \cap \Delta_n \neq \emptyset, T^1 \cap \Delta_n \neq \emptyset; T^1 = i + ||k||_{\infty} [-1, 1]^d \}|
\]

\[
\leq N_n(k) + (3||k||_{\infty})^d |\{ i \in \mathbb{Z}^d : T^1 \cap \Delta_n \neq \emptyset, T^1 \cap \Delta_n \neq \emptyset; T^1 = (\lambda_{n}^{\max})^{-1}(i + [0, 1]^d) \}|
\]

\[
\leq N_n(k) + C \cdot (||k||_{\infty})^d (\lambda_{n}^{\max})^{d-1},
\]

by the boundary condition on \( R_0 \). Also, by Lemma 8.1 and stationarity, for each \( k \neq 0 \in \mathbb{Z}^d \):

\[
|\sigma(k)| \leq 8||V||^2 \cdot \left[ E(||Z(t)||(2r+5)/r) \right]^{2r/(2r+5)} \alpha(||k||_{\infty}, 1)^{d/(2r+5)}
\]

\[
\leq C \cdot \alpha_1(||k||_{\infty})^{d/(2r+5)}, k \in \mathbb{Z}^d.
\]

Using \(|\{ k \in \mathbb{Z}^d : ||k||_{\infty} = x \}| \leq 2(2x + 1)^{d-1} \), the covariances are absolutely summable over \( \mathbb{Z}^d \):

\[
\sum_{k \in \mathbb{Z}^d} |\sigma(k)| \leq |\sigma(0)| + C \cdot \sum_{x=1}^{\infty} 2(2x + 1)^{d-1} \alpha_1(x)^{d/(2r+4)} < \infty.
\]

Using the set and covariance bounds above and (40), we find

\[
N_n \text{Var}(\hat{Z}_{N_n}) = \frac{1}{N_n} \sum_{k \in \mathbb{Z}^d} N_n(k) \sigma(k) = \tau^2 + I_n;
\]

\[
\tau^2 = \sum_{k \in \mathbb{Z}^d} \sigma(k)
\]

where with (38) and (40):

\[
|I_n| \leq \frac{1}{N_n} \sum_{k \in \mathbb{Z}^d} |N_n - N_n(k)| \cdot |\sigma(k)|
\]

\[
\leq C \cdot \frac{(\lambda_{n}^{\max})^{d-1}}{N_n} \sum_{x=1}^{\infty} x^{2d-1} \alpha_1(x)^{d/(2r+4)} = O \left( (\det(\Delta_n))^{-1/4} \right).
\]
By elementary moment bounds and Lemma 8.2,

\[ N_n \text{Var}(Q_{N_n}) \leq N_n E(Q_{N_n}^2) \leq C \cdot N_n \left[ E(\|\mathcal{Z}_{N_n}\|^4) + E(\|\mathcal{Z}_{N_n}\|^{4+2a}) \right] = O(\det(\Delta_n)^{-1}). \]

Finally, with the above variance bounds for \( \hat{Y}_{N_n} \) and \( Q_{N_n} \), we apply the Cauchy-Schwartz inequality to the covariance

\[ N_n |\text{Cov}(\hat{Y}_{N_n}, Q_{N_n})| \leq \left[ N_n \text{Var}(\hat{Y}_{N_n}) \right]^{1/2} \left[ N_n \text{Var}(Q_{N_n}) \right]^{1/2} \leq C \cdot \det(\Delta_n)^{-1/2}. \]

The order of the difference \( |N_n \text{Var}(\hat{Q}_n) - \tau^2| \) is set by the bound on the covariance \( \text{Cov}(\hat{Y}_{N_n}, Q_{N_n}) \).

To establish Lemma 10.1(b), we can apply the Bolthausen (1982) CLT to a normalized form of \( \hat{Y}_{N_n} = \nabla' (\mathcal{Z}_{N_n} - \mu) \). We note also that the boundary condition for the CLT is satisfied because

\[ \partial[R_n \cap Z^d] = \{ i \in R_n \cap \mathbb{Z}^d : \text{there exists some } m \in R_n \cap \mathbb{Z}^d, \text{dis}(i, m) = 1 \}, \]

\[ |\partial[R_n \cap Z^d]| \leq \left| \{ i \in Z^d : T^i \cap \Delta_n \mathcal{R}_0 \neq \emptyset, T^i \cap \Delta_n \mathcal{R}^c_0 \neq \emptyset; T^i = i + [-1, 1]^d \} \right| \]

\[ \leq C \cdot (\lambda^{-a})^{d-1} = o(N_n). \]

Hence, by Lemma 10.1(a) and \( E(\hat{Y}_{N_n}) = 0, n \geq 1 \), we apply the CLT under Assumptions A.1-A.5 and Condition \( M_r \):

\[ \sqrt{N_n} \cdot \hat{Y}_{N_n} \xrightarrow{d} \mathcal{N}(0, \tau^2). \]

In the proof of Lemma 10.1(a), we established that \( N_n \text{E}(\hat{Q}_{N_n}^2) \longrightarrow 0 \), so that \( \sqrt{N_n} \cdot \hat{Q}_{N_n} \) converges to zero in probability. Because \( \theta = H(\mu) \), the distributional result in (b) above follows from Slutsky's theorem. □

We first establish a few lemmas which help compute the bias of the estimators \( \hat{\tau}_{n,OL}^2 \) and \( \hat{\tau}_{n,NOOL}^2 \).

**Lemma 10.2** Let \( \tilde{A}_{i,n} = \Delta_n (i + \mathcal{R}_0) \), \( sN_{i,n} = |Z^d \cap \tilde{A}_{i,n}| \), \( \tilde{Y}_{i,n} = [sN_{i,n}]^{-1} \sum_{z \in Z^d \cap \tilde{A}_{i,n}} \nabla'(Z(z) - \mu) \)

for \( i \in \mathbb{Z}^d \). Suppose \( d \geq 2 \) and that, in addition to Assumptions A.1 - A.5, Conditions D2 and M_{2+a} hold. Then,

\[ E(\hat{\tau}_{n,OL}^2) - \tau^2 = O \left( \{\text{det}(\Delta_n)\}^{-1/2} + o \left( \{\text{det}(\Delta_n)\}^{-1/4} \right) \right). \]

\[ E(\hat{\tau}_{n,NOOL}^2) - |J_{n,OL}|^{-1} \sum_{i \in J_{n,OL}} sN_{i,n} \cdot E(\tilde{Y}_{i,n}^2) \]

Proof: We shall begin by considering \( E(\hat{\tau}_{n,OL}^2) \) and adopt the same notation used in (13) and (14) \( (Z_{i,n}, Y_{i,n}, \text{etc.}) \). By stationarity and the algebraic expansion in (15):

\[ E(\hat{\tau}_{n,OL}^2) = sN_n \left[ E(\tilde{Y}_{i,n}^2) + E(Q_{i,n}^2) + 2E(\tilde{Y}_{i,n} Q_{i,n}) - E(\tilde{Y}_{i,n}^2) - E(Q_{i,n}^2) - 2E(\tilde{Y}_{i,n} Q_{i,n}) \right]. \]
With the moment arguments based on Lemma 8.2 and Condition \( D_r \), we have

\[
\begin{align*}
\varsigma N_n E(Y_{0,n}^2) &\leq C, \\
\varsigma N_n E(Q_{0,n}^2) &\leq C \cdot (\varsigma N_n)^{-1}, \\
\varsigma N_n E(Y_{1,n}^2) &\leq C \cdot \varsigma N_n \cdot (N_n)^{-1}, \\
\varsigma N_n E(Q_{1,n}^2) &\leq \varsigma N_n E(Q_{0,n}^2),
\end{align*}
\]

where bound on \( \varsigma N_n E(\hat{Y}_{1,n}^2) \) follows from (21). By Holder's inequality and Assumption A.2:

\[
E(\hat{Y}_{0,n}^2) = \varsigma N_n \cdot E(Y_{0,n}^2) + O\left((\varsigma N_n)^{-1/2}\right) + O\left(\varsigma N_n \cdot (N_n)^{-1}\right).
\]

Note that \( E(Y_{0,n}^2) = E(Y_{0,n}^2)\), \( \varsigma N_n = \varsigma N_{0,n} \). Hence, applying Lemma 8.3 and \( \det(\varsigma \Delta_n)/\det(\Delta_n) = o\left(\det(\varsigma \Delta_n)\right) \) by Assumption A.2, we establish Lemma 10.2 for \( \tilde{\tau}_{n,OL}^2 \).

We now show Lemma 10.2 for \( \tilde{\tau}_{n,NO}^2 \). An algebraic expansion of \( \tilde{\tau}_{n,NO}^2 \) is given in (31). We now use the growth rates in Lemma 8.3 and also (28). By the moment bound arguments in (32) and (34), it follows that

\[
E|\tilde{A}_{1,n}| \leq \max_{k \in J_{NOL}} \max_{l \in J_n(k)} \varsigma N_{1,n} \cdot \varsigma N_{l+k,n} \left[ E(\hat{Y}_{l,n}^2) E(Q_{l,n}^2) \right]^{1/2} \leq C \cdot \{\det(\varsigma \Delta_n)\}^{-1/2},
\]

\[
E|\tilde{A}_{2,n}| \leq \max_{i \in J_{NOL}} \varsigma N_{1,n} \cdot E(Q_{i,n}^2) \leq C \cdot \{\det(\varsigma \Delta_n)\}^{-1}.
\]

Likewise, from elementary expectation inequalities, (28), and Lemma 8.2,

\[
E|\tilde{A}_{4,n}| \leq \left[ \max_{i \in J_{NOL}} \varsigma N_{1,n} \right] \left[ \max_{i \in J_{NOL}} E(Q_{i,n}^2) \right] \leq C \cdot \{\det(\varsigma \Delta_n)\}^{-1},
\]

\[
E|\tilde{A}_{8,n}| \leq \left( \max_{i \in J_{NOL}} \varsigma N_{1,n} E(Q_{i,n}^2) \right)^{1/2} \leq C \cdot \{\det(\varsigma \Delta_n)\}^{-1}.
\]

With (35) and (36) (and the other corresponding variance bounds) and Holder's inequality, we have

\[
E|\tilde{A}_{3,n}|, E|\tilde{A}_{5,n}| \leq C \cdot |J_{NOL}|^{-1}; \quad E|\tilde{A}_{6,n}|, E|\tilde{A}_{7,n}| \leq C \cdot \{\det(\Delta_n)\}^{-1/2}.
\]

Hence, taking expectations in (31), we now find

\[
E(\tilde{\tau}_{n,NO}^2) = |J_{NOL}|^{-1} \sum_{i \in J_{NOL}} \varsigma N_{1,n} \cdot E(\hat{Y}_{i,n}^2) + O\left(\{\det(\varsigma \Delta_n)\}^{-1/2}\right) + O\left(|J_{NOL}|^{-1}\right).
\]

Lemma 10.2 for \( \tilde{\tau}_{n,NO}^2 \) then follows from Lemma 8.3 and Assumption A.2. \( \square \)

We establish the following lemma that provides a small refinement to Lemma 10.2 made possible when the function \( H(\cdot) \) is smoother. We shall make use of this lemma in bias expansions of \( \tilde{\tau}_{n,OL}^2 \) and \( \tilde{\tau}_{n,NO}^2 \) in lower sampling dimensions, namely \( d = 1 \) or \( 2 \).
Lemma 10.3 Assume $d = 1$ or $2$. In addition to Assumptions A.1 - A.5, suppose that Conditions D_3 and M_{3+a}. Then,

\[
E(\hat{\tau}_{n,\text{OL}}^2) = s_N n \cdot E(\hat{Y}_{0,n}^2) = \begin{cases} \frac{O\left(\left\{\det\left(s_{\Delta_n}\right)\right\}^{-1}\right)}{\sigma\left(\left\{\det\left(s_{\Delta_n}\right)\right\}^{-1/2}\right)} & \text{if } d = 1, \\
\end{cases}
\]

Proof: Consider first $\hat{\tau}_{n,\text{OL}}^2$. For $i \in J_{\text{OL}}$, we use a third-order Taylor's expansion of each subsample statistic around $\mu$: \( \tilde{\theta}_{i,n} = H(\mu) + Y_{i,n} + Q_{i,n} + C_{i,n} \), where \( Y_{i,n} = \nabla'(Z_{i,n} - \mu) \);

\[
Q_{i,n} = \sum_{\|a\| = 2} \frac{c_{\alpha}}{a!} (Z_{i,n} - \mu)^a + C_{i,n} = 3 \sum_{\|a\| = 3} \frac{c_{\alpha}}{a!} (Z_{i,n} - \mu)^a \int_0^1 (1 - \omega)^a D^a H(\mu + \omega(Z_{i,n} - \mu)) d\omega.
\]

Here \( C_{i,n} \) denotes the remainder term in the Taylor's expansion and \( Q_{i,n} \) is defined a little differently.

Write the sample means for the Taylor terms: \( \tilde{Y}_n = |J_{\text{OL}}|^{-1} \sum_{i \in J_{\text{OL}}} Y_{i,n} \); \( \tilde{Q}_n = |J_{\text{OL}}|^{-1} \sum_{i \in J_{\text{OL}}} Q_{i,n} \); \( \tilde{C}_n = |J_{\text{OL}}|^{-1} \sum_{i \in J_{\text{OL}}} C_{i,n} \). As in (15), we find after some algebra and taking expectations:

\[
E(\hat{\tau}_{n,\text{OL}}^2) = s_N n \cdot E\left[ \tilde{Y}_{0,n}^2 + Q_{0,n}^2 + C_{0,n}^2 + 2\tilde{Y}_{0,n} Q_{0,n} + 2\tilde{Y}_{0,n} C_{0,n} + 2Q_{0,n} C_{0,n} \right]
\]

by stationarity. The moment inequalities in (43) are still valid and, by Lemma 8.2 and Condition D_3, we can produce bounds:

\[
s_N n E(C_{0,n}^2) \leq C \cdot (s_N n)^{-2}, \quad s_N n E(C_{0,n}^2) \leq s_N n E(C_{0,n}^2),
\]

using the appropriate moments of \( \|Z_{0,n} - \mu\| \). By Holder's inequality and the scaling conditions from Assumptions A.1-A.2, we then have

\[
E(\hat{\tau}_{n,\text{OL}}^2) = s_N n \left[ E(\tilde{Y}_{0,n}^2) + 2E(Y_{0,n} Q_{0,n}) \right] + \begin{cases} \frac{O\left(\left\{\det\left(s_{\Delta_n}\right)\right\}^{-1}\right)}{\sigma\left(\left\{\det\left(s_{\Delta_n}\right)\right\}^{-1/2}\right)} & \text{if } d = 1, \\
\end{cases}
\]

Since \( s_N n E(Y_{0,n}^2) = s_N n E(\tilde{Z}_{0,n}^2) \), it now suffices to show

\[
s_N n E(Y_{0,n} Q_{0,n}) = O(\{\det(s_{\Delta_n})\}^{-1})
\]

to establish Lemma 10.3 for $\hat{\tau}_{n,\text{OL}}^2$.

To this end, we assume WLOG that $\mu = 0$ and write

\[
E(Y_{0,n} Q_{0,n}) = \sum_{i=1}^p \sum_{j=1}^p \sum_{k=1}^p c_{i,j,k} Z_{i,0,n} Z_{j,0,n} Z_{k,0,n}, \quad \text{where} \quad (48)
\]

\[
c_{i,j,k} = H(\mu) / \partial x_i, \quad a_{i,j,k} = \begin{cases} \frac{\partial^2 H(\mu)}{\partial x_j \partial x_k} & \text{if } j \neq k, \\
1/2 \cdot \frac{\partial^2 H(\mu)}{\partial^2 x_j} & \text{if } j = k \end{cases}
\]

for \( i, j, k \in \{1, \ldots, p\} \).
\[ Z_{0,n} = (Z_{(1,0,n)}, \ldots, Z_{(p,0,n)})' \in \mathbb{R}^p \] is the vector of coordinate sample means. Fix \( i, j, k \in \{1, \ldots, p\} \).

For \( s \in \mathbb{Z}^d \), denote the observation \( Z(s) = (Z_1(s), \ldots, Z_p(s))' \in \mathbb{R}^d \). Then,

\[
E(Z_i(t)Z_j(u)Z_k(w)) = E(Z_i(u)Z_j(v))E(Z_k(w)),
\]

where \( t, u, v, w \in \mathbb{Z}^d \).

By stationarity, Lemma 8.1, Assumption A.3, and Condition \( M_r \), we have:

\[
(sN_n)^{-2}|L_{1,n}| \leq C \cdot (sN_n)^{-1}E(||z(t)||^2),
\]

\[
(sN_n)^{-2}|L_{2,n}| \leq \frac{C}{(sN_n)^2} \sum_{x=1}^{\infty} sN_n \cdot \{|k \in \mathbb{Z}^d : ||k||_{\infty} = x\}|(x, 1)^{\beta/(2r+\delta)}
\]

\[
\leq \frac{C}{sN_n} \sum_{x=1}^{\infty} x^{d-1} \alpha_1(x)^{\beta/(2r+\delta)}
\]

\[
\leq C \cdot (sN_n)^{-1}.
\]

The next step follows from Lahiri (1999a). For \( y_1, y_2, y_3 \in \mathbb{R}^d \), define

\[
\text{dis}_3({y_1, y_2, y_3}) = \max_{1 \leq i \leq 3} \text{dis}(\{y_1\}, \{y_1, y_2, y_3\} \setminus \{y_i\}).
\]

If \( z \geq 1 \in \mathbb{Z}, \) then \([[y_1, y_2] \in (\mathbb{Z}^d)^2 : \text{dis}_3({y_1, y_2, 0}) = x]) \leq Cz^{2d-1} \) from Lahiri (1999a, Theorem 4.1). Thus,

\[
(sN_n)^{-2}|L_{3,n}| \leq \frac{C}{(sN_n)^2} \sum_{x=1}^{\infty} sN_n \cdot \{|(y_1, y_2) \in (\mathbb{Z}^d)^2 : \text{dis}_3({y_1, y_2, 0}) = x\}|(x, 2)^{\beta/(2r+\delta)}
\]

\[
\leq \frac{C}{sN_n} \sum_{x=1}^{\infty} x^{2d-1} \alpha_1(x)^{\beta/(2r+\delta)}
\]

\[
\leq C \cdot (sN_n)^{-1},
\]

using Lemma 8.1, Assumption A.3, and Condition \( M_r \) again. Thus, by (48):

\[
(sN_n)^{-1}E(Y_{2,n}Q_{0,n}) \leq \frac{C}{(sN_n)^2} \sum_{i=1}^{\infty} |L_{i,n}| \leq C \cdot (sN_n)^{-1}.
\]

We have now shown (47) and completed the proof of Lemma 10.3 for \( \hat{\theta}_{n,OL}^2 \).
We briefly outline the proof for $\tilde{f}_{n,\text{NOL}}$, which involves the same essential techniques presented above with some minor counting modifications based on the set size inequalities in (36). Taylor's Theorem is used to expand the subsample statistics on a NOL subregion: $\hat{\theta}_{i,n} = H(\tilde{Z}_{i,n}) = H(\mu) + \tilde{Y}_{i,n} + \tilde{Q}_{i,n} + \tilde{C}_{i,n}$, $i \in J_{\text{NOL}}$, up to a third order remainder term (analogous to the case for $\tilde{f}_{n,\text{OL}}$ but using $\tilde{Z}_{i,n}$ instead of $Z_{i,n}$). Similar to the handling of the quantities $\tilde{A}_{i,n}, i = 1, \ldots, 8$, in the proof of Lemma 10.2, one can use (45) [which are still valid though $\tilde{Q}_{i,n}$ is not a remainder term] and also inequalities resulting from Lemma 8.2, Condition $D_r$, and (28): 

$$\max_{i \in J_{\text{NOL}}} s N_{i,n} E(\tilde{Y}_{i,n}^2) \leq C, \quad \max_{i \in J_{\text{NOL}}} s N_{i,n} E(\tilde{Q}_{i,n}^2) \leq C \cdot |P_n|^{-1}, \quad \max_{i \in J_{\text{NOL}}} s N_{i,n} E(\tilde{C}_{i,n}^2) \leq C \cdot |P_n|^{-2} \quad (50)$$

to show (after algebraically expanding $\tilde{f}_{n,\text{NOL}}$):

$$E(\tilde{f}_{n,\text{NOL}}^2) = \frac{1}{|J_{\text{NOL}}|} \sum_{i \in J_{\text{NOL}}} s N_{i,n} \left[ E(\tilde{Y}_{i,n}^2) + 2E(\tilde{Y}_{i,n} \tilde{Q}_{i,n}) \right] + \begin{cases} O \left(\{\text{det}(s \Delta_n)\}^{-1}\right) & \text{if } d = 1, \\ o \left(\{\text{det}(s \Delta_n)\}^{-1/2}\right) & \text{if } d = 2, \end{cases}$$

by using Holder's inequality and Assumptions A.1-A.2. For each $i \in J_{\text{NOL}}$, redefine the previous quantities $L_{i,n}, i = 1, 2, 3$, by substituting $s N_{i,n}, \tilde{A}_{i,n}$ in place of $s N_{i,n}, R_n$; by repeating the same arguments, we can bound each $|L_{i,n}| \leq C \cdot (s N_{i,n})^{-1}$ and $(s N_{i,n})^{-1} \leq |P_n|^{-1}$ from (28). Then,

$$\max_{i \in J_{\text{NOL}}} s N_{i,n} |E(\tilde{Y}_{i,n} \tilde{Q}_{i,n})| \leq \frac{C}{|P_n|} = O \left(\{\text{det}(s \Delta_n)\}^{-1}\right),$$

from which the rest of Lemma 10.3 follows. □

We use the next lemma in the proof of Theorem 4.2. It allows us to approximate lattice point counts with Lebesgue volumes, in $\mathbb{R}^2$ or $\mathbb{R}^3$, to a "sufficient" degree of accuracy. In fact, we obtain a fairly sharp bound on relevant approximation error with "nearly convex" templates $R_0$ in $\mathbb{R}^2$. [Namely, if $d = 2$ and $R_0^2$ contains a closed ball around the origin of radius $\epsilon$, then we can take $C_\epsilon = \epsilon^{-1/2}$ in Lemma 10.4(a); see Nordman (2002) for more details.]

**Lemma 10.4 (a)** Let $d = 2, 3$ and $R_0 \subset (-1/2, 1/2)^d$ be "nearly convex" as defined in Theorem 4.2. Let $\{b_n\}_{n=1}^\infty$ be a sequence of positive real numbers such that $b_n \to \infty$. If $k \in \mathbb{Z}^d$, then there exists $N_k \in \mathbb{Z}^+$ such that for $n \geq N_k$, $i \in \mathbb{Z}^d$,

$$\left| \left( |b_n R_0| - |b_n(i + R_0) \cap \mathbb{Z}^d| \right) - \left( |b_n R_0 \cap (k + b_n R_0)| - |b_n(i + R_0) \cap (k + b_n(i + R_0)) \cap \mathbb{Z}^d| \right) \right| \leq \begin{cases} C_k \|k\|_\infty^2 & \text{if } d = 2 \\ C_k \left( b_n^{5/3} + C_{k,n} b_n^2 \right) & \text{if } d = 3, \end{cases}$$

where $C_k$ is a constant dependent on $k$. 

Proof: (Sketch) Consider the difference between the volumes of two sets, $|b_n(i + R_0) \cap \mathbb{Z}^d|$ and $|b_n R_0 \cap (k + b_n R_0)|$. The term $\{\text{det}(s \Delta_n)\}^{-1}$ in (50) bounds the approximation error up to third order in $N_{i,n}$.
where $C_k > 0$ and $\{\xi_{k,n}\}_{n=1}^{\infty} \subset \mathbb{R}$ is a nonnegative sequence (possibly dependent on $k$) such that $\xi_{k,n} \to 0$.

(b) Let $R_0 = R_1 \cup R_2 \subset (-1/2,1/2]^3$. Suppose there exist convex sets $A_1, A_2$ such that $A_j^0 \neq \emptyset$, $A_j^0 \subset A_j$ for $j = 1, 2$. Assume that $\partial A_1 \cap \partial A_2 = (A_1 \setminus A_2^0) \cap (A_2 \setminus A_2^0)$ is empty or finite. Then, the bound in Lemma 10.4(a) for $d = 2$ is valid and, in addition,

$$\left| \left( |b_n R_0| - |b_n R_0 \cap (k + b_n R_0)| \right) - \left( \sum_{j=1}^{2} \Gamma_{j,n}(k) - \Gamma_{n}(k) \right) \right| \leq C_k,$$

$$\Gamma_{n}(k) = |b_n (R_1 \cap R_2)| - |b_n (R_1 \cap R_2) \cap (k + b_n (R_1 \cap R_2))|; \quad \Gamma_{j,n}(k) = |b_n R_j| - |b_n R_j \cap (k + b_n R_j)|, \quad j = 1, 2.$$


10.2 Proofs: Lemma 4.1; Corollaries 4.1, 6.1; Theorems 4.1, 4.2, 5.1, 6.1

To establish Lemma 4.1, we require some additional notation. For $i, k \in \mathbb{Z}^d$, let

$$s N_{i,n}(k) = |\hat{R}_{i,n} \cap (k + \hat{R}_{i,n}) \cap \mathbb{Z}^d|,$

the number of sampling sites or lattice points in the intersection of a NOL subregion with its $k$-translate.

Proof of Lemma 4.1: Note, for $i, k \in \mathbb{Z}^d$, $s N_{i,n}(k)$ is a subsample version of $N_n(k)$ in (38). Then,

$$s N_{i,n}(k) \leq s N_{i,n} \leq \hat{N}_{i,n}(k)$$

$$s N_{i,n}(k) + |\{j \in \mathbb{Z}^d : T^j \cap s \Delta_n(i + \hat{R}_0) \neq \emptyset, T^j \cap s \Delta_n(i + \hat{R}_0) \neq \emptyset; \ T^j = j + \|k\|_\infty [-1, 1]^d\}|$$

$$\leq s N_{i,n}(k) + |\{j \in \mathbb{Z}^d : T^j \cap s \Delta_n \hat{R}_0 \neq \emptyset, T^j \cap s \Delta_n \hat{R}_0 \neq \emptyset; \ T^j = j + \|k\|_\infty [-2, 2]^d\}|$$

$$\leq s N_{i,n}(k) + 5\|k\|_\infty d |\{j \in \mathbb{Z}^d : T^j \cap \hat{R}_0 \neq \emptyset, T^j \cap \hat{R}_0 \neq \emptyset; \ T^j = (\lambda_n^{\text{max}})^{-1}(j + [0, 1]^d)\}|$$

$$\leq s N_{i,n}(k) + C \cdot (\|k\|_\infty)^d (\lambda_n^{\text{max}})^{-d-1}, \quad (51)$$

by the boundary condition on $R_0$ and $\inf_{i \in \mathbb{Z}^d} \|s \Delta_n \hat{i} - j\|_\infty \leq 1/2$.

Modify (41) by replacing $N_n, N_n(k), \hat{Y}_{N_n}$ with $s N_{i,n}, s N_{i,n}(k), \hat{Y}_{i,n} = \nabla'(\hat{Z}_{i,n} - \mu)$ (i.e. use a NOL subsample version in place of the sample one); and replace $N_n, \Delta_n, \lambda_n^{\text{max}}$ with the subsample analogs $s N_{i,n}, s \Delta_n, s \lambda_n^{\text{max}}$ in (42). We then find [using (40)] for each $i \in \mathbb{Z}^d$

$$s N_{i,n} \cdot E(\hat{Y}_{i,n}^2) - \tau^2 = \frac{1}{s N_{i,n}} \sum_{k \in \mathbb{Z}^d} (s N_{i,n}(k) - s N_{i,n}) \sigma(k) \equiv \hat{I}_{i,n}, \quad (52)$$
where
\[
\sup_{l \in \mathbb{Z}^d} |s_{l,n}| \leq \sup_{l \in \mathbb{Z}^d} \left\{ \frac{1}{s_{n,l,n}} \sum_{k \in \mathbb{Z}^d} |s_{N_{l,n}(k)} - s_{N_{l,n}}| \cdot |\sigma(k)| \right\}
\leq C \cdot \left( \frac{(\omega \lambda_n^{\max})^{d-1}}{|s_{P_n}|} \sum_{x=1}^{\infty} x^{2d-1}\alpha_1(x)^{d/(2r+d)} \right)
\leq C \cdot (\det(s_{A_n}))^{-1/d},
\] (53)

using (51), (28) and Assumption A.1. Now applying Lemma (10.1)(a) and Assumption A.2; Lemma 10.2 for \(d \geq 2\) and Lemma 10.3 for \(d = 1\); and Assumption A.1, Lemma 4.1 follows. □

Proof of Theorem 4.1: Here \(s_{N_{l,n}} = s_{N_n}, s_{N_{l,n}(k)} = C_n(k), E(\hat{Y}_{l,n}) = E(\hat{Y}_{0,n})\) for each \(i, k \in \mathbb{Z}^d\) (since \(s_{\lambda_n} \in \mathbb{Z}_+\) and \(\det(s_{A_n}) = (s_{\lambda_n})^d\). Applying Lemma 10.2 for \(d \geq 3\) and Lemma 10.3 for \(d = 2\), Lemma 10.1(a), and Assumption A.2,
\[
E(\hat{r}_n) - r_n^2 = s_{N_n}E(\hat{Y}_{0,n}) - r^2 + o((s_{\lambda_n})^{-1})
= \frac{-1}{s_{\lambda_n}|R_0|} \sum_{k \in \mathbb{Z}^d} g_n(k) + o((s_{\lambda_n})^{-1}),
\]
g_n(k) \equiv \frac{s_{N_n} - C_n(k)}{(s_{\lambda_n})^{d-1} \cdot s_{N_n}} \cdot \sigma(k)

using (53) as well. By (53) and Lemma 8.3, note
\[
\sum_{k \in \mathbb{Z}^d} |g_n(k)| \leq C, n \in \mathbb{Z}_+,
\]
g_n(k) \rightarrow C(k)\sigma(k), k \in \mathbb{Z}^d.

Hence, by the Lebesgue Dominated Convergence Theorem (LDCT), the proof of Theorem 4.1 is complete. □

To establish Theorem 4.2, we require some additional notation. For \(i, k \in \mathbb{Z}^d\), let
\[
D_{i,n}(k) = \left( |\hat{r}_{i,n}| - s_{N_{i,n}} \right) - \left( \left| \hat{r}_{i,n} \cap (k + \hat{r}_{i,0}) \right| - s_{N_{i,n}(k)} \right)
= \left( |s_{\lambda_n}R_0| - s_{N_{i,n}} \right) - \left( |s_{\lambda_n}R_0 \cap (k + s_{\lambda_n}R_0) - s_{N_{i,n}(k)}| \right)
\]
denote the difference between two Lebesgue volumes-for-lattice-point-count approximations.

Proof of Theorem 4.2: We handle here the cases \(d = 2\) or \(3\) and defer the proof for \(d = 1\) to Section 12. We note first that if \(V(k)\) exists for each \(k \in \mathbb{Z}^d\), then Lemma 10.4(a) implies \(C(k) = V(k)\).
Consider first $\tau_{n,NOL}^2$. Applying Lemma 10.2 for $d = 3$, and Lemma 10.3 for $d = 2$, with (52) gives

$$E(\tau_{n,NOL}^2) - \tau_n^2 = \left| J_{n,NOL} \right|^{-1} \sum_{i \in J_{n,NOL}} s_i n_i + o\left( (s\lambda_n)^{-1} \right)$$

$$= \left| J_{n,NOL} \right|^{-1} \sum_{i \in J_{n,NOL}} s_i n_i + o\left( (s\lambda_n)^{-1} \right),$$

by (28) and (53). Then, using (40), we can arrange terms to write

$$\left| J_{n,NOL} \right|^{-1} \sum_{i \in J_{n,NOL}} s_i n_i = \Psi_n + \sum_{k \in \mathbb{Z}^d} G_n(k)$$

$$G_n(k) = \sum_{i \in J_{n,NOL}} s_i n_i + o\left( (s\lambda_n)^{-1} |J_{n,NOL}| \right),$$

for $\Psi_n = -\sum_{k \in \mathbb{Z}^d} |s R_n|^{-1} (\sum_{k \in \mathbb{Z}^d} s R_n (k) \cap (k + s R_n)) |\sigma(k)|$. Since $R_0$ is convex, the boundary condition is valid and it holds that for all $i, k \in \mathbb{Z}^d$

$$\left| s N_{n,i} (k) - [s \lambda_n R_0 \cap (k \cap s \lambda_n R_0)] \right| = \left| s N_{n,i} (k) - [i R_{i,n} \cap (k \cap i R_{i,n})] \right|$$

$$\leq \left| \{ j \in \mathbb{Z}^d : T^j \cap s \Delta_n (i + R_0) \neq \emptyset \}, T^j \cap s \Delta_n (i + R_0) \neq \emptyset \} \right| 1$$

$$\leq C \cdot (s\lambda_n)^{d-1}, \quad (54)$$

following the same argument establishing Lemma 8.3(c); and

$$\left| |s R_n| - |s R_n \cap (k + s R_n)| \right| \leq C \cdot (||k||_{\infty})^{d} (s\lambda_n)^{d-1} \quad (55)$$

applying (51) and (54). Then (40), Lemma 10.4(a), (54), and (55) give

$$\sum_{k \in \mathbb{Z}^d} |G_n(k)| \leq C, \quad n \in \mathbb{Z}^+; \quad G_n(k) \rightarrow 0, \quad k \in \mathbb{Z}^d; \quad s\lambda_n \cdot \Psi_n = O(1).$$

By the LDCT, we establish

$$\sum_{k \in \mathbb{Z}^d} G_n(k) = o\left( (s\lambda_n)^{-1} \right); \quad E(\tau_{n,NOL}^2) - \tau_n^2 = \Psi_n \left( 1 + o(1) \right),$$

which is the formulation of Theorem 4.2 in terms of $\Psi_n$. If $V(k)$ exists for each $k \in \mathbb{Z}^d$, then (40) and (55) imply that we can use the LDCT again to produce

$$\Psi_n = \frac{-1}{s\lambda_n |R_0|} \left( \sum_{k \in \mathbb{Z}^d} V(k) |\sigma(k)| \right) (1 + o(1)). \quad (56)$$

Hence, we have established Theorem 4.2 for $\tau_{n,NOL}^2$.

Consider now $\tilde{\tau}_{n,NOL}^2$. We can repeat the same steps as above to produce:

$$E(\tilde{\tau}_{n,NOL}^2) - \tau_n^2 = \Psi_n + \sum_{k \in \mathbb{Z}^d} G_n^*(k)$$

$$G_n^*(k) = \frac{D_0(n,k) \cdot |\sigma(k)|}{(s\lambda_n)^{d-1}}.$$
The same arguments above for $G_n$ apply to $G_n^*$ and the arguments in (56) also remain valid when each $V(k)$ exists, $k \in \mathbb{Z}^d$, which establishes Theorem 4.2 for $\mathbb{T}_{n,OL}^2$. Note as well that if $V(k)$ exists for each $k \in \mathbb{Z}^d$, then Lemma 10.4(a) and Lemma 4.1 also imply the second formulation of the bias in Theorem 4.2. The proof of Theorem 4.2 is finished. □.

**Proof of Theorem 5.1:** Follows from Theorems 3.1 and 4.1 and simple arguments from Calculus involving minimization of a smooth function of a real variable. □.

**Proof of Corollary 6.1:** The same proof for Theorem 4.2 applies, using Lemma 10.4(b) in place of Lemma 10.4(a). □

**Proof of Theorem 6.1:** To reiterate, here $R_0^d = \prod_{j=1}^d(-\ell_j/2, \ell_j/2)$, $\overline{R_0}^d = \prod_{j=1}^d[-\ell_j/2, \ell_j/2]$ for $\{\ell_1, \ldots, \ell_d\} \subset (0, 1]$.

We now develop some tools to facilitate the counting arguments, required for expanding the biases of $\mathbb{T}_{n,OL}^2$ and $\mathbb{T}_{n,NOL}^2$. We first define a count for the number of $\mathbb{Z}^d$ lattice points lying in a "face" of a $d$-dimensional rectangle (with "faces" parallel to the coordinate axes). For a rectangle or cube $T = \prod_{j=1}^d[c_j, \tilde{c}_j] \subset \mathbb{R}^d$, $c_j, \tilde{c}_j \in \mathbb{R}$, define the following "border" point set (in terms of the closure $\overline{T}$):

$$B(T) = \bigcup_{j=1}^d \left\{ \mathbf{s} = (s_1, \ldots, s_d) \in \mathbb{Z}^d : s_j \in \{c_j, \tilde{c}_j\} \right\}.$$ 

We note now that, for $i, k \in \mathbb{Z}^d$,

$$\left(\mathbb{Z}^d \cap (k + s\lambda_n(i + R_0)) \right) \setminus \left(\mathbb{Z}^d \cap (k + s\lambda_n(i + R_0)) \right) = B\{k + s\lambda_n(i + R_0)\} \setminus B\{k + s\lambda_n(i + R_0)\}.$$  

It also holds that, for each $k, i \in \mathbb{Z}^d, d \geq 2$,

$$|B\{s\lambda_n(i + R_0)\}| - |B\{s\lambda_n(i + R_0) \cap (k + s\lambda_n(i + R_0))\}| = |B\{s\lambda_n(i + R_0)\} \setminus \left( (k + s\lambda_n(i + R_0)) \cup (-k + s\lambda_n(i + R_0)) \right)|$$

and there exists $N_k \in \mathbb{Z}_+$ such that: for $n \geq N_k$, any $i \in \mathbb{Z}^d$ (and $t \in [-1/2, 1/2]^d$),

$$|B\{s\lambda_n(i + R_0)\} \setminus \left( (k + s\lambda_n(i + R_0)) \cup (-k + s\lambda_n(i + R_0)) \right)| \leq 3^{2(d+1)}||k||_\infty (s\lambda_n)^{d-2}.  

We give a sketch of the proof of (58), which involves an induction argument. Assume WLOG that $s\lambda_n \min_{1 \leq j \leq d} \ell_j \geq 2||k||_\infty$. For $d = 2$, the inequality in (58) is straightforward to check. Assume the
inequality holds for $d$ in the induction step. For $s \lambda_n(i + \mathbb{R}_0) \subset \mathbb{R}^{d+1}$, pick one of the $d + 1$ coordinate directions and select a corresponding "face" of this rectangle, say

$$F_{j,m,n} \equiv \{ s = (s_1, \ldots, s_{d+1})' \in s \lambda_n(i + \mathbb{R}_0) : s_j = s \lambda_n(i + (-1)^m \cdot \ell_j/2) \}$$

for $j = 1, \ldots, d + 1$ and $m = 1, 2$. Define also the vectors

$$k_j = (k_1, \ldots, k_{j-1}, 0, k_{j+1}, \ldots, k_{d+1})' \in \mathbb{Z}^{d+1}, \quad 1 \leq j \leq d + 1.$$ 

Note that if $s$ is an element of $s \lambda_n(i + \mathbb{R}_0)$ but $s \pm k$ are not, then it holds that there is some $j, m$ and $x \in F_{j,m,n} \setminus [(k_j + F_{j,m,n}) \cup (-k_j + F_{j,m,n})\},$ where $x_p = s_p$ for $p \in \{1, \ldots, d+1\} \setminus \{j\}$. That is, components of $s$ (except the $j$th) match those of a point on an edge of some "face" $F_{j,m,n}$ ($s$ might not lie on the face itself). Treating the face $F_{j,m,n}$ as a $d$-dimensional rectangle and using the induction hypothesis, we can bound the maximum potential number of $\mathbb{Z}^d$ lattice points on the edges of this face that do not remain on the same face when translated by $\pm k_j$; for a given point $x$ in $F_{j,m,n}$, we then bound the maximum number of possible boundary lattice points for $s \lambda_n(i + \mathbb{R}_0)$, which share the same values as $x$ in all coordinate directions except the $j$th by $3s \lambda_n$; we may repeat the same tactic for each "face." Hence,

$$|B(s \lambda_n(i + \mathbb{R}_0)) \setminus \left( (k + s \lambda_n(i + \mathbb{R}_0)) \cup (-k + s \lambda_n(i + \mathbb{R}_0)) \right)|$$

$$\leq \sum_{m=1}^{2} \sum_{j=1}^{d+1} \left| \mathbb{Z}^d \cap F_{j,m,n} \setminus \left( (k_j + F_{j,m,n}) \cup (-k_j + F_{j,m,n})\right) \right| \cdot (3s \lambda_n)$$

$$\leq 3^{2(d+1)} \cdot 3^{2(d+1)} ||k||_{\infty} (s \lambda_n)^{d-1},$$

completing the sketched proof of (58).

Fix $k \neq 0 \in \mathbb{Z}$ and assume $n$ is large enough so that $s \lambda_n \min_{1 \leq j \leq d} \ell_j > 3 + ||k||_{\infty}$. Then we write, for $i \in \mathbb{Z}^{d}$:

$$|D_{i,n}(k)| = \left| \sum_{j=1}^{4} (P_{2j-1,i,n} - P_{2j,i,n}) \right| ;$$

$$P_{1,i,n} = \prod_{j=1}^{d} s \lambda_n(\ell_j), \quad P_{2,i,n} = \prod_{j=1}^{d} (s \lambda_n(\ell_j) - |k_j|), \quad P_{3,i,n} = |B(s \lambda_n(i + \mathbb{R}_0))|, \quad P_{4,i,n} = |B(s \lambda_n(i + \mathbb{R}_0))|,$$

$$P_{5,i,n} = \prod_{j=1}^{d} s \lambda_n(\ell_j + |k_j|), \quad P_{6,i,n} = \prod_{j=1}^{d} (s \lambda_n(i + \mathbb{R}_0) - \ell_j \cap \mathbb{Z}), \quad P_{7,i,n} = \prod_{j=1}^{d} (s \lambda_n(i + \mathbb{R}_0) - \ell_j \cap \mathbb{Z} - |k_j|),$$

$$P_{8,i,n} = |B(s \lambda_n(i + \mathbb{R}_0) \cap [k + s \lambda_n(i + \mathbb{R}_0)]|, \quad P_{9,i,n} = |B(s \lambda_n(i + \mathbb{R}_0) \cap [k + s \lambda_n(i + \mathbb{R}_0)]|).$$

We use (57) above to express

$$sN_{i,n} = P_{5,i,n} - P_{4,i,n} + P_{3,i,n}, \quad sN_{i,n}(k) = P_{6,i,n} - P_{7,i,n} + P_{8,i,n}.$$ 

We show now that, for $k \neq 0$, there exists $N_k \in \mathbb{Z}^+, C_k > 0$, such that: $n \geq N_k, i \in \mathbb{Z}^{d}$

$$|D_{i,n}(k)| = \left| \sum_{j=1}^{4} (P_{2j-1,i,n} - P_{2j,i,n}) \right| \leq C_k \cdot (s \lambda_n)^{d-2}, \quad i \in \mathbb{Z}^{d}.$$ (59)
By (58), for large \( n \geq N_0 \),

\[
0 \leq (P_{3,i,n} - P_{4,i,n}) \leq (P_{4,i,n} - P_{7,i,n}) \leq C_k \cdot (s_\lambda n)^{d-2}.
\]

Note now that: for \( i \in Z, j \in \{1, \ldots, d\} \),

\[
\left| Z \cap (s_\lambda n (i + \ell_j [-1/2, 1/2]) - \ell_j) - s_\lambda n \cdot \ell_j \right| \leq 3.
\]

Hence, for \( s_\lambda n \min_{1 \leq j \leq d} \ell_j > 3 + \|k\|_\infty \),

\[
\left| (P_{3,i,n} - P_{4,i,n}) - (P_{4,i,n} - P_{7,i,n}) \right| 
\leq \sum_{A \subseteq \{1, \ldots, d\} \atop |A| \leq d} \left( \|k\|_\infty \right)^{|A|} \left| \prod_{j \in A} s_\lambda n \ell_j - \prod_{j \in A} s_\lambda n (i_j + \ell_j [-1/2, 1/2]) - \ell_j \cap Z \right|
\leq \sum_{A \subseteq \{1, \ldots, d\} \atop |A| \leq d} 3d \cdot (\|k\|_\infty)^{|A|} (s_\lambda n)^{|A| - 1}
\leq d \cdot 3^d \cdot (\|k\|_\infty)^{d-1} (s_\lambda n)^{d-2},
\]

which holds uniformly in \( i \in Z^d \). With these bounds, we have (59).

Applying (59) [in place of Lemma 10.4(a)], the same proof used for Theorem 4.2 establishes Theorem 6.1. \( \square \)

Remark. For creating rectangular subsamples, if \( s_\lambda j^{(\ast)} \) denotes the scaling factor in the \( j \)th direction, \( j = 1, \ldots, d \) (rather than using a common directional scaling factor), it is not difficult to show that the estimators \( \hat{\tau}_{n,j}^2 \) and \( \hat{\tau}_{n,k\times k}^2 \) have asymptotic bias:

\[
\sum_{k \in \mathbb{Z}^d} \left( \sum_{j=1}^d \frac{|\ell_j|}{\lambda_j^{(\ast)}} \ell_j \right) \sigma(k) + o \left( \max_{1 \leq j \leq d} (s_\lambda j^{(\ast)})^{-1} \right),
\]

under the Assumptions and Conditions of Theorem 6.1.

11 Appendix 1: Proof of Theorem 9.1

11.1 Preliminaries

The next lemma is due to some measure theoretic considerations in the proof of Theorem 9.1.

Lemma 11.1 For a \( d \times d \) positive diagonal matrix, \( \Delta_0 \),

(a) The function \( \psi : \mathbb{R}^d \to \mathbb{R} \) defined as \( \psi(x) = 1[(x + \Delta_0 R_0) \cap \Delta_0 R_0] \) is Borel measurable and bounded.

(b) Sets of the form \( \{ x \in \mathbb{R}^d : \psi(x) = \epsilon \} \), for some \( \epsilon \geq 0 \), can have positive Lebesgue measure for at
most a countable collection.

(c) For $x \in \mathbb{R}^d$, $\psi(x) = |\Delta_0||R_0|$ if and only if $x = 0$.

Proof: (a) Let $g(x, y) : \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{R}$ defined as $g(x, y) = I_{\{y \in \Delta_0 R_0\}} f_{\{(y-x) \in \Delta_0 R_0\}}$. By Fubini's theorem, $\psi(x) = \int_{\mathbb{R}^d} g(x, y) dy = |(\Delta_0 R_0 + x) \cap \Delta_0 R_0|$ is a measurable function of $x$. Clearly, $|\psi(x)| \leq |\Delta_0 R_0| < \infty$ for all $x \in \mathbb{R}^d$.

(b) By (a) above, the sets ${x \in \mathbb{R}^d : \psi(x) = \epsilon}$, $\epsilon \geq 0$, are Borel measurable and for a Borel set $A \subset \mathbb{R}^d$, the measure $\nu(A) = \int_A \psi(x) dx$ is finite. From Billingsley (1986, Theorem 10.4), the set ${\epsilon \geq 0 : |\{x \in \mathbb{R}^d : \psi(x) = \epsilon\}| > 0}$ is at most countable.

(c) Let $x = (x_1, \ldots, x_d)' \neq 0 \in \mathbb{R}^d$. Suppose, WLOG, $\tilde{a} = |x_1|/2 > 0$. For $z \in \mathbb{R}^d$, define the first coordinate projection $\pi_1(z) = z_1 \in \mathbb{R}$ and let $Q_x = \{z \in \mathbb{R}^d : (-\pi_1(x) \cdot \pi_1(z))/|\pi_1(x)| \in [0, \infty]\}$. Let $m \in \mathbb{R}^d$ be the maximum of the continuous function $|\pi_1(\cdot)|$ on the compact set $R_0^d \cap Q_x$. For an open ball of radius $\tilde{a}$ around $m$, say $B(m, \tilde{a})$, we have that $B(m, \tilde{a}) \cap \Delta_0^d \neq \emptyset$ is open ($m$ is a closure point of $R_0^d$). This implies that $|B(m, \tilde{a}) \cap \Delta_0^d| > 0$. For $z \in B(m, \tilde{a}) \cap \Delta_0^d$, $|\pi_1(z) - \pi_1(m)| < \tilde{a}$ and for $z \in x + \Delta_0^d$, $|\pi_1(z) - \pi_1(m)| \geq |x_1|$; these facts together imply $(B(m, \tilde{a}) \cap \Delta_0^d) \cap (x + \Delta_0^d) = \emptyset$. By the boundary condition on $R_0$, $|R_0^d| = |R_0| = |\Delta_0^d|$ so that

$$|R_0| - |R_0 \cap (x + \Delta_0^d)| = |R_0^d \setminus (x + \Delta_0^d)| \geq |B(m, \tilde{a}) \cap \Delta_0^d| > 0.$$ 

Because $\Delta_0$ is nonsingular, we have that $|R_0| > |R_0 \cap ((\Delta_0)^{-1}x + \Delta_0)|$ for all $x \neq 0 \in \mathbb{R}^d$. $\square$

11.2 Proof of Theorem 9.1

WLOG we assume that $\mu = 0$; that $\text{det}(\lambda \Delta_n)/(\lambda^{n+1})^d \geq 1/2 \cdot \text{det}(\Delta_0)$ by Assumption A.1; and that each diagonal component of $\lambda \Delta_n$ is greater than 1. Note that, for $k \in Z^d$,

$$(\lambda \Delta_n)^2 \text{Cov}(Y_{k,n}^2, Y_{k,n}^2) = (\lambda \Delta_n)^2 E(Y_{k,n}^2 Y_{k,n}^2) - \lambda \Delta_n E(Y_{k,n}^2)^2,$$

by stationarity of the r.f. $Z(\cdot)$. For a set $T \subset \mathbb{R}^d$, define the function $\nabla (\cdot)$ as

$$\nabla (T) = \sum_{s \in Z^d \cap T} \nabla Z(s).$$

Let $\ell = (\lambda \Delta_n)^{\tilde{a}}$ with $\tilde{a} = 1/2 + (\kappa \cdot \tilde{a})/(2 \cdot [2r - 1 - 1/d] \cdot [2r + \tilde{a}])$. Now, for any $k \in Z^d$, we define a set,

$$B_k^\ell = \{i \in Z^d : i \in k + \lambda \Delta_n R_0, (i + \ell(-1, 1)^d) \cap (k + \lambda \Delta_n R_0^d) \neq \emptyset\}$$
of lattice points "near" the border of the translate \( R_{k,n} = k + \xi R_n \). With the boundary condition on \( R_0 \),

\[
|B^k_n| = |B^0_n| \leq |\{i \in \mathbb{Z}^d : T^4 \cap \xi \Delta_n R_0 \neq \emptyset, T^4 \cap \xi \Delta_n R_0 \neq \emptyset; T^4 = i + 2t(-1, 1)^d\}| \\
\leq C \cdot 2t \cdot (\xi \lambda_{\text{max}}^d)^{-1}
\]

(60)

uniformly in \( k \in \mathbb{Z}^d \) [following the same arguments from (12)].

Next define the intersection, as a function of \( k \in \mathbb{Z}^d \), \( R^{(i)}_{k,n} = R_n \cap (k + \xi R_n) \) and write

\[
H_{1,n}(k) = \mathbb{E}(R_{k,n} \setminus R^{(i)}_{k,n}) \quad \text{(a sum over sites in } R_{k,n} \text{ but not } R_{0,n} = \xi R_n),
\]

\[
H_{2,n}(k) = \mathbb{E}(R_{0,n} \setminus R^{(i)}_{k,n}) \quad \text{(a sum over sites in } R_{0,n} \text{ but not } R_{k,n}),
\]

\[
H_{3,n}(k) = \mathbb{E}(R^{(i)}_{k,n}) \quad \text{(a sum over sites in both } R_{0,n} \text{ and } R_{k,n}).
\]

Define a function \( h_n(\cdot) : \mathbb{Z}^d \rightarrow \mathbb{R} \) as:

\[
h_n(k) = \mathbb{E}[\xi f^2_n(k)] \mathbb{E}[\xi f^2 B_n(k)] + \mathbb{E}[\xi f_n(k)] \mathbb{E}[\xi f_n(k)] + \mathbb{E}[\xi f^3_n(k)] - (\xi N_n)^4 \mathbb{E}[\xi f^4_n(k)].
\]

We will now use the following proposition; the proof is given separately in Section 11.3.

**Proposition 11.1** Under the Assumptions and Conditions of Theorem 3.1,

\[
\max_{k \in E_n} |(\xi N_n)^2 \text{Cov}(Y^2_{0,n}, Y^2_{k,n}) - (\xi N_n)^{-3} h_n(k)| = o(1).
\]

(61)

By (61) and \( |E_n| = O(\xi N_n) \), we have

\[
\left| \xi N_n \sum_{k \in E_n} \text{Cov}(Y^2_{0,n}, Y^2_{k,n}) - (\xi N_n)^{-3} \sum_{k \in E_n} h_n(k) \right| = o(1).
\]

(62)

Consequently, we need only focus on \((\xi N_n)^{-3} \sum_{k \in E_n} h_n(k)\) to complete the proof.

For measurability reasons to be encountered later, we create the following set defined in terms of the \( \mathbb{R}^d \) Lebesgue measure:

\[
E^+ = \left\{ 0 < \epsilon < \min\{1, (\det(\Delta_0)|R_0|)/2\} : \left| \{x \in \mathbb{R}^d : |(x + \Delta_0 R_0) \cap \Delta_0 R_0| = \epsilon \text{ or } (\det(\Delta_0)|R_0| - \epsilon) \right| = 0 \right\}.
\]

By Lemma 11.1(b), the set \( \{0 < \epsilon < \min\{1, (\det(\Delta_0)|R_0|)/2\} : \epsilon \notin E^+ \} \) is at most countable. For \( \epsilon \in E^+ \), define a new set as a function of \( \epsilon \) and \( n \):

\[
\tilde{R}_{\epsilon,n} = \{k \in \mathbb{Z}^d : |R^{(i)}_{k,n}| > \epsilon (\xi \lambda_{\text{max}}^d), |R_n \setminus R^{(i)}_{k,n}| > \epsilon (\xi \lambda_{\text{max}}^d)\}.
\]

Note also \( \tilde{R}_{\epsilon,n} \subset E_n \), because \( k \notin E_n \) implies \( R^{(i)}_{k,n} = \emptyset \).

We now further expand \((\xi N_n)^{-3} \sum_{k \in E_n} h_n(k)\) using the following proposition involving the newly defined \( \tilde{R}_{\epsilon,n} \).
Proposition 11.2 There exist \( N \in \mathbb{Z}_+ \) and a function \( b(\cdot) : \mathbb{E}^+ \rightarrow \mathbb{R}^+ \cong (0, \infty) \) such that \( b(\epsilon) \downarrow 0 \) as \( \epsilon \downarrow 0 \) and

\[
(\epsilon N_n)^{-3} \left| \sum_{k \in \mathcal{E}_n} h_n(k) - \sum_{k \in \mathcal{E}_n} h_n(k) \right| \leq C \left( \epsilon + (\epsilon \lambda_1^{(\ast)} - 1) + [b(\epsilon)]^d \right),
\]

(63)

where \( C > 0 \) does not depend on \( \epsilon \in \mathbb{E}^+ \) or \( n \geq N \).

The proof of (63) is tedious and given in Section 11.4. The argument involves bounding the sum of \( h_n(\cdot) \) over two separate sets in \( \mathcal{E}_n \): those integers in \( \mathcal{E}_n \) that are either "too large" or "too small" to be included in \( \tilde{R}_{\epsilon,n} \).

To finish the proof, our approach will be to write (for an arbitrary \( \epsilon \in \mathbb{E}^+ \)) \( (\epsilon N_n)^{-3} \sum_{k \in \tilde{R}_{\epsilon,n}} h_n(k) \) as an integral of a step function with respect to the Lebesgue measure of a step function, say \( f_{\epsilon,n}(x) \), then show \( \lim_{n \to \infty} f_{\epsilon,n}(x) \) exists a.e. on \( \mathbb{R}^d \), and apply the Lebesgue Dominating Convergence Theorem (LDCT). By letting \( \epsilon \downarrow 0 \), we will obtain the limit of \( \epsilon N_n \sum_{k \in \mathcal{E}_n} \text{Cov}(Y_{0,n}^2, Y_{\epsilon,n}^2) \).

Pick and fix \( \epsilon \in \mathbb{E}^+ \). For \( k \in \tilde{R}_{\epsilon,n} \),

\[
|R_{k,n}^{(0)} \cap \mathbb{Z}^d| \geq |R_{k,n}^{(0)}| - 2 \cdot |\{i \in \mathbb{Z}^d : T^i \cap s \Delta_n T_0 \neq \emptyset, T^i \cap s \Delta_n T_0 \neq \emptyset; T^i = i + (-1/2, 1/2)^d| > \epsilon (\lambda_1^{(\ast)})^{d-1} - C \cdot 2(\lambda_1^{(\ast)})^{d-1},
\]

by counting arguments based on the boundary condition of \( R_0 \) and the definition of \( \tilde{R}_{\epsilon,n} \). Likewise,

\[
\epsilon N_n - |R_{k,n}^{(0)} \cap \mathbb{Z}^d| \geq |s \mathcal{R}_n \setminus R_{k,n}^{(0)}| - C \cdot 2(\lambda_1^{(\ast)})^{d-1} > \epsilon (\lambda_1^{(\ast)})^{d} - C \cdot 2(\lambda_1^{(\ast)})^{d-1}.
\]

For some \( N_\epsilon \in \mathbb{Z}_+ \), it holds that for all \( k \in \tilde{R}_{\epsilon,n}, n \geq N_\epsilon \):

\[
|R_{k,n}^{(0)} \cap \mathbb{Z}^d| \geq 1, \quad \epsilon N_n - |R_{k,n}^{(0)} \cap \mathbb{Z}^d| \geq 1.
\]

Assume \( n \geq N_\epsilon \), then we can rewrite \( (\epsilon N_n)^{-2} h_n(k), k \in \tilde{R}_{\epsilon,n} \), in the well-defined form:

\[
\frac{h_n(k)}{(\epsilon N_n)^2} = \mathbb{E} \left[ \frac{H_2, n_1(k)}{[s N_n - |R_{k,n}^{(0)} \cap \mathbb{Z}^d|] \cdot [s N_n - |R_{k,n}^{(0)} \cap \mathbb{Z}^d|]} \right] \mathbb{E} \left[ \frac{H_2, n_1(k)}{[s N_n - |R_{k,n}^{(0)} \cap \mathbb{Z}^d|] \cdot [s N_n - |R_{k,n}^{(0)} \cap \mathbb{Z}^d|]} \left( 1 - \frac{|R_{k,n}^{(0)} \cap \mathbb{Z}^d|}{s N_n} \right)^2 \right] + \sum_{j=1}^2 \mathbb{E} \left[ \frac{H_2, n_1(k)}{[s N_n - |R_{k,n}^{(0)} \cap \mathbb{Z}^d|] \cdot [s N_n - |R_{k,n}^{(0)} \cap \mathbb{Z}^d|]} \right] \mathbb{E} \left[ \frac{H_2, n_1(k)}{[s N_n - |R_{k,n}^{(0)} \cap \mathbb{Z}^d|] \cdot [s N_n - |R_{k,n}^{(0)} \cap \mathbb{Z}^d|]} \left( 1 - \frac{|R_{k,n}^{(0)} \cap \mathbb{Z}^d|}{s N_n} \right)^2 \right] - [s N_n \cdot \mathbb{E}(Y_{0,n}^2)]^2
\]

For \( x = (x_1, \ldots, x_d) \in \mathbb{R}^d \), define \( \lfloor x \rfloor = ([x_1], \ldots, [x_d]) \in \mathbb{Z}^d \), where \( \lfloor \cdot \rfloor \) is the "floor" function applied to each component of \( x \). Let \( f_{\epsilon,n}(x) : \mathbb{R}^d \rightarrow \mathbb{R} \) be the step function defined as

\[
f_{\epsilon,n}(x) = (\epsilon N_n)^{-2} f_{\{\lambda_1^{(\ast)}, x \in \tilde{R}_{\epsilon,n}\}} h_n([\lambda_1^{(\ast)} \cdot x]).
\]
We have then that (with the same fixed $\epsilon \in \mathbb{E}^+$):

$$
\frac{1}{sN_n} \sum_{k \in R_{s,n}} (sN_n)^{-1} h_n(k) = \frac{(s\lambda_{1}^{(\ast)})^d}{sN_n} \cdot \int_{\mathbb{R}^d} f_{s,n}(x) \, dx.
$$

(64)

We will now prove the following:

$$
\lim_{n \to \infty} f_{s,n}(x) = f_s(x) \equiv I_{\{x \in \bar{R}_s\}} \cdot (2\pi^d)^{-\frac{1}{2}} \cdot \left| \frac{(x + \Delta_0 R_0) \cap \Delta_0 R_0}{\det(\Delta_0) \cdot |R_0|} \right|^2 \quad a.e. \ x \in \mathbb{R}^d,
$$

(65)

with $\bar{R}_s = \{x \in \mathbb{R}^d : |x + \Delta_0 R_0| > \epsilon, |\Delta_0 R_0 \setminus (x + \Delta_0 R_0)| > \epsilon\}$, a Borel measurable set.

For $x \in \mathbb{R}^d$, write $x_n = [s\lambda_{1}^{(\ast)} \cdot x]$ to ease the notation. To establish (65), we begin by showing (a.e. $\mathbb{R}^d$ Lebesgue measure) convergence of indicator functions:

$$
I_{\{x_n \in \bar{R}_{s,n}\}} \to I_{\{x \in \bar{R}_s\}} \quad a.e. \ x \in \mathbb{R}^d.
$$

(66)

Define the set $A_n(x) = ([s\lambda_{1}^{(\ast)}]^{-1}(x_n + s\Delta_n R_0) \cap ([s\lambda_{1}^{(\ast)}]^{-1} s\Delta_n R_0)$, as a function of $x \in \mathbb{R}^d$. Note that, a.e. $y \in \mathbb{R}^d$,

$$
I_{\{y \in ([s\lambda_{1}^{(\ast)}]^{-1}(x_n + s\Delta_n R_0) \cap ([s\lambda_{1}^{(\ast)}]^{-1} s\Delta_n R_0) \cap \mathbb{R}^d\}} = I_{\{y \in \Delta_n\}},
$$

by Assumption A.1 and $|\Delta_0 \bar{R}_0^d \setminus \Delta_0 R_0^d| = 0$. Also, it holds for some $C > 0$,

$$
\int_{\mathbb{R}^d} \left( I_{\{y \in ([s\lambda_{1}^{(\ast)}]^{-1}(x_n + s\Delta_n R_0) \cap ([s\lambda_{1}^{(\ast)}]^{-1} s\Delta_n R_0) \cap \mathbb{R}^d\}} - I_{\{y \in \Delta_n\}} \right) \, d\mathbb{R}^d = |A_n(x)| \leq \det(s\Delta_n)/(s\lambda_{1}^{(\ast)})^d \leq C.
$$

We have now justified an application of the LDCT: for each $x \in \mathbb{R}^d$,

$$
|A_n(x)| \to |(x + \Delta_0 R_0^d) \cap \Delta_0 R_0^d| = |(x + \Delta_0 R_0) \cap \Delta_0 R_0|.
$$

Let $\hat{A}_n(x) = ([s\lambda_{1}^{(\ast)}]^{-1} \Delta_n R_0 \setminus A_n(x)$ so that (using the above): $|\hat{A}_n(x)| \to |\Delta_0 R_0 \setminus (x + \Delta_0 R_0)|$. Thus, if $x \in \bar{R}_s$,

$$
|A_n(x)| \to |(x + \Delta_0 R_0) \cap \Delta_0 R_0| > \epsilon; \quad |\hat{A}_n(x)| \to |\Delta_0 R_0 \setminus (x + \Delta_0 R_0)| > \epsilon.
$$

(67)

The above result implies further that for some $N_{\epsilon,x} \in \mathbb{Z}_+$ dependent on $x \in \bar{R}_s, \epsilon \in \mathbb{E}^+$:

$$
n \geq N_{\epsilon,x} \implies |A_n(x)| > \epsilon, \quad |\hat{A}_n(x)| > \epsilon
$$

so that $x_n \in \bar{R}_{s,n}$ by definition and it holds eventually that

$$
1 = I_{\{[s\lambda_{1}^{(\ast)} \cdot x] \in \bar{R}_{s,n}\}} \to I_{\{x \in \bar{R}_s\}} = 1.
$$
Now consider values in $\mathbb{R}^t$. If $x \in \mathbb{R}^t$ such that $|(x + \Delta_0 R_0) \cap \Delta_0 R_0| < \epsilon$ (or $\Delta_0 R_0 \setminus (x + \Delta_0 R_0)| < \epsilon$), then $|A_n(x)| < \epsilon$ (or $|\hat{A}_n(x)| < \epsilon$) eventually for large $n$ and so

$$0 = I_{\{x \in \mathbb{R}^t \mid |A_n(x)| < \epsilon \}} \rightarrow I_{\{x \in \mathbb{R}^t \}} = 0$$

in this case. Finally, $\epsilon \in E^+$ implies that a last possible subset of $\mathbb{R}^t$ has Lebesgue measure zero; namely,

$$\left\{ x \notin \mathbb{R}^t : |(x + \Delta_0 R_0) \cap \Delta_0 R_0| = \epsilon \text{ or } |\Delta_0 R_0 \setminus (x + \Delta_0 R_0)| = \epsilon \right\} = 0.$$

We have now proven (66).

We next establish a limit for $(sN_n)^{-2}h_n(x_n), x \in \mathbb{R}_t$. We wish to show:

$$\frac{|R_{x_n,n}^{(1)} \cap \mathbb{Z}^d|}{sN_n} \rightarrow \frac{|(x + \Delta_0 R_0) \cap \Delta_0 R_0|}{\det(\Delta_0)|R_0|}, \ x \in \mathbb{R}_t. \quad (68)$$

Using the bound $|R_{x_n,n}^{(1)}| - |R_{x_n,n}^{(1)} \cap \mathbb{Z}^d| \leq C \cdot (s\lambda_1^{(1)})^{-d}$ using the $R_0$-boundary condition as in the proof of Lemma 8.3 and noting the limit in (67), we find

$$(s\lambda_1^{(1)})^{-d}|R_{x_n,n}^{(1)} \cap \mathbb{Z}^d| \rightarrow |(x + \Delta_0 R_0) \cap \Delta_0 R_0|, \ x \in \mathbb{R}_t.$$

By this and $(s\lambda_1^{(1)})/sN_n \rightarrow (\det(\Delta_0)|R_0|)^{-1}$, (68) follows.

We would now like to show: for each $x \in \mathbb{R}_t$, $j = 1$ or 2,

$$\mathbb{E}\left[ \frac{H_j(x_n)}{sN_n - |R_{x_n,n}^{(1)} \cap \mathbb{Z}^d|} \right] \rightarrow \mathbb{E}[(\mathbb{V}^T \mathbb{Z}_{\infty})^2], \ sN_n \mathbb{E}(Y_{\Delta_0,n}^2) \rightarrow \mathbb{E}[(\mathbb{V}^T \mathbb{Z}_{\infty})^2], \quad (69)$$

where $\mathbb{V}^T \mathbb{Z}_{\infty}$ is a normal $\mathcal{N}(0, r^2)$ random variable and so it follows $\mathbb{E}[(\mathbb{V}^T \mathbb{Z}_{\infty})^{2j}] = (2j - 1)r^{2j}, j = 1, 2$. The limits in (69) will follow essentially from the Central Limit Theorem (CLT) of Bolthausen (1982).

We first verify that the CLT can be applied, starting with checking a required boundary condition. Let

$$T \in \left\{ R_{x_n,n}^{(1)}, (x_n + sR_0) \setminus R_{x_n,n}^{(1)}, sR_0, sR_0 \setminus R_{x_n,n}^{(1)} \right\}$$

and define $\partial[T \cap \mathbb{Z}^d] \equiv \{ i \in T \cap \mathbb{Z}^d : \text{there exists some } m \in T^c \cap \mathbb{Z}^d, \text{ dist}(i, m) = 1 \}$, so that the cardinality $|\partial[T \cap \mathbb{Z}^d]|$ counts the number of $\mathbb{Z}^d$ lattice points near the boundary of $T$. Then,

$$|\partial[T \cap \mathbb{Z}^d]| \leq 2|B_0^0|, \quad |B_0^0|/|T \cap \mathbb{Z}^d| = o(1),$$

follows from (60), (67), and Lemma 8.3. Hence, the boundary requirement of the Bolthausen (1982) CLT is satisfied. For brevity, let $\{U_n\}, n \geq N_\epsilon$, denote any of the following collections of random
variables

\[ \left\{ \sqrt{n} \mathbf{N} \cdot \mathbf{Y}_n \right\}, \quad \left\{ \frac{H_{3,n}(\mathbf{x}_n)}{|R_{d,n} \cap \mathbb{Z}^d|^{1/2}} \right\}, \]

or

\[ \left\{ \frac{H_{3,n}(\mathbf{x}_n)}{(N_n - |R_{d,n} \cap \mathbb{Z}^d|)^{1/2}} \right\} \]

for any given \( \mathbf{x} \in \tilde{R}_e, j = 1, 2 \). The condition \( n \geq N_e \), for the fixed \( \varepsilon \in \mathbb{E}^+ \), again ensures that all lattice counts in the denominators above (for any \( \mathbf{x} \in \tilde{R}_e \)) are positive. Under Assumptions A.3-A.4 and Condition \( M_r \), we may now apply the CLT:

\[ U_n \xrightarrow{d} \mathcal{N}(0, \tau^2). \]

Also, note that the collections \( \{U_n\}, \{U_n^2\} \) are uniformly integrable from the moment bound

\[ \mathbb{E}(U_n^2) \leq C, \]

by Condition \( M_r \) and Lemma 8.2, where \( C \) does not depend on \( \mathbf{x} \in \tilde{R}_e, n \geq N_e \). Uniform integrability and the distributional limit from the CLT ensure (69).

Putting (66), (68), and (69) together, we have now shown the (a.e.) convergence of the univariate functions \( f_{\varepsilon,n}(\mathbf{x}) \) as in (65). For all \( k \in \mathbb{E}_n \) and \( n \geq N_e \), Lemma 8.2 ensures: \( (sN_n)^{-2}|h_n(k)| \leq C \), implying that for all \( \mathbf{x} \in \mathbb{R}^d \),

\[ |f_{\varepsilon,n}(\mathbf{x})| \leq C \cdot I_{\{x \in [-c,c]^d \}} \leq C \cdot I_{\{x \in [-c,d]^d \}}, \quad \varepsilon > 0, \]

by Assumption A.1. With this uniform bound on \( f_{\varepsilon,n}(\cdot) \) and the limits in (65), we can apply the LDCT to get

\[ \lim_{m \to \infty} \int_{\mathbb{R}^d} f_{\varepsilon,n}(\mathbf{x})d\mathbf{x} = \int_{\mathbb{R}^d} f_{\varepsilon}(\mathbf{x})d\mathbf{x}, \quad \varepsilon \in \mathbb{E}^+. \quad (70) \]

Let \( \{\varepsilon_m\}_{m=1}^\infty \subset \mathbb{E}^+ \) such that \( \varepsilon_m \downarrow 0 \). Define the set

\[ \tilde{R}_0 = \left\{ \mathbf{x} \in \mathbb{R}^d : 0 < |(\mathbf{x} + \Delta_0 R_0) \cap \Delta_0 R_0| < \det(\Delta_0)|R_0| \right\}. \]

Then, \( \tilde{R}_\varepsilon \subset \Delta_0[-1,1]^d \) and Lemma 11.1(c) implies, for all \( \mathbf{x} \neq 0 \in \mathbb{R}^d \),

\[ \lim_{m \to \infty} I_{\{\mathbf{x} \in \tilde{R}_\varepsilon\}} \rightarrow I_{\{\mathbf{x} \in \tilde{R}_0\}} \]

so that by the LDCT again

\[ \lim_{m \to \infty} \int_{\mathbb{R}^d} f_{\varepsilon,m}(\mathbf{x})d\mathbf{x} = \int_{\mathbb{R}^d} f_0(\mathbf{x})d\mathbf{x}, \quad f_0(\mathbf{x}) \equiv I_{\{\mathbf{x} \in \tilde{R}_0\}} \cdot |2\tau^4| \cdot \left( \frac{|(\mathbf{x} + \Delta_0 R_0) \cap \Delta_0 R_0|}{\det(\Delta_0)|R_0|} \right)^2. \quad (71) \]
We sequentially use (62), (63), (64), (70), (71) and that \((\lambda^{(n)})^d/sN_n \to (\det(\Delta_0) |R_0|)^{-1}\) to show:

\[
\lim_{n \to \infty} \left| sN_n \sum_{k \in \mathbb{Z}^d} \text{Cov}(Y_{0,n,k}^2, Y_{1,n,k}^2) - \frac{1}{\det(\Delta_0) |R_0|} \int_{\mathbb{R}^d} f_0(x) dx \right| \\
\leq \lim_{n \to \infty} \left| sN_n \sum_{k \in \mathbb{Z}^d} \text{Cov}(Y_{0,n,k}^2, Y_{1,n,k}^2) - \frac{(\lambda^{(n)})^d}{sN_n} \int_{\mathbb{R}^d} f_{s,n}(x) dx \right| \\
+ \lim_{n \to \infty} \left| \frac{(\lambda^{(n)})^d}{sN_n} \int_{\mathbb{R}^d} f_{s,n}(x) dx - \frac{1}{\det(\Delta_0) |R_0|} \int_{\mathbb{R}^d} f_0(x) dx \right| \\
+ \frac{1}{\det(\Delta_0) |R_0|} \int_{\mathbb{R}^d} f_{s,n}(x) dx - \int_{\mathbb{R}^d} f_0(x) dx \leq C \cdot (\epsilon_m + |b(\epsilon_m)|^d) + \frac{1}{\det(\Delta_0) |R_0|} \int_{\mathbb{R}^d} f_{s,n}(x) dx - \int_{\mathbb{R}^d} f_0(x) dx \to 0 \text{ as } \epsilon_m \downarrow 0.
\]

Finally,

\[
\frac{1}{\det(\Delta_0) |R_0|} \int_{\mathbb{R}^d} f_0(x) dx = \frac{2r^d}{\det(\Delta_0) |R_0|} \int_{\mathbb{R}^d} \frac{|x + \Delta_0 R_0 \cap \Delta_0 R_0|^2}{\det(\Delta_0) |R_0|} dx = 2r^d \int_{|R_0|} \frac{|(y + R_0) \cap |R_0|^2}{|R_0|^2} dy,
\]

using a change of variables \(y = \Delta_0^{-1} x\). This completes the proof of Theorem 9.1. \(\square\)

### 11.3 Proof of Proposition 11.1

This is justified through points (i.)-(iii.) below. We adopt the notation from the proof of Theorem 9.1 and continue to assume \(\mu = 0\) WLOG.

(i.) Let \(k \in \mathbb{Z}^d\). For \(j = 0\) or \(k \in \mathbb{Z}^d\), decompose the sum \(sN_n Y_{j,n} = S_{j,k,n} + S_{(1),k,n} + \tilde{S}_{j,k,n}\), where

\[
S_{j,k,n} = \Sigma(R_{j,n} \setminus \{B_{n}^{k} \cup B_{n}^{0}) \cup B_{n}^{1(k,k,n)}\}), \\
S_{(1),k,n} = \Sigma(R_{0,n}^{k} \setminus \{B_{n}^{k} \cup B_{n}^{0})\}), \\
\tilde{S}_{j,k,n} = \Sigma(R_{0,n} \setminus \{B_{n}^{k} \cup B_{n}^{0})\}).
\]

We show the following difference goes to zero, uniformly \(k \in \mathbb{Z}^d\), allowing us to "throw out" the border sums \(\tilde{S}_{0,k,n}, \tilde{S}_{k,n}\) which contribute negligibly to the expectation \(E(Y_{0,n,k}^2, Y_{1,n,k}^2)\).

\[
\left| (sN_n)^2 E(Y_{0,n,k}^2, Y_{1,n,k}^2) - (sN_n)^{-2} E[(S_{k,k,n} + S_{(1),k,n})^2 \cdot (S_{0,k,n} + S_{(1),k,n})^2] \right| \\
\leq 36(sN_n)^{-2} \max_{x \in \mathbb{Z}^d} \left\{ |E[S_{k,k,n} + S_{(1),k,n}]|^{1/4} |E[S_{0,k,n} + S_{(1),k,n}]|^{1/4} \right\} \\
\leq C \cdot (sN_n)^{-1/4} \left| \ell \cdot (\lambda^{(n)})^{d-1} \right|^{1/4} = o(1),
\]

by first applying Holder's inequality and then using Lemma 8.1 and (60). Lemma 8.3 implies the final inequality with \(\ell^{-d} \det(\Delta_n) \to 0\).
(ii.) After cutting out the border observations, the remaining sums of interest in evaluating $E(Y_{0,n}^2 Y_{k,n}^2)$ are "essentially" independent. We establish the claim:

$$
\begin{align*}
&\frac{1}{(s N_n)^{-2}} \cdot \max_{k \in \mathbb{Z}^d} \left| E \left[ (S_{k,k,n} + S_{0,k,n} + S_{0,k,n})^2 \right] \right| \\
&= \left[ E(S_{0,k,n}^2) E(S_{0,k,n}^2) + E(S_{k,k,n}^2) E(S_{0,k,n}^2) + E(S_{0,k,n}^2) E(S_{1,k,n}^2) + E(S_{0,k,n}^2) E(S_{1,k,n}^2) \right] = o(1).
\end{align*}
$$

We begin with the portion:

$$
\begin{align*}
&\max_{j \in \mathbb{Z}^d} \frac{(s N_n)^{-2}}{N_n} \left| \text{Cov}(S_{j,k,n}^2, S_{0,k,n}^2) \right| \\
&\leq \max_{j \in \mathbb{Z}^d} \left\{ I\left( Z^d \cap R_{j,n} \setminus (R_{k,n}^0 \cup B_n^k \cup B_n^0) \neq \emptyset \right) \right\} \cdot \frac{(s N_n)^{-2}}{N_n} \left[ \text{Cov}(S_{j,k,n}^2, S_{0,k,n}^2) \right] \\
&\cdot 8(s N_n)^{-2} \left\{ E\left[ (S_{j,k,n}^2) E\left[ (S_{0,k,n}^2) \right] \right] \right\}^{r/(2r+\delta)} \\
&\cdot \alpha \left( \text{dis}\left[ Z^d \cap R_{j,n} \setminus (R_{k,n}^0 \cup B_n^k \cup B_n^0), Z^d \cap R_{k,n}^{(1)} \setminus (B_n^k \cup B_n^0) \right] \right)^{\delta/(2r+\delta)},
\end{align*}
$$

using Lemma 8.1. Since the "border" lattice points of $R_{k,n}, R_{0,n}$ have been removed, we have that in the $\alpha$-mixing coefficient from (73)

$$
\text{dis}\left[ Z^d \cap R_{j,n} \setminus (R_{k,n}^0 \cup B_n^k \cup B_n^0), Z^d \cap R_{k,n}^{(1)} \setminus (B_n^k \cup B_n^0) \right] \geq \ell, \quad j = 0, k \in \mathbb{Z}^d,
$$

assuming the sets involved in the distance measure are nonempty. To see why, WLOG suppose $j = 0$ and suppose $p \in Z^d \cap R_{0,n} \setminus (R_{k,n}^0 \cup B_n^k \cup B_n^0)$, and $q \in Z^d \cap R_{k,n}^{(1)} \setminus (B_n^k \cup B_n^0)$ such that $\text{dis}(p, q) < \ell$. Then, $p \in q + \ell(-1, 1)^d \subset R_{k,n}$ by construction, implying $p \in R_{k,n}^{(1)}$, a contradiction.

Applying Lemma 8.2 to (73),

$$
\begin{align*}
&\max_{j \in \mathbb{Z}^d} \frac{(s N_n)^{-2}}{N_n} \left| \text{Cov}(S_{j,k,n}^2, S_{0,k,n}^2) \right| \\
&\leq C \cdot \alpha \left( \left\lfloor \ell \right\rfloor \right) \left| (s N_n) \right|^{\delta/(2r+\delta)} \\
&\leq C \cdot \left( \alpha \left| (\ell) \right| \right) \left( s N_n \right)^{\delta/(2r+\delta)} \\
&\leq C \cdot \left( \left\lfloor \text{det}(s \Delta_n) \right\rfloor \right)^{\delta/(2r+\delta)} \\
&= o(1),
\end{align*}
$$

with inequalities following from Assumption A.3, Condition $M_r$, the growth rate of $s N_n$ from Lemma 8.1, and $(\kappa - \delta)/(2r + \delta) < e(2r - 1 - 1/d)$. The constant $C$ above does not depend on $k \in E_n$.

By an argument similar to (74), it can be shown that

$$
\begin{align*}
&\max_{k \in E_n} \frac{(s N_n)^{-2}}{N_n} \left| E(S_{k,k,n}^2 S_{0,k,n}^2) - E(S_{k,k,n}^2) \cdot E(S_{0,k,n}^2) \right| = o(1).
\end{align*}
$$
In light of (74) and (75), to finish the proof of (72), it is sufficient to show that for any real-integers \( u, u_i, x \) such that \( 0 < u, u_i, x < 2 \), \( l < x < 3 \), and \( u + u_i - x - 4 \) (with the exception that \( u = u_i = 1 \) if \( x = 2 \)):

\[
\max_{k \in E_n} (sN_n)^{-2} \left| \mathbb{E}(S_{k,k,n}^u S_{0,k,n}^u S_{i(1),k,n}^r) \right| = o(1).
\]

These can be deduced from covariance bounds analogous to those in (73) and (74), along with using the fact that either \( \mathbb{E}(S_{k,k,n}^u) = 0 \) or \( \mathbb{E}(S_{0,k,n}^u) = 0 \) in any case considered. We handle one case for illustration.

For \((u, w, x) = (1, 2, 1),\):

\[
\max_{k \in E_n} (sN_n)^{-2} \left| \mathbb{E}(S_{k,k,n}^u S_{0,k,n}^u S_{i(1),k,n}) \right| = \max_{k \in E_n} (sN_n)^{-2} \left| \text{Cov}(S_{k,k,n}^u, S_{0,k,n}^u S_{i(1),k,n}) \right| \leq \frac{8}{(sN_n)^{1/2}} \cdot \left\{ \mathbb{E}\left[ |S_{k,k,n}^u|^{(2r+d)/r} \right] \cdot \max \left\{ \mathbb{E}\left[ (S_{0,k,n}^u)^{(4r+4d)/r} \right], \mathbb{E}\left[ (S_{i(1),k,n})^{(4r+24d)/r} \right] \right\} \right\}^{1/2} \cdot \frac{r/(2r+d)}{\alpha \left( \left\{ \det(s\Delta_n) \right\}^{2(2r+d)/r} \right)}
\]

\[
\leq C \cdot \left\{ \det(s\Delta_n) \right\}^{r/(2r+d)} = o(1),
\]

by first applying the covariance bound from Lemma 8.1 and Holder's inequality and then using Lemma 8.2, Assumption A.3, and Condition M.; we used the distance

\[
\text{dis}\{Z^d \cap R_{k,n} \setminus (R_{k,n}^U \cup B_n^k \cup B_n^0), Z^d \cap R_{0,n} \setminus (B_n^0 \cup B_n^0) \} \geq \ell
\]

above as well, assuming these sets are nonempty. Thus, (72) is established.

(iii.) Because the number of lattice points \( Z^d \) in the sets \( B_n^0, B_n^k \) (for \( k \in \mathbb{Z}^d \)) is negligible compared to \( sN_n \), we can add these points back into the sums \( S_{0,k,n}, S_{k,k,n}, S_{i(1),k,n} \) defined in (ii.) (for convenience) to get new sums

\[
\overline{S}(R_{0,n} \setminus R_{k,n}^U), \overline{S}(R_{k,n} \setminus R_{k,n}^U), \overline{S}(R_{k,n}^U),
\]

respectively. We aim now to show:

\[
(sN_n)^{-2} \cdot \max_{k \in E_n} \left| \mathbb{E}(S_{k,k,n}^u) \cdot \mathbb{E}(S_{0,k,n}^u) \cdot \mathbb{E}(S_{i(1),k,n}^r) + \mathbb{E}(S_{k,k,n}^u) \cdot \mathbb{E}(S_{0,k,n}^u) \cdot \mathbb{E}(S_{i(1),k,n}^r) + \mathbb{E}(S_{0,k,n}^u) \cdot \mathbb{E}(S_{i(1),k,n}^r) + \mathbb{E}(S_{i,1,k,n}^r) - \left( h_n(k) + (sN_n)^{-2} \mathbb{E}(Y_{0,n}^2) \right) \right| = o(1).
\]

Using Holder's inequality, Lemma 8.2, and (60), we can first show the following difference in sums
is of sufficiently small order.

\[
\max_{j \in \mathbb{N}_0} (sN_n)^{-1} \left| E(S_{j,k,n}^2) - E[\Sigma^2(R_{j,n} \setminus R_{j,n}^{(n)})] \right|
\leq \max_{j \in \mathbb{N}_0} (sN_n)^{-1} \left\{ \max_{0 < v < 1} \left( E[\Sigma^2(R_{j,n} \setminus R_{j,n}^{(n)})] \right)^{v/2} \left( E[\Sigma^2(R_{j,n} \setminus R_{j,n}^{(n)})] \right)^{1/2-v} \right\}
\leq C \cdot \left( \frac{\ell \cdot (\lambda_n^{max})^{d-1}}{sN_n} \right)^{1/2} = o(1).
\]  

(77)

Likewise, for \( v = 1 \) or 2,

\[
\max_{k \in \mathbb{N}_0} (sN_n)^{-v} \left| E(S_{\iota k,n, n}^{2v}) - E[\Sigma^2(R_{\iota k,n}^{(n)})] \right| \leq C \cdot \left( \frac{\ell \cdot (\lambda_n^{max})^{d-1}}{sN_n} \right)^{1/2} = o(1).
\]  

(78)

Lastly, we observe that, by Lemma 8.2,

\[
\max_{j \in \mathbb{N}_0} (sN_n)^{-1} \left\{ E(S_{\iota k,n}^4), E(S_{\iota k,n}^2), E[\Sigma^2(R_{\iota k,n}^{(n)})], E[\Sigma^2(R_{\iota k,n}^{(n)})] \right\} \leq C,
\]  

(79)

Using moment bounds directly above and the difference bounds in (77) and (78), we obtain (76).

Then, (61) follows easily from (i.)-(iii.) above; the proof of Proposition 11.1 is now finished.

11.4 Proof of Proposition 11.2

We retain the same notation from the proof of Theorem 9.1 and WLOG assume \( \mu = 0 \). For clarity of exposition, we show Proposition 11.2 by dividing the proof into two parts. In part (I.), we formulate a way to handle integers in \( E_n \) that are "too large" to be in \( \hat{R}_{\epsilon,n} \); in part (II.), the integers in \( E_n \) which are "too small" to be included in \( \hat{R}_{\epsilon,n} \). After dealing with each, we put the pieces together and complete the proof of the claim in (63).

(I.) We show here that:

\[
(sN_n)^{-3} \sum_{k \in \mathbb{N}_0} I_{\{ |k| \leq (s\lambda_n^{max})^d \}} |h_n(k)| \leq C \cdot (s\lambda_n^{max})^{-1}.
\]  

(80)

To this end, it is sufficient to prove

\[
\max_{k \in \mathbb{N}_0} I_{\{ |k| \leq (s\lambda_n^{max})^d \}} \cdot (sN_n)^{-2} \cdot |h_n(k)| \leq C \cdot (s\lambda_n^{max})^{-1},
\]  

(81)

in which case (80) follows from \(|E_n| = O(\det \{s\Delta_n\})\).
Pick and fix \( \epsilon \in \mathbb{E}^+ \). We first must bound the number of \( \mathbb{Z}^d \) lattice points inside the intersection of the subsamples \( \frac{R_n}{\Delta_n} \) and \( R_{\Delta_n} \). Namely, if \( k \in \mathbb{E}_n \) and \( |R_{\Delta_n}(k)| \leq \epsilon (\lambda_1^{(n)})^d \), then

\[
|R_{\Delta_n}(k) \cap \mathbb{Z}^d| \leq |R_{\Delta_n}(k)| + \sum_{m=0, k \in \mathbb{Z}^d} |\{i \in \mathbb{Z}^d : T_i \cap (m + \Delta_n R_0) \neq \emptyset, T_i \cap (m + \Delta_n R_0') \neq \emptyset; T_i = 1 + [-1/2, 1/2]^d\}|
\]

\[
\leq \epsilon (\lambda_1^{(n)})^d + C \cdot (\lambda_n^{(n)})^{d-1}
\]

with the inequalities following the \( R_0 \) boundary condition (see the proof of Lemma 8.3) and Assumption A.1. This bound on \( |R_{\Delta_n}(k)| \), \( k \in \mathbb{E}_n \), is important for moment inequalities (under Lemma 8.2) for sums of the r.f. over the intersections \( R_{\Delta_n}(k) \). In particular, we use this bound to produce:

\[
\max_{k \in \mathbb{E}_n} \left\{ I(|R_{\Delta_n}(k)| \leq \epsilon (\lambda_1^{(n)})^d) \cdot \max_{j=1, 2} \left\{ (sN_n)^{-j} E[H_{\Delta_n}^j(k)] \right\} \right\} \leq C \cdot |\{ \epsilon + (\lambda_1^{(n)})^{-1} \}|
\]

\[
\max_{k \in \mathbb{E}_n} \left\{ I(|R_{\Delta_n}(k)| \leq \epsilon (\lambda_1^{(n)})^d) \cdot (sN_n)^{-1} \cdot \max \left\{ E[H_{\Delta_n}^2(k)], E[H_{\Delta_n}^3(k)] \right\} \right\} \leq C,
\]

(82)

with applications of Lemma 8.1, Lemma 8.3, and Assumption A.1 Noting that \( (sN_n)^2 E[Y_0^2] = E[(H_{\Delta_n}(k) + H_{\Delta_n}(k))^2] \) by stationarity, we obtain that

\[
\max_{k \in \mathbb{E}_n} \left\{ I(|R_{\Delta_n}(k)| \leq \epsilon (\lambda_1^{(n)})^d) \cdot (sN_n)^{-2} \cdot |h_n(k)| \right\} \leq C \cdot \max_{|R_{\Delta_n}(k)| \leq \epsilon (\lambda_1^{(n)})^d} \left\{ \prod_{j=1}^{\infty} \left( \max_{k \in \mathbb{E}_n} (sN_n)^{-1} \cdot (E[H_{\Delta_n}^j(k)])^{n/2} (E[H_{\Delta_n}^3(k)]^{n/2} \right) + \epsilon + (\lambda_1^{(n)})^{-1} \right\}
\]

\[
\leq C \cdot (\epsilon + (\lambda_1^{(n)})^{-1})
\]

following by sequentially applying from Holder's inequality, Lemma 8.2, and the moment bounds in (82); the constant \( C > 0 \) does not depend on \( \epsilon \) or \( n \). This establishes (81) and hence (80). The step (I.) is complete. Note as well that we relied only on moment bounds to show (81) and we did not need to account for the number of integers \( k \in \mathbb{E}_n \) satisfying \( |R_{\Delta_n}(k)| \leq \epsilon (\lambda_1^{(n)})^d \). The approach taken in part (II.) will differ.

(II.) We now consider \( k \in \mathbb{E}_n \) such that \( |R_{\Delta_n}(k)| \leq |\Delta_n R_0 \setminus (k + \Delta_n R_0)| \leq \epsilon (\lambda_1^{(n)})^d \). We will work to construct the function \( b(\cdot) \) from the claim in (63) and build a corresponding set, say \( D_{\epsilon, n} \), such that:

\[
\{k \in \mathbb{E}_n : |R_{\Delta_n}(k)| \leq \epsilon (\lambda_1^{(n)})^d \} \subset \mathbb{E}_n \cap D_{\epsilon, n}, \quad |D_{\epsilon, n}|/sN_n \leq C \cdot (b(\epsilon))^d + (\epsilon (\lambda_1^{(n)})^{-1})
\]

(83)
where $C$ is independent of $n \geq N$ (some $N \in \mathbb{Z}_+$) and $\epsilon \in E^+$; this is the goal of (IL).

We begin with defining the function $b(\cdot) : E^+ \rightarrow \mathbb{R}^+$. For $x = (x_1, \ldots, x_d)' \in \mathbb{R}^d$, let $\pi_j(x) = x_j$ be the projection of the $j$th coordinate of $x$. For each coordinate $j = 1, \ldots, d$, let $m_j^+$ and $m_j^-$ be maximums of the continuous function $|\pi_j(\cdot)|$ on the compact sets $\{x \in \mathbb{R}^d : x_j \geq 0\} \cap \Delta_0 R_0^d$ and $\{x \in \mathbb{R}^d : x_j \leq 0\} \cap \Delta_0 R_0^d$, respectively; $m_j^+, m_j^- \in \mathbb{R}^d$ are, of course, functions of each coordinate component $j = 1, \ldots, d$.

For each $j = 1, \ldots, d$ and fixed $\epsilon \in E^+$, define the sets
\[
D_{\epsilon j}^+ = \left\{ \omega \geq 0 : \left| \left\{ x \in \mathbb{R}^d : x_j \in [\pi_j(m_j^+) - \omega, \pi_j(m_j^+)] \right\} \cap \Delta_0 R_0^d \right| > 2\epsilon \right\},
\]
\[
D_{\epsilon j}^- = \left\{ \omega \geq 0 : \left| \left\{ x \in \mathbb{R}^d : x_j \in [\pi_j(m_j^-), \pi_j(m_j^-) + \omega] \right\} \cap \Delta_0 R_0^d \right| > 2\epsilon \right\},
\]
where above $[\pi_j(m_j^+) - \omega, \pi_j(m_j^+)]$ and $[\pi_j(m_j^-), \pi_j(m_j^-) + \omega]$ are closed intervals in $\mathbb{R}$.

Define the corresponding set infimums, for $j = 1, \ldots, d$
\[
\Omega_{\epsilon j}^+ = \inf D_{\epsilon j}^+, \quad \Omega_{\epsilon j}^- = \inf D_{\epsilon j}^-.
\]

We note two important properties of $\Omega_{\epsilon j}^+$ and $\Omega_{\epsilon j}^-$, for each $j = 1, \ldots, d$:

1. For all $\epsilon \in E^+$, $\Omega_{\epsilon j}^+ > 0$ and $\Omega_{\epsilon j}^- > 0$.

2. Treated as functions on $E^+$, both $\Omega_{\epsilon j}^+$ and $\Omega_{\epsilon j}^-$ are decreasing in $\epsilon$.

Furthermore, if $\{\epsilon_m\}_{m=1}^\infty \subset E^+$, $\epsilon_m \downarrow 0$, then $\Omega_{\epsilon_m j}^+ \downarrow 0$ and $\Omega_{\epsilon_m j}^- \downarrow 0$.

We briefly justify each of these statements:

1.): WLOG focus on $\Omega_{\epsilon j}^+$ and suppose $\Omega_{\epsilon j}^+ = 0$. Then there exists a sequence $\{\omega_m\}_{m=1}^\infty \subset D_{\epsilon j}^+$, $\omega_m \downarrow 0$, so that
\[
2\epsilon \leq \lim_{m \to \infty} \left| \left\{ x \in \mathbb{R}^d : x_j \in [\pi_j(m_j^+) - \omega_m, \pi_j(m_j^+)] \right\} \cap \Delta_0 R_0^d \right| \leq \left| \left\{ x \in \mathbb{R}^d : x_j = \pi_j(m_j^+) \right\} \cap \Delta_0 R_0^d \right| = 0 \quad (\text{zero } \mathbb{R}^d \text{ Lebesgue measure}),
\]
a contradiction. This convergence in measure follows from $|\Delta_0 R_0^d| < \infty$ and the fact that the sets are nested and decreasing (cf. Royden, 1988). Hence, it must be that $\Omega_{\epsilon j}^+, \Omega_{\epsilon j}^- > 0$ for each $j = 1, \ldots, d$.

2.): As functions of $\epsilon$, both $\Omega_{\epsilon j}^+$ and $\Omega_{\epsilon j}^-$ are clearly decreasing. Denote an open ball of radius $\omega > 0$ around $x \in \mathbb{R}^d$ as $B(x, \omega)$. Since $m_j^+, m_j^- \in \Delta_0 R_0^d$, the Lebesgue measures $|B(m_j^+, \omega) \cap \Delta_0 R_0^d| > 0$ and
$|B(m_j^+, \omega) \cap \Delta_0 R_0^3| > 0$ for all $\omega > 0$, $j = 1, \ldots, d$; this follows because both intersections are nonempty and open ($m_j^+$, $m_j^-$ are closure points). We have then for any $\omega > 0$ and $j = 1, \ldots, d$:

$$\left| \left\{ x \in \mathbb{R}^d : x_j \in [\pi_j(m_j^+) - \omega, \pi_j(m_j^-)] \right\} \cap \Delta_0 R_0^3 \right| \geq |\Delta_0 R_0^3 \cap B(m_j^+, \omega)| > 0,$$

$$\left| \left\{ x \in \mathbb{R}^d : x_j \in [\pi_j(m_j^-), \pi_j(m_j^-) + \omega] \right\} \cap \Delta_0 R_0^3 \right| \geq |\Delta_0 R_0^3 \cap B(m_j^-, \omega)| > 0,$$

using that $x \in \Delta_0 R_0^3 \cap B(m_j^+, \omega)$ implies $\pi_j(m_j^+) - \omega < x_j \leq \pi_j(m_j^-)$. To now establish the second property claimed for $\Omega_{+,j}$, suppose there is a sequence $e_m \in E^+, e_m \downarrow 0$, and $\Omega_{+,m,j} \downarrow C > 0$. Then, it follows that eventually, $2e_m < |\Delta_0 R_0^3 \cap B(m_j^+, C/2)|$ so that $C/2 \in D_{+,m,j}$, implying $\Omega_{+,m,j} \leq C/2$, and then the contradiction: $\lim_{m \to \infty} \Omega_{+,m,j} = C \leq C/2$. Hence, we must have $e_m \downarrow 0$ implies $\Omega_{+,m,j} \downarrow 0$. The same argument holds for $\Omega_{-,j}$, justifying statement (2.) above.

Continuing with the construction of $b(\cdot)$, we now define the function $b(\cdot)$ on $E^+$ as

$$b(\epsilon) = \max_{1 \leq j \leq d} \left( \max\{\Omega_{+,j}, \Omega_{-,j}\} \right)$$

and let

$$D_{\epsilon,n} = \left\{ k = (k_1, \ldots, k_d)' \in \mathbb{Z}^d : |k_j| \leq b(\epsilon) \cdot (s\lambda_j^{(\omega)})/\lambda_j, j = 1, \ldots, d\right\},$$

where $\lambda_1, \ldots, \lambda_d$ denote the positive diagonal entries of $\Delta_0$. Using the definition of the set, we bound the ratio

$$\frac{|D_{\epsilon,n}|}{s \lambda_{\epsilon,n}^d} \leq C \cdot \prod_{j=1}^d \left( \frac{2b(\epsilon)}{\lambda_j} + \frac{1}{s\lambda_j^{(\omega)}} \right) \leq C \cdot (s\lambda_j^{(\omega)})^d + (s\lambda_j^{(\omega)} - 1)^{-1},$$

by Assumption A.1 and Lemma 8.3.

Note that $\det(s\Delta_0)/(s\lambda_j^{(\omega)})^d \geq 1/2 \cdot \det(\Delta_0)$ when $n \geq N$, for some $N \in \mathbb{Z}_+$. We show, in the following, that for $n \geq N$, $\epsilon \in E^+$:

$$\kappa \in E_n, \kappa \notin D_{\epsilon,n} \implies |s \Delta_n \setminus R_{\epsilon,n}^0(\kappa)| > \epsilon (s\lambda_j^{(\omega)})^d. \quad (85)$$

If $\kappa \notin D_{\epsilon,n}$, then there exists $j \in \{1 \ldots d\}$ such that $|k_j| > b(\epsilon) \cdot (s\lambda_j^{(\omega)})/\lambda_j$. WLOG suppose $k_j < 0$ (the other case $k_j > 0$ is handled similarly). Then, $(k_j \cdot \lambda_j)/(s\lambda_j^{(\omega)}) < -b(\epsilon) \leq -\Omega_{+,j}$ by construction; and if $x \in \Delta_0(\Delta_0^{-1}k + R_0^3)$, then $x_j \leq \pi_j(m_j^+) + (k_j \cdot \lambda_j)/(s\lambda_j^{(\omega)})$. With these two observations and $n \geq N$, we have:

$$\left| (s\lambda_j^{(\omega)})^{-1} \Delta_n R_0^3 \setminus (s\lambda_j^{(\omega)})^{-1} (k + s \Delta_n R_0^3) \right| = \det((s\lambda_j^{(\omega)})^{-1} \Delta_n \Delta_0^{-1}) \cdot |\Delta_0 R_0^3 \setminus \Delta_0 \Delta_0^{-1} k + \Delta_0 R_0^3| \geq 1/2 \cdot \left| \left\{ x \in \mathbb{R}^d : x_j \in [\pi_j(m_j^+) + (k_j \cdot \lambda_j)/(s\lambda_j^{(\omega)}), m_j^+] \right\} \cap \Delta_0 R_0^3 \right| > 1/2 \cdot (2\epsilon) = \epsilon.$$
Note, as well, that \(|(\lambda_1^{(\ast)})^{-1} s \Delta_n R_0^c \setminus (\lambda_1^{(\ast)})^{-1} (k + s \Delta_n R_0^c)| > \epsilon\) if and only if
\[ |s \Delta_n R_0^c \setminus (k + s \Delta_n R_0^c)| > \epsilon (\lambda_1^{(\ast)})^d,\]
because the boundary condition on \(R_0\) implies \(|R_0^c| = |R_0| = |\overline{R}_0|\). We have now shown (85).

With (84) and (85), we have proven (83) and constructed the function \(b(\cdot)\). We proceed now complete the proof of Proposition 11.2 using the tools built in steps (I) and (II) above.

If \(k \in E_n \setminus \hat{R}_{e,n}\) (for \(e \in E^+\)), there are two disjoint possibilities: \(|\lambda_1^{(\ast)}(k)| < \epsilon (\lambda_1^{(\ast)})^d\) or \(|\lambda_1^{(\ast)}(k)| > \epsilon (\lambda_1^{(\ast)})^d\) and \(k \in D_{e,n}\), which follows from (85). We finally arrive at: for \(n \geq N\),
\[
(sN_n)^{-3} \sum_{k \in E_n \setminus \hat{R}_{e,n}} h_n(k) \leq (sN_n)^{-3} \sum_{k \in E_n} \left( I_{\{|\lambda_1^{(\ast)}(k)| \leq \epsilon (\lambda_1^{(\ast)})^d\}} + I_{\{k \in D_{e,n}\}} \right) |h_n(k)| \\
\leq C : (\epsilon + (\lambda_1^{(\ast)})^{-1} + |D_{e,n}|/\alpha N_n) \\
\leq C : (\epsilon + (\lambda_1^{(\ast)})^{-1} + [b(e)]^d),
\]
using (80) to handle the sum over integers satisfying \(|\lambda_1^{(\ast)}(k)| \leq \epsilon (\lambda_1^{(\ast)})^d\); and for the sum on \(D_{e,n}\), using (83) with \((sN_n)^{-2}|h_n(k)| \leq C\), for all \(k \in \mathbb{Z}^d\), by Lemma 8.2 [see (79) as well]. Hence, Proposition 11.2 and (83) are now established. \(\square\)

12 Appendix 2: Proof of Theorem 4.2, \(d = 1\)

By assumption, there exists \(b_1, b_2 \in \mathbb{R}\) such that \(b_1 < 0 < b_2\) and \((b_1, b_2) \subset R_0 \subset [b_1, b_2]\). To ease the following counting arguments, define a concrete version \(sP_n = \mathbb{Z} \cap (s\lambda_n b_1 + 1/2, s\lambda_n b_2 - 1/2)\) from (24); for \(i \in J_{\|\cdot\|_{\infty}},\) let \(P_{i,n} = v_{i,n} + sP_n \subset s\lambda_n (i + R_0)\), where \(v_{i,n} \in \mathbb{Z}^d\) such that \(\|v_{i,n} - s\lambda_n i\|_{\infty} \leq 1/2\).

Then, when \(n\) is large (so that \(s\lambda_n b_1 + 1 < 0 < s\lambda_n b_2 - 1\)): \(i \in J_{\|\cdot\|_{\infty}},\) \n\[|sP_n| = |P_{i,n}| \leq sN_n(i) \leq |sP_n| + 4. \tag{86}\]

We first work to expand both \(E(\hat{t}_{n,o,l}^2)\) and \(E(\hat{t}_{n,o,l}^2)\) and find the important terms for determining the sizes of these expectations.

Consider \(\hat{t}_{n,o,l}^2\). For \(i \in J_{o,l}\), we use a fourth-order Taylor’s expansion of each subsample statistic around \(\mu\):
\[
\hat{\delta}_{i,n} = H(\mu) + \sum_{j=1}^{3} \left( \sum_{\|a\|_{L^2} = 2} c_a (Z_{i,n} - \mu)^a \right) + M_{i,n} \equiv H(\mu) + Y_{i,n} + Q_{i,n} + C_{i,n} + M_{i,n},
\]
\[
M_{i,n} = 4 \sum_{\|a\|_{L^2} = 4} c_a (Z_{i,n} - \mu)^a \int_0^1 (1 - w)^3 D^a H(\mu + \omega (Z_{i,n} - \mu)) d\omega.
\]
Write the sample means for the Taylor terms: 
\[ \hat{Y}_n = \left| J_{\text{set}} \right|^{-1} \sum_{i \in J_{\text{set}}} Y_{i,n}; \quad \hat{Q}_n = \left| J_{\text{set}} \right|^{-1} \sum_{i \in J_{\text{set}}} Q_{i,n}; \]
\[ \hat{C}_n = \left| J_{\text{set}} \right|^{-1} \sum_{i \in J_{\text{set}}} C_{i,n}; \quad \hat{M}_n = \left| J_{\text{set}} \right|^{-1} \sum_{i \in J_{\text{set}}} M_{i,n}. \]
From an algebraic expansion of \( \hat{\tau}_{n,\text{set}}^2 \) and stationarity, we find the expectation:
\[
E(\hat{\tau}_{n,\text{set}}^2) = sN_n E \left( (Y_{0,n} + Q_{0,n} + C_{0,n} + M_{0,n})^2 - (\hat{Y}_n + \hat{Q}_n + \hat{C}_n + \hat{M}_n)^2 \right). \tag{87}
\]
Although the terms are defined slightly differently, the moment inequalities in (43) and (31) still are valid (and follow from Lemma 8.2 and Condition D3) as well as, bounds:
\[
sN_n E(M_{0,n}^2) \leq C \cdot (sN_n)^{-3}; \quad sN_n E(M_{0,n}^2) \leq sN_n E(M_{0,n}^2).
\]
Then, by Holder's inequality and Assumption A.2, the terms in (87) which do not immediately have \( o(1/s^n) \) expectations can be listed as: 
\[ sN_n Y_{0,n}^2, sN_n Y_{0,n} Q_{0,n}, sN_n Q_{0,n}^2, sN_n Q_{0,n} C_{0,n}, sN_n Q_{0,n}^2. \]
Note then that, by subtracting \( sN_n \{E(Q_{0,n}^2)\}^2 \),
\[
sN_n \{E(\hat{Q}_n^2) - (E(Q_{0,n}))^2\} = sN_n \text{Var}(\hat{Q}_n) \leq \frac{C}{N_n},
\]
which can be shown with counting and moment arguments similar to those showing (22) in the proof of Theorem 3.1(a). Hence, we can express (87) as
\[
E(\hat{\tau}_{n,\text{set}}^2) = sN_n \left( Y_{0,n}^2 + Q_{0,n}^2 + 2Y_{0,n}Q_{0,n} + 2Y_{0,n}C_{0,n} \right) - sN_n \{E(Q_{0,n})\}^2 + o(1/s^n). \tag{88}
\]
Now we examine \( E(\hat{\tau}_{n,\text{set}}^2) \). We again use a Taylor's expansion to write \( H(\tilde{Z}_{i,n}) = H(\mu) + \tilde{Y}_{i,n} + \tilde{Q}_{i,n} + \tilde{C}_{i,n} + \tilde{M}_{i,n} \) and define analogous sample means over \( J_{\text{set}} \): 
\[ \tilde{Y}_n, \tilde{Q}_n, \tilde{C}_n, \tilde{M}_n. \]
Using moment inequalities in (45) and (50) valid by Lemma 8.2 and Condition D3 [though \( \tilde{Q}_{i,n}, \tilde{C}_{i,n} \) are not remainder terms in the Taylor expansion] and
\[
E(\tilde{Q}_{i,n}^2) \leq \max_{i \in J_{\text{set}}} E(\tilde{Q}_{i,n}^2); \quad E(\tilde{C}_{i,n}^2) \leq \max_{i \in J_{\text{set}}} E(\tilde{C}_{i,n}^2); \quad E(\tilde{M}_{i,n}^2) \leq \max_{i \in J_{\text{set}}} E(\tilde{M}_{i,n}^2). \tag{89}
\]
we find
\[
E(\hat{\tau}_{n,\text{set}}^2) = |J_{\text{set}}|^{-1} \sum_{i \in J_{\text{set}}} sN_{i,n} \left( E(\tilde{Y}_{i,n}^2) + 2E(\tilde{Y}_{i,n} \tilde{Q}_{i,n}) + 2E(\tilde{Y}_{i,n} \tilde{C}_{i,n}) + E(\tilde{Q}_{i,n}^2) \right)
\]
\[-2|J_{\text{set}}|^{-1} \sum_{i \in J_{\text{set}}} sN_{i,n} E(\tilde{Q}_{i,n} \tilde{Q}_{i,n}) + |J_{\text{set}}|^{-1} \left( \sum_{i \in J_{\text{set}}} sN_{i,n} \right) E(\tilde{Q}_{i,n}^2) + o(1/s^n). \]
Applying moment bounds from (50) and (89) with Holder's inequality and (86) produces
\[
-2|J_{\text{set}}|^{-1} \sum_{i \in J_{\text{set}}} sN_{i,n} E(\tilde{Q}_{i,n} \tilde{Q}_{i,n}) + |J_{\text{set}}|^{-1} \left( \sum_{i \in J_{\text{set}}} sN_{i,n} \right) E(\tilde{Q}_{i,n}^2) = -sN_n E(\tilde{Q}_{n}^2) + O\left(|sP_n|^{-2}\right). \tag{90}
\]
We next define, for each \( i \in J_{\text{NOI}} \), new random variables: \( \hat{Y}_{i,n}^*, \hat{Q}_{i,n}^*, \hat{C}_{i,n}^* \) as

\[
|P_i,n|^{-j} \sum_{\|a\| = j} \frac{c_\alpha}{\alpha!} \left( \sum_{s \in P_i,n} (Z(s) - \mu) \right)^\alpha
\]

for \( j = 1, 2, 3 \), respectively. We can expand the following quantities as the difference of two linear combinations (of products between coordinate subsample means) with some algebra, cancel common terms, and apply (87) and Holder's inequality to the expectations to find moment bounds:

\[
\begin{align*}
\max_{i \in J_{\text{NOI}}} \max_{s} \left\{ E\left( |\hat{Y}_{i,n}^* - \hat{C}_{i,n}^*| \right) \right\} &\leq C \cdot |s| P_n^{-1/2}, \\
\max_{i \in J_{\text{NOI}}} \max_{s} |E\left( |\hat{Y}_{i,n}^* - \hat{Q}_{i,n}^*| \right)| &\leq C \cdot |s| P_n^{-3/2}, \\
\max_{i \in J_{\text{NOI}}} \max_{s} |E\left( |\hat{Q}_{i,n}^* - \hat{Q}_{i,n}^2| \right)| &\leq C \cdot |s| P_n^{-3/2}.
\end{align*}
\]

(91)

With counting and moment arguments similar to those showing (37) [bounding the variance of \( \hat{A}_{4,n}^* \)], we may show

\[
\begin{align*}
\max_{i \in J_{\text{NOI}}} \left\{ E\left( |\hat{Y}_{i,n}^* - \hat{C}_{i,n}^*| \right) \right\} &\leq C \cdot |s| P_n^{-1/2}, \\
\max_{i \in J_{\text{NOI}}} \max_{s} |E\left( |\hat{Y}_{i,n}^* - \hat{Q}_{i,n}^*| \right)| &\leq C \cdot |s| P_n^{-3/2}, \\
\max_{i \in J_{\text{NOI}}} \max_{s} |E\left( |\hat{Q}_{i,n}^* - \hat{Q}_{i,n}^2| \right)| &\leq C \cdot |s| P_n^{-3/2}.
\end{align*}
\]

(91)

Then, by (92) and (93),

\[
\begin{align*}
sN_n \left( E(\hat{Q}_{i,n}^2) - \left( E(\hat{Q}_{0,n}^2) \right)^2 \right) &= \max_{i \in J_{\text{NOI}}} \left( \left| \sum_{i \in J_{\text{NOI}}} |E(\hat{Q}_{i,n})| - \left( E(\hat{Q}_{0,n}) \right)^2 \right| \right) \\
\end{align*}
\]

which with (90) and (91) further implies

\[
\begin{align*}
\max_{i \in J_{\text{NOI}}} \left[ \left( \sum_{i \in J_{\text{NOI}}} \left| E(\hat{Q}_{i,n}) \right| - \left( E(\hat{Q}_{0,n}) \right)^2 \right) \right] = o(1/\sqrt{n}).
\end{align*}
\]

(94)

We next establish that for each \( i \in J_{\text{NOI}} \),

\[
sN_i,n E(\hat{Y}_{i,n}^2) - \tau_i^2 = - \frac{1}{s_n|R_0|} \sum_{k \in \mathbb{Z}} |k| \sigma(k) + E_i,n, \quad \max_{i \in J_{\text{NOI}}} |E_i,n| = o(1/\sqrt{n}).
\]

(95)
Because \( N_{1,n}(k) = (N_{1,n} - |k|) I_{\{N_{1,n} \geq |k|\}} \) for \( i, k \in \mathbb{Z} \), we apply (40):

\[
N_{1,n} E(\hat{Y}_{1,n}^2) - \tau^2 = -\frac{1}{s N_{1,n}} \sum_{k \in \mathbb{Z}} (s N_{1,n} - s N_{1,n}(k)) \sigma(k) = -\frac{1}{s \lambda_n |R_0|} \sum_{k \in \mathbb{Z}} |k| \sigma(k) + E_{i,1,n}^r;
\]

\[
E_{i,1,n}^r = \frac{1}{s N_{1,n}} \sum_{k \in \mathbb{Z} : |k| \geq s N_{1,n}} |k| \sigma(k) + \left( \frac{1}{s \lambda_n |R_0|} - \frac{1}{s N_{1,n}} \right) \sum_{k \in \mathbb{Z}} |k| \sigma(k)
\]

where using (13), Condition \( M_r \), (87), and Lemma 8.3:

\[
\sup_{i \in J_{\text{NOL}}} |E_{i,1,n}^r| \leq C \cdot |s P_n|^{-2} \sum_{z=1}^{\infty} x \alpha_1(z)^{4/(2r+4)} = o(1/s \lambda_n).
\]

Applying Lemma 10.1(a) and Assumption A.2, we establish (20).

We then show that for \( i \in J_{\text{NOL}} \),

\[
2(s N_{1,n})^3 E(\hat{Y}_{1,n} \hat{Q}_{1,n}) = \sum_{k_1, k_2 \in \mathbb{Z}} (1 + f_{(k_1, k_2) \neq 0}) \sigma^*(k_1, k_2) + E_{i,2,n}, \quad \max_{i \in J_{\text{NOL}}} |E_{i,2,n}| = o(1/s \lambda_n)
\]

where the “lagged” expectations:

\[
\sigma^*(k_1, k_2) = 2 \cdot \mathbb{E} \left( \sum_{z \in \mathbb{Z} : |z| = 1} \frac{C \alpha_1(z)}{|z|} \left[ Z(t) - \mu \right] \sigma \left[ Z(t + k_1) - \mu \right]^\beta \left[ Z(t + k_2) - \mu \right]^\gamma \right)
\]

appear as in the statement of the constant \( C_0 \) from Theorem 4.2. For \( s \in \mathbb{Z} \), write \( Z(s) = (Z_1(s), \ldots, Z_p(s))' \in \mathbb{R}^p \) and

\[
\hat{Y}_{i,n} = \sum_{s \in \mathbb{Z} : |s| \leq R_{i,n}} (Z(s) - \mu)/s N_{1,n} = (\hat{Y}_{i,1,n}, \ldots, \hat{Y}_{i,p,n})' \in \mathbb{R}^p.
\]

WLOG assume \( \mu = 0 \). Fix \( i, j, k \in \{1, \ldots, p\} \). Then,

\[
(s N_{1,n})^3 E(\hat{Y}_{i,1,n} \hat{Y}_{j,1,n} \hat{Y}_{k,1,n}) = s N_{1,n} E(Z_i(t)Z_j(t)Z_k(t)) + \sum_{x \in \mathbb{Z} : s \leq x \leq s N_{1,n}} (s N_{1,n} - |x|) f_{i,j,k}(x) + \sum_{s \leq x \leq s N_{1,n}} (s N_{1,n} - x) \sum_{1 \leq s \leq x, 1 \leq s < x} h_{i,j,k}(x, y),
\]

where the functions \( f_{i,j,k}(\cdot) \) and \( h_{i,j,k}(\cdot, \cdot) \) are defined as:

\[
f_{i,j,k}(x) = \mathbb{E} \left( Z_i(t)Z_j(t)Z_k(t + x) + Z_i(t)Z_j(t + x)Z_k(t) + Z_i(t + x)Z_j(t)Z_k(t) \right)
\]

\[
h_{i,j,k}(x, y) = \sum_{(i', j', k') \in P(i,j,k)} \mathbb{E} \left( Z_{i'}(t)Z_{j'}(t + y)Z_{k'}(t + x) + Z_{i'}(t)Z_{j'}(t + y)Z_{k'}(t - x) \right),
\]

with \( P(i,j,k) \) denoting the set of all 6 permutations of the vector \((i,j,k)\). We then write

\[
(s N_{1,n})^3 E(\hat{Y}_{i,1,n} \hat{Y}_{j,1,n} \hat{Y}_{k,1,n}) = \mathbb{E}(Z_i(t)Z_j(t)Z_k(t)) + \sum_{x \in \mathbb{Z} \setminus \{0\}} f_{i,j,k}(x) + \sum_{s \in \mathbb{Z} : s \geq 0} h_{i,j,k}(x, y) + E_{ijk,1,2,n},
\]
with

\[
|E_{ij,k,l,n}| \leq \left(\alpha N_{l,n}\right)^{-1} \sum_{x \in \mathbb{Z} \setminus \{0\}} |x f_{i,j,k}(x)| + \sum_{s \in \mathbb{Z},\, \ell \geq 2n} |f_{i,j,k}(x)|
\]

\[
+ \left(\alpha N_{l,n}\right)^{-1} \sum_{s \in \mathbb{Z},\, \ell \leq 2n} |x h_{i,j,k}(x,y)| + \sum_{s \in \mathbb{Z},\, \ell \leq 2n} |h_{i,j,k}(x,y)|. \tag{98}
\]

We will show now that \(\max_{i \leq j 
\text{We turn our attention now to those sums of } h_{i,j,k}(x,y) \text{ in (98). With the distance metric } \text{dis}_3(\cdot) \text{ from (49), we below define a useful set (as a function on } \mathbb{Z}_+ \text{) for counting purposes and also bound its cardinality:}

\[
G(d_0) \equiv \{(x,y) \in \mathbb{Z}^2 : 0 < |y| \leq |x|, \text{dis}_3(\{0,x,y\}) = d_0\}, \quad |G(d_0)| \leq C \cdot d_0.
\]

(The bound on the set size follows easily from \(|y| \leq d_0 \) and \(\min\{|x|,|x-y|\} \leq d_0 \) in \(G(d_0)\).) Applying this set bound, Lemma 8.1, Condition \(M_r\), and using \(|x| \leq 2d_0 \) when \(\text{dis}_3(\{0,x,y\}) = d_0 \), we find

\[
\sum_{x \in \mathbb{Z}\setminus\{0\}} |x h_{i,j,k}(x,y)| \leq \sum_{x \in \mathbb{Z}\setminus\{0\}} |x f_{i,j,k}(x)| + |P_n|^{-1} \sum_{x \in \mathbb{Z}\setminus\{0\}} |x h_{i,j,k}(x,y)|.
\]

Hence, we have now established that \(\max_{i \leq j \leq l \leq n} |E_{ij,k,l,n}| = o(1) \) in (23) for any \(1 \leq i,j,k \leq p\). Note that \(2\alpha N_{l,n}^2 \text{E}(\tilde{Y}_{1,n} \tilde{Q}_{1,n})\) is a linear combination of expectations of products between coordinate sample means as in (23). It also holds that

\[
\sum_{s \in \mathbb{Z}, \, \ell \leq 2n} h_{i,j,k}(x,y) = \sum_{s \in \mathbb{Z}, \, \ell \leq 2n} \text{E}(Z_i(t)Z_j(t+k_1)Z_k(t+k_2))
\]

by stationarity. (That is, terms in each of the above sums can be placed in a one-to-one correspondence by matching \(\text{E}(Z_i(t)Z_j(t+k_1)Z_k(t+k_2))\) from the second sum above with \(\text{E}(Z_i(t - m_{k_1,k_2})Z_j(t +\)
\( k_1 - m_{k_1,k_2} Z_k(t + k_2 - m_{k_1,k_2}) \) in the first sum, where

\[
m_{k_1,k_2} = \begin{cases} 
0 & \text{if sign}(k_1) = \text{sign}(k_2), \\
 k_1 + (k_2 - k_1)I(|k_2| \geq |k_1|) & \text{otherwise}.
\end{cases}
\]

Each of the matched expectations is the same by stationarity. A little algebraic manipulation of the above sums of \( f_{i,j,k} \) and \( h_{i,j,k} \) provides the desired result in (96).

Hence, by (88), (94), (20), and (96) (and using \( Z_{0,n} = \hat{Z}_{0,n} \)), we can write

\[
E(\tilde{r}_n^2) - r_n^2 = - \frac{1}{\lambda_n |R_0|} \sum_{k \in \mathbb{Z}} |k| \sigma(k) + \sum_{k_1,k_2 \in \mathbb{Z}} \sigma^*(k_1,k_2)
\]

\[
+ 2\lambda_n N \cdot Y_{0,n} C_{0,n} + \lambda_n \left( Q_{0,n}^2 - \{ E(Q_{0,n}) \}^2 \right) + o(1/\lambda_n).
\]

where \( \tilde{r}_n^2 \) denotes either \( r_{n,DL}^2 \) or \( r_{n,NOL}^2 \).

By the boundary property of the template \( R_0 \),

\[
|\{ i \in \mathbb{Z} \cap \lambda_n R_0 : x \in \mathbb{Z} \cap \lambda_n R_0^c ; |i - x| = 1 \}| \leq 2 = o(\lambda_n),
\]

so that, under Assumptions A.3.-A.4, and Condition \( M_r \), we have that: for all \( b \in \mathbb{R}^p \)

\[
(\lambda_n)^{-1/2} \cdot b'(Z_{0,n} - \mu) \xrightarrow{d} b' Z_{\infty},
\]

a normal \( N(0,b'\Sigma Z_{\infty} b) \) random variable by the Bolthausen (1982) Central Limit Theorem. By Lemma 8.2 and Condition \( D_3 \), we have that for all \( n \geq 1 \)

\[
(\lambda_n)^3 E(|Q_{0,n}|^3), (\lambda_n)^4 E(|Y_{0,n} C_{0,n}|^4) \leq C \cdot \max \left\{ (\lambda_n)^3 E(\|Z_{0,n} - \mu\|^3), (\lambda_n)^4 E(\|Z_{0,n} - \mu\|^4) \right\} \leq C.
\]

The above moment bounds ensure the uniform integrability necessary to obtain the following normal moment limits:

\[
(\lambda_n)^3 \left( E(Q_{0,n}^2) - (E(Q_{0,n}))^2 \right) \rightarrow \text{Var} \left( \sum_{\|\alpha\|_1 = 2} c_{\alpha} Z_{\alpha}^2 \right),
\]

\[
(\lambda_n)^2 E(Y_{0,n} C_{0,n}) \rightarrow E \left[ \left( \sum_{\|\alpha\|_1 = 1} c_{\alpha} Z_{\alpha}^2 \right) \left( \sum_{\|\alpha\|_1 = 2} c_{\alpha} Z_{\alpha}^2 \right) \right].
\]

By (99) and (100) above, the proof of Theorem 4.2 for the \( d = 1 \) case is now complete! \( \square \)

References


On the approximation of differenced lattice point counts with application to statistical bias expansions

A paper to be submitted to Transaction of the American Mathematical Society

Daniel John Nordman and Soumendra N. Lahiri

Abstract

This paper formulates a lattice point counting problem, which is important to statistical theory with spatial lattice data. The goal is to approximate two subtracted lattice point counts, where the counts correspond to a set of increasing domain and an intersection of this set with a vector translate. It is well-known that, even in the plane, volumes can poorly measure lattice point counts, producing approximation errors of the same order as the number of near-boundary lattice points (e.g. polygonal regions). However, it is shown here that Lebesgue volumes can still adequately estimate differences between lattice point counts. The results are valid for sets and potential sampling regions of a variety of shapes, including all convex regions in 2- and 3-dimensional Euclidean space. The approximation tools permit more bias derivations for spatial statistics. New variance expansions for spatial sample means are provided for non-rectangular sampling regions.

Key Words: Asymptotic expansions, convex sets, lattice points, O-estimates, spatial statistics, sample mean, volume

1 Introduction

Lattice point theory is generally concerned with estimating the number of integers \( \mathbb{Z}^d \) (or some other lattice points) which lie inside large bounded bodies in \( \mathbb{R}^d \) Euclidean space. Historically lattice point counting has focused on the plane \( \mathbb{R}^2 \) and one important question: When the inside of a curve with area \( A \) is blown-up by a scaling factor \( b \), how big is the discrepancy between the area \( b^2 A \) and
the number of $\mathbb{Z}^2$ integer points inside the new curve? Writing the lattice point count $\#(b)$ of the new curve-delineated set as
\[ \#(b) = b^2A + \Delta(b), \quad (1) \]
the search for the best possible estimates of the remainder $\Delta(b)$ is known as the “$O$-problem” in number theory [cf. Krätzel (1988)]. Usually the curves considered are sufficiently “smoothly winding” and estimation tools for exponential sums are applied [cf. Huxley (1992), Chapter 2]. Better approximations of exponential sums often in turn lead to sharper “$O$-estimates.” For example, for convex sets with a “nice” boundary, van der Corput (1920)’s answer to the posed question above is $O(b^{46/39+\epsilon})$, while the best answer, based on the latest analytic methods for exponential sums, is $O(b^{48/73+\epsilon})$ for appropriately smooth curves [Huxley (1993, 1996)]. The $O$-problem and exponential sum estimation are highly active areas of research, with increasing attention as well in higher dimensional extensions [cf. Krätzel and Nowak (1991, 1992)].

In this paper, we introduce a variation on the $O$-problem which has significant application in statistics. Asymptotic developments with spatial statistics for lattice data often require an approximation for the difference between two lattice point counts. Suppose $R_n \subset \mathbb{R}^d$ represents a sampling region with available observations located at sampling sites $\mathbb{Z}^d \cap R_n$. Then, bias expansions for statistics computed on $R_n$ often depend crucially on subtracted counts:
\[ \#Z^d \cap R_n - \#Z^d \cap R_n \cap (k + R_n), \quad k \in \mathbb{Z}^d, \quad (2) \]
where $\#B$ denotes the cardinality of a finite set $B$. Because the computation of volumes is usually more tractable than counts, Lebesgue volumes could in principle be used to approximate lattice point counts, as in the $O$-problem. However, there is a catch: the required accuracy of the approximation to (2) must typically be of smaller order than the number of lattice points near the boundary of the sampling region $R_n$. Serious complications then arise with “volume-for-count” estimation and, even in the plane, many non-trivial $O$-estimates are in fact set by the number of bordering lattice points [cf. Krätzel (1988), Theorem 1.7; Huxley (1996), Lemma 2.1.1, Theorem 2.3.3]. That is, for $\mathbb{R}^2$ regions without nice smooth borders, analytical methods for exponential sums become inapplicable and the remainder $\Delta(b)$ in (1) is often exactly of order $O(b)$, corresponding to both the perimeter length and the number of $\mathbb{Z}^2$ lattice points near the boundary of the inflated curve. This unfortunately holds true for many useful polygonal-shaped sampling regions in $\mathbb{R}^2$ (eg. triangles, trapezoids).

The purpose of this paper is two-fold. We first wish to describe the need in statistics for approximations with differenced lattice point counts, as in (2). We then produce some advances in estimating
such differences for (almost) convex sets, or sampling regions, in \( \mathbb{R}^2 \) or \( \mathbb{R}^3 \) space. It is shown that differenced volumes can well approximate subtracted lattice point counts, even though the Lebesgue volume can fail terribly to measure the number of lattice points in either set appearing in (2).

The rest of the paper is organized as follows. In Section 2, we outline the importance of (subtracted) lattice point counts in determining a spatial statistic's expectation or bias. A precise statement of a new lattice point counting problem, with close connections to the \( O \)-problem but with more relevance to statistics, is given in Section 3. In Section 4, main results on lattice point count approximations are presented. We apply our approximation tools to derive new statistical bias expansions which are valid for non-rectangular sampling regions in \( \mathbb{R}^2 \) and \( \mathbb{R}^3 \) in Section 5. Through several examples, we illustrate the influence of sampling region's geometry on the variance of the spatial sample mean. For clarity of exposition, all proofs are given separately and divided among Sections 6, 7, and 8.

## 2 A counting problem in statistics

Statisticians often wish to quantify the performance of an estimator, defined on some spatial sampling region, by determining the difference (or bias) between the estimator's expected value and some population parameter value targeted for inference. Expectation and bias expansions involving spatial lattice data often require difficult lattice point counts.

To make the discussion concrete, we describe a common sampling formulation from spatial statistics [cf. Sherman and Carlstein (1994), Lahiri (1999ab)]. Suppose \( \{Y_s : s \in \mathbb{Z}^d\} \) is a stationary collection of real-valued random variables; \( R_0 \) is a Borel subset of \((-1/2, 1/2)^d\) containing the origin as an interior point; \( \{b_n\} \) is a positive, real sequence which goes to infinity; and a sampling region \( R_n \) is obtained by "inflating" the template \( R_0 \) by \( b_n \): \( R_n = b_n R_0 \subset \mathbb{R}^d \). Then, we treat those integer points (sampling sites) located inside the sampling region \( R_n \) as the available observations for inference. Namely, the data are

\[
\{Y_s : s \in \mathbb{Z}^d \cap R_n\}.
\]

This sampling scenario leads to an "increasing domain" asymptotic framework for studying statistics of spatial lattice data [cf. Cressie (1993)]. For practical reasons, \( R_0 \) is usually equipped with a mild boundary condition which guarantees that the number of lattice points near the border of \( R_n \) is of smaller order \( O(b_n^{d-1}) \) than the totality of observations in \( R_n \). This avoids pathological sampling regions and implies that \( \# \mathbb{Z}^d \cap R_n / |R_n| \to 1 \) as \( n \to \infty \), using \( |B| \) for the Lebesgue volume of an uncountable set. Such boundary conditions are typically satisfied by many convex and non-convex (e.g. star-shaped)
sets.

The expectation of a spatial statistic depends intricately on the determination of set sizes like \( \#Z^d \cap R_n \cap (k + R_n), \ k \in Z^d \). These lattice point counts arise, for example, in the leading order component of an estimator's bias, which is often of the form:

\[
\frac{1}{\#Z^d \cap R_n} \sum_{k \in Z^d} \left\{ \#Z^d \cap R_n \cap (k + R_n) - \#Z^d \cap R_n \right\} c(k) \\
= \frac{1}{b_n |R_0|} \lim_{n \to \infty} b_n^{(d-1)} \sum_{k \in Z^d} \left\{ \#Z^d \cap R_n \cap (k + R_n) - \#Z^d \cap R_n \right\} c(k) + o(b_n^{-1}) \tag{3}
\]

for some scaling constants \( \{c(k) : k \in Z^d\} \) satisfying a summability condition. In one common scenario in statistics, each constant \( c(k) = \text{Cov}(Y_{i+k}, Y_{i}) \) represents the autocovariance function for a mixing random field, with \( \sum_{k \in Z^d} ||k||_{\infty} c(k) < \infty, ||k||_{\infty} = \max_{1 \leq i \leq d} |k_i| \) [cf. Doukhan (1994)].

For instance, consider a time series stretch \( Y_1, \ldots, Y_n \) with mean \( E(Y_j) = 0 \) and autocovariance function \( c(k) = \text{Cov}(Y_j, Y_{j+k}) = E(Y_j Y_{j+k}), k \in Z \) (an equivalent size \( n \) sequence is obtained by \( R_0 = (-1/2,1/2], b_n = n \) in our sampling formulation). The expectation of a simple statistic based on the sample mean \( \bar{Y}_n = \frac{1}{n} \sum_{j=1}^{n} Y_j/n \) is given by

\[
E(n\bar{Y}_n^2) = \frac{1}{n} \sum_{k=-n}^{n} (n-|k|) \sigma(k), \tag{4}
\]

an elementary (variance) result from time series [cf. Fuller (1996), Corollary 6.1.1.2]. However, the expectation in (4) involves a natural lattice point count: the number of observations in the sample which are exactly \( k \) time lags apart for each \( k \in Z, |k| \leq n \); this is equivalently the size of the set \( Z \cap [1, n] \cap (k + [1, n]) \). The difference between (4) and its limiting value, the population parameter \( \nu_\infty = \sum_{k \in Z} \sigma(k) \), can be expressed:

\[
nE(\bar{Y}_n^2) - \nu_\infty = n^{-1} \sum_{k \in Z} \left\{ (n-|k|) - n \right\} \sigma(k) + O(n^{-1} \sum_{|k| \geq n} |k| \sigma(k)) \\
= -n^{-1} \sum_{k \in Z} |k| \sigma(k) + o(n^{-1}), \tag{5}
\]

a special case of (3). We have now expanded the bias the estimator \( n\bar{Y}_n^2 \) for \( \nu_\infty \) and isolated its well-known leading order \( O(b_n^{-1}) \) term which is required, for example, to determine optimal block sizes for statistical resampling methods such as the block bootstrap [cf. Hall, Horowitz and Jing (1995); Lahiri (1996)] and subsample-jackknife [Künsch (1989), Politis and Romano (1993)]. The expansions in (3) are analogously needed in optimizing spatial subsampling methods (using the spatial sample mean \( \bar{Y}_n = \sum_{e \in \mathcal{C}^d \cap R_n} Y_e/\#R_n \cap Z^d \) and appear in bias expansions involving spectrographic estimators [cf. Guyon (1995), Chapter 4; Nordman and Lahiri (2002)].
The asymptotic study of spatial statistics for lattice data is often confined to rectangular sampling regions $R_n = [\prod_{j=1}^d [1, n\ell_j]]$, with $\ell_j > 0$ (e.g., subsample-based estimation [Possolo (1991), Politis and Romano (1994), Hall and Jing (1996)]; Whittle estimation [Heyde and Gay (1993)]; bias expansions for autocovariance estimators and other spectral means [cf. Guyon (1995), Theorem 4.1.2]). These sampling regions have application, for instance, in agricultural field experimentation and image analysis, but also fail to encompass many real sampling shapes [cf. Cressie (1993)]. Note as well that the lattice point counts required in (3) for rectangular sampling regions are a fairly straightforward generalization of the time series case in (4):

$$\# \{\mathbb{Z}^d \cap \prod_{j=1}^d [1, n\ell_j] \cap (k + \prod_{j=1}^d [1, n\ell_j])\} = \prod_{j=1}^d ([n\ell_j] - |k_j|), \quad k \in \mathbb{Z}^d, |k_j| \leq n\ell_j.$$  

This counting aspect of rectangular regions greatly simplifies the theoretical development of associated spatial estimators.

The lack of better lattice point counting techniques does hinder statistical theory in some regards. Without the capacity to isolate and explicitly compute expansions as in (3), statisticians cannot determine the effect of a sampling region's shape or dimension on the properties of a spatial statistic. The need in statistics exists for better lattice count approximation tools, and estimation of differenced counts from (2) are especially useful in bias studies.

3 Problem statement

We first give a precise formulation of the counting problem of interest. Suppose $R_n = b_n R_0$ is a sampling region as described in Section 2, where the number of $\mathbb{Z}^d$ points near the boundary of $R_n$ is $O(b_n^{-1})$. For $t \in (-1/2, 1/2]^d$, define the translated integer lattice $\mathbb{Z}^d = t + \mathbb{Z}^d$. We will count $\mathbb{Z}^d$ points, rather than use the integer lattice $\mathbb{Z}^d$, to allow for a more general embedding of $R_n$ in $\mathbb{R}^d$ Euclidean space and avoid forcing $R_n$ to expand around a potential lattice point at the origin. The interest is approximating the difference in counts

$$\#Z^d \cap R_n - \#Z^d \cap (k + R_n), \quad k \in Z^d. \quad (6)$$

We now add an important twist which makes the approximation problem non-trivial and also helps to determine the main statistical bias component in (3): for each $k \neq 0 \in \mathbb{Z}^d$, the order of an approximation to (6) should satisfy

$$o(\ell_n^{d-1}). \quad (7)$$
We note that, by substituting a difference between Lebesgue volumes $|R_n| - |R_n \cap (k + R_n)|$ in place of (6), a uniform bound on the approximation error

$$C \cdot ||k||_\infty (b_n)^{d-1}, \quad C > 0, \ k \in \mathbb{Z}^d$$

(8)

follows from a very mild (and fairly uninformative) boundary condition on $R_0$: for any positive sequence $a_n \to 0$ as $n \to \infty$, the number of cubes of the lattice $a_n \mathbb{Z}^d$ intersecting both $R_0$ and $R_n$ is $O((a_n^{-1})^{d-1})$. This particular condition is satisfied by many convex (e.g. ellipsoids, polygons) and non-convex sets (e.g. star-shaped) [cf. Lahiri (1999a)].

For $k \in \mathbb{Z}^d$, define the discrepancy between volume and lattice point count of the intersection $R_n \cap (k + R_n)$:

$$\Omega_n(k) = |R_n \cap (k + R_n)| - \#\mathbb{Z}^d \cap R_n \cap (k + R_n).$$

We have mentioned that, even with seemingly simple bodies in space (e.g. convex polygons without smooth borders), volumes may not well approximate lattice point counts. Specifically, the magnitude of $\Omega_n(k)$ may be of exact order $O(b_n^{-d-1})$ (set also in (8) by the unrestrictive boundary condition on $R_0$) and not to the precision required in (7).

In this paper, we do approximate the number of lattice points in $R_n$ and $R_n \cap (k + R_n)$ from (6) by set volumes, though the Lebesgue volume may not adequately capture the lattice point count in either set. However, when subtracted, the errors incurred in the approximation of both sets can cancel each other to a great extent. That is, the difference between the two discrepancies $\Omega_n(0)$ and $\Omega_n(k)$ can be shown to satisfy (7) for a broad range of sets (sampling regions) in $\mathbb{R}^d$, $d \leq 3$, which are “nearly convex.”

**Definition.** A set $R \subset \mathbb{R}^d$ is called nearly convex if there exists a convex set $B$ such that $B^o \subset R \subset \overline{B}$, where $B^o$ and $\overline{B}$ denote the interior and closure of $B$, respectively.

The considered (sampling) region $R_n$ differs from a convex set possibly only at its boundary, but $\mathbb{Z}^d$ lattice points on the border of $R_n$ may be arbitrarily included or excluded. Statistically speaking, this sampling framework allows for missing observations (sampling sites) near the edges of the sampling region. Hence, the expanding sets $R_n$ are not necessarily closed, which also differs from the usual set-up in the traditional $O$-problem [cf. Krätzel (1988), Chapter 3].
4 Main results

4.1 Approximations in $\mathbb{R}^2$

We now give a theorem establishing a fairly sharp bound on the resulting approximation error when using volumes in (6) rather than counts, valid for regions based on large class of "nearly convex" templates $R_0$ in $\mathbb{R}^2$. In the following, $\mathbb{Z}^+ = \{0, 1, 2, \ldots\}$ denotes the nonnegative integers.

**Theorem 1** Let $d = 2$ and $R_0$ be nearly convex. Suppose $R_0^n$ contains a closed ball of radius $\varepsilon$ around the origin. Then, for $k \in \mathbb{Z}^2$, there exists $N_k \in \mathbb{Z}^+$ such that for $n \geq N_k$,

$$|\Omega_n(0) - \Omega_n(k)| \leq \varepsilon^{-1}28||k||^2_{\infty}.$$  

The proof of Theorem 1 is given in Section 6.1 and involves placing the volume-for-count approximation errors of $R_n \cap (k + R_n)$ and a subset of $R_n$ into a one-to-one correspondence, for each $k \in \mathbb{Z}^2$; the size of the volume and lattice point count of the remaining portion of $R_n$ may be bounded by $||k||^2_{\infty}$. The argument exploits properties of convex sets to pinpoint the subset of $R_n$ of interest for each $k \in \mathbb{Z}^2$.

Approximation results analogous to Theorem 1 may be possible for some templates representable as a finite union of convex sets. Some fairly complicated non-convex (sampling) regions could then be described. In this paper, we do not consider this possibility greatly but we provide a small extension in the next theorem, treating templates $R_0 = S_1 \cup S_2$ formed by the union of two almost convex sets $S_1, S_2$ with borders that are at most finitely intersecting. The assumption that $S_1, S_2$ share (no more than) finitely many common border points permits some control over the size of $b_n(S_j \setminus S_j - j) \cap (k + b_n(S_3 - j \setminus S_j))$ and allows a determination of the amount of cancellation between errors $\Omega_n(0)$ and $\Omega_n(k)$.

In the following, write $\partial B = B \setminus B^o$ to denote the boundary of a set $B \subset \mathbb{R}^2$.

**Theorem 2** Let $d = 2$ and $R_0 = S_1 \cup S_2$, where each $S_j$ is nearly convex with a nonempty interior. Suppose $\partial S_1 \cap \partial S_2$ is empty or finite. Then, for $k \in \mathbb{Z}^2$, there exists $C > 0$ and $N_k \in \mathbb{Z}^+$ such that for $n \geq N_k$,

$$|\Omega_n(0) - \Omega_n(k)| \leq C \cdot ||k||^2_{\infty},$$

$$\left(|R_n| - |R_n \cap (k + R_n)|\right) - \left(V_{1,n}(k) + V_{2,n}(k) - V_{3,n}(k)\right) \leq C \cdot ||k||^2_{\infty},$$

where $V_{j,n}(k) = |R_{j,n}| - |R_{j,n} \cap (k + R_{j,n})|$, $R_{j,n} = b_n S_j$, $j = 1, 2, 3$; $S_3 = S_1 \cap S_2$.

Theorem 2 says that we can even find a computationally more feasible estimate of count difference in (6). Namely, (6) can be approximated fairly well with a linear combination of separately computed volume differences $V_{j,n}(k)$, each corresponding to an individual (nearly) convex set. The proof of Theorem 2 is given in Section 6.2.
4.2 An IR² extension

Some spatial resampling methods in statistics involve tiling a subset of IR² space with disjoint congruent sampling regions [cf. Politis and Romano (1994), Sherman and Carlstein (1994), Lahiri (1999a), Nordman (2002)]. For this purpose, a commonly used collection of expanding, non-overlapping regions is given by:

\[ R_{i,n} = b_n(i + R_0), \quad i \in Z^2, \]

and, analogous to (6), the need also arises to approximate the count differences

\[ \#Z^n \cap R_{i,n} - \#Z^n \cap (k + R_{i,n}), \quad k \in Z^2, \]

for an arbitrary \( i \in Z^2 \). We can provide an extension of Theorems 1 and 2 for bounding the approximation error incurred by estimating (9) with the corresponding difference in volumes: \( |R_{i,n}| - |R_{i,n} \cap (k + R_{i,n})| = |R_n| - |R_n \cap (k + R_n)| \). Let

\[ \Omega_{i,n}(k) = |R_n \cap (k + R_n)| - \#Z^n \cap R_{i,n} \cap (k + R_{i,n}), \quad i, k \in Z^2. \]

We can now bound the volume-for-count approximation error uniformly across each potential non-overlapping (sampling) region, \( i \in Z^2 \).

**Corollary 1** Under the conditions of Theorem 1 or Theorem 2, for each \( k \in Z^2 \), there exists \( C > 0 \) and \( N_k \in Z^+ \) such that for \( n \geq N_k \),

\[ |\Omega_{i,n}(0) - \Omega_{i,n}(k)| \leq C \cdot ||k||^2, \quad i \in Z^2. \]

A bound with \( C = e^{-1.28} \) holds under Theorem 1 assumptions.

4.3 Results for IR³

Our final theorem shows that corresponding Lebesgue volumes can estimate a difference of lattice point counts for a broad range of IR³ templates (to a sufficient degree of accuracy as in (7)). We formulate the theorem to be available for application to non-overlapping regions in IR³: \( R_{i,n}, i \in Z^3 \).

**Theorem 3** Let \( R_0 \) be nearly convex for \( d = 3 \) and \( k \in Z^3 \). Then, there exists \( N_k \in Z^+ \) and \( C_k > 0 \) such that, for \( n \geq N_k \),

\[ |\Omega_{i,n}(0) - \Omega_{i,n}(k)| \leq C_k \left(b_n^{5/3} + \xi_{k,n} b_n^2\right), \quad i \in Z^2, \]

where \( \{\xi_{k,n}\}_{n=1}^\infty \subset [0, \infty) \) is a sequence (possibly dependent on \( k \)) such that \( \xi_{k,n} \to 0 \).
The bound on the approximation error may certainly be improved with more knowledge of the specific geometry of the template \( R_0 \). The proof of Theorem 3 in Section 7 shows that the null sequence \( \{\Omega_{i,n}(0) - \Omega_{i,n}(k)\} \) depends on a continuous function of the boundary of \( R_0 \) (a surface), but not necessarily the curvature of the surface. More information about \( R_0 \) may then sharpen the bound on \( |\Omega_{i,n}(0) - \Omega_{i,n}(k)| \).

5 Applications to statistical bias expansions

The bias expansion in (5) is well-known in statistical time series. However, the same expansion has yet been undetermined for the spatial counterpart involving spatial sampling regions \( R_n = b_n R_0 \) in \( \mathbb{R}^d \) for a real-valued random field \( \{Y_s : s \in \mathbb{Z}^d\} \), \( E(Y_s) = 0 \), \( c(k) = \text{Cov}(Y_s, Y_{s+k}) \), \( k \in \mathbb{Z}^d \); a spatial sample mean \( \hat{Y}_n = \sum_{s \in \mathbb{Z}^d \cap R_n} Y_s / N_n \), \( N_n = \# R_n \cap \mathbb{Z}^d \); and the spatial population parameter \( \nu_{d,\infty} = \sum_{k \in \mathbb{Z}^d} c(k) \) for \( d \geq 2 \). Assuming the spatial autocovariances to be appropriately summable \( \sum_{k \in \mathbb{Z}^d} ||k||_{\infty}^d |c(k)| < \infty \), we can expand and explicitly determine the leading order bias term \( B_0 \) of

\[
N_n E(Y_n^2) - \nu_{d,\infty} = \frac{B_0}{b_n} + o(b_n^{-1})
\]

from (3) by using Theorems 1-3. (Specifically, we apply these theorems with (8) and the Lebesgue Dominated Convergence Theorem.)

Table 1 provides this bias component \( B_0 \) for several differently shaped templates \( R_0 \) to demonstrate the influence of the geometry of the sampling region \( R_n \) on the bias of \( N_n \hat{Y}_n^2 \) for \( \nu_{d,\infty} \); note (10) equivalently represents the difference between the variance of the standardized sample mean \( N_n^{1/2} \hat{Y}_n \) and its limiting variance.

6 Proofs

We need some additional notation before beginning the proofs. Vectors in \( \mathbb{R}^d \) are denoted with bold font \( \mathbf{x} = (x_1, \ldots, x_d)' \), \( \mathbf{k} = (k_1, \ldots, k_d)' \). For \( \mathbf{x} \in \mathbb{R}^d \), \( ||\mathbf{x}|| \) denotes the Euclidean norm, while \( ||\mathbf{x}||_{\infty} = \max_{1 \leq i \leq d} |x_i| \) remains the \( l^\infty \) norm. Define the dot product \( \langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}' \mathbf{y} \). Denote the greatest common divisor of positive integers \( i, j \) as \( \text{gcd}(i, j) \).
Table 1  Sampling templates $R_0 \subset (-1/2,1/2)^d$, $d = 2,3$, and associated bias $B_0$ from (10). For $x \in \mathbb{R}^d$, $||x||$ denotes the Euclidean norm; $||x||_1 = \sum_{i=1}^d |x_i|$; $||x||_\infty = \max_{1 \leq i \leq d} |x_i|$.

<table>
<thead>
<tr>
<th>$R_0$</th>
<th>$B_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Circle of radius 1/2 at origin</td>
<td>$4/\pi \sum_{k \in \mathbb{Z}^d}</td>
</tr>
<tr>
<td>“Diamond” in Figure 1(i)</td>
<td>$2 \sum_{k \in \mathbb{Z}^d}</td>
</tr>
<tr>
<td>Triangle in Figure 1(ii)</td>
<td>$\sum_{</td>
</tr>
<tr>
<td>“Cross” in Figure 1(iii)</td>
<td>$4/3 \sum_{k \in \mathbb{Z}^d}</td>
</tr>
<tr>
<td>Rectangle $(-1/2,1/2]^2$</td>
<td>$\sum_{k \in \mathbb{Z}^d}</td>
</tr>
<tr>
<td>Sphere of radius 1/2 at origin</td>
<td>$3/2 \sum_{k \in \mathbb{Z}^d}</td>
</tr>
</tbody>
</table>

We often use some elementary, but important, properties of convex sets which can be found in Kelly and Weiss (1979), p. 111-116; one such feature of a convex set $B$ is that $x \in B^o$, $y \in \overline{B}$ implies that $B^o$ contains the open line segment between $x$ and $y$. By assumption, there exists a convex set $B \subset \mathbb{R}^d$ such that $B^o \subset R_0 \subset \overline{B}$. By convexity, $B^o = R_0^o$ and $\overline{B} = \overline{R_0}$ are convex sets; $R_0^o = \overline{R_0}^o$ and $\overline{R_0} = \overline{R_0}^o$; the “borders” of $R_0$ and $\overline{R_0}$ are the same (i.e. $\overline{R_0} \setminus R_0^o = R_0^o \setminus \overline{R_0}$).

6.1 Proof of Theorem 1

Note that there is nothing to prove when $k = 0 \in \mathbb{Z}^2$. Fix $k = (k_1,k_2)^t \neq 0 \in \mathbb{Z}^2$ and define a new vector (a function of $k$):

$$\hat{k} = \frac{1}{g(k)}(-k_2,k_1)^t \in \mathbb{Z}^2, \quad g(k) = \begin{cases} \|k\|_\infty & \text{if } k_1 = 0 \text{ or } k_2 = 0, \\ \gcd(|k_1|,|k_2|) & \text{otherwise.} \end{cases} \quad (11)$$

Write $M = \sup\{\langle x,\hat{k} \rangle : x \in \overline{R_0}\}$, $m = \inf\{\langle x,\hat{k} \rangle : x \in \overline{R_0}\}$ and let $\nu_\text{M}, \nu_\text{m} \in \overline{R_0}$ such that $M = \langle \nu_\text{M},\hat{k} \rangle$, $m = \langle \nu_\text{m},\hat{k} \rangle$. (The existence of $\nu_\text{M}, \nu_\text{m}$ is clear by compactness and continuity). Note also that $B(0,\varepsilon) \subset R_0^o$ implies $M > 0$ and $0 > m$.

Because the origin $0 \in R_0^o$, we have that for all $\omega \in [0,1)$, $\omega \nu_\text{M}, \omega \nu_\text{m} \in R_0^o$. This implies, in turn, that for all $c \in (m,M)$, the set $\overline{R_0} \cap \{x \in \mathbb{R}^2 : \langle x,\hat{k} \rangle = c\}$ is a closed line segment and that the sets $\overline{R_0} \cap \{x \in \mathbb{R}^2 : \langle x,\hat{k} \rangle = M\}$, $\overline{R_0} \cap \{x \in \mathbb{R}^2 : \langle x,\hat{k} \rangle = m\}$ are each either a single point or a closed line segment [cf. (Kelly and Weiss) 1977, p. 112].
We will first define the boundary points of $R_0$.

**Proposition 1** There exist continuous functions $f_\nu, f_\omega : [m, M] \rightarrow \overline{R}_0 \setminus R_0^2$ such that, for all $c \in [m, M]$,

$$f_\nu(c) = y_c + \frac{\bar{p}_c - k}{||k||}, \quad f_\omega(c) = y_c - \frac{\bar{p}_c - k}{||k||},$$

where $y_c = \begin{cases} \frac{(c/M) \cdot v_m}{||k||}, & \text{if } c \geq 0, \\ \frac{(c/m) \cdot v_m}{||k||}, & \text{if } c < 0, \end{cases}$ and $\bar{p}_c = \sup\{\rho : y_c + ||k||^{-1} \rho \cdot k \in \overline{R}_0\}$.

The proof of Proposition 1 is given in Section 8.

For $c \in [m, M]$, define the diameter of the closed "line" segment $\{x \in \overline{R}_0 : (x, \hat{k}) = c\}$ as

$$d_m(c) = \sup\{||x_i - x_2|| : x_i \in \overline{R}_0, (x_i, \hat{k}) = c; i = 1, 2\} = ||f_\nu(c) - f_\omega(c)|| = \bar{p}_c - p_c.$$

From the continuity of $f_\nu$ and $f_\omega$, $d_m(\cdot)$ is also continuous on $[m, M]$ so that there exists $\bar{c} \in [m, M]$ such that $d_m(\bar{c}) = \max\{d_m(c) : c \in [m, M]\}$ by compactness. The function $d_m(\cdot)$ is also concave on $[m, M]$. We make the following claims based on the continuity of $d_m(\cdot)$ and the convexity of $\overline{R}_0$:

**Proposition 2** (a) The set $A_\epsilon = \{c \in [m, M] : d_m(c) = d_m(\bar{c})\}$ is a closed interval.

(b) If $m \leq c_1 < c_2 < \inf A_\epsilon$, then $d_m(c_1) < d_m(c_2) < d_m(\bar{c})$. If $\sup A_\epsilon < c_1 < c_2 \leq M$, then $d_m(c_2) < d_m(c_1) < d_m(\bar{c})$.

Proposition 2 is proven in Section 8.

We now develop a systematic way of locating and counting the $\mathbb{Z}^2$ lattice points within the set $R_n$ so that we can track (and subsequently count) points in the intersection $R_n \cap (k + R_n)$. In particular, we desire to know which lattice points exit $R_n$ upon translation by $k$. We begin with a partition of the space $\mathbb{Z}^2$ by lines (hyperplanes) which are parallel to the vector $k$ and thereby the relevant direction of translation.

For $\hat{k} = (\hat{k}_1, \hat{k}_2)'$, we have by construction that $\gcd(|\hat{k}_1|, |\hat{k}_2|) = 1$ (even if one of the coordinates of $\hat{k}$ is zero). Hence, there exist integers $w_1, w_2 \in \mathbb{Z}$ such that $w_1 \hat{k}_1 + w_2 \hat{k}_2 = 1$ [cf. Gallian (1994), p. 6]. Letting $w = (w_1, w_2)'$, we have for all $c \in \mathbb{Z}$: $cw + t \in \mathbb{Z}^2$ and, furthermore,

$$x, y \in \mathbb{Z}^2, (x, \hat{k}) = (y, \hat{k}) \quad \Rightarrow \quad y = q||k||\frac{k}{||k||} + x,$$

for some $q \in \mathbb{Z}$. We briefly justify (12). If $(x, \hat{k}) = (y, \hat{k})$, then $x - y = \omega k$ for some $\omega \in \mathbb{R}$ and also $|x_1 - y_1||\hat{k}_1| = |y_2 - x_2||\hat{k}_2|$ with $|x_1 - y_1|, |y_2 - x_2| \in \mathbb{Z}_+$. Assume $x \neq y$. Then there are three possibilities: 1.) If $|\hat{k}_1| > 0$, $|\hat{k}_2| > 0$, then $|\hat{k}_1|$ divides $|y_2 - x_2|$ and $|\hat{k}_2|$ divides $|x_1 - y_1|$, implying that...
\[ q|k_1| = |y_2 - x_2|, q|k_2| = |x_1 - y_1| \text{ for some } q \geq 1 \in \mathbb{Z}_+; 2.) \text{ If } k_2 = 0, \text{ then } |k_1| = 1 \text{ and } |y_2 - x_2| = 0; 3.) \text{ If } k_1 = 0, \text{ then } |k_2| = 1 \text{ and } |y_2 - x_2| = 0. \text{ In all cases, } ||x - y|| = q||k|| \text{ for some } q \in \mathbb{Z}_+ \text{ and we have (12).} \]

Hence, we may express \( \mathbb{Z}^2 \) in terms of lines parallel to \( k \):

\[ \mathbb{Z}^2 = \left\{ cw + q||k|| \frac{k}{||k||} + t : c \in \mathbb{Z}, q \in \mathbb{Z}_+ \right\}. \tag{13} \]

We now develop some notation for matching "border" points of \( \mathbb{R}_n \) (falling on certain parallel lines/hyperplanes of interest) with points of \( \mathbb{R}_n \cap \mathbb{Z}^2 \) near the boundary of \( \mathbb{R}_n \). Note that, if \( x \in \mathbb{R}_n \), then \( \langle x, k \rangle \in [m_n, M_n] \) for \( m_n = b_n m, M_n = b_n M \). Define functions \( f_{u,n}, f_{-n} : [m_n, M_n] \rightarrow \mathbb{R}_n \setminus \mathbb{R}_n^* \) such that, for all \( c \in [m_n, M_n] \),

\[ f_{u,n}(c) = b_n f_u(c/b_n), \quad f_{-n}(c) = b_n f_{-c}(c/b_n). \]

Let \( t_0 = (t, k) \). Note, if \( x \in \mathbb{R}_n \cap \mathbb{Z}^2 \), then \( \langle x, k \rangle \in [m_n, M_n] \cap (\mathbb{Z} + t_0) \). We also define functions \( f^*_{u,n}, f^*_{-n} : [m_n, M_n] \cap (\mathbb{Z} + t_0) \rightarrow \mathbb{Z}^2 \) such that

\[ f^*_{u,n}(c) = f_{u,n}(c) - \frac{k}{||k||}, \quad f^*_{-n}(c) = f_{-n}(c) + \frac{k}{||k||}, \quad c \in [m_n, M_n] \cap (\mathbb{Z} + t_0), \tag{14} \]

where \( \beta_{c,n} = 0 \) if \( f_{u,n}(c) \in \mathbb{R}_n \cap \mathbb{Z}^2 \) and \( 0 < \beta_{c,n} \leq ||k|| \), otherwise; and \( \beta_{c,n} = 0 \) if \( f_{-n}(c) \in \mathbb{R}_n \cap \mathbb{Z}^2 \), and \( 0 < \beta_{c,n} \leq ||k|| \), otherwise. Note that \( f^*_{u,n}, f^*_{-n} \in \mathbb{Z}^2 \) are well-defined functions by (12). Write functions \( e_{1,n}, e_{2,n} : [m_n, M_n] \cap (\mathbb{Z} + t_0) \rightarrow [-1/2, 1/2] \) such that for \( c \in [m_n, M_n] \cap (\mathbb{Z} + t_0) \),

\[ e_{1,n}(c) = 1/2 - \frac{||k||^{-1}||f_{u,n}(c) - f^*_{u,n}(c)||}{||k||}, \quad e_{2,n}(c) = 1/2 - \frac{||k||^{-1}||f_{-n}(c) - f^*_{-n}(c)||}{||k||}. \tag{15} \]

The functions in (15) incorporate the distance between the border of \( \mathbb{R}_n \) and the "closest" \( \mathbb{Z}^2 \) lattice points (defined by \( f^*_{u,n}, f^*_{-n} \)) on lines (parallel to \( k \)) cutting through \( \mathbb{R}_n \).

To ease the counting arguments, we define a function that counts the number of \( \mathbb{Z}^2 \) points lying simultaneously in \( \mathbb{R}_n \) and on lines (parallel to \( k \)) slicing through \( \mathbb{R}_n \). Define the function \( F_n : \mathbb{R} \rightarrow [0, \infty) \) such that

\[ F_n(c) = \#(\mathbb{Z}^2 \cap \{x \in \mathbb{R}_n : \langle x, k \rangle = c\}, \quad c \in \mathbb{R}. \]

We will make use of the following equality.

**Proposition 3.** Let \( c_{M,n} = [M_n - t_0] + t_0 \), \( c_{m,n} = [m_n - t_0] + t_0 \). For \( c \in [m_n, M_n] \cap (\mathbb{Z} + t_0) \),

\[ F_n(c) = \||k||^{-1}||f_{u,n}(c) - f^*_{u,n}(c)|| + e_{1,n}(c) + e_{2,n}(c) + I_{(e_{1,n}(c) = 0)} I_{m,n} + I_{(e_{2,n}(c) = 0)} I_{m,n}. \tag{16} \]

where \( I_{M,n} \), \( I_{m,n} \) denote evaluations of \( I_{(f_{u,n}(c) = f_{-n}(c)) \in \mathbb{Z}^2 \setminus \mathbb{R}_n} \) at \( c_{M,n} \) and \( c_{m,n} \), respectively.
We prove (16) in Section 8.

Now suppose, for \( n > N^* \in \mathbb{Z^+} \), it holds that \( 4\|k\|_\infty < \varepsilon^2 \), implying
\[
m_n - t_0 < -1, \quad M_n - t_0 > 1, \quad c_{M,n} - c_{m,n} > 2,
\]
(17)
because \( R^n_\varepsilon \) contains an open ball of radius \( \varepsilon \leq 1/2 \) around the origin. We use (13) and (16) to write the \( \mathbb{Z}^2 \) lattice point count of \( R_n \) as a sum over \( \mathbb{Z} + t_0 \):
\[
\#\mathbb{Z}^2 \cap R_n = \sum_{i=cm,n} F_n(i)
= \sum_{i=cm,n} \left( \frac{\|k\|^{-2}}{1} \left| f_{u,n}(i) - f_{L,n}(i) \right| + e_{1,n}(i) + e_{2,n}(i) \right) + I_{M,n} + I_{m,n}.
\]
(18)

We then re-write portions of the sums in (18) to obtain a "trapezoidal" approximation of the area \( |R_n| \). For \( i \in [t_0, cm,n] \cap (\mathbb{Z} + t_0) \), let \( T_{i,1,n} \) denote the closed trapezoid defined by the corners \( f_{u,n}(i), f_{L,n}(i), f_{u,n}(i+1), f_{L,n}(i+1) \); for \( i \in \{cm,n, t_0\} \cap (\mathbb{Z} + t_0) \), \( T_{i,2,n} \) is the trapezoid with corners \( f_{u,n}(i), f_{L,n}(i), f_{u,n}(i-1), f_{L,n}(i-1) \); let \( T_{c,1,n} \) denote the trapezoid formed by the points \( f_{u,n}(c), f_{L,n}(c) \) and \( f_{u,n}(M_n), f_{L,n}(M_n) \) for \( c = cm,n, j = 1 \) \( c = cm,n, j = 2 \).

Note that the distance between the hyperplanes \( (x, \hat{k}) = c_1, (x, \hat{k}) = c_2 \) for \( c_1, c_2 \in \mathbb{R} \) is \( |c_2 - c_1|/\|\hat{k}\| \) [cf. Kelly and Weiss (1979), p. 99]. Hence, the area of the trapezoid \( T_{i,1,n} \), \( i \in (cm,n, cm,n) \cap (\mathbb{Z} + t_0) \), is:
\[
|T_{i,1,n}| = \frac{1}{2\|k\|} \left( \frac{\|f_{u,n}(i) - f_{L,n}(i)\| + \|f_{u,n}(i - (-1)^j) - f_{L,n}(i - (-1)^j)\|}{\|k\|} \right),
|T_{cm+1,n}| = \frac{M_n - cm,n}{2\|k\|} \left( \frac{\|f_{u,n}(cm,n) - f_{L,n}(cm,n)\| + \|f_{u,n}(M_n) - f_{L,n}(M_n)\|}{\|k\|} \right),
|T_{cm+2,n}| = \frac{cm,n - m_n}{2\|k\|} \left( \frac{\|f_{u,n}(cm,n) - f_{L,n}(cm,n)\| + \|f_{u,n}(m_n) - f_{L,n}(m_n)\|}{\|k\|} \right).
\]

Hence, using \( cm,n < t_0 < cm,n \) by (17), we can write in (18):
\[
\sum_{i=cm,n} \frac{\|k\|^{-2}}{1} |f_{u,n}(i) - f_{L,n}(i)| = \sum_{i=cm,n} |T_{i,1,n}| + \sum_{i=cm,n} |T_{i,2,n}| + E_{M,n} + E_{m,n};
\]
(19)
\[
E_{M,n} = -(2\|k\|^{-2}) (M_n - cm,n) |f_{u,n}(M_n) - f_{L,n}(M_n)|
+ (2\|k\|^{-2}) (1 - M_n + cm,n) |f_{u,n}(cm,n) - f_{L,n}(cm,n)|,
E_{m,n} = -(2\|k\|^{-2}) (cm,n - m_n) |f_{u,n}(m_n) - f_{L,n}(m_n)|
+ (2\|k\|^{-2}) (1 - m_n + cm,n) |f_{u,n}(cm,n) - f_{L,n}(cm,n)|.
\]

We note that for \( i \in \{t_0, M_n\} \cap (\mathbb{Z} + t_0) \),
\[
\left| \left\{ x \in R_n : i \leq (x, \hat{k}) \leq \min\{M_n, i + 1\} \right\} \right| - |T_{i,1,n}| = a_{u,1,n}(i) + a_{L,1,n}(i);
\]
\[ a_{v,1,n}(i) = \|\hat{k}\|^{-1} \int_0^1 \left| f_{v,n}(i + \omega \min\{b_n M, i + 1\} - i) - \left( f_{v,n}(i) + \omega\left( f_{v,n}(\min\{M_n, i + 1\}) - f_{v,n}(i)\right)\right) \right| \, d\omega, \quad \gamma = L, U, \]

by convexity of \( R_n \). The integrands in \( a_{u,1,n} \) and \( a_{L,1,n} \) are continuous and hence integrable. Likewise, for \( i \in [m_n, t_0] \cap (Z + t_0) \),

\[ \left| \{ x \in R_n : \max\{m_n, i - 1\} \leq \langle x, \hat{k} \rangle \leq i \} \right| - |T_{v,2,n}| = a_{u,2,n}(i) + a_{L,2,n}(i), \]

where the functions \( a_{u,2,n}(i) \) and \( a_{L,2,n}(i) \) are scaled integrals obtained by substituting \( \max\{m_n, i - 1\} \) for \( \min\{A_n, i + 1\} \) in the definitions of \( a_{u,1,n}(i) \) and \( a_{L,1,n}(i) \), respectively.

Then, using (18) and (19), we write

\[
\Omega_n(0) = \\
= \sum_{i=0}^{c_{M,n}} \left( a_{u,1,n}(i) + a_{L,1,n}(i) - e_{1,n}(i) - e_{2,n}(i) \right) - I_{M,n} - E_{M,n} \\
+ \sum_{i=e_{m,n}}^{t_0} \left( a_{u,2,n}(i) + a_{L,2,n}(i) - e_{1,n}(i) - e_{2,n}(i) \right) - I_{m,n} - E_{m,n} + (e_{1,n}(t_0) + e_{2,n}(t_0)) \\
\equiv P_{1,n} + P_{2,n} + (e_{1,n}(t_0) + e_{2,n}(t_0)). \tag{20}
\]

We now consider the discrepancy: \( \Omega_n(k) \). We first define some functions of the border of the convex set \( R_n \cap (R_n + k) \) analogous to the ones defined with respect to \( R_n \). Let

\[ \hat{m}_n = \inf\{\langle x, \hat{k} \rangle : x \in \overline{R_n} \cap (k + \overline{R_n})\}, \quad \hat{M}_n = \sup\{\langle x, \hat{k} \rangle : x \in \overline{R_n} \cap (k + \overline{R_n})\}. \]

For \( c \in [\hat{m}_n, \hat{M}_n] \) and \( y_{c,n} \in \overline{R_n} \cap (k + \overline{R_n}) \) such that \( \langle y_{c,n}, \hat{k} \rangle = c \), define

\[ \hat{f}_{u,n}(c) = y_{c,n} + k \cdot \sup\{\beta \geq 0 : y_{c,n} + \beta \cdot k \in \overline{R_n} \cap (k + \overline{R_n})\}, \]
\[ \hat{f}_{L,n}(c) = y_{c,n} - k \cdot \sup\{\beta \geq 0 : y_{c,n} - \beta \cdot k \in \overline{R_n} \cap (k + \overline{R_n})\}. \]

The functions \( \hat{f}_{u,n}, \hat{f}_{L,n} \in \overline{R_n} \cap (k + \overline{R_n}) \) are analogous to \( f_{u,n} \) and \( f_{L,n} \), but created with respect to the intersection \( R_n \cap (k + R_n) \). If \( c \in [\hat{m}_n, \hat{M}_n] \cap (Z + t_0) \), write \( \hat{f}_{u,n}(c), \hat{f}_{L,n}(c) \in \mathbb{Z}^2 \) similar to \( f^\ast_{u,n}(c), f^\ast_{L,n}(c) \) substituting \( \hat{f}_{u,n} \) for \( f_{u,n} \) (or \( \hat{f}_{L,n} \) for \( f_{L,n} \)) and \( R_n \cap (k + R_n) \) for \( R_n \) in the definition from (14). Likewise, for \( c \in [\hat{m}_n, \hat{M}_n] \cap (Z + t_0) \), define the border “errors” \( e_{1,n}, e_{2,n} \) with respect to \( R_n \cap (k + R_n) \) by replacing \( \{f_{u,n}, f^\ast_{u,n}, f_{L,n}, f^\ast_{L,n}\} \) with their counterparts \( \{\hat{f}_{u,n}, \hat{f}^\ast_{u,n}, \hat{f}_{L,n}, \hat{f}^\ast_{L,n}\} \) in the definition of \( e_{1,n} \) and \( e_{2,n} \) from (15).

Define \( \hat{e}_{m,n} = [\hat{M}_n - t_0] + t_0, \hat{e}_{m,n} = [\hat{m}_n - t_0] + t_0 \). Assuming that

\[ [\hat{M}_n - t_0] \geq 1, \quad [\hat{m}_n - t_0] \leq -1, \tag{21} \]
(which we later show follows from \( n \geq N_k \), we can write as in (20):

\[
\Omega_n(k) = \sum_{i=0}^{N_k} \left( \hat{a}_{i,1,n}(i) + \hat{a}_{i,1,n}(i) - \hat{e}_{i,1,n}(i) - \hat{e}_{i,2,n}(i) \right) - I_{M,n} = \tilde{E}_{M,n}
\]

\[
+ \sum_{i=0}^{N_k} \left( \hat{a}_{i,2,n}(i) + \hat{a}_{i,2,n}(i) - \hat{e}_{i,1,n}(i) - \hat{e}_{i,2,n}(i) \right) - I_{M,n} = \tilde{E}_{M,n} + (\hat{e}_{i,1,n}(t) + \hat{e}_{i,2,n}(t))
\]

\[
= \hat{P}_{1,n} + \hat{P}_{2,n} + (\hat{e}_{i,1,n}(t) + \hat{e}_{i,2,n}(t)),
\]

\[\text{(22)}\]

where \( \tilde{E}_{M,n}, \tilde{E}_{M,n}, \tilde{I}_{M,n}, \tilde{I}_{M,n} \), the functions \( \hat{a}_{i,1,n}, \hat{a}_{i,1,n} \) (defined on \([t_0, \tilde{M}_n] \cap (Z + t_0)\)), and \( \hat{a}_{i,2,n}, \hat{a}_{i,2,n} \) (defined on \([m_n, t_0] \cap (Z + t_0)\)) are analogous to their \( R_n \) counterparts; these are constructed on \( R_n \cap (k + R_n) \) by substituting for \( f_{u,n} \) for \( f_{u,n} \), \( f_{l,n} \) for \( f_{l,n} \), \( M_n \) for \( M_n \), and \( m_n \) for \( m_n \) in the definition of each item in the collection \{\( E_{M,n}, E_{M,n}, I_{M,n}, I_{M,n}, f_{u,n}, f_{u,n} \)\}.

We now show that

\[
n \geq N_k \implies |\tilde{M}_n - t_0| > 1, \quad |P_{1,n} - \hat{P}_{1,n}| \leq \varepsilon^{-1}14||k||^2\alpha.
\]

\[\text{(23)}\]

We consider two possible cases: (I) \( \text{dim}(M) > 0 \); (II) \( \text{dim}(M) = 0 \). Note that

\[
\{x \in R_n \cap (R_n + k) : (x, \hat{k}) \geq t_0\} = \bigcup_{x \in \tilde{E}_{M,n} \cap \tilde{E}_{M,n}} \left\{ f_{l,n}(c) + \omega||k||^{-1}k : ||k|| \leq \omega \leq ||f_{u,n}(c) - f_{l,n}(c)|| \right\}.
\]

(I). If \( \text{dim}(M) > 0 \), then eventually \( ||f_{u,n}(M_n) - f_{l,n}(M_n)|| > 2||k|| \) or equivalently \( \text{dim}(M) > b_n^{-1}2||k|| \).

Because \( ||t||/b_n < \varepsilon \), we have \( t \in R_n^0 \) and the distance between the hyperplanes \( (x, \hat{k}) = (t, \hat{k}) = t_0 \) and \( (x, \hat{k}) = 0 \) is

\[
||\hat{k}||^{-1}||t_0|| \leq 1 < \varepsilon b_n.
\]

\[\text{(24)}\]

Using \( ||f_{u,n}(0) - f_{l,n}(0)|| \geq 2\varepsilon b_n \geq 8||k||_\infty \), we have

\[
||f_{u,n}(t_0) - f_{l,n}(t_0)|| \geq 2((\varepsilon b_n)^2 - 1)^{1/2} \geq 2\varepsilon b_n - 2 \geq 4||k||_\infty > 2||k||,
\]

or equivalently \( \text{dim}(t_0 / b_n) > b_n^{-1}2||k|| \). By Proposition 2(b) and the continuity of the diameter function \( \text{dim}(-) \) on \([m, M]\) (by Proposition 1), we have that for all \( c \in [t_0 / b_n, M] \):

\[
\text{dim}(c) > b_n^{-1}2||k|| \iff b_n \text{dim}(c) = ||f_{u,n}(bc_n) - f_{l,n}(bc_n)|| > 2||k||.
\]

This implies that \( \tilde{M}_n = M_n \) and so \( (21) \) for \( \tilde{M}_n \) follows from \( (17) \); and also for \( c \in [t_0, M_n] \),

\[
\{x \in R_n \cap (k + R_n) : (x, \hat{k}) = c\} = \left\{ f_{l,n}(c) + \omega||k||^{-1}k : ||k|| \leq \omega \leq ||f_{u,n}(c) - f_{l,n}(c)|| \right\}.
\]
Hence, we have \( f_{u,n}(c) = f_{u,n}(c) \) and \( f_{u,n}(c) = k + f_{u,n}(c) \) with \( \|f_{u,n}(c) - f_{u,n}(c)\| > \|k\| \) (by Proposition 2). It then follows that, by construction, for \( i \in [t_0, M_n] \cap (Z + t_0) \):

\[
a_{\theta,1,n}(i) = a_{\theta,1,n}(i), \quad a_{\lambda,1,n}(i) = a_{\lambda,1,n}(i).
\]

Also, \( I_{M,n} = 0 = I_{M,n} \); and for \( i \in [t_0, M_n] \cap (Z + t_0) \), it follows that

\[
\int_{u,n}(i) = \int_{u,n}(i), \quad \int_{u,n}(i) = \int_{u,n}(i) + k,
\]

and so \( e_{1,n}(i) = e_{1,n}(i), e_{2,n}(i) = e_{2,n}(i), i \in [t_0, M_n] \cap (Z + t_0) \). Thus, for large \( n \geq N_k \),

\[
|P_{1,n} - \tilde{P}_{1,n}| = |E_{M,n} - \tilde{E}_{M,n}|
\]

\[
= (2\|k\|)^{-1}\|k\|(M_n - c_{M,n})
\]

\[
\leq \|k\| \infty.
\]

(II) Assume \( dm(M) = 0 \). As shown in (24), for \( n \geq N_k \), we have \( M > t_0/b_n \) and \( dm(t_0/b_n) > b_n^{-1}\|k\| \).

By these inequalities, Proposition 2, and the continuity of \( dm(\cdot) \) from Proposition 1, there exists a unique \( \bar{c}_n \in (t_0, b_n M) \) such that: \( dm(\bar{c}_n/b_n) = b_n^{-1}\|k\| \) or equivalently \( \|f_{u,n}(\bar{c}_n) - f_{u,n}(\bar{c}_n)\| = \|k\| \);

and for \( c \in [t_0, \bar{c}_n) \), \( dm(c/b_n) > b_n^{-1}\|k\| \) or equivalently \( \|f_{u,n}(c) - f_{u,n}(c)\| > \|k\| \); while for \( c \in (\bar{c}_n, b_n M) \), \( dm(c/b_n) < b_n^{-1}\|k\| \) or equivalently \( \|f_{u,n}(c) - f_{u,n}(c)\| < \|k\| \).

We further note that \( \varepsilon b_n\|k\|^{-1} k, -\varepsilon b_n\|k\|^{-1} k \in R_n^\infty \) and \( f_{u,n}(M_n) \in R_n^\infty \) so that, for each \( \omega \in [0, 1] \), the points

\[
\begin{align*}
\int_{u,n}(M_n) + \omega(\varepsilon b_n\|k\|^{-1} k - f_{u,n}(M_n)) &= \int_{u,n}(M_n) + \omega(\varepsilon\|k\|^{-1} k - f_u(M)), \\
\int_{u,n}(M_n) - \omega(\varepsilon b_n\|k\|^{-1} k + f_{u,n}(M_n)) &= \int_{u,n}(M_n) - \omega(\varepsilon\|k\|^{-1} k + f_u(M))
\end{align*}
\]

are in \( R_n^\infty \) [cf. Kelly and Weiss (1979)]. In particular, we have \( y_{1,n}, y_{2,n} \in R_n^\infty \) for

\[
y_{1,n} = f_{u,n}(M_n) + k - \varepsilon^{-1}\|k\| f_u(M), \quad y_{2,n} = f_{u,n}(M_n) - k + \varepsilon^{-1}\|k\| f_u(M).
\]

Then, \( (y_{1,n}, \bar{k}) = (y_{2,n}, \bar{k}) = M(b_n - \varepsilon^{-1}\|k\|) \) and \( \|y_{1,n} - y_{2,n}\| = 2\|k\| \) by construction. From the convexity of \( R_n^\infty \), \( \omega y_{1,n} + (1 - \omega)y_{2,n} \in R_n^\infty \), \( \omega \in [0, 1] \) and, hence,

\[
dm(M(1 - (\varepsilon b_n)^{-1}\|k\|)) \geq b_n^{-1}\|y_{1,n} - y_{2,n}\| \geq b_n^{-1}2\|k\|.
\]

By this and Proposition 2(b), it must be the case that \( M(b_n - \varepsilon^{-1}\|k\|) < \bar{c}_n < M_n \); then \( n \geq N_k \) (and \( \varepsilon \leq 1/2 \)) implies

\[
M\left(1 - (\varepsilon b_n)^{-1}\|k\|\right) - b_n^{-1}(2 + \|k\|_\infty) > 0
\]
so that $2 + t_0 < M(b_n - e^{-1}||k||) < \bar{\epsilon}_n < b_n M$. We now have

$$\{x \in R_n \cap (k + \bar{R}_n) : (x, \tilde{k}) \geq t_0\} = \bigcup_{c \in [t_0, \tilde{c}_n]} \{f_{u,n}(c) + \omega||k||^{-1}k : ||k|| \leq \omega \leq ||f_{u,n}(c) - f_{l,n}(c)||\}$$

and $\tilde{M}_n = \bar{\epsilon}_n$ with $[\tilde{M}_n - t_0] > 1$ (establishing the relevant part of (21), (23)).

Analogous to the case with $dm(M) > 0$, for $i \in (t_0, \tilde{c}_M, n) \cap (\mathbb{Z} + t_0)$, we have

$$f_{u,n}(i) = f_{u,n}(i) + f_{l,n}(i) = f_{l,n}(i) + k$$
$$a_{u,1,n}(i) = a_{u,1,n}(i) + a_{c,1,n}(i)$$
$$\tilde{e}_{1,n}(i) = e_{1,n}(i) + \tilde{e}_{2,n}(i) = e_{2,n}(i).$$

After cancelling terms common to both sums $P_{1,n}$ and $\tilde{P}_{1,n}$, we may write

$$P_{1,n} - \tilde{P}_{1,n} = \sum_{i=\tilde{c}_M}^{c_{M,n}} (a_{u,1,n}(i) + a_{c,1,n}(i) - e_{1,n}(i) - e_{2,n}(i)) - E_{M,n} - I_{M,n}$$
$$e_{1,n}(i) - e_{2,n}(i) + \tilde{e}_{1,n}(c_{M,n}) + \tilde{e}_{2,n}(c_{M,n}) + \tilde{e}_{1,n}(c_{M,n}) + \tilde{e}_{2,n}(c_{M,n}) + \tilde{E}_{M,n} + \tilde{I}_{M,n}.$$

Because $c_{M,n} - \tilde{c}_{M,n} \leq M - \tilde{c}_n + 1 \leq e^{-1}||k||M + 1, M \leq ||\tilde{k}||_{\infty}$, and $|e_{j,n}(i)| \leq 1/2, i \in [t_0, M_n] \cap (\mathbb{Z} + t_0)$, $j = 1, 2$:

$$\left|\sum_{i=\tilde{c}_M}^{c_{M,n}} e_{1,n}(i) + e_{2,n}(i)\right| \leq e^{-1}8||k||^2_{\infty}$$

(25)

Also,

$$\sum_{i=\tilde{c}_M}^{c_{M,n}} a_{u,1,n}(i) + a_{c,1,n}(i) \leq \left|\{x \in \overline{R}_n : (x, \tilde{k}) \in [\tilde{c}_n, M_n]\}\right|$$
$$\leq ||\tilde{k}||^{-1}(M_n - \tilde{c}_n) \sup\{||f_{u,n}(c) - f_{l,n}(c)|| : c \in [\tilde{c}_n, M_n]\}$$
$$\leq ||\tilde{k}||^{-1}\left(e^{-1}||k||M\right)||k||$$
$$\leq e^{-1}2||k||^2_{\infty}.$$

(26)

By construction, we have

$$0 \leq I_{M,n}, |\tilde{e}_{1,n}(c_{M,n}) + \tilde{e}_{2,n}(c_{M,n}) + \tilde{I}_{M,n} - e_{1,n}(c_{M,n}) - e_{2,n}(c_{M,n})| \leq 1.$$

(27)

In handing the remaining terms in the difference $P_{1,n} - \tilde{P}_{1,n}$, we note the following three cases:

Case 1. If $\tilde{c}_n \in \mathbb{Z} + t_0$, then $\tilde{c}_n = \tilde{c}_{M,n}$. It follows that $\tilde{E}_{M,n} = 0; a_{u,1,n}(\tilde{c}_n) = a_{c,1,n}(\tilde{c}_n) = 0$ because $||f_{u,n}(\tilde{c}_n) - f_{l,n}(\tilde{c}_n)|| = 0$; and $a_{u,1,n}(\tilde{c}_n) + a_{c,1,n}(\tilde{c}_n) \leq e^{-1}2||k||^2_{\infty}$ as in (26). Also, $\tilde{c}_n \leq c_{M,n}$ implies $|E_{M,n}| \leq 1/2 \cdot ||\tilde{k}||^{-1}||k|| \leq ||k||_{\infty}$. 

Case 2. If $\tilde{c}_n \not\in \mathbb{Z} + t_0$ and $\tilde{c}_{M,n} + 1 \leq M_n$, then $\tilde{c}_{M,n} + 1 \leq c_{M,n}$. Here again $|E_{M,n}| \leq ||k||_\infty$ because $\tilde{c}_n \leq c_{M,n}$. Note that $\tilde{c}_{M,n} \leq \tilde{c}_n \leq \tilde{c}_{M,n} + 1$ and, for $c \in [\tilde{c}_{M,n}, \tilde{c}_n]$, 

$$\bar{f}_{L,n}(c) = f_{L,n}(c), \quad \bar{f}_{U,n}(c) = f_{L,n}(c) + k, \quad ||f_{U,n}(\tilde{c}_n) - f_{L,n}(\tilde{c}_n)|| = ||k||.$$ 

Using this, we split the volume, $|\{x \in \mathcal{R}_n : (x, \hat{k}) \in [\tilde{c}_{M,n}, \tilde{c}_{M,n} + 1]\}|$, into a sum of two volumes corresponding to regions separated by the hyperplane, $(x, \hat{k}) = \tilde{c}_n$:

$$||\hat{k}|| \left( a_{U,1,n}(\tilde{c}_{M,n}) + a_{L,1,n}(\tilde{c}_{M,n}) \right) + \frac{1}{2} \sum_{j=0}^{1} ||f_{U,n}(\tilde{c}_{M,n} + j) - f_{L,n}(\tilde{c}_{M,n} + j)||/2$$

$$= ||\hat{k}|| \left| \left\{ x \in \mathcal{R}_n : (x, \hat{k}) \in [\tilde{c}_{M,n}, \tilde{c}_{M,n} + 1] \right\} \right|$$

$$= 1/2 \cdot (\tilde{c}_n - \tilde{c}_{M,n}) \left( ||f_{U,n}(\tilde{c}_{M,n}) - f_{L,n}(\tilde{c}_{M,n})|| + ||f_{U,n}(\tilde{c}_n) - f_{L,n}(\tilde{c}_n)|| \right) + ||\hat{k}|| V_{\tilde{c}_n}$$

$$+ 1/2 \cdot (\tilde{c}_{M,n} - \tilde{c}_n + 1) \left( ||f_{U,n}(\tilde{c}_{M,n} + 1) - f_{L,n}(\tilde{c}_{M,n} + 1)|| + ||f_{U,n}(\tilde{c}_n) - f_{L,n}(\tilde{c}_n)|| \right)$$

$$+ \sum_{Y \in L, U} \int \frac{1}{2} \left[ ||f_{U,n}(\tilde{c}_{M,n} + \omega(\tilde{c} - \tilde{c}_{M,n})) - f_{U,n}(\tilde{c}_n) + \omega(f_{U,n}(\tilde{c}_n) - f_{U,n}(\tilde{c}_{M,n})) \right] ||d\omega$$

$$= 1/2 \cdot (\tilde{c}_n - \tilde{c}_{M,n}) \left( ||f_{U,n}(\tilde{c}_{M,n}) - f_{L,n}(\tilde{c}_{M,n})|| + ||\hat{k}|| \right)$$

$$+ 1/2 \cdot (\tilde{c}_{M,n} - \tilde{c}_n + 1) \left( ||f_{U,n}(\tilde{c}_{M,n} + 1) - f_{L,n}(\tilde{c}_{M,n} + 1)|| + ||\hat{k}|| \right)$$

$$+ ||\hat{k}|| \left( a_{U,1,n}(\tilde{c}_{M,n}) + a_{L,1,n}(\tilde{c}_{M,n}) \right) + ||\hat{k}|| V_{\tilde{c}_n}, \quad (28)$$

Case 3. If $\tilde{c}_n \not\in \mathbb{Z} + t_0$ and $\tilde{c}_{M,n} + 1 > M_n$, then $\tilde{c}_{M,n} = c_{M,n} < \tilde{c}_n < M_n$. Similar to Case 2, we write

$$(2||\hat{k}||)^{-1}(M_n - \tilde{c}_{M,n}) \left( ||f_{U,n}(\tilde{c}_{M,n}) - f_{L,n}(\tilde{c}_{M,n})|| + a_{U,1,n}(\tilde{c}_{M,n}) + a_{L,1,n}(\tilde{c}_{M,n}) \right)$$

$$= \left| \left\{ x \in \mathcal{R}_n : (x, \hat{k}) \in [\tilde{c}_{M,n}, M_n] \right\} \right|$$

$$= (2||\hat{k}||)^{-1}(\tilde{c}_n - \tilde{c}_{M,n}) \left( ||f_{U,n}(\tilde{c}_{M,n}) - f_{L,n}(\tilde{c}_{M,n})|| + ||\hat{k}|| \right)$$

$$+ a_{U,1,n}(\tilde{c}_{M,n}) + a_{L,1,n}(\tilde{c}_{M,n}) + (2||\hat{k}||)^{-1}||\hat{k}|| \left( M_n - \tilde{c}_n \right) + V_{\tilde{c}_n}.$$
\[ ||\hat{\mathbf{k}}|| V_{\alpha}^j = \sum_{Y \in \mathbb{L} U \mathbb{Q}} \int_{Y} \left( f_{Y,n}(\tilde{c}_n + \omega(M_n - \tilde{c}_n)) - \left( f_{Y,n}(\tilde{c}_n) + \omega(f_{Y,n}(M_n) - f_{Y,n}(\tilde{c}_n)) \right) \right) d\omega. \]

Because \( \tilde{E}_{M,n} - E_{M,n} = (2||k||)^{-1}\{ (M_n - \tilde{c}_n) ||f_{M,n}(\tilde{c}_M,n) - f_{L,n}(\tilde{c}_M,n)|| - (\tilde{c}_M,n + 1 - \tilde{c}_n)||k|| \} \) and \( V_{\alpha}^j \leq 2||k||_{\infty} \) (as with \( V_{\alpha}^j \)), we have

\[ |\tilde{E}_{M,n} - E_{M,n} + a_{u,n}(\tilde{c}_M,n) + a_{l,n}(\tilde{c}_M,n) - a_{u,n}(\tilde{c}_M,n) - a_{l,n}(\tilde{c}_M,n)| \]
\[ = (2||k||)^{-1}(1 + \tilde{c}_M,n - M_n)||k|| + V_{\alpha}^j \leq 4||k||_{\infty}. \]

With (25), (26), (27), and Cases 1-3, we establish part (II) and find that (23) holds as well when \( d \mathbf{m}(M) = 0. \)

In showing (23) [through steps (I) and (II)], we also proved: for \( n \geq N_k, \), \( e_{j,n}(t_0) = e_{j,n}(t_0), j = 1,2. \)

By repeating symmetrical arguments on the half space \{ \mathbf{x} \in \mathbb{R}^2 : (\mathbf{x},k) \leq t_0 \}, it follows that for \( n \geq N_k, \), \( [\tilde{m}_n - t_0] \leq -1 \) and \( |P_{2,n} - \tilde{P}_{2,n}| \leq \varepsilon^{-1} ||k||_{\infty}. \) By (20) and (22), the proof of Theorem 1 is then complete. \( \square \)

### 6.2 Proof of Theorem 2

We can assume that it is not the case that \( S_1 \cap S_2 = \overline{S}_j \) for \( j = 1 \) or 2. (If so, the Theorem 1 for \( d = 2 \) already establishes this case.)

Let \( S_{j,n} = b_n S_j, j = 1,2; S_3 = S_1 \cap S_2; \) and \( S_{3,n} = b_n S_3. \) Denote the discrepancy between the volume and \( \mathbb{Z}^2 \) lattice count of \( S_{j,n} \cap (k + S_{j,n}): \)

\[ \Omega_{j,n}(k) = |S_{j,n} \cap (k + S_{j,n})| - \#S_{j,n} \cap (k + S_{j,n}), \quad k \in \mathbb{Z}^2, \]

for \( j = 1,2,3. \) Since each \( S_j, j = 1,2 \) is almost convex with a non-empty interior, it holds that: for each \( k \in \mathbb{Z}^2, \) there exists \( N_k \in \mathbb{Z}, \) \( C > \) such that for \( n \geq N_k, \)

\[ |\Omega_{j,n}(0) - \Omega_{j,n}(k)| \leq C||k||^2_{\infty}, \quad j = 1,2,3. \] (29)

We can justify (29), following the arguments from the proof of Corollary 1. First fix \( j \in \{1,2\}. \) Pick a point \( s_j \in S_j^0 \) and find a closed ball around \( s_j \) in \( S_j^0 \) of radius \( \varepsilon_j \leq 1/2. \) For each \( n, \) there exists \( t_{j,n} \in (-1/2,1/2)^2 \) such that \( t = b_n s_j - t_{j,n} \in \mathbb{Z}^2. \) Define the lattice \( \mathbb{Z}^{2*} = t_{j,n} + \mathbb{Z}^2. \) Then for each \( z \in \mathbb{Z}^2, \) \( \Omega_{j,n}(k) = \Omega_{j,n}^*(k) \) where

\[ \Omega_{j,n}^*(k) = |(S_{j,n} - b_n s_j) \cap (k + S_{j,n} - s_j)| - \#\mathbb{Z}^{2*} \cap (S_{j,n} - b_n s_j) \cap (k + S_{j,n} - b_n s_j). \]
Since $S_{j,n} - b_n s_j$ expands around the origin ($0$ is an interior point), the same arguments in the proof of Theorem 1 apply to bounding $|\Omega_{i,n}(k)|$ for large $n$ guaranteeing that $\varepsilon_i^2 b_n \geq 4||k||_\infty$. Hence, (29) follows for $j = 1, 2$.

The result in (29) also applies to $\Omega_{3,n}(k)$ if the intersection $S_3$ has a nonempty interior. Then the set $S_3$ is also almost convex, because $S_1^i \cap S_2^i \neq \emptyset$ and $S_1 \cap S_2$ are convex. If $S_1^i \cap S_2^i = \emptyset$, then it must be the case that $S_1 \cap S_2 = \{v\}$, $v \in \mathbb{R}^2$. (This follows from the fact that $S_1^i, S_2^i$ are (bounded) disjoint open convex sets, so that by the Separating Hyperplane Theorem [cf. Kelly and Weiss (1979)], there exists $b \neq 0 \in \mathbb{R}^2$, $c \in \mathbb{R}$ such that

$$\langle x, b \rangle > c \quad \text{if} \quad x \in S_1^i, \quad \langle x, b \rangle < c \quad \text{if} \quad x \in S_2^i.$$  

Thus, $S_1 \cap S_2 \cap \{x \in \mathbb{R}^2 : \langle x, b \rangle = c\} = \{v\}$, because $\partial S_1 \cap \partial S_2$ is finite. Hence, if $S_1^i \cap S_2^i = \emptyset$, then $|\Omega_{3,n}(k)| \leq 2$ for $k \neq 0$ so that (29) holds trivially.

We make note, in the following, of some useful partitions of counts and volumes. For $k \neq 0$, write:

\[
\begin{align*}
\#\mathbb{Z}^2 \cap R_n \cap (k + R_n) &= \sum_{j=1}^{2} \#\mathbb{Z}^2 \cap S_{j,n} \cap (k + S_{j,n}) - \#\mathbb{Z}^2 \cap S_{3,n} \cap (k + S_{3,n}) + \Phi_n(k), \\
|R_n \cap (k + R_n)| &= \sum_{j=1}^{2} |S_{j,n} \cap (k + S_{j,n})| - |S_{3,n} \cap (k + S_{3,n})| + \Phi_n(k), \\
\Phi_n(k) &= \sum_{j=1}^{2} \#\mathbb{Z}^2 \cap (S_{3-j,n} \setminus S_{j,n}) \cap (k + S_{j,n} \setminus S_{3-j,n}), \\
\Phi_n(k) &= \sum_{j=1}^{2} |(S_{3-j,n} \setminus S_{j,n}) \cap (k + S_{j,n} \setminus S_{3-j,n})|.
\end{align*}
\]

Then we can write

\[
\Omega_{0,n}(0) - \Omega_{0,n}(k) = \sum_{j=1}^{2} \left( \Omega_{j,n}(0) - \Omega_{j,n}(k) \right) - \left( \Omega_{3,n}(0) - \Omega_{3,n}(k) \right) - \Phi_n(k) - \Phi_n(k).
\]

By (29) and (30), it suffices to show: for each $k \neq 0 \in \mathbb{Z}^2$, there exist $N_k \in \mathbb{Z}_+$ and $C > 0$ such that for $n \geq N_k$,

\[
\Phi_n(k), \Phi_n(k) \leq C \cdot ||k||^2,
\]

to complete the proof of Theorem 2. To show (31), we will consider two cases: (I) $S_1^i \cap S_2^i \neq \emptyset$; or (II) $S_1^i \cap S_2^i = \emptyset$.

Some additional notation is necessary. We write $\text{ray}(x, y)$, $\text{seg}(x, y)$, and $\text{segdist}(x, y)$ for the ray, open,
and closed line segment between the points \( x, y \in \mathbb{R}^2 \); that is,

\[
\text{ray}(x, y) = \{ x + \omega(y - x) \in \mathbb{R}^2 : 0 \leq \omega \},
\]

\[
\text{seg}(x, y) = \{ x + \omega(y - x) \in \mathbb{R}^2 : 0 < \omega < 1 \},
\]

\[
\overline{\text{seg}}(x, y) = \text{seg}(x, y) \cup \{x, y\}.
\]

We will also often use the following properties of a convex set \( A \) (cf. Kelly and Weiss, 1979, p. 99):

- If \( x \in A^o \) and \( y \in A \), then \( \text{seg}(x, y) \subseteq A^o \); the border of \( A \), \( \partial A = A \setminus A^o \), is a simple closed curve;
- If \( x \in A^o \) and \( y \neq x \in \mathbb{R}^2 \), then \( \text{ray}(x, y) \) intersects \( \partial A \) exactly once.

(I). Suppose that \( S_j^o \setminus \overline{S}_{3-j} \neq \emptyset \) for \( j = 1, 2 \). By assumption, we then have common border points \( \partial S_1 \cap \partial S_2 = \{v_1, \ldots, v_m\}, m \geq 2 \), where each \( v_i \) is distinct. The fact that \( m \geq 2 \) is intuitively simple, but we will outline the proof of this in the following parenthetical argument.

(Note that there exists \( y_0 \in S_j^o \cap S_k^o, y_j \in S_j^o \setminus \overline{S}_{3-j} \) for \( j = 1, 2 \). Then, for each \( j = 1, 2 \), \( \text{ray}(y_0, y_j) \) intersects \( \partial S_1 \) and \( \partial S_2 \) exactly once by convexity of the sets. Because \( y_1 \notin \mathbb{R}^2 \), \( \text{ray}(y_0, y_1) \) must intersect \( \partial S_1 \) at a point, say \( y_1^* \), on \( \text{seg}(y_0, y_1) \); this implies further that \( y_1^* \in S_1^o \cap \partial S_2 \). Also, \( \text{ray}(y_0, y_2) \) intersects \( \partial S_2 \) at the point \( y_2^* \) somewhere "beyond" \( y_2 \) (and the border of \( S_1 \)) on the same ray so that \( y_2^* \in \partial S_2 \setminus \overline{S}_{1} \). Note both \( \partial S_1 \) and \( \partial S_2 \) are simple closed curves. If an interior point of \( S_2^o \) lies on \( \text{seg}(y_1^*, y_2^*) \), the line between \( y_1^* \) and \( y_2^* \) separates \( \partial S_2 \) into two parts, each lying on an opposite side (open half-space) of this line; then, by the Jordan Curve Theorem, \( \partial S_2 \) must intersect/cross \( \partial S_1 \) at least once on each open side of \( \partial S_2 \). That is, the border \( \partial S_2 \) curves away from \( y_1^* \) on each side of the line between \( y_1^* \) and \( y_2^* \), winding toward \( y_2^* \) from two different directions; and on the same line, by construction and convexity, \( \partial S_1 \) and \( \partial S_2 \) cannot intersect. On the other hand, if \( \overline{\text{seg}}(y_1^*, y_2^*) \subseteq \partial S_2 \), then by the Jordan Curve Theorem again: \( \overline{\text{seg}}(y_1^*, y_2^*) \cap \partial S_1 \neq \emptyset \) and \( (\partial S_2 \setminus \overline{\text{seg}}(y_1^*, y_2^*)) \cap \partial S_1 \neq \emptyset \). Hence, we have that \( 2 \leq m < \infty \).)

Assume that \( m = 2 \) for the moment. We can express the line defined by \( v_1 \) and \( v_2 \) as a hyperplane involving \( b \in \mathbb{R}^2 \):

\[
\left\{ x \in \mathbb{R}^2 : \langle x, b \rangle = \langle v_1, b \rangle \right\}, \quad b \neq 0 \in \mathbb{R}^2.
\]

By convexity and \( \partial S_1 \cap \partial S_2 = \{v_1, v_2\} \), one of the following must be true: \( \text{seg}(v_1, v_2) = S_1^o \cap \partial S_2 \), \( \text{seg}(v_1, v_2) = S_2^o \cap \partial S_1 \), or \( \text{seg}(v_1, v_2) \subseteq S_1^o \cap S_2^o \). In any case, we have for some \( j^* \in \{1, 2\} \):

\[
\text{seg}(v_1, v_2) \cap (S_j^o \cap \partial S_{3-j^*}) = \emptyset.
\]

By convexity, this implies that \( S_{3-j^*} \setminus \overline{S}_{j^*} \) lies in one open half-space determined by \( \langle x, b \rangle = \langle v_1, b \rangle \) and \( \overline{S}_{j^*} \setminus \overline{S}_{3-j^*} \) is a subset of the opposite (possibly closed) half-space.
Fix $k \neq 0 \in \mathbb{Z}^2$ and define $\hat{k}$ as in (35). If $(v_2 - v_1, k) = 0$, we may take $b = \hat{k}$. In which case, we easily have (31): for all $n:$

$$\Phi_n(k) = 0, \quad \Psi_n(k) \leq \# \{ b_n v_1, b_n v_2 \} \leq 2$$

because both sets $\mathcal{S}_{1,n} \setminus \mathcal{S}_{2,n}$ and $\mathcal{S}_{2,n} \setminus \mathcal{S}_{1,n}$ are parallel as they "slide" upon translation by $k$.

Suppose $(v_2 - v_1, k) \neq 0$. Let $\hat{m} = 1/2 \cdot (v_1 + v_2)$. Then there exists a $\rho \neq 0 \in \mathbb{R}$ and $m_k \in \mathbb{R}^2$ such that

$$m_k = \hat{m} + \rho \|k\|^{-1}k \in \mathcal{S}_1 \cap \mathcal{S}_2.$$

The above follows from the fact that the diameter of $\{ \mathcal{S}_i \cap \mathcal{S}_j : (x,k) = (v,k) \}$ cannot be zero by Propositions 1 and 2, because $(\hat{m},k)$ lies strictly between $(v_1, \hat{k})$ and $(v_2, \hat{k})$. Then, we can make a closed ball of radius $0 < \eta < \rho$ around $m_k$, say $B(m_k, \eta)$, such that $B(m_k, \eta) \subset \mathcal{S}_1 \cap \mathcal{S}_2$.

We will exploit the mentioned properties of convex sets to build $V$-shaped "cones," originating from both $v_1$ and $v_2$, that lie outside $\mathcal{S}_1 \cup \mathcal{S}_2$; we use these cones to quantify the number of $\mathbb{Z}^2$ points in $\mathcal{S}_{j,n} \setminus \mathcal{S}_{3-j,n}$ that move over to $\mathcal{S}_{3-j,n} \setminus \mathcal{S}_{j,n}$ when translated by $k$ (for each $j = 1, 2$).

Note, for $j = 1, 2$,

$$\text{seg}(v_j, m_k + \omega \|k\|^{-1}k) \subset \mathcal{S}_1 \cap \mathcal{S}_2, \quad \omega \in [-\eta, \eta]$$

and for the "cones": $\omega \in [-\eta, \eta], \beta > 0,

$$v_j + \beta (v_j - m_k - \omega \|k\|^{-1}k) \not\in \overline{R}_1 \cup \overline{R}_2.$$ 

This implies, for $\omega \in [-\eta, \eta], j \in \{1, 2\},

$$b_n v_j + \beta (m_k + \omega \|k\|^{-1}k - v_j) \in \mathcal{S}_{2,n} = b_n (\mathcal{S}_1 \cap \mathcal{S}_2), \quad 0 < \beta < b_n$$

$$b_n v_j + \beta (v_j - m_k - \omega \|k\|^{-1}k) \not\in \mathcal{S}_{1,n} \cup \mathcal{S}_{2,n} = b_n (\mathcal{S}_1 \cup \mathcal{S}_2), \quad \beta > 0. \quad (32)$$

Suppose $b_n \eta > \|k\|$ and $w_n \in (\mathcal{S}_{j,n} \setminus \mathcal{S}_{3-j,n}) \cap (k + (\mathcal{S}_{3-j,n} \setminus \mathcal{S}_j))$ for some selected $j \in \{1, 2\}$. Then, $w_n$ and $w_n - k$ lie in opposite (not necessarily open, but complementary) half-spaces defined by the line or hyperplane $(x, b) = b_n (v_1, b)$, and also

$$(w_n, \hat{k}) = (w_n - k, \hat{k}) = (\tilde{w}_n, \hat{k}); \quad \tilde{w}_n = b_n \hat{m} + c_n (v_2 - v_1), \quad c_n = \frac{(w_n - b_n \hat{m}, \hat{k})}{(v_2 - v_1, \hat{k})}.$$

Because $\tilde{w}_n$ lies on the line $(x, b) = b_n (v_1, b)$ (by definition), we must have $\|w_n - \tilde{w}_n\| \leq \|k\|$. Furthermore, it must be the case that

$$2 |c_n| \not\in [0, b_n \eta^{-1} \|k\|].$$
which we briefly justify with a proof by contradiction. A little algebra gives:

\[ \mathbf{w}_n = b_n \nu' + (\mathbf{m}_n - \nu')(b_n - 2|c_n|), \quad j' = \begin{cases} 2 & \text{if } c_n > 0 \\ 1 & \text{if } c_n < 0 \end{cases} \]

Let

\[ \nu_{c,n} = b_n \nu' + (b_n - 2|c_n|)(\mathbf{m}_n - \eta\|\mathbf{k}\|^{-1}\mathbf{k} - \nu'), \]

\[ \nu_{u,n} = b_n \nu' + (b_n - 2|c_n|)(\mathbf{m}_n + \eta\|\mathbf{k}\|^{-1}\mathbf{k} - \nu'). \]

Then, we have \( \nu_{c,n}, \nu_{u,n}, \mathbf{w}_n, \mathbf{w}_n - \mathbf{k} \) all lie on the same line with \( \nu_{c,n}, \nu_{u,n} \in b_n(S_1 \cap S_2) \) by (32) and \( \|\nu_{c,n} - \nu_{u,n}\| \geq 2\|\mathbf{k}\| \) by construction. The points \( \mathbf{w}_n, \mathbf{w}_n - \mathbf{k} \) cannot lie between \( \nu_{c,n} \) and \( \nu_{u,n} \) (because \( \mathbf{w}_n, \mathbf{w}_n - \mathbf{k} \notin b_n(S_1 \cap S_2) \)) and \( \nu_{c,n}, \nu_{u,n} \) cannot lie between \( \mathbf{w}_n \) and \( \mathbf{w}_n - \mathbf{k} \). Hence, \( \mathbf{w}_n, \mathbf{w}_n - \mathbf{k} \) lie on either \( \text{ray} (\nu_{c,n}, \nu_{u,n} - \mathbf{k}) \) or \( \text{ray} (\nu_{u,n}, \nu_{u,n} + \mathbf{k}) \), implying further that \( \mathbf{w}_n, \mathbf{w}_n - \mathbf{k} \in S_1,n \) or \( \mathbf{w}_n, \mathbf{w}_n - \mathbf{k} \in S_2,n \), a contradiction. We have now established (33).

We can show also that

\[ 2|c_n| \notin [b_n + \eta^{-1}\|\mathbf{k}\|, \infty], \quad (34) \]

as we will explain in the following. Assume (34) is false. Defining \( \nu_{c,n}, \nu_{u,n} \) as above, we have again the points \( \nu_{c,n}, \nu_{c,n}, \mathbf{w}_n, \mathbf{w}_n - \mathbf{k}, \mathbf{w}_n \) fall on the same line, \( \|\nu_{c,n} - \nu_{u,n}\| \geq 2\|\mathbf{k}\| \), but now \( \text{seg}(\nu_{c,n}, \nu_{u,n}) \subset b_n(S_1 \cup S_2)^c \) by (32). Because \( (\nu_2 - \nu_1, \mathbf{k}) \neq 0 \) and \( (\nu_2 - \nu_1, b) = 0 \), we must have \( (b, \mathbf{k}) \neq 0 \) and so

\[ (\nu_{c,n} - b_n \nu_1, b) = (\rho - \eta)(b_n - 2|c_n|)(b_n, \mathbf{k})/\|\mathbf{k}\|, \]

\[ (\nu_{u,n} - b_n \nu_1, b) = (\rho + \eta)(b_n - 2|c_n|)(b_n, \mathbf{k})/\|\mathbf{k}\|, \]

have the same, nonzero sign \( (\eta < |\rho|) \). Hence, \( \nu_{c,n}, \nu_{u,n} \) both lie in one open half-space or side of the hyperplane \( (\mathbf{x}, b) = b_n(\nu_1, b) \). Furthermore,

\[ b_n^{-1}\mathbf{w}_n \in (\partial S_1 \setminus S_2) \cup (\partial S_2 \setminus S_1) \cup (S_1 \cup S_2)^c. \]

[This follows from the fact that because \( b_n^{-1}\mathbf{w}_n \notin \overline{\text{seg}}(\nu_1, \nu_2) \): if \( b_n^{-1}\mathbf{w}_n \in S_j^o \) for some \( j \in \{1, 2\} \), then \( \text{ray}(b_n^{-1}\mathbf{w}_n, \nu_1) \) intersects \( \partial S_j \) at \( \nu_1 \) and \( \nu_2 \), a contradiction of a basic property of convex sets; if \( b_n^{-1}\mathbf{w}_n \in \partial S_j \) for \( j = 1 \) or \( 2 \), then the closed line segment including the points \( b_n^{-1}\mathbf{w}_n, \nu_1, \nu_2 \) is a subset of \( \partial S_j \), implying \( b_n^{-1}\mathbf{w}_n \notin \partial S_{3-j} \), else \( \partial S_1 \cap \partial S_2 \) would be an uncountable set.] We can now argue that the closed line segment containing \( \nu_{c,n}, \nu_{u,n}, \mathbf{w}_n \) must be a subset of \( (S_1,n)^c \) or \( (S_2,n)^c \). (In brief, the points \( \nu_{u,n} \) (or \( \nu_{c,n} \)) and \( b_n \mathbf{m}_k \) lie in opposite open half-spaces defined by \( (\mathbf{x}, b) = b_n(\nu_1, b) \), while \( \mathbf{w}_n \) lies on this line. Also, \( \text{seg}(\mathbf{m}_k, \mathbf{m}_k) \subset S_1 \cap S_2 \) so that

\[ \text{ray} (\gamma \mathbf{m}_k + (1 - \gamma) \mathbf{m}_k, \nu_j) \setminus \overline{\text{seg}} (\gamma \mathbf{m}_k + (1 - \gamma) \mathbf{m}_k, \nu_j) \subset (S_1 \cup S_2)^c, \quad \gamma \in (0, 1), \ j \in \{1, 2\} \]
from convexity. Hence, we can conclude that all the rays of the form
\[ \text{ray}(\lambda_0 (v_
u + i), \gamma v + (1 - \gamma) \tilde{w}_n), \quad \gamma \in (0, 1), \quad v = v_{L,n}, v_{U,n} \]
are not in \( S_{j,n} \), \( j = 1 \) or \( 2 \). Now we have that \( w_n \) and \( w_n - k \), which lie on complementary half-spaces created by \( (x, b) = b_n(v_1, b) \), must be separated by the open line segment containing \( v_{L,n}, v_{U,n}, \tilde{w}_n \), as all these points \( \{v_{L,n}, v_{U,n}, \tilde{w}_n, w_n, w_n - k\} \) fall on a common line. This implies a contradiction, because again \( ||v_{L,n} - v_{U,n}|| \geq 2||k|| \). We have now established (34).

Essentially, by (33) and (34), we can bound the length of the section on \( (x, b) = b_n(v_1, b) \) at which a point from \( S_{j,n} \setminus S_{3-j,n} \) can cross over to \( S_{j,n} \setminus S_{3-j,n} \) when translated by \( k \). For an arbitrary \( w_n \in (S_{j,n} \setminus S_{3-j,n}) \cap (k + (S_{2-j,n} \setminus S_{j,n})), j \in \{1, 2\} \), we can pull together (13), (33), (34), and the fact that \( ||w_n - w_n|| \leq ||k|| \) to show: for \( b_n \eta > ||k|| \),
\[
\Phi_n(k) \leq 4||k|| \cdot \frac{2n^{-1}||k|| \cdot ||v_2 - v_1, \hat{k}||}{||k||} \leq 16n^{-1}||k||^2, \\
\Psi_n(k) \leq 4\eta^{-1}||k|| \cdot ||(v_2 - v_1, \hat{k})|| + 2(||\hat{k}|| - 1||k|| + 1) \leq 48\eta^{-1}||k||^2.
\]
This proves (31) for Case (I) when \( \partial S_1 \cap \partial S_2 = \{v_1, v_2\} \) (ie. \( m = 2 \)).

Suppose Case (I) holds and \( \partial S_1 \cap \partial S_2 = \{v_1, \ldots, v_m\} \), \( m > 2 \). The border of \( S_1 \cap S_2 \) can be expressed as a continuous simple closed curve \( C(\theta), \theta \in [0, 2\pi] \) and WLOG we may assume there exists \( \theta_1 < \ldots < \theta_m \in [0, 2\pi] \) such that \( v_j = C(\theta_j) \) for \( j = 1, \ldots, m \). We can then create hyperplanes/lines between consecutive pairs of elements in \( \partial S_1 \cap \partial S_2 \):
\[ v_j + \omega(v_j - v_{j+1}), \quad \omega \in \mathbb{R}. \]
Note that the closed line segments \( \overline{seg}(v_j, v_{j+1}), j = 1, \ldots, m \), form sides of a polygon in \( \overline{S}_1 \cap \overline{S}_2 \). If we pick and fix \( w_n \in (S_{j,n} \setminus S_{3-j,n}) \cap (k + (S_{2-j,n} \setminus S_{j,n})), \) for some \( k \neq 0 \in \mathbb{Z}^2, j \in \{1, 2\} \), then there exists \( q \in \{1, \ldots, m\} \) such that the line between \( b_n v_q \) and \( b_n v_{q+1} \) separates \( w_n \) and \( w_n - k \). That is, \( w_n \) and \( w_n - k \) must lie in opposite (complementary) half-spaces defined by the line between \( b_n v_q \) and \( b_n v_{q+1} \) (possibly one point among \( w_n, w_n - k \) may fall on this line). Thus, when \( \partial S_1 \cap \partial S_2 \) is finite, we can control the sizes of \( \Phi_n(k) \) and \( \Psi_n(k) \) as in the proof of Case (I) for \( m = 2 \) and establish (31).

(II). For sets \( T_1, T_2 \subset \mathbb{R}^2 \), define the set distance \( \text{dis}(E_1, E_2) = \inf\{||x_1 - x_2|| : x_i \in E_i\} \).

If \( \overline{S}_1 \cap \overline{S}_2 = \emptyset \), then \( \text{dis}(\overline{S}_1, \overline{S}_2) = \delta > 0 \). If \( k \neq 0 \) and \( b_n \delta > ||k|| \), then for each \( j = 1, 2 \) we have:
\[ \text{dis}(\overline{S}_{j,n}, k + \overline{S}_{3-j,n}) > 0 \implies \Phi_n(k) = \Psi_n(k) = 0, \]
and so (31) follows easily.

If $\mathbb{R}_1 \cap \mathbb{R}_2 \neq \emptyset$, then it must be the case that

$$\mathbb{S}_1 \cap \mathbb{S}_2 = \{x \in \mathbb{S}_1 \cap \mathbb{S}_2 : (x, b) = c\} = \{v\},$$

for some $b \neq 0, v \in \mathbb{R}^2, c \in \mathbb{R}$. Let $v^* \neq v \in \mathbb{R}^2, (v^*, b) = c$.

If $k \neq 0 \in \mathbb{Z}^2$ and $(v - v^*, k) = 0$, then we can pick $b = k$ and obtain that

$$\Phi_n(k) = 0, \quad \Psi_n(k) \leq 1,$$

establishing (31).

If $(v - v^*, k) \neq 0$, then define the set

$$H(\beta) \equiv \{x \in \mathbb{R}^2 : (x, k) = (v, k) + \beta(v - v^*, k)\} \quad \beta \in \mathbb{R},$$

and the function $h : \mathbb{R} \rightarrow \mathbb{R}^+ \cup \{\infty\}$ as

$$h(\beta) = \begin{cases} 
\text{dis}(\mathbb{S}_1 \cap H(\beta), \mathbb{S}_2 \cap H(\beta)) & \text{if } \mathbb{S}_j \cap H(\beta) \neq \emptyset \text{ for } j = 1, 2 \\
\infty & \text{otherwise.}
\end{cases}$$

We sketch the remainder of the proof for this situation, which relies on elementary properties of convex sets and is very similar in spirit to the proof of Case (I), $m = 2$. One can show that $h(\beta)$ is positive and increasing on $(0, \infty)$ (and positive and decreasing on $(-\infty, 0)$) by the convexity of $\mathbb{S}_1, \mathbb{S}_2$. If there exists a $\beta \in (0, \infty)$ such that $h(\beta) < \infty$, then we can choose $y_1, y_2 \in H(\beta), y_1 \neq y_2, (y_1, b) = c$, such that $y_1, y_2$ lie on a line segment "between" $\mathbb{S}_1$ and $\mathbb{S}_2$ and $\text{seg}(v, y_i) \subset (\mathbb{S}_1 \cup \mathbb{S}_2)^c$ for each $i = 1, 2$.

We can do the same if there exists a $\beta \in (-\infty, 0)$ such that $h(\beta) < \infty$. We can then use arguments similar to the proof of Case (I), $m = 2$ (although the arguments are geometrically simpler here): the line $(x, b) = b_n c$ separates the bodies $S_{j,n} \setminus \{b_n v\}$ and $S_{3-j,n} \setminus \{b_n v\}$ ($b_n v$ is on the line); we find the segment of this line (including the point $b_n v$) which must be crossed for a point in $S_{j,n} \setminus S_{3-j,n}$ to be in $(k + (S_{3-j,n} \setminus S_{j,n}))$ upon translation by $k$; and we can eventually bound the length of this line segment using the fact that the distance $b_n ||y_1 - y_2||$ will be large (or equivalently bound those $\beta \in \mathbb{R}$, where $h(\beta) \leq ||k||/b_n$). We then obtain that for $k \neq 0 \in \mathbb{Z}^2$ and large enough $n$

$$\max\{\Phi_n(k), \Psi_n(k)\} \leq C \cdot ||k||^2_\infty$$

establishing (31). The proof of Theorem 2 is now complete. $\square$

6.3 Proof of Corollary 1

Fix $i \neq 0 \in \mathbb{Z}^2$. (The $i = 0$ case follows directly from Theorems 1 and 2.) The set $b_n(i + R_0)$ grows outward in all directions and "slides" through $\mathbb{R}^2$ space as $n$ increases. However, for each $n$, there exists
$t_{i,n} \in (-1/2, 1/2]^2$ such that $t - b_n i - t_{i,n} \in \mathbb{Z}^2$. The $\mathbb{Z}^2 = \mathbb{t} + \mathbb{Z}^2$ lattice points in $R_{i,n} = b_n (t + R_0)$ can be placed in a one-to-one correspondence with the $\mathbb{Z}^2\ast = t_{i,n} + \mathbb{Z}^2$ lattice points in $R_n = b_n R_0$; i.e. $s \in \mathbb{Z}^2 \cap R_{i,n}$ if and only if $s - b_n i \in \mathbb{Z}^2\ast \cap R_n$. That is, for each $k \in \mathbb{Z}^2$,

$$\# \mathbb{Z}^2 \cap R_{i,n} \cap (k + R_{i,n}) = \# \mathbb{Z}^2\ast \cap R_n \cap (k + R_n).$$

Since we estimate each count $\# \mathbb{Z}^2 \cap R_{i,n} \cap (k + R_{i,n})$ by the volume $|R_n \cap (k + R_n)|$, the result follows by applying the same arguments from the proof of Theorem 1, or Theorem 2, (with the same $n_0$) to the $\mathbb{Z}^2\ast$ lattice and the region $R_n$. □

7 Proof of Theorem 3

From the arguments establishing Corollary 1 (applied to $d = 3$), we need only consider the case that $i = 0 \in \mathbb{Z}^3$, i.e. $R_{i,n} = R_n$. Fix $k = (k_1, k_2, k_3) \neq 0 \in \mathbb{Z}^3$. WLOG assume that $R_0$ contains the closed ball of radius $\varepsilon \leq 1/2$ around the origin (say $\overline{B}(0, \varepsilon) \subset R_0$) and $n$ is large enough so that $\varepsilon^2 k_1^{1/3} > 4||k||_\infty$.

To facilitate counting, we first determine a systematic method for describing the $\mathbb{Z}^3$ lattice points within $R_n$ as a function of the selected $k$. Since $k \neq 0$, say WLOG component $k_1 \neq 0$. Define $\hat{k}, \hat{k} \in \mathbb{Z}^3$ as:

$$\hat{k} = \hat{g}^{-1} k \equiv (\hat{k}_1, \hat{k}_2, \hat{k}_3)^t, \quad \hat{k} = \hat{g}^{-1}(\hat{k}_2, -\hat{k}_1, 0)^t;$$

$$\hat{g} = \max\{|i|, |j|, |k|, |j|/i \in \mathbb{Z}_+, j = 1, 2, 3\}, \quad \hat{g} = \max\{|i|, |j|/i \in \mathbb{Z}_+, j = 1, 2\}.$$

The divisors $\hat{g}$ and $\hat{g}$ are each essentially the greatest common divisor of a set of nonnegative integers. As in (12), there exists $w_1, w_2 \in \mathbb{Z}$ such that, for $w^* = (w_1, w_2, 0)^t \in \mathbb{Z}^3$, we have $(w^*, \hat{k}) = 1$. Let $v^* = (\hat{g}||k||^2)^{-1}(-\hat{k}_1 \hat{k}_3, -\hat{k}_2 \hat{k}_3, \hat{g}||k||^2)^t \in \mathbb{R}^3$.

We gather some useful properties of $w^*$ and $v^*$ in the following proposition; the proof is provided in Section 8.

**Proposition 4** Let $i, j \in \mathbb{Z}^3$.

(a) $(i, \hat{k}) = (j, \hat{k})$, $(i, v^*) = (j, v^*)$ if and only if $i - j = m \hat{k}$ for some $m \in \mathbb{Z}$.

(b) $(i, \hat{k})$, $(i - (i, \hat{k}) \cdot w^*, v^*) \in \mathbb{Z}$.

(c) For $c_1, c_2 \in \mathbb{Z}$, there exists $i \in \mathbb{Z}^3$ such that $(i, \hat{k}) = c_1$ and $(i - c_1 w^*, v^*) = c_2$.

Let $v_m, v_* \in \overline{R_0}$ such that

$$(v_m, \hat{k}) = \min_{x \in \overline{R_0}} \langle x, \hat{k} \rangle \equiv m \leq -\varepsilon, \quad (v_*, \hat{k}) = \max_{x \in \overline{R_0}} \langle x, \hat{k} \rangle \equiv M \geq \varepsilon,$$
using $B(0, e) \subset R^n_0$. We need to define some functions for counting purposes. For $c \in [m, M]$, let

$$h(c) = \inf\{(x - cw^*, v^*) : x \in R^n_0, (x, \hat{k}) = c\}$$

$$H(c) = \sup\{(x - cw^*, v^*) : x \in R^n_0, (x, \hat{k}) = c\}$$

For $y \in [h(c), H(c)]$, define

$$D^*(c, y) = \sup\{||x_i - x_j|| : x_j \in R^n_0, (x_j, \hat{k}) = c, (x_j - cw^*, v^*) = y, j = 1, 2\},$$

a collection of "diameters" with respect to (convex) slices of $R^n_0$ (created by an intersecting plane). For $c \in [m, M]$, define

$$D(c) = \sup\left\{D^*(c, y) : y \in [h(c), H(c)]\right\}$$

to denote the maximum of each collection of set diameters. Using convexity of $R^n_0$, it can be shown that $D(\cdot), H(\cdot)$, and $-h(\cdot)$ are concave and continuous on $[m, M]$ (similar to the proof of Proposition 1); likewise, $D^*(c, \cdot)$ is concave and continuous on $[h(c), H(c)]$, $c \in [m, M]$.

Let $M_n = b_n M$ and $m_n = b_n m$. We define now "scaled-up" versions of the functions $D(\cdot), H(\cdot), h(\cdot)$ for $R_n$. Create functions $D_n, H_n, h_n : [m_n, M_n] \rightarrow R$:

$$D_n(c) = b_n D(c/b_n), \quad H_n(c) = b_n H(c/b_n), \quad h_n(c) = b_n h(c/b_n), \quad c \in [m_n, M_n].$$

Define $t_0 = (t, \hat{k})$. We can then write

$$\#Z^3 \cap R_n = \#Z^3 \cap (R_n - t)$$

$$= \sum_{i \in Z \cap ([m_n, M_n] - t_0)} \sum_{j = -\infty}^{\infty} \#Z^3 \cap \{x \in R_n - t : (x, \hat{k}) = i, (x - iw^*, v^*) = j\}$$

$$= \sum_{i \in Z \cap ([m_n, M_n] - t_0)} \sum_{j \in S_{i,n}} \#Z^3 \cap \{x \in R_n - t : (x, \hat{k}) = i, (x - iw^*, v^*) = j\};$$

$$S_{i,n} = \{(x - iw^*, v^*) : x \in R_n - t, (x, \hat{k}) = i\}, \quad i \in Z \cap ([m_n, M_n] - t_0)$$

$$= [h_n(i + t_0), H_n(i + t_0)] + t_0 \cdot (w^*, v^*) - (t, v^*).$$

We will restrict our attention to hyperplanes intersecting $R_n - t$ which have a "significant" number if integer points. Let

$$\Upsilon^+_n = \max\{c \in Z \cap ([m_n, M_n] - t_0) : H_n(c + t_0) - h_n(c + t_0) \geq 2\},$$

$$\Upsilon^-_n = \min\{c \in Z \cap ([m_n, M_n] - t_0) : H_n(c + t_0) - h_n(c + t_0) \geq 2\},$$

$$\Gamma^+_n = \sup\{c \in [m_n, M_n] - t_0 : D_n(c + t_0) \geq 2b_n^{1/3}\},$$

$$\Gamma^-_n = \inf\{c \in [m_n, M_n] - t_0 : D_n(c + t_0) \geq 2b_n^{1/3}\}. $$
Note that either $T_n^+ + t_0 = M_n$, or $T_n^- + t_0 < M_n$ and $D_n(T_n^- + t_0) = 2b_{n/3}^1$, by the continuity of $D(\cdot)$ and the definition of $D_n(\cdot)$; likewise, either $T_n^- + t_0 = m_n$, or $T_n^- + t_0 > m_n$ and $D_n(T_n^- + t_0) = 2b_{n/3}^1$.

Let

$$C_n^- = \max\{T_n^-, \lfloor T_n^- \rfloor\}, \quad C_n^+ = \min\{T_n^+, \lceil T_n^+ \rceil\}.$$ 

We make a few comments to help frame the relative positions of $C_n^-$, $C_n^+$ in $\mathbb{R}$, using convex geometry and the assumption that $\epsilon^2 b_{n/3}^1 > 4\|k\|_{\infty}$. For $\tilde{k} = (\tilde{k}_1, \tilde{k}_2, 0)'$, note that $\|\tilde{k}\|^{-1} \omega \tilde{k} \cos(\theta) + \omega(0, 0, \sin(\theta))' \in R_3^0$ for each $\omega \in [0, \epsilon], \theta \in [0, 2\pi]$; this implies that

$$b_n v + \beta \left(\|\tilde{k}\|^{-1} \omega \tilde{k} \cos(\theta) + \omega(0, 0, \sin(\theta))' - v\right) \in R_n^0, \quad \omega \in [0, \epsilon], \theta \in [0, 2\pi], \beta \in (0, b_n]$$

with $v = v_m, v_m'$. It follows then that: $H_n(c) - h_n(c) \geq 2$ if $m(b_n - \epsilon^{-2}) \leq c \leq M(b_n - \epsilon^{-2})$; $D_n(c) \geq 2b_{n/3}$ if $m(b_n - \epsilon^{-2} b_{n/3}) \leq c \leq M(b_n - \epsilon^{-2} b_{n/3})$. Thus,

$$\begin{align*}
T_n^+ + t_0 &\geq M(b_n - \epsilon^{-2}) > 1 + 2\|k\|_{\infty}, \\
T_n^+ + t_0 &\geq M(b_n - \epsilon^{-2} b_{n/3}) > 1 + 2\|k\|_{\infty}, \\
T_n^- - t_0 &\geq m(b_n - \epsilon^{-2}) < -1 - 2\|k\|_{\infty}, \\
T_n^- + t_0 &\leq m(b_n - \epsilon^{-2} b_{n/3}) < -1 - 2\|k\|_{\infty}.
\end{align*} \quad (37)$$

We will approximate the number of lattice points in the intersection of $R_n$ and the plane $(x, \tilde{k}) = i + t_0$ (a "slice") with the area (Lebesgue measure in $\mathbb{R}^2$) of the same set for each $i \in Z \cap [C_n^-, C_n^+]$. This process is analogous to the approach in the proof of Theorem 1. For the planar set $S_{i,n}$ in (36), $i \in \{C_n^-, \ldots, C_n^+\}$, we define boundary functions $f_{u,i,n}(c), f_{l,i,n}(c)$ for $c \in S_{i,n}$ and $e_{1,i,n}(c), e_{2,i,n}(c)$ for $c \in Z \cap S_{i,n}$, analogous to the corresponding used in the proof of Theorem 1. (However, here we choose to work with $R_n - t$ rather than $R_n$.) To make these functions precise, we start with that, for $c \in S_{i,n}$, there exists

$$v_{c,i,n} \in R_n - t, \quad (v_{c,i,n}, \tilde{k}) = i, \quad (v_{c,i,n} - i\omega^*, v^*) = c.$$ 

Then define functions $f_{u,i,n}, f_{l,i,n} : S_{i,n} \rightarrow \{x \in R_n \setminus R_n^0 : (x, \tilde{k}) = c\}$; $f_{u,i,n}, f_{l,i,n} : Z \cap S_{i,n} \rightarrow \{x \in Z^2 : (x, \tilde{k}) = c\}$; and $e_{1,i,n}, e_{2,i,n} : Z \cap S_{i,n} \rightarrow [-1/2, 1/2]$ such that

$$\begin{align*}
f_{u,i,n}(c) &= v_{c,i,n} + \sup\{\beta \geq 0 : v_{c,i,n} + \beta\|\tilde{k}\|^{-1} \tilde{k} \in R_n - t\} \cdot \|\tilde{k}\|^{-1} \tilde{k}, \\
f_{l,i,n}(c) &= v_{c,i,n} - \sup\{\beta \geq 0 : v_{c,i,n} - \beta\|\tilde{k}\|^{-1} \tilde{k} \in R_n - t\} \cdot \|\tilde{k}\|^{-1} \tilde{k}, \\
f_{u,i,n}^*(c) &= f_{u,i,n}(c) - \tilde{e}_{c,i,n}\|\tilde{k}\|^{-1} \tilde{k}, \\
f_{l,i,n}^*(c) &= f_{l,i,n}(c) + \tilde{e}_{c,i,n}\|\tilde{k}\|^{-1} \tilde{k}, \\
e_{1,i,n}(c) &= 1/2 - \|\tilde{k}\|^{-1}\|f_{u,i,n}(c) - f_{l,i,n}(c)\|, \\
e_{2,i,n}(c) &= 1/2 - \|\tilde{k}\|^{-1}\|f_{l,i,n}^*(c) - f_{u,i,n}^*(c)\|.
\end{align*} \quad (38)$$
where \( p_{e,n} = 0 \) if \( f_{u,i,n}(c) \in (R_n - t) \cap Z^3 \), and \( 0 < p_{e,n} \leq \|k\| \), otherwise; \( p_{e,n} = 0 \) if \( f_{u,i,n}(c) \in (R_n - t) \cap Z^3 \), and \( 0 < p_{e,n} \leq \|k\| \), otherwise.

Similar to (18) from the proof of Theorem 1, we write the number of \( Z^3 \) lattice points in the planar slice \( \{x \in R_n : (x, \hat{k}) = i + t_0\} \) (or equivalently the \( Z^3 \) integer points in the slice \( \{x \in R_n - t : (x, \hat{k}) = i\} \)) as a sums of scaled distances between border points of the convex slice. Using Proposition 4,

\[
\#Z^3 \cap \{x \in R_n : (x, \hat{k}) = i + t_0\} = \#Z^3 \cap \{x \in R_n - t : (x, \hat{k}) = i\} = \sum_{j \in Z \cap S_{i,n}} |Z^3 \cap \{x \in R_n - t : (x, \hat{k}) = i, (x - iw^*, v^*) = j\}| = \|k\|^{-1} \sum_{j \in Z \cap S_{i,n}} \|f_{u,i,n}(j) - f_{u,i,n}(j)\| + E_{i,n,1} + I_{i,n},
\]

where, for \( i \in \mathbb{Z} \cap [C_n^+, C_n^-] \), \( E_{i,n,1} = \sum_{j \in Z \cap S_{i,n}} (e_{i,i,n}(j) + e_{i,i,n}(j)) \) and \( I_{i,n} \) denotes the sum of two indicator functions (similar to the \( I_{M,n} \), \( f_{m,n} \) terms from (16) but we will use only the fact that \( |I_{i,n}| \leq 2 \).

Fix \( i \in \mathbb{Z} \cap [C_n^+, C_n^-] \) and write \( a_{i,n} = \inf S_{i,n}, \bar{a}_{i,n} = \sup S_{i,n} \). For each \( j, j + 1 \in \mathbb{Z} \cap S_{i,n} \), we create a trapezoid, say \( T_{j,n} \), formed by the points in the plane \( (x, \hat{k}) = i \): \( f_{u,i,n}(j), f_{u,i,n}(j + 1), f_{L,i,n}(j), f_{L,i,n}(j + 1) \), which has area (or \( \mathbb{R}^2 \) Lebesgue measure)

\[
A[T_{j,n}] \equiv \frac{1}{2\|v^*\|} \left( \|f_{u,i,n}(j) - f_{u,i,n}(j)\| + \|f_{u,i,n}(j + 1) - f_{u,i,n}(j + 1)\| \right).
\]

We can similarly create two other trapezoids formed by two collection of points:

\[
f_{u,i,n}(\bar{a}_{i,n}), f_{u,i,n}(\bar{a}_{i,n}), f_{u,i,n}([\bar{a}_{i,n}]), f_{u,i,n}([\bar{a}_{i,n}]);
\]

\[
f_{u,i,n}(a_{i,n}), f_{u,i,n}(a_{i,n}), f_{u,i,n}([a_{i,n}]), f_{u,i,n}([a_{i,n}])
\]

say \( T_{i,n-1,n} \) and \( T_{i,n-1,n} \), respectively, with corresponding areas

\[
A[T_{i,n-1,n}] = \frac{c - |c|}{2\|v^*\|} \left( \|f_{u,i,n}(c) - f_{u,i,n}(c)\| + \|f_{u,i,n}([c]) - f_{u,i,n}([c])\| \right), \quad c = \bar{a}_{i,n};
\]

\[
A[T_{i,n-1,n}] = \frac{|c| - c}{2\|v^*\|} \left( \|f_{u,i,n}(c) - f_{u,i,n}(c)\| + \|f_{u,i,n}([c]) - f_{u,i,n}([c])\| \right), \quad c = a_{i,n}.
\]

We now examine \( \mathbb{R}^2 \) areas associated with the intersections formed between \( R_n - t \) and some "planes" of interest (determined by \( \hat{k} \)). To this end, we introduce a little notation: for \( c \in \mathbb{R} \), write

\[
A_{c,n} \equiv \text{area (\( \mathbb{R}^2 \) Lebesgue measure) of} \{x \in R_n - t : (x, \hat{k}) = c\}
\]

We shall focus on the particular collection of areas: \( A_{i,n}, i \in \mathbb{Z} \cap [C_n^+, C_n^-] \).
Following steps analogous to those used in (19) and noting \([\alpha_{i,n}] < [\pi_{i,n}]\) by (37), \(\|v^*\| = \|k\|^{-1}\|k\|\)
(by construction), we can rewrite the last sum of norms in (39) as a sum of areas of trapezoids:
\[
\|k\|^{-1} \sum_{j \in Z \cap S_{i,n}} \|f_{i,n,j}(j) - f_{i-1,n,j}(j)\| = \sum_{j = \lfloor \alpha_{i,n} \rfloor - 1}^{\lfloor \pi_{i,n} \rfloor} \|k\|^{-1} A[T_{i,n}] + E_{i,n,2},
\]
(41)
\[
2\|k\| E_{i,n,2} - \sum_{c = \alpha_{i,n} \pm \alpha_{i,n}} \|f_{i,n,c} - f_{i-1,n,c}\|
= \sum_{(c_1, c_2) = (\alpha_{i,n}, \alpha_{i,n} \pm 1)} (1 - |c_1 - c_2|) \left(\|f_{i,n,c_1} - f_{i-1,n,c_1}\| + \|f_{i,n,c_2} - f_{i-1,n,c_2}\|\right).
\]

The areas of the potentially "irregular" trapezoids \(T_{i,n,i-1,n}\) and \(T_{i,n,1-1,n}\) (formed at opposite \(k\)-parallel "ends" of the two-dimensional region \(\{x \in R_n : (x, k) = i\}\)) contribute to \(E_{i,n,2}\) above
(which resemble the "error" quantities \(E_{M,n}, E_{m,n}\) from the trapezoidal area estimate in (19)).
The sum of trapezoidal areas in (41) approximate \(\|k\|^{-1} A_{i,n}, i \in Z \cap [C^-_n, C^+_n]\). Write
\[
E_{i,n,3} = \|k\|^{-1} A_{i,n} - \sum_{j = \lfloor \alpha_{i,n} \rfloor}^{\lfloor \pi_{i,n} \rfloor} \|k\|^{-1} A[T_{j,n}] \geq 0
\]
(42)
to denote error incurred by subtracting the trapezoidal approximation from \(A_{i,n}\); we note that \(E_{i,n,3}\) is
nonnegative by the convexity of \(R_n - t\).

We next comment that for \(i, i + 1 \in Z \cap [C^-_n, C^+_n]\),
\[
\frac{1}{2\|k\|}(A_{i,n} + A_{i+1,n})
\]
(43)
approximates the volume (\(R^3\) Lebesgue measure) of the section of \(R_n - t\) enclosed between the planes
\(\langle x, \hat{k} \rangle = i, \quad \langle x, \hat{k} \rangle = i + 1\);

note the distance between these planes is \(\|k\|^{-1}\). Write the difference between the actual volume
\(\{x \in R_n : (x, \hat{k}) \in [i, i + 1]\}\) and the approximation in (43) as
\[
E_{i,n,4} = \|k\|^{-1} \int_0^1 \left[ A_{i+\omega,n} - (1 - \omega)A_{i,n} - \omega A_{i+1,n} \right] d\omega \geq 0, \quad i \in Z \cap [C^-_n, C^+_n].
\]
Above we used the definition of \(R^2\) "area" from (40) and convexity, which guarantees that \(E_{i,n,4}\) is
nonnegative. Thus, for \(i < j \in Z \cap [C^-_n, C^+_n]\),
\[
\{x \in R_n - t : (x, \hat{k}) \in [i, j]\} - \|k\|^{-1} \sum_{i \in Z \cap [i, j]} A_{i,n} = \sum_{i \in Z \cap [i, j-1]} E_{i,n,4} - (2\|k\|)^{-1}(A_{i,n} + A_{j,n}).
\]
(44)

We next consider estimating the \(Z^3\) lattice point count of \(R_n \cap (k + R_n)\) (or equivalently the \(Z^3\)
integer count of \((R_n \cap (k + R_n)) - t\) by areas and volumes based on minor reformulations of our
previously developed quantities. For $c \in [C_n^-, C_n^+]$, let

$$h_n(c) = \inf \left\{ (x - c w^*, v^*) : x \in \overline{R_n} \cap (k + R_n), (x, \hat{k}) = c \right\},$$

$$H_n(c) = \sup \left\{ (x - c w^*, v^*) : x \in \overline{R_n} \cap (k + R_n), (x, \hat{k}) = c \right\},$$

which are analogous to the functions $h_n, H_n$ defined with respect to $R_n \cap (k + R_n)$. We remark that $h_n(c), H_n(c)$ are well-defined for $c \in [C_n^-, C_n^+]$ because:

$$D_n(c + t_0) \geq 2b_n^{1/3} > \|k\| \implies \{x \in \overline{R_n} \cap (k + R_n) : (x, \hat{k}) = c + t_0 \neq \emptyset. \quad (45)$$

Also define

$$\hat{T}_n^+ = \max \{ c \in Z \cap [C_n^-, C_n^+] : H_n(c + t_0) - \hat{h}_n(c + t_0) \geq 2 \},$$

$$\hat{T}_n^- = \min \{ c \in Z \cap [C_n^-, C_n^+] : H_n(c + t_0) - \hat{h}_n(c + t_0) \geq 2 \}.$$

Because a circle of radius $2\|k\|$ can be embedded in the open slice $\{x \in R_n : (x, \hat{k}) = c\}, m \leq (b_n - \epsilon^{-2}2\|k\|^{-1})c \leq M$, it follows that

$$\hat{T}_n^+ + 1 + t_0 \geq M(b_n - \epsilon^{-2}2\|k\|) > 1 + 2\|k\|,$$

$$\hat{T}_n^- - 1 + t_0 \leq m(b_n - \epsilon^{-2}2\|k\|) < -1 - 2\|k\|.$$

Let $M_n^* = \max\{C_n^-, \hat{T}_n^-\}$ and $M_n^+ = \min\{C_n^+, \hat{T}_n^+\}$.

Fix $i \in Z \cap [M_n^-, M_n^+]$. We will now create counterparts to the set $S_{i,n}$, and the $S_{i,n}$-defined functions $f_{i,n}(-), f_{i,n}(\cdot), e_{i,n}(\cdot)$, and $e_{2,i,n}(\cdot)$, with respect to $
\{x \in (R_n \cap (k + R_n)) - t : (x, \hat{k}) = i\}$

as follows: define $\hat{S}_{i,n}$ by replacing $\overline{R_n} - t$ with $((\overline{R_n} \cap (k + R_n)) - t$ in (36); define $\hat{f}_{i,n}(-), \hat{f}_{i,n}(\cdot), \hat{e}_{i,n}(\cdot)$, and $\hat{e}_{2,i,n}(\cdot)$ on $\hat{S}_{i,n}$ by replacing $\overline{R_n} - t$ with $(\overline{R_n} \cap (k + R_n)) - t$ and $R_n$ with $R_n \cap (k + R_n)$ in (38). Note that (47) and $\hat{S}_{i,n}$ are nonempty by (45).

We can approximate the number of $Z^3$ integer points within (47) by following the same steps used for $\{x \in R_n - t : (x, \hat{k}) = i\}$.

1. Express the integer point count of (47) as the sum over counts:

$$\#Z^3 \cap \{x \in (R_n \cap (k + R_n)) - t : (x, \hat{k}) = i, (x - iw^*, v^*) = j\}, \quad j \in Z,$$

which can be further written as a sum of a scaled distances $||\hat{k}||^{-1}(||\hat{f}_{i,n}(j)|| - ||\hat{f}_{i,n}(j)||)$ and "boundary" errors $\hat{I}_{i,n,1} + \hat{E}_{i,n,1}$; here $\hat{E}_{i,n,1}$ consists of a sum of $(\hat{e}_{i,n}(j) + \hat{e}_{2,i,n}(j))$ over $Z \cap \hat{S}_{i,n}$ and $\hat{I}_{i,n,1}$ is the sum of two indicator functions. [Thus, we repeat steps analogous to those in (39) and produce error quantities analogous to $I_{i,n}, E_{i,n,1}$ but defined on $(R_n \cap (k + R_n)) - t.$]
2. Write this sum of scaled line segment distances as a sum of trapezoidal areas which approximate:

\[ \|k\|^{-1} \cdot \tilde{A}_{i,n} \equiv \|k\|^{-1} \cdot \text{area of } \{ x \in (R_n \cap (k + R_n)) - t : \langle x, \tilde{k} \rangle = i \}. \]  

(48)

In this approximation of \( \|k\|^{-1} \tilde{A}_{i,n} \), an additional error \( \tilde{E}_{i,n,2} - \tilde{E}_{i,n,3} \) results, which is a difference of terms analogous to \( E_{i,n,2}, E_{i,n,3} \) from (41) and (42) [but again defined on \( (R_n \cap (k + R_n)) - t \) and using \( \tilde{S}_{i,n} \) in \( \tilde{E}_{i,n,2} \)].

In summary, we may then write

\[ \# \{ x \in (R_n \cap (k + R_n)) - t : \langle x, \tilde{k} \rangle = i \} = \|k\|^{-1} A_{i,n} + \tilde{I}_{i,n} + \tilde{E}_{i,n,1} + \tilde{E}_{i,n,2} - \tilde{E}_{i,n,3}. \]  

(49)

with all quantities defined with respect to \( (R_n \cap (k + R_n)) - t \) but functionally analogous their counterparts on \( R_n - t \).

We now focus on establishing that for all \( i \in \mathbb{Z} \cap [M^{-1}, M^+], \)

\[ |E_{i,n,1} - \tilde{E}_{i,n,1}| \leq 36\|k\|_\infty b_n^{2/3}, \]

\[ |(E_{i,n,2} - E_{i,n,3}) - (\tilde{E}_{i,n,2} - \tilde{E}_{i,n,3})| \leq 72\|k\|_\infty b_n^{2/3}, \]

(50)

which we will use shortly after to finish the proof of Theorem 3.

Fix \( i \in \mathbb{Z} \cap [M^{-1}, M^+]. \) Then there exists \( y_1, y_2, y_3 \in \{ x \in \overline{R_n} : \langle x, \tilde{k} \rangle = i + t_0 \}, \) such that

\[ \omega \left( y_1 + b_n^{1/3}(2\|k\|)^{-1}k \right) + (1 - \omega) \left( y_1 - b_n^{1/3}(2\|k\|)^{-1}k \right) \in R_n^3, \quad \omega \in [0,1] \]

(because \( D_n(i + t_0) \geq 2b_n^{1/3} \) and the function \( D^*(b_n^{-1}(i + t_0), \cdot) \) is continuous, concave on the \( [h(b_n^{-1}(i + t_0)), H(b_n^{-1}(i + t_0))] \) interval) and

\[ \langle y_2 - t - iw^*, v^* \rangle = a_{i,n}, \quad \langle y_3 - t - iw^*, v^* \rangle = \tilde{a}_{i,n}. \]

We are suppressing here the dependence of \( y_1, y_2, y_3 \) on \( i \) and \( n. \)

The open line segment from \( y_2 \) or \( y_3 \) to any point between \( y_1 + b_n^{1/3}(2\|k\|)^{-1}k \) and \( y_1 - b_n^{1/3}(2\|k\|)^{-1}k \) lies in \( R_n^3. \) In particular,

\[ \psi_j(\beta), \quad \psi_j^*(\beta) \in R_n^3, \quad \beta \in (0,b_n], \quad j \in \{2,3\}, \]

\[ \psi_j(\beta) = y_j + \beta \cdot b_n^{-1} \left( y_1 + b_n^{1/3}(2\|k\|)^{-1}k - y_j \right), \quad \psi_j^*(\beta) = y_j + \beta \cdot b_n^{-1} \left( y_1 - b_n^{1/3}(2\|k\|)^{-1}k - y_j \right), \]

where the above \( \mathbb{R}^3 \) vectors are functions of \( \beta \) and \( j. \) For \( \beta = 2\|k\|b_n^{2/3}, \) it holds that

\[ \|\psi_j(\beta) - \psi_j^*(\beta)\| = 2\|k\|, \quad \langle \psi_j(\beta), v^* \rangle = \langle \psi_j^*(\beta), v^* \rangle \quad j \in \{2,3\}. \]

By this, we have: if \( \langle \psi_2(\beta) - (i + t_0)w^*, v^* \rangle \leq c \leq \langle \psi_3(\beta) - (i + t_0)w^*, v^* \rangle, \) then \( \|f_{i,n}(c) - f_{i,n}(c)\| \geq 2\|k\| \) by the concavity of \( D^*(b_n^{-1}(i + t_0), \cdot) \) on \( [h(b_n^{-1}(i + t_0)), H(b_n^{-1}(i + t_0))] \); this result implies that
if \(c \in \mathbb{Z}\) and \(\langle \psi_0(\vec{\beta}) - t - i \vec{w}^*, \vec{v}^* \rangle \leq c \leq \langle \psi_0(\vec{\beta}) - t - i \vec{w}^*, \vec{v}^* \rangle\) (assuming the interval is nonempty),
then
\[
\epsilon_{1,i,n}(c) = \tilde{\epsilon}_{1,i,n}(c), \quad \epsilon_{2,i,n}(c) = \tilde{\epsilon}_{2,i,n}(c)
\]
[see also (36)]. By the above, \(|\epsilon_{j,i,n}|, |\tilde{\epsilon}_{j,i,n}| \leq 1/2\) for \(j = 1, 2\), and the fact that
\[
|\langle y_j, \vec{v}^* \rangle - \langle \psi_0(\vec{\beta}), \vec{v}^* \rangle| \leq 6\|\vec{k}\|^2 b_n^{2/3}, \quad j = 2, 3,
\]
we reach:
\[
|E_{1,i,n} - \tilde{E}_{1,i,n}| \leq 12\|\vec{k}\|^2 b_n^{2/3}.
\]
We have established the first inequality in (50). We can also obtain
\[
|E_{1,i,n,2} - E_{1,i,n,3} - (\tilde{E}_{1,i,n,2} - \tilde{E}_{1,i,n,3})| \leq 12\|\vec{k}\|^2 b_n^{2/3},
\]
which is achieved through bounding the Lebesgue \(\mathbb{R}^2\) measure of the set
\[
\{x \in \mathbb{R}^n : (x, \hat{\vec{k}}) = i + t_0, \quad D^*(b_n^{-1} (i + t_0), b_n^{-1} (x - (i + t_0)w^*, \vec{v}^*)) < b_n^{-1}\|\vec{k}\|\}
\]
with a \(\|\vec{k}\|\) multiple of the bounds on \(\{(y_j, \vec{v}^*) - \langle \psi_j(\vec{\beta}), \vec{v}^* \rangle, j \in \{2, 3\}\}\) from (51). The underlying arguments closely resemble those presented in proof Theorem 1 for handling the quantities \(E_{M,n}, \tilde{E}_{M,n}, a_{1,1,n, a_{1,1,n}}, a_{1,1,n, a_{1,1,n}}\) (see (26) and the subsequent discussion of \(|\tilde{P}_{j,n} - \tilde{P}_{j,n}|, j = 1, 2\)).

This now sets the second bound in (50).

Using an approach analogous to (44), the sum
\[
\|\hat{\vec{k}}\|^{-1} \sum_{i \in \mathbb{Z} \cap [M_n^-, M_n^+]} \tilde{A}_{i,n}
\]
can be used to approximate \(|\{x \in (R_n \cap (\vec{k} + R_n)) - t : (x, \hat{\vec{k}}) \in [M_n^-, M_n^+]\}|\). Namely,
\[
\left|\left\{x \in (R_n \cap (\vec{k} + R_n)) - t : (x, \hat{\vec{k}}) \in [M_n^-, M_n^+]\right\}\right| - \|\hat{\vec{k}}\|^{-1} \sum_{i \in \mathbb{Z} \cap [M_n^-, M_n^+]} \tilde{A}_{i,n}
\]
\[
= \sum_{i \in \mathbb{Z} \cap [M_n^-, M_n^+]} \tilde{E}_{i,n,i} - (2\|\hat{\vec{k}}\|^{-1})(\tilde{A}_{M_n^-,n} + \tilde{A}_{M_n^+,n}),
\]
where for \(i, i + 1 \in \mathbb{Z} \cap [M_n^-, M_n^+]\),
\[
\tilde{E}_{i,n,i} = \|\hat{\vec{k}}\|^{-1} \int_0^1 \left[\tilde{A}_{i+\omega,n} - (1 - \omega)\tilde{A}_{i,n} - \omega\tilde{A}_{i+1,n}\right] d\omega \geq 0.
\]

From (39), (41), (44), (49), (52), and \(#\mathbb{Z}^2 \cap R_n \cap (\vec{k} + R_n) = \#\mathbb{Z}^2 \cap \{(R_n \cap (\vec{k} + R_n) - t\}, we can write:
\[
|\Omega_{0,n}(0) - \Omega_{0,n}(\vec{k})| \leq \sum_{j=1}^{N} W_{j,n},
\]
where, letting $S_n = \{ x \in \mathbb{R}^3 : (x, \hat{k}) \in [M_n^- \cdots M_n^+] \}$,

\[
W_{1,n} = \left| |R_n - t \cap S_n^| - |(R_n \cap (k + R_n)) - t \cap S_n^| \right|,
\]

\[
W_{2,n} = 2 \cdot \#Z^3 \cap \left\{ x \in R_n - t : (x, \hat{k}) \notin [C_n^-, C_n^+] \right\},
\]

\[
W_{3,n} = 2 \cdot \#Z^3 \cap \left\{ x \in R_n - t : (x, \hat{k}) \in [C_n^-, M_n^-] \cup (M_n^+, C_n^+) \right\},
\]

\[
W_{4,n} = \left| \left( |R_n - t \cap S_n^| - \#Z^3 \cap (R_n - t) \cap S_n^ \right) - \left( |(R_n \cap (k + R_n)) - t \cap S_n^| - \#Z^3 \cap (R_n \cap (k + R_n) - t) \cap S_n^ \right) \right|.
\]

We bound each $W_{j,n}, j \in \{1, 2, 3, 4\}$, in the following.

We note first that, for $c \in [m_n, M_n]$, we can easily bound the difference between the $\mathbb{R}^2$ Lebesgue measures (areas) of the sets

\[
\{ x \in R_n - t : (x, \hat{k}) = c \}, \quad \{ x \in (R_n \cap (k + R_n)) - t : (x, \hat{k}) = c \},
\]

which correspond to planar "slices" of $R_n - t$ and $(R_n \cap (k + R_n)) - t$, respectively. That is, from (40) and (48):

\[
A_{c,n} - \tilde{A}_{c,n} \leq \|v^*\|^{-1} \|k\| \left( H_n(c) - h_n(c) \right) \leq 4\|k\|^2 b_n,
\]

using $|H_n(c)|, |h_n(c)| \leq 2\|k\|b_n$. With this bound, the inequalities from (37) and (46), and $|m|, |M| \leq \|k\|_\infty$, we find

\[
W_{1,n} \leq \left( 2e^{-2}(\|m| + M)b_n^{1/3} \right) \cdot (4\|k\|^2 b_n) \leq \epsilon^{-2} 48\|k\|^3 \cdot b_n^{4/3}.
\]

It follows from inequalities in (37), the concavity of $D(\cdot), H(\cdot), -h(\cdot)$, and Proposition 4 that

\[
\#Z^3 \cap \left\{ x \in R_n - t : (x, \hat{k}) \notin (\max\{Y_n^-, \hat{T}_n^- \} - 1, \min\{Y_n^+, \hat{T}_n^+ \} + 1) \right\} \leq 2e^{-2}(\|m| + M) \left( \sup_{c \in [m_n, M_n]} \|\hat{k}\|^{-1} D_n(c) + 1 \right)
\]

\[
\leq 16e^{-2}\|k\|_\infty b_n;
\]

\[
\#Z^3 \cap \left\{ x \in R_n - t : (x, \hat{k}) \notin (\Gamma_n^-, \Gamma_n^+) \right\} \leq 2e^{-2}(\|m| + M)b_n^{1/3} \left( 2b_n^{1/3}\|\hat{k}\|^{-1} + 1 \right) \left( \sup_{c \in [m_n, M_n]} (H_n(c) - h_n(c)) + 1 \right)
\]

\[
\leq 2^8 \epsilon^{-3}\|k\|_\infty b_n^{5/3}.
\]
Hence, the two count inequalities above provide bounds to show that, for large $n$

$$W_{2,n} \leq 2^9 e^{-2||k||_\infty} \left( b_n + b_n^{5/3} \right).$$

Note that for integer $i \in [C_n^-, M_n^-) \cup (M_n^+, C_n^+],$

$$\tilde{h}_n(i + t_0) - \tilde{h}_n(i + t_0) \geq 2, \quad H_n(i + t_0) \geq \tilde{h}_n(i + t_0) \geq \tilde{h}_n(i + t_0) \geq h_n(i + t_0),$$

implying, with Proposition 4, that

$$\# \mathbb{Z}^3 \cap \left\{ x \in R_n - t : \langle x, k \rangle = i \right\}$$

$$\leq 2 \left( \|k\|^{-1} D_n(i + t_0) + 1 \right) + \left[ \left( H_n(i + t_0) - h_n(i + t_0) + 1 \right) \left( \|k\|^{-1} \|k\| + 1 \right) \right]$$

$$\leq 32 \|k\|^2 b_n.$$

Applying this bound with the one from (46), we have

$$W_{3,n} \leq 2 e^{-2 \|k\|} \left( M \cdot I_{(M_n^+ < C_n^-)} + |m| \cdot I_{(M_n^+ > C_n^-)} \right) \max \left\{ \# \mathbb{Z}^3 \cap \{ x \in R_n - t : \langle x, k \rangle = i \} : i \in \mathbb{Z} \cap \left( [C_n^-, M_n^-) \cup (M_n^+, C_n^+) \right) \right\}$$

$$\leq 64 e^{-2 \|k\|^4} b_n.$$

We now handle $W_{4,n}$ by subdividing the quantity into more manageable components:

$$W_{4,n} \leq W_{5,n} + W_{6,n} + W_{7,n};$$

$$W_{5,n} = (2 \| k \|)^{-1} \left( A_{M_n^+, n} - \tilde{A}_{M_n^+, n} + A_{M_n^-, n} - \tilde{A}_{M_n^-, n} \right)$$

$$W_{6,n} = \sum_{i \in \mathbb{Z} \cap [M_n^-, M_n^+] \setminus \{i \}} \left( I_{i,n} + \tilde{I}_{i,n} + \sum_{j=1}^3 |E_{i,n,j} - \tilde{E}_{i,n,j}| \right)$$

$$W_{7,n} = \sum_{i \in \mathbb{Z} \cap [M_n^-, M_n^+] \setminus \{i \}} |E_{i,n,4} - \tilde{E}_{i,n,4}|.$$

By the same arguments in (53) (ie. examining differences between areas), we have

$$W_{5,n} \leq \|k\|^{-1} \|k\| \|v^*\|^{-1} \sup_{c \in [m_n, M_n]} (H_n(c) - h_n(c)) \leq 4 \|k\|^2 b_n.$$

From (50) and $0 \leq I_{i,n}, \tilde{I}_{i,n} \leq 2$, it follows that

$$W_{6,n} \leq 3 \left( b_n (M - m) + 1 \right)^2 \left( 72 \|k\|_{\infty} \|k\|_{\infty}^{5/2} \right) \leq 64 \|k\|^{4/5} b_n^{5/3}.$$

To handle $W_{7,n}$, we first create a "difference" function: for $i \in \mathbb{Z} \cap [M_n^-, M_n^+]$, $\omega \in [0, 1],$

$$\chi_{i+\omega,n} = \left( A_{i+\omega,n} - \tilde{A}_{i+\omega,n} \right) - \|v^*\|^{-1} \|k\| \left( H_n(i + \omega + t_0) - h_n(i + \omega + t_0) \right),$$

$$W_{7,n} \leq \sum_{i \in \mathbb{Z} \cap [M_n^-, M_n^+] \setminus \{i \}} \chi_{i+\omega,n}.$$
with the "area" definitions of \( A_{i+\omega,n}, \tilde{A}_{i+\omega,n} \) from (40) and (48). Define also the "scaled" set:

\[
K_n(i + \omega + t_0) =
\]

\[
b_n \cdot \left\{ c : h(b_n^{-1}(i + \omega + t_0)) \leq c \leq H(b_n^{-1}(i + \omega + t_0)), D^*(b_n^{-1}(i + \omega + t_0), c) \geq b_n^{-1}
\mid \|k\| \right\}.
\]

We can then bound the function \( \chi_{c,n} \): for \( c \in [M_n^-, M_n^+] \),

\[
|\chi_{c,n}| = \|\nu^*\|^{-1} \|k\| \left| h_n(c + t_0) - \inf K_n(c + t_0) + \sup K_n(c + t_0) - H_n(c + t_0) \right|
\leq \|\nu^*\|^{-1} \|k\| \cdot 2 \left( 6 \|k\|^2 b_n^{3/2} \right),
\]

which follows from \( D_n(c + t_0) \geq 2b_n^{1/3} > \|k\| \) so that we can bound the distances \( |h_n(c + t_0) - \inf K_n(c + t_0)|, |\sup K_n(c + t_0) - H_n(c + t_0)| \) by using the properties of \( \psi_j(\bar{\beta}), \psi_j^*(\bar{\beta}) \) and the bound on \( |\langle y_j, \nu^* \rangle - (\psi_j(\bar{\beta}), \nu^*)| \) from (51). Hence, we reach:

\[
\frac{|E_{i,n} - \tilde{E}_{i,n}^n|}{\|k\| - \|\nu^*\|} \leq \int_0^1 \left( [H_n(i + \omega + t_0) - H_n(i + t_0)] + |h_n(i + t_0) - h_n(i + \omega + t_0)| \right) \, d\omega
\leq 12 \|k\|^3 \left( b_n \cdot \xi_{k,n} + b_n^{2/3} \right)
\]

where

\[
\xi_{k,n} = \sup \left\{ \max \{|H(c_1) - H(c_2)|, |h(c_1) - h(c_2)| \} : c_1, c_2 \in [m, M], |c_1 - c_2| \leq b_n^{-1} \right\};
\]

we used, in the last inequality above,

\[
H_n(c) = b_n \cdot H(b_n^{-1} c), \quad h_n(c) = b_n \cdot h(b_n^{-1} c), \quad c \in [m_n, M_n],
\]

by definition. We can now set a bound on \( W_{k,n} \):

\[
W_{k,n} \leq 36 \|k\|^4 \left( b_n^2 \cdot \xi_{k,n} + b_n^{5/3} \right).
\]

Note that, because \( h(\cdot), H(\cdot) \) are continuous on \([m, M], \xi_{k,n} \to 0 \) as \( n \to \infty \).

In conclusion, for large \( n \) (such that \( e^2 b_n^{-1/3} > 4\|k\|_\infty \)), we have

\[
\sum_{j=1}^4 W_{j,n} \leq C_k \cdot b_n^2 \left( \xi_{n,k} + b_n^{-1/3} \right)
\]

for some \( C_k > 0 \). This finishes the proof of the Theorem 3. \( \Box \)
8 Appendix: Supplementary proofs

Proof of Proposition 1. Define an invertible linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$, where for $x = (x_1, x_2)' \in \mathbb{R}^2$,

$$T(x) = \frac{1}{||k||}(-k_2x_1 + k_1x_2, -k_1x_1 - k_2x_2)' .$$

Applying the transformation to $R_0$, we have the new set $T(R_0)$ is closed and convex and, if $x \in R_0$ such that $(x, k) = c \in [m, M]$, then $T(x) = \frac{1}{||k||}(c, -k_1x_1 - k_2x_2)'$. The function $h(c) = \inf\{y \in \mathbb{R} : (c||k||^{-1}, y)' \in T(R_0)\}$ is finite and convex on $[m, M]$ [cf Lay (1982), p. 200]. Note now that, for $c \in [m, M]$ and $\rho \geq 0$,

$$T(y_c + ||k||^{-1}\rho \cdot k) = \frac{1}{||k||}(c||k||^{-1}, -||k||^{-1}(y_c, k) - \rho)' \implies h(c) = \frac{1}{||k||}(c||k||^{-1}, k_1c, k_2h(c))' = f_u(c).$$

(54)

To show that $f_u$ is continuous on $[m, M]$, it suffices to establish the continuity of $h : [m, M] \to \mathbb{R}$ by (54). The convexity of $h(-)$ on $[m, M]$ ensures the function is absolutely continuous on $(m, M)$ [cf. Royden (1988), p. 113]. We prove that $h(-)$ is continuous at $M$; continuity of $h(-)$ at $m$ follows by symmetry. If for $c \in (m, M)$, $[h(M) - h(c)](M - c)^{-1} \leq 0$, then

$$[h(M) - h(m)](M - m)^{-1} \leq [h(M) - h(c)](M - c)^{-1} \leq 0, \quad c \in (m, M)$$

by the convexity of $h(-)$. Letting $c \uparrow M$, we have:

$$0 \leq [h(M) - h(c)] \leq (h(M) - h(m))(M - m)^{-1}(M - c) \to 0.$$

If there exists a $c \in (m, M)$ such that $[h(M) - h(c)](M - c)^{-1} > 0$, then for all $z \in [c, M)$, $h(M) - h(z) > 0$. In particular, if $x_n \in [c, M]$ such that $x_n \to M$, then $[h(x_n)] \leq 1$ implies there is a convergence subsequence $h(x_{n_j}) \to L_M$ for some $L_M \in \mathbb{R}$. Furthermore, $(x_n||k||^{-1}, h(x_n))' \in T(R_0)$ implies $(M||k||^{-1}, L_M) \in T(R_0)$, because the latter set is closed. From the definition of $h(-)$ and the construction of the subsequence, we have now

$$h(M) \leq L_M = \lim_{j \to \infty} h(x_{n_j}) \leq h(M).$$

Thus, every sequence $h(x_n)$, with $x_n \in [c, M]$ and $x_n \to M$, has a convergent subsequence and continuity of $h(-)$ at $M$ follows. We have now established the continuity of $h$ and, hence $f_u$, on $[m, M]$.

A similar argument shows that the function $\tilde{h}(c) = -\sup\{y \in \mathbb{R} : (c||k||^{-1}, y)' \in T(R_0)\}$ is convex and continuous on $[m, M]$ and

$$f_u(c) = \frac{1}{||k||}(\frac{-k_2c}{||k||} + k_1\tilde{h}(c), k_1c, k_2\tilde{h}(c))'.$$

(54)
is thereby continuous on \([m, M]\). The proof of Proposition 1 is complete. □

**Proof of Proposition 2.** Proposition 2 follows from the fact that if \(c_1 < c_2 \in [m, M]\) such that \(d\mathbf{m}(c_1) = d\mathbf{m}(c_2) = d\mathbf{m}(\bar{c})\) and \(\omega \in (0, 1)\), then

\[
\omega f_u(c_1) + (1 - \omega)f_u(c_2), \quad \omega f_c(c_1) + (1 - \omega)f_c(c_2) \in \overline{F_0} \cap \{x \in \mathbb{R}^2 : \langle x, \hat{\mathbf{k}} \rangle = \omega c_1 + (1 - \omega)c_2\}
\]

and so, by convexity and linearity,

\[
d\mathbf{m}(\bar{c}) \geq d\mathbf{m}(\omega c_1 + (1 - \omega)c_2) \\
\geq \left\|\left|k\right|^{-1}(\omega\|f_u(c_1) - f_u(c_1)\| + (1 - \omega)\|f_u(c_2) - f_u(c_2)\|)k\right\| = d\mathbf{m}(\bar{c}).
\]

Thus, \(A_\varnothing\) is an interval and must be closed by the continuity of \(d\mathbf{m}(\cdot)\). To show Proposition 2, there exists \(\omega \in (0, 1)\) such that \(c_2 = \omega \cdot c_1 + (1 - \omega) \cdot \inf A_\varnothing\) so that, as above,

\[
d\mathbf{m}(c_2) \geq \left\|\left|k\right|^{-1}(\omega\|f_u(c_1) - f_u(c_1)\| + (1 - \omega)\|f_u(\inf A_\varnothing) - f_u(\inf A_\varnothing)\|)k\right\| \\
= \omega d\mathbf{m}(c_1) + (1 - \omega)d\mathbf{m}(\bar{c}) > d\mathbf{m}(c_1).
\]

The Proposition 2(b) can be shown similarly. □

**Proof of Proposition 3.** We establish the proof by considering two cases.

**Case 1.** If \(c \in b_n(m, M) \cap (\mathbb{Z} + t_0)\), then by linearity: \(\|f_{u,n}(c) - f_{\varnothing,n}(c)\| > 0\) by Proposition 2.

1.a If \(F_n(c) \neq 0\), then \(F_n(c) = \left\|k\right|^{-1}\|f^*_{u,n}(c) - f^*_{\varnothing,n}(c)\| + 1\) by (12) and, because the \(\mathbb{R}^2\) points fall on the same line,

\[
\|f_{u,n}(c) - f_{\varnothing,n}(c)\| = \|f_{u,n}(c) - f^*_{u,n}(c)\| + \|f^*_{u,n}(c) - f^*_{\varnothing,n}(c)\| + \|f^*_{\varnothing,n}(c) - f_{\varnothing,n}(c)\|.
\]

Note also that \(\|f_{u,n}(c) - f_{\varnothing,n}(c)\| > \left\|k\right\|\) implies \(F_n(c) \neq 0\).

1.b If \(F_n(c) = 0\) and \(\|f_{u,n}(c) - f_{\varnothing,n}(c)\| < \left\|k\right\|\) then

\[
\|k\| = \|f^*_{u,n}(c) - f_{\varnothing,n}(c)\| + \|f_{u,n}(c) - f_{\varnothing,n}(c)\| + \|f_{\varnothing,n}(c) - f_{\varnothing,n}(c)\|,
\]

\[
\|f^*_{u,n}(c) - f_{\varnothing,n}(c)\| = \|f^*_{u,n}(c) - f_{\varnothing,n}(c)\| + \|f_{u,n}(c) - f_{\varnothing,n}(c)\|,
\]

\[
\|f^*_{\varnothing,n}(c) - f_{\varnothing,n}(c)\| = \|f^*_{\varnothing,n}(c) - f_{\varnothing,n}(c)\| + \|f_{u,n}(c) - f_{\varnothing,n}(c)\|.
\]

Hence, \(F_n(c) = 0 = \|k\|^{-1}\|f_{u,n}(c) - f_{\varnothing,n}(c)\| + \varepsilon_{1,n}(c) + \varepsilon_{2,n}(c)\).

1.c If \(F_n(c) = 0\) and \(\|f_{u,n}(c) - f_{\varnothing,n}(c)\| = \|k\|\), then \(f_{u,n}(c) = f^*_{u,n}(c)\), \(f_{\varnothing,n}(c) = f^*_{\varnothing,n}(c) \in \mathbb{Z}^2 \cap b_nR_0^\varnothing\) and \(F_n(c) = 0 = \|k\|^{-1}\|f_{u,n}(c) - f_{\varnothing,n}(c)\| + \varepsilon_{1,n}(c) + \varepsilon_{2,n}(c)\).
Case 2. Consider now $c = c_{M,n}$ (the treatment of $c_{m,n}$ is analogous).

2.a If $I_{M,n} = 0$, the arguments in Case 1 remain valid.

2.b If $I_{M,n} = 1$, then

$$f_{0,n}(c_{M,n}) = f_{0,n}(c_{M,n}) + ||\vec{k}|| \cdot ||\vec{k}||^{-1} \vec{k}, \quad f_{0,n}(c_{M,n}) = f_{0,n}(c_{M,n}) - ||\vec{k}|| \cdot ||\vec{k}||^{-1} \vec{k}$$

so that $F_n(c_{M,n}) = 0 = e_{1,n}(c_{M,n}) + e_{2,n}(c_{M,n}) + I_{M,n}$.

Cases 1 and 2 above establish Proposition 3. □

Proof of Proposition 4. Proposition 4(a): It follows from the same arguments used to show (12) that

$$(i_1, i_2, 0)' = (i, \vec{k}) \cdot w^* + m_i ||\vec{k}|| \vec{k} ||\vec{k}||^{-1}, \quad (j_1, j_2, 0)' = (j, \vec{k}) \cdot w^* + m_j ||\vec{k}|| \vec{k} ||\vec{k}||^{-1},$$

for some $m_i, m_j \in \mathbb{Z}$, where $\vec{k} = (\vec{k}_1, \vec{k}_2, 0)'$. Then, the inner product assumptions imply that $||\vec{k}_3|| = m_i - m_j = \|g\| i_1 - j_2$. If $||\vec{k}_3|| \neq 0$ and $|i_1 - j_2| = 0$, then $|m_i - m_j| = 0$ and $i = j$; if $||\vec{k}_3|| = 0$ and $|i_1 - j_2| \neq 0$, then $|m_i - m_j| \neq 0$ and so $g$ must divide $|m_i - m_j|$ (since if $g \neq 1$, then $g$ cannot divide $||k||_3$ by definition of $g$); and if $||\vec{k}_3|| = 0$, then $g = 1$ by construction. In any event, $g = (m_i - m_j) \in \mathbb{Z}$. Upon subtraction, we find $i - j = m\vec{k}$ for $m = g^{-1}(m_i - m_j) \in \mathbb{Z}$.

Proposition 4(b): The latter set claim follows from the fact that, expanding $(i_1, i_2, 0)'$ as above (in terms of $m_i$ and $\vec{k}$),

$$\langle i - (i, \vec{k}) \cdot w^*, v' \rangle = (||\vec{k}||^2 g)^{-1} (-m_i ||\vec{k}|| ||\vec{k}||^3 + ||g\vec{k}||^2 i_3)$$

$$= -m_i k_3 + g i_3 \in \mathbb{Z},$$

because $g ||\vec{k}|| = ||\vec{k}||$.

Proposition 4(c): There exists $s_1, s_2 \in \mathbb{Z}$ such that $-k_3 s_1 + g s_2 = 1$. (To see this, note for $\vec{k}_3 = 0$, we have $g = 1$ and so we can pick $s_1 = 0, s_2 = 1$; if $\vec{k}_3 \neq 0$ and $g = 1$, we can choose $s_1 = 1, s_2 = k_3 + 1$; finally, if $\vec{k}_3 \neq 0$ and $g > 1$, then $gcd(\vec{k}, [\vec{k}_3]) = 1$ so that the result follows as well.) Then, let $i = c_1 w^* + g^{-1} s_2 c_2 k + (0,0, s_2 c_2)' \in \mathbb{Z}$. □

References


Abstract

This paper introduces blockwise empirical likelihood confidence intervals for the mean of a long-range dependent process. Both linear and non-linear processes are considered. Empirical likelihood is shown to provide valid nonparametric inference in instances where the block bootstrap is known to fail. The confidence levels of the empirical likelihood intervals are empirically calibrated using the sampling window method of Hall, Jing and Lahiri (1998). The consistency of the sampling window estimator for distribution of the sample mean is also extended to linear, long-range dependent processes.

Key Words: Bootstrap, empirical likelihood, long-range dependence, sampling window method, subsamples

1 Introduction

Empirical likelihood (EL), proposed by Owen (1988, 1990), makes possible likelihood-based statistical inference without a specified distribution for the data. That is, the method generates a nonparametric likelihood. To name a few attractive properties of EL which are well-known for independent and identically distributed (iid) observations [cf. Hall and La Scala (1990)], EL confidence regions respect parameter ranges and have shapes naturally reflecting the data constitution; logarithms of EL ratios have limiting chi-square distributions, like the parametric likelihood version [cf. Owen (1990), Qin and Lawless (1994)]; and Barlett corrections can, at times, reduce the coverage error of confidence regions.

Empirical likelihood confidence intervals for the mean of a long-range dependent process

A paper to be submitted to the Annals of Statistics

Daniel J. Nordman and Soumendra N. Lahiri
[cf. DiCiccio, Hall and Romano (1990)]. Many applications of EL are provided in Owen (2001). Kitamura (1997) recently proposed a framework for EL with stationary, weakly dependent (mixing) time processes. The method applies the Owen (1990) formulation of EL to observational "blocks" rather than individual observations and relies on blocking techniques from the bootstrap and subsampling literature [cf. Carlstein (1986), Künsch (1989), Liu and Singh (1992), Politis and Romano (1992)]. With short-range dependent processes, the properties of blockwise EL are comparable to those (mentioned above) from the iid setting [Kitamura (1997)]. The aim of this paper is to investigate the asymptotic behavior of blockwise EL when the data exhibit long-range dependence (LRD). We consider EL estimation of the process mean $E(\gamma) = \mu$.

An important finding of this paper is that the blockwise EL procedure in Kitamura (1997) fails under LRD. The reason for this, simply stated, is that the usual standardized version of the sample mean, $n^{1/2}(\bar{Y}_n - \mu)$, cannot be trusted to produce a normal limit under strong dependence. Wrong "scaling" is partially to blame because slowly decaying covariances among strongly dependent observations cause the variance of the sample mean $\text{Var}(\bar{Y}_n)$ to converge to 0 at a slower, unknown rate (larger than $n^{-1}$). However, a more serious complication is that the properly standardized sample mean can have a non-normal limit under LRD. This characteristic in fact ruins the applicability of the moving block bootstrap (MBB) for inference on the mean of a strongly dependent process, as shown by Lahiri (1993). With EL methods, a basic feature from the iid or weakly dependent EL settings can collapse under LRD: the logarithm of EL ratios may not have chi-square limits for constructing confidence regions.

This paper shows however that, in situations where the MBB breaks down for inference on the process mean $\mu$, a modified blockwise EL procedure can still yield valid confidence intervals under LRD. The results are applicable to two entirely different kinds of long-range dependent (LRD) processes. To build confidence intervals for $\mu$, we combine EL techniques with "sampling window" method of Hall, Jing, and Lahiri (1998) (hereafter HJL) for estimating the sampling distribution of the standardized sample mean.

The rest of the paper is organized as follows. Section 2 formulates the strongly dependent processes considered and briefly reviews some known results on LRD. The construction of a blockwise EL ratio for the mean and its limiting distribution are given in Section 3. Section 4 provides a method for setting EL interval estimates with empirically calibrated confidence and addresses the issue of block length selection. The sampling window method of HJL (1998) is explained and applied; its consistency properties are extended. Section 5 contains the proofs of the main results.
2 Preliminaries

2.1 Process assumptions

Let \( \mathbb{Z} \) denote the set of integers and \( r_n \sim s_n \) denote \( r_n/s_n \to 1 \) as \( n \to \infty \). We suppose that the observed data \( Y_1, \ldots, Y_n \) represent a realization from a stationary, real-valued LRD process \( \{Y_t\}, t \in \mathbb{Z} \) which satisfies one of the following assumptions.

**Assumption L.** For iid innovations \( \{\varepsilon_t\}, t \in \mathbb{Z} \) with \( E(\varepsilon_t) = 0 \), \( E(\varepsilon_t^2) = 1 \), and \( E|\varepsilon_t|^k < \infty \) for \( k \geq 3 \),

\[
Y_t = \mu + \sum_{j \in \mathbb{Z}} c_{j-t} \varepsilon_j, \quad t \in \mathbb{Z}, \quad \sum_{j \in \mathbb{Z}} c_j^2 < \infty,
\]

and the autocovariance function \( r(k) = \text{Cov}(Y_t, Y_{t+k}) \) is regularly varying at infinity;

\[
r(k) \sim k^{-\alpha_1} L_1(k), \quad k \to \infty
\]

for some \( 0 < \alpha_1 < 1 \) and function \( L_1 \) slowly varying at infinity, ie. \( L_1 : (0, \infty) \to \mathbb{R} \) is positive, integrable over every finite interval, and \( \lim_{x \to \infty} L_1(\lambda x)/L_1(x) = 1 \) for all \( \lambda > 0 \) [cf. Appendix 1, Ibragimov and Linnik (1971) for other basic properties of \( L_1 \)].

**Assumption G.** The process \( \{Y_t\}, t \in \mathbb{Z} \) is a function of a long-range dependent, stationary Gaussian process \( \{Z_t\}, E(Z_t) = 0, E(Z_t^2) = 1 \): for \( t \in \mathbb{Z} \),

\[
Y_t = G(Z_t), \quad E(Y_t) = \mu,
\]

where \( G : \mathbb{R} \to \mathbb{R} \) is a Borel-measurable function. The autocovariances \( r_Z(k) = \text{Cov}(Z_t, Z_{t+k}) \) are of the form in (1) with respect to \( 0 < \alpha_Z < q^{-1} \) and a slowly varying function \( L_Z(\cdot) \) where

\[
q = \inf \left\{ k \geq 1 : E[H_k(Z_t)(G(Z_t) - \mu)] \neq 0 \right\}
\]

and \( H_k(x) = (-1)^k \exp(x^2/2)(d^k/dx^k)(\exp(-x^2/2)) \), \( x \in \mathbb{R} \) represents the \( k \)th Hermite polynomial. The function \( G(\cdot) \) satisfies \( E|G(Z_t)|^{2c} < \infty \) for some \( c > (q\alpha_Z)^{-1} \).

The processes in Assumptions L and G represent two of the most common formulations of strong dependence, which arise naturally in astronomy, economics, hydrology, and geophysics [cf. Mandelbrot and van Ness (1968), Granger and Joyeux (1980), Beran (1994)]. Unlike with weakly dependent observations which satisfy mixing conditions with a rapidly decaying mixing coefficient [cf. Künsch (1989), Politis and Romano (1992), Kitamura (1997)], the sum of the covariances diverges to infinity for LRD
processes satisfying (1) [e.g., the processes \{Y_t\} and \{Z_t\} under Assumptions L and G, respectively]. The representation of slowly decaying covariances in (1) is also equivalent to the frequent, alternative definition of LRD where the spectral density has a pole at the origin [cf. Beran (1994)].

Assumption L allows for a LRD linear process and encompasses two popular models for strong dependence: the fractional Gaussian processes of Mandelbrot and van Ness (1968) and the fractional autoregressive integrated moving average (FARIMA) models of Adenstedt (1974), Granger and Joyeux (1980) and Hosking (1981).

We adopt the LRD treatment of Taqqu (1975) in Assumption G, in which the variance of \( \bar{Y}_n \) depends heavily on the Hermite rank \( q \) of \( G(\cdot) - \mu \). HJL (1998) considered these types of processes in developing their sampling window method. Roughly speaking, treating \( r(k) = \text{Cov}(Y_t, Y_{t+k}) \) like \( r^q_Z(k) \) intuitively explains the rate of decay in the process covariances \( r(k) \) and effectively determines the size of \( \text{Var}(\bar{Y}_n) \) [cf. Taqqu (1975), Theorem 3.1, Corollary 3.1].

In the next section, we characterize two important properties of the sample mean \( \bar{Y}_n \) which can differ substantially from the short-range dependence case: the growth rate of \( \text{Var}(\bar{Y}_n) \) and the limiting distribution of standardized sample mean.

### 2.2 Asymptotic distribution of the sample mean

We describe some distributional aspects of the sample mean under LRD which will help frame our work with EL confidence intervals for \( \mu = E(Y_t) \). For independent or weakly dependent data generating processes, confidence intervals for \( \mu \) based on a normal-approximation of \( \bar{Y}_n \) are known to be asymptotically equivalent to EL confidence intervals [cf. Owen (1988)]. In some sense, EL confidence intervals for \( \mu \) can be anticipated to inherit asymptotic distributional properties from \( \bar{Y}_n \).

For both types of process admissible under Assumptions L or G, the rate of decay in autocovariances prescribed by (1) [for the process \{Z_t\} under Assumption G or the process \{Y_t\} directly under Assumption L] implies that \( \text{Var}(\bar{Y}_n) \) does not converge to 0 at the \( n^{-1} \) rate. Instead,

\[
\text{Var}(\bar{Y}_n) \sim n^{-\alpha}L(n),
\]  

for some \( 0 < \alpha < 1 \) and slowly varying function \( L \). To be precise, (2) holds for Assumption L or G processes with:

\[
\alpha = \begin{cases} 
\alpha_1 & \text{under L} \\
q_\alpha & \text{under G} 
\end{cases} \quad L(\cdot) = \begin{cases} 
((2 - \alpha_1)(1 - \alpha_1))^{-1}L_1(\cdot) & \text{under L} \\
((2 - q_\alpha)(1 - q_\alpha))^{-1}2C^2_ZL_2^g(\cdot) & \text{under G}
\end{cases}
\]
where $C_q = \mathbb{E}[H_q(Z_l)(G(Z_l) - \mu)]/\sqrt{q!}$ [cf. Taqqu (1975), Lemma 3.1, Theorem 3.1]. To ease our discussion, we will use the common notational standard set in (3) and refer simply to the relationship in (2) for either Assumption L or G processes.

The correct scaling for $a_n(\bar{Y}_n - \mu)$ to have a limit distribution is uncertain under LRD, depending on the unknown $\alpha$ and $L(\cdot)$ through (2). In contrast, $a_n = n^{1/2}$ is the known, correct scaling with short-range dependence. A more serious problem with inference based on $\bar{Y}_n$ under LRD is that, even if correctly standardized, the sample mean may have a non-normal limit. The following result, due to Davydov (1970) [Assumption L] and Taqqu (1975, 1979)-Dobrushin and Major (1979) [Assumption G], makes this precise. Let $d_n = \{n^{-\alpha}L(n)\}^{1/2}$.

**Theorem 1** Let $\{Y_t\}, t \in \mathbb{Z}$ be a process which satisfies Assumption L or G.

a) Under L, $(\bar{Y}_n - \mu)/d_n \overset{d}{\rightarrow} Z_1$, a standard normal random variable.

b) Under G, $(\bar{Y}_n - \mu)/d_n \overset{d}{\rightarrow} W_q$, where the random variable $W_q$ is defined by a multiple Wiener-Itô integral with respect to the random spectral measure $W$ of the Gaussian white-noise process:

$$W_q = \{A^q \hat{A}\}^{-1/2} \int \frac{\exp\{i(x_1 + \cdots + x_q)\} - 1}{i(x_1 + \cdots + x_q)} \prod_{k=1}^{q} |z_k|^{(\alpha_z - 1)/2} dW(z_1) \cdots dW(z_q)$$

for $A = 2\Gamma(\alpha_z) \cos(\alpha_z \pi/2)$, $\hat{A} = 2/((2 - q\alpha_z)(1 - q\alpha_z))$.

When $q = 1$, $W_q$ is equal in distribution to a standard normal $Z_1$. However, the distribution of $W_q$ is non-normal for $q \geq 2$ [cf. Taqqu (1975)]. For further details on the stochastic representation of $W_q$ and the concept of a multiple Wiener-Itô integral with respect to the random spectral measure of stationary process, see Dobrushin and Major (1979) and Dobrushin (1979), respectively.

The moving block bootstrap (MBB) approximation of $\bar{Y}_n$ can fail terribly for LRD processes specifically because the bootstrap sample mean is always asymptotically normal while the true sample mean $\bar{Y}_n$ might not have a normal limit [Lahiri (1993)]. (In contrast, the MBB works very well for weakly dependent processes in which a central limit theorem holds.) Given the strong parallels often drawn between the bootstrap and EL [cf. Owen (1990), Kitamura (1997)], one might expect a non-normal asymptotic distribution for $\bar{Y}_n$ to also complicate standard EL techniques, based on normal limit theory and chi-square distributions. We show that this is indeed the case in the next section.
3 Asymptotic distribution of blockwise empirical likelihood

3.1 Construction of blockwise empirical likelihood for the mean

We now construct an EL function and associated EL profile ratio for the process mean \( E(Y_i) = \mu \). Our EL formulation uses observational blocks as in Kitamura (1997) to capture the process dependence structure.

Let \( 1 \leq \ell \leq n \) be the block length; let \( B_i = (Y_i, \ldots, Y_{i+\ell-1}) \) denote the \( i \)th data block for \( 1 \leq i \leq N = n - \ell + 1 \); and let \( M_i \) be the sample mean of the elements in \( B_i \). We use the maximal possible number \( N \) of length \( \ell \) blocks among the observed time sequence \( Y_1, \ldots, Y_N \). Valid EL confidence intervals are possible with a non-overlapping block scheme as well, where we use only the blocks \( B_{(i-1)+1}, \ldots, B_i \), \( 1 \leq i \leq n/\ell \). However, we focus our presentation on the maximal block version.

We consider assigning probabilities \( \{p_i\}_{i=1}^N \) to each block sample mean \( \{M_i\}_{i=1}^N \) under a process mean-based restriction and examine the corresponding (multinomial) likelihood function: \( \prod_{i=1}^N p_i \). The profile blockwise EL function for \( \mu \) is given by

\[
L_n(\mu) = \sup \left\{ \prod_{i=1}^N p_i : p_i > 0, \sum_{i=1}^N p_i = 1, \sum_{i=1}^N p_i M_i = \mu \right\}
\]

where we maximize the product of probabilities satisfying a "mean \( \mu \)" linear constraint. If the conditioning set is empty for some \( \mu \in \mathbb{R} \), we can define \( L_n(\mu) = -\infty \). When positive, \( L_n(\mu) \) is a maximum realized at unique weights \( p_i = N^{-1} \{1 + \lambda_\mu(M_i - \mu)\}^{-1} \) where \( \lambda_\mu \) is determined by

\[
\sum_{i=1}^N (M_i - \mu) \{1 + \lambda_\mu(M_i - \mu)\}^{-1} = 0;
\]

these \( p_i \) values result from using Lagrange multipliers to find the constrained extrema in (4). Owen (1988, 1990) discusses these and further computational aspects of EL.

Without constraints, the product \( \prod_{i=1}^N p_i \) is maximized when each \( p_i = N^{-1} \), corresponding to the empirical distribution of \( \{M_i\}_{i=1}^N \). The profile empirical likelihood ratio for the mean \( \mu \) is then

\[
R_n(\mu) = L_n(\mu)/N^{-N} = \prod_{i=1}^N \{1 + \lambda_\mu(M_i - \mu)\}^{-1}.
\]

A confidence interval for \( \mu \) is then determined by those values of \( \mu \) with relatively high EL, ie. sets of the form

\[
\{\mu : R_n(\mu) \geq A\},
\]

where \( A > 0 \) is chosen by some distributional "calibration" to set a desired confidence level. It can be verified that the set in (6) is convex [cf. Hall and La Scala (1990), Theorem 2.2] and therefore represents a true confidence interval. We next discuss the asymptotic distribution of log-EL ratio under LRD.
3.2 Limit distribution of empirical likelihood ratio

One highly celebrated feature of EL with independent (or even weakly dependent but blocked) data is that it allows a nonparametric casting of Wilk's theorem [Wilks (1938)] for constructing confidence regions. That is, the logarithm of the EL ratio has an asymptotic chi-square distribution when evaluated at a true parameter value [cf. Owen (1988, 1990, 1991) for iid sampling; Kitamura (1997) for mixing processes].

We now give a substantially different result on the asymptotic distribution of the blockwise EL ratio in (5) under LRD. As in the EL framework of Kitamura (1997), we require a correction factor to adjust for the blocks used in place of individual observations in (4) and ensure our statistics have correct (non-degenerate) large-sample distributional properties. Write \( B_n = \left( \frac{d^2}{d^2 n} \right) N^{-1} \) for the block adjustment term.

**Theorem 2** Suppose \( \{Y_t\}, t \in \mathbb{Z} \) satisfies either Assumption L or G. Let \( E(Y_t) = \mu_0 \) denote the true process mean and suppose \( \ell^{-1} + n^{-(1-2\delta)/2} = o(1) \) for some \( \delta \in (0, 1/2) \). Then,

\[
-2B_n \log R_n(\mu_0) \xrightarrow{d} Y^2_\infty,
\]  

(7)

where \( Y^2_\infty \) is the limiting distribution of \( (\bar{Y}_n - \mu_0)/\sqrt{n} \).

If the process \( \{Y_t\} \) satisfies Assumption L or G with \( q = 1 \), then the "standard" Wilks distributional result follows in (7) because \( Y^2_\infty \) is a chi-square distribution \( \chi^2_1 \) by Theorem 1. The 1 degree of freedom owes to the fact that the parameter \( \mu \in \mathbb{R} \). However, the log-EL ratio in Theorem 2 will have a non-chi-square limit distribution precisely whenever the sample mean \( \bar{Y}_n \) is asymptotically non-normal. In practice, the exact limit distribution \( Y^2_\infty \) is uncertain.

In its mechanics, the EL ratio in (7) asymptotically uses a subsample-based estimator of \( \text{Var}(\bar{Y}_t) \) to estimate \( \text{Var}(\bar{Y}_n) \) and requires a correction \( B_n \) involving \( \text{Var}(\bar{Y}_t)/\text{Var}(\bar{Y}_n) \). Lahiri (1993) demonstrates that a similar problem arises with the MBB sample mean under LRD and produces a degenerate bootstrap approximation of \( \bar{Y}_n \) unless corrected by rescaling. The block adjustment for the log-EL ratios in Kitamura (1997) [see p. 2089] involves \( n/\ell \) since \( n/\ell \sim \text{Var}(\bar{Y}_t)/\text{Var}(\bar{Y}_n) \) under short-range dependence. That is, if the process autocovariances are absolutely summable under mixing conditions (or even independence), the variance of \( \text{Var}(\bar{Y}_n) \) is known to be asymptotically proportional to \( n^{-1} \) [Fuller (1996), Corollary 6.1.1.2]. However, (2) implies that the \( n/\ell \)-based block correction from the weak dependence case is inappropriate under LRD, requiring a different correction \( B_n \) in (7).

In contrast to the iid or weakly dependent data scenarios, both \( B_n \) and \( Y^2_\infty \) are unknown under LRD. Of course, if we knew the exact block adjustment and limit law in (7), then we could readily...
obtain EL confidence intervals for $\mu$ with many of the same desirable properties mentioned in Hall and La Scala (1990), but perhaps without using a chi-square distribution to calibrate confidence levels. Likelihood-based inference on $\mu$, involving regions like (6) of high confidence, has not been invalidated. We just require a way to pick $A$ in (6) to establish a desired confidence level.

As it turns out, we can set the confidence level of EL confidence intervals for $\mu$ by using Theorem 2 with empirical (nonparametric) tools from the "sampling window" method of HJL (1998). Even in the presence of strong dependence, we can still apply EL principles to obtain asymptotically correct confidence intervals for $\mu$.

4 Empirical calibration of empirical likelihood confidence

4.1 Sampling window method

We briefly present the "sampling window" estimator of the sampling distribution $F_n(x)$ of the normalized sample mean $(\bar{Y}_n - \mu)/d$, proposed by HJL (1998). Those authors formulated the subsample-based estimator for strongly dependent processes prescribed by Assumption G. We will use some of their subsampling devices for distribution and variance estimation to next formulate EL confidence intervals for $\mu$.

The sampling window estimator of the distribution $F_n(x)$ is given by

$$\hat{F}_n(x) = N^{-1} \sum_{i=1}^{N} I(\{(M_i - \bar{Y}_n)/d_i < x\}),$$

where $I(\cdot)$ denotes the indicator function. The estimator treats block analogs $(M_i - \bar{Y}_n)/d_i$ as scaled-down replicates of the standardized sample mean $(\bar{Y}_n - \mu)/d$ to create a sample proportion: the number of block evaluations $(M_i - \bar{Y}_n)/d_i$ not exceeding $x$.

There is, of course, the complication that $\hat{F}_n$ involves $d_n$ and thereby depends on unknown quantities $\alpha$ and $L(\cdot)$ under LRD. To handle this problem, HJL (1998) introduce a device to consistently estimate both $d_n$ (and $d_\ell$) via subsampling.

To describe the method, write the length $m$ sample mean $M_{m_i} = m^{-1} \sum_{j=i}^{i+m-1} Y_j$ for integers $m_i$, $i \geq 1$; define a block-based estimator $d_m^2 = (n - m + 1)^{-1} \sum_{i=1}^{n-m+1} (M_{m_i} - \bar{Y}_n)^2$ of $d_n^2$. (To avoid confusion, we note that $d_m^2$ here differs slightly from HJL (1998) because we use subsample means, not sums.) Let $m_{1n}, m_{2n} \in [1,n]$ and $m_{1\ell}, m_{2\ell} \in [1,\ell]$ denote sequences of integers such that for some $\epsilon \in (0,1)$: as $n \to \infty$,

$$m_{1n}^2/m_{2n} \sim n, \quad m_{1\ell}^2/m_{2\ell} \sim \ell, \quad m_{1n}, m_{1\ell} = O(n^{1-\epsilon}), \quad i = 1,2.$$  \hspace{1cm} (8)
Define estimators $\hat{d}_n^2 = d_{n_{11}}^2 / d_{n_{22}}^2$ and $\hat{d}_t^2 = d_{m_{11}}^2 / d_{m_{22}}^2$.

Following HJL (1998), we next define the sampling window estimator of the studentized sample mean: $(\hat{Y}_n - \mu) / \hat{d}_n$. Let $\hat{d}_t^{(i)}$ denote the analog of $\hat{d}_t$ defined on the $i$th block $B_i$ by using the smoothing parameters $m_{1t}$ and $m_{2t}$ specified in (8). Then, each $B_i$ yields a studentized block sample mean $(M_i - \bar{Y}_n) / \hat{d}_t^{(i)}$ for inference on the sampling distribution, say $F_{1n}$, of $(\hat{Y}_n - \mu) / \hat{d}_n$. Write the sampling window estimator of $F_{1n}(x)$ as:

$$\hat{F}_{1n}(x) = N^{-1} \sum_{i=1}^{N} I\left\{ (M_i - \bar{Y}_n) / \hat{d}_t^{(i)} \leq x \right\}.$$ 

The consistency of sampling distribution estimation via subsampling is well-known in the independent and weakly dependent (mixing) sample scenarios where $d_n = Cn^{-1/2}$ [cf. Politis and Romano (1994), Hall and Jing (1996), Lahiri (1996)]. HJL (1998) prove that their sampling window estimators successfully approximate the distributions $F_n$ and $F_{1n}$ for long-memory processes under Assumption G. We now extend the consistency of $\hat{F}_n$ and $\hat{F}_{1n}$ to include LRD non-Gaussian, linear processes [under Assumption L]. Although our goal is EL confidence intervals, the consistency result for the sampling window estimators is important in its own right because it shows the general applicability of the "sampling window" method. For completeness, we include the results of HJL (1998).

Theorem 3 Let $\{Y_t\}, t \in \mathbb{Z}$ satisfy either Assumption L or G and $\ell^{-1} + n^{-(1 - \delta)} = o(1)$ for some $\delta \in (0, 1)$.

a) Then, under Assumption L or G with $q = 1, 2$,

$$\sup_{x \in \mathbb{R}} |\hat{F}_{1n}(x) - F_{1n}(x)| \xrightarrow{p} 0, \quad \text{as } n \to \infty; \quad (9)$$

if in addition conditions in Theorem 2.2, HJL (1998) hold, then (9) follows for all $q \geq 1$.

b) Suppose $m_{1n}, m_{2n}, m_{1t}, m_{2t}$ satisfy (8) and

$$L^2(xy)/\{L(x^2)L(y^2)\} \to 1 \quad \text{as } x, y \to \infty. \quad (10)$$

Then, $\hat{d}_n / d_n, \hat{d}_t / d_t \xrightarrow{p} 1$ as $n \to \infty$; under Assumption L or G with $q = 1, 2$,

$$\sup_{x \in \mathbb{R}} |\hat{F}_{1n}(x) - F_{1n}(x)| \xrightarrow{p} 0, \quad \text{as } n \to \infty; \quad (11)$$

if in addition conditions in Theorem 2.2, HJL (1998) hold, then (11) follows for all $q \geq 1$.

We remark that Theorem 2.2 of HJL (1998) involves assumptions on the process dependence strength for $\{Z_t\}$, steeped in terms of the spectral density $f_Z$ rather than through the covariances $r_Z(t)$. The
representation of covariances $r_Z(\cdot)$ which satisfy (1) [with $\alpha_Z$, $L_Z(\cdot)$] is equivalent to one implying $f_Z$ has a pole at zero [cf. Zygmund (1968), Beran (1994)].

The condition in (10), due to HJL (1998), implies that $\tilde{d}_m$ can consistently estimate $d_m$ under LRD. For fractional Gaussian and FARIMA long-range dependent processes, the function $L(\cdot) = C_\alpha$ is constant in (2) and so easily satisfies (10) [cf. Fox and Taqqu (1986)].

4.2 Confidence intervals for the mean

By modifying components from the sampling window method, we now give a construction for EL confidence intervals of $\mu$ with LRD processes. We first replace $B_n$ with a function of the data which requires no knowledge of $\alpha$ and $L(\cdot)$: $\tilde{B}_n = (d_t^2/d^2_n)N^{-1}$, using the smoothing parameters from (8). We can also define an alternative estimator of $B_n$ with: $\tilde{B}_2n = (d_t^2/d^2_n)N^{-1}$.

To make EL confidence intervals with approximate confidence level $\beta \in (0,1)$, we estimate the appropriate quantile from the distribution of $Y_n^2$ in (7). Let $F(x)$ denote the continuous distribution of $Y_n$. Write $Q_\beta = \inf\{x > 0 : F(x) - F(-x) \geq \beta\}$ so that $[Q_\beta]^2$ is the $\beta$-quantile of $Y_n^2$, ie. $P(Y_n^2 \leq [Q_\beta]^2) = \beta$.

By Theorem 3, $\tilde{F}_n(x)$ [free of unknown quantities] consistently estimates $F_n(x)$ which in turn will be close to $F(x)$ as $n$ increases. Hence, define $\tilde{Q}_{1n\beta} = \inf\{x > 0 : \tilde{F}_n(x) - \tilde{F}_n(-x) \geq \beta\}$ to estimate $Q_\beta$. We can also define different, but completely data-based, estimators of $Q_{\beta}$ through new sampling window estimators of $F_n(x)$ (and $F(x)$): let

$$\tilde{F}_{jn}(x) = N^{-1}\sum_{i=1}^{N} I\{(M_i - \tilde{Y}_n)/\tilde{d}_j \leq x\}, \quad \tilde{d}_j = \begin{cases} d_t^2 & \text{if } j = 2 \\ \tilde{d}_j & \text{if } j = 3 \end{cases}$$

and define $\tilde{Q}_{jn\beta} = \inf\{x > 0 : \tilde{F}_{jn}(x) - \tilde{F}_{jn}(-x) \geq \beta\}$ for $j = 2$ or $3$.

Choose $B_n \in \{\tilde{B}_n : j = 1, 2\}$ and $Q_{n\beta} \in \{\tilde{Q}_{jn\beta} : j = 1, 2, 3\}$. We can now write an EL confidence interval for the mean $\mu$ of a LRD process, which is of the form in (6):

$$C_{\mu,n}(\beta) = \left\{ \mu : -2\tilde{B}_n \log R_n(\mu) \leq [\tilde{Q}_{n\beta}]^2 \right\}.$$ (12)

The above EL interval is empirically calibrated with approximate confidence level $\beta$ and has asymptotically correct coverage. Unlike with the MBB [cf. Lahiri (1993)], EL allows valid estimation of $\mu$ under LRD even when the sample mean $\tilde{Y}_n$ has a non-normal limit. We show this in the subsequent theorem.

Theorem 4 Assume the conditions of Theorem 1, that $q \leq 2$, that $m_{jn}$, $m_{jt}$ satisfy (8), and that (10) holds. Then as $n \to \infty : \tilde{B}_n/B_n \to 1$; $\tilde{Q}_{n\beta} \to Q_\beta$; and denoting the true process mean $E(Y_t) = \mu_0$,

$$P(\mu_0 \in C_{\mu,n}(\beta)) \to \beta.$$
If in addition the conditions in Theorem 2.2, HJL (1998) hold for Assumption G processes, the convergence results follow for all $q \geq 1$.

We remark that the EL confidence intervals in (19) are also valid under weak dependence [with the same mixing conditions and summable covariances in the EL set-up of Kitamura (1997)]. Short-range dependence can be accommodated in our EL formulation by letting $\alpha = 1$ and $L(n) = \sigma_\infty \equiv \sum_{k \in \mathbb{Z}} r(k) > 0$ in (2). This choice provides the same block adjustment $B_n$ from Kitamura (1997), depending on $n/\ell$, and also implies that $\text{Var}(\bar{Y}_n)$ is asymptotically proportional to $n^{-1}\sigma_\infty$ in (2) and that $a_n = n^{1/2}\sigma_\infty^{-1/2}$ is used to scale $a_n(\bar{Y}_n - \mu)$ to attain a proper limit distribution... which are correct! We note that Politis and Romano (1994) have already shown the validity of the subsample-based sampling window estimator under mixing. Hence, our approach to setting EL estimate intervals can also side-step classification of strong vs. weak process dependence.

4.3 Empirical block selection

The EL confidence intervals for $\mu$ in (19) require the selection of block lengths for $\ell, m_{1n},$ and (possibly) $m_{jt}$. Choices for $m_{1n}, m_{jt}$ are fairly straightforward to handle, given a block length $\ell$. As HJL (1998) state, plausible values are $m_{1n} = n^{(1+\theta)/2}$ and $m_{2n} = n^{\theta}$ (or $m_{1t} = \ell^{(1+\theta)/2}, m_{2t} = \ell^\theta$) for $\theta \in (0, 1)$. Picking $\theta$ near 1 will keep the block sizes $m_{1n}, m_{jt}$ large and sharpen the accuracy of $B_n$ (as large $m$ values reduce the bias of $d_n^2$ [pointed out in HJL (1998)]).

Recommendations for selecting $\ell$ to determine the EL ratio for $\mu$ in (5) are more difficult. Even with mixing processes, where the (known) adjustment factor $B_n$ and (5) depend only on $\ell$, the effect of block size is not well-understood with blockwise EL [cf. Kitamura (1997)]. Experience with blocking in subsampling and bootstrap applications indicate that $\ell \sim cn^d, d \leq 1/3$ is usually appropriate (even optimal) for purposes of variance, bias, or sampling distribution estimation under short-range dependence [cf. Künsch (1989), Hall, Horowitz, and Jing (1995), Hall and Jing (1996), Lahiri (1996)]. However, intuition would indicate that blocks should be longer under strong dependence, which HJL (1998) propose. We can take $\ell = cn^{d/2}$ with large $\phi \in (0, 1)$ to produce large blocks under Theorem 4 conditions. HJL (1998) make a similar suggestion finding block lengths such as $cn^{1/2}$ to be effective in their “sampling window” estimator with Assumption G processes.
5 Proofs

5.1 Proof of Theorem 3

We consider only the linear processes under Assumption L. HJL (1999) establish the case involving Assumption G processes. Before proving Theorem 3, we state some useful lemmas concerning moments of the sample mean $Y_n$ from a LRD linear process.

Lemma 1 (from Theorem 18.6.5, Ibragimov and Linnik (1971)) For a sequence of iid random variables $\{e_t\}$ with $E(e_t) = 0$, $E(e_t^2) = 1$, let $Y_t = \mu + \sum_{j \in \mathbb{Z}} c_{j-t} e_j$, $t \in \mathbb{Z}$ with $\sum_{j \in \mathbb{Z}} c_j^2 < \infty$. If $\sigma^2_n = n^2 \text{Var}(\tilde{Y}_n) \to \infty$ as $n \to \infty$, then for each $k \in \mathbb{Z}$ and $n \geq 1$:

$$\sigma^{-1}_n \left| \sum_{j=1}^{n} c_{k-j} \right| \leq a_n = \left\{ \sigma^{-1}_n \left( 4 + 2\sigma^{-1}_n \right) \sum_{j \in \mathbb{Z}} c_j^2 \right\}^{1/2}.$$  \hspace{1cm} (13)

Lemma 2 (Lemma 4, Davydov (1970)) Let $\{Y_t\}$, $t \in \mathbb{Z}$, be the linear process in Lemma 1 and $k \geq 1$. If $E(|e_t|^{2k}) < \infty$, then for all $n \geq 1$ and some $A_k > 0$

$$E\left\{ n(\tilde{Y}_n - \mu) \right\}^{2k} \leq A_k \left( \sigma_n^2 \right)^k.$$  \hspace{1cm} (14)

Proof of Theorem 3 a). We use the same notational standard from (3). The rate of covariance decay in (1) implies that

$$\sigma^2_n = n^2 \text{Var}(\tilde{Y}_n) \sim n^2d^2_n$$ \hspace{1cm} (15)

and Theorem 1 implies $(\tilde{Y}_n - \mu)/d_n$ has a limiting standard normal distribution function $F(x) = \Phi(x)$.

Let $F^*_n(x) = N^{-1} \sum_{i=1}^{N} \mathbb{I}\{ (M_i - \mu)/d_i \leq x \}$. Denote the supremum norm $\|g\|_{\infty} = \sup\{|g(x)| : x \in \mathbb{R}\}$ for a real valued function $g : \mathbb{R} \to \mathbb{R}$. For each $\epsilon > 0$, we can write

$$\|F_n - F\|_{\infty} \leq \|\hat{F}_n - F\|_{\infty} + \|F_n - \hat{F}\|_{\infty}.$$ \hspace{1cm} (16)

Because $L$ is positive and $x^\gamma L(x) \to \infty$, $x^{-\gamma} L(x) \to 0$ as $x \to \infty$ for any $\gamma > 0$ [cf. Appendix 1, Ibragimov and Linnik (1971)], we have for large $n$:

$$(d^2_n)/(d^2_n) \leq C(n/\ell)^{-a(L(n)/L(\ell))} \leq Cn^{-ad}L(n)/L(\ell) = o(1).$$ \hspace{1cm} (17)

From this, $P(|\tilde{Y}_n - \mu|/d_n > \epsilon) = o(\epsilon)$ by Chebychev's inequality and (17); the continuity of $F$ and Theorem 1 imply $\|F_n - F\|_{\infty} = o(1)$ and also $\sup_{x \in \mathbb{R}} |F(x + \epsilon) - F(x)| = o(1)$ as $\epsilon \to 0$; and hence it suffices now to show

$$E\left\{ \sup_{x \in \mathbb{R}} |F^*_n(x) - F(x)| \right\} \to 0.$$ \hspace{1cm} (18)
However, note that: $F = \Phi$ is continuous; a standard normal $Z$ is determined by its moments; (15) holds; \( \sigma_n/(nd_n) \to 1 \) by (17); and Theorem 1 and Lemma 2 imply that, for all $k \geq 1$,

$$
E \left\{ \left( (Y_n - \mu)/(dn) \right)^k \right\} \to E(Z^k) = \begin{cases} 0 & k \text{ odd} \\ \frac{(k-1)(k-3) \cdots (1)}{k} & k \text{ even}. \end{cases} \tag{17}
$$

Hence, it stands that (16) will follow by establishing [from Theorem 2.4, HJL (1998)]: for any positive integers $a, b$ and $0 < \epsilon < 1$

$$
\max_{n \geq 1} \left| E \left[ (S_1)^a (S_i)^b \right] - E(Z^a)E(Z^b) \right| = o(1) \quad \left( S_i = \ell(M_i - \mu)/\sigma_i \right). \tag{18}
$$

To prove (18), we consider 3 cases.

Some additional notation is first required. Write for $i \geq 1$,

$$
S_i = \sum_{j \in Z} d_{j(i)} \epsilon_j, \quad d_{j(i)} = \sigma_i^{-1} \sum_{k=1}^\ell c_{j-(i-1)-k}.
$$

We suppress the dependence of $d_{j(i)}$ on $\ell$ above. Write

$$
r_{\ell}(i) \equiv \sigma_i^{-2} \text{Cov}(S_1, S_i) = \sum_{j \in Z} d_{j(i)} d_{j(i)}' \quad i \geq 1 \quad \left( r_{\ell}(1) = 1 \right). \tag{19}
$$

We point out that it cannot be assumed that $\sum_{j=1}^\infty |c_j| < \infty$ or $\sum_{j=1}^\infty |d_{j(i)}| < \infty$, which complicates the arguments to follow.

For $Z^2 = \{0, 1, \ldots\}$, define an ordering on $(r, s), (u, v) \in Z^2$ as follows: $(r, s) \succeq (u, v)$ if and only if $r + s \geq u + v$; or $r + s = u + v$, max\{r, s\} $> \text{max}\{u, v\}$; or $r + s = u + v$, max\{r, s\} $= \text{max}\{u, v\}$, $r \geq u$. Letting $\lambda_m = (r_1, s_1, \ldots, r_m, s_m)$, define the set

$$
P_{ab} = \bigcup_{m=1}^{[a+b/2]} \left\{ \lambda_m \in Z_m^2 : r_j + s_j \geq 2, (r_j, s_j) \succeq (r_{j+1}, s_{j+1}), \sum_{j=1}^m r_j = a, \sum_{j=1}^m s_j = b \right\},
$$

where $a, b \geq 1$ are integers and $[\cdot]$ denotes the integer part function. For $a, b, i \geq 1$ we can write

$$
E \left[ (S_i)^a (S_i)^b \right] = \sum_{\lambda_m \in P_{ab}} K_1(\lambda_m) K_2(\lambda_m) \Delta_i(\lambda_m),
$$

$$
\Delta_i(\lambda_m) = \sum_{k_1, \ldots, k_m \in Z} \prod_{j=1}^m (d_{k_1(i)})^{r_j} (d_{k_2(i)})^{s_j}, \quad K_1(\lambda_m) = \prod_{j=1}^m E(\epsilon_i^{r_j+s_j})
$$

where $K_2(\lambda_m)$ denotes the number of partitions $(W_1, \ldots, W_m)$ of the set $\{y_{k,i} : k = 1, 2; 1 \leq i \leq a; 1 \leq j \leq b\}$ such that $\sum_{j=1}^a I\{y_{1j} \in W_k\} = r_k$, $\sum_{j=1}^b I\{y_{2j} \in W_k\} = s_k$ for each $1 \leq k \leq m$. We used above that $E(\epsilon_i) = 0$ to formulate $r_j + s_j \geq 2$ in $P_{ab}$.
**Case 1: a + b is odd.** Note \(E(Z^a)E(Z^b) = 0\). We now proceed with an induction argument, involving the count \(c(\lambda_m) = \sum_{j=1}^m f(r_j = s_j = 1)\), to show that for any \(\lambda_m \in \mathcal{P}_{ab}\),

\[
\max_{n \leq s \leq \max t} |\Delta_i(\lambda_m)| = o(1). \tag{20}
\]

Because \(K_1, K_2\) are bounded and \(\mathcal{P}_{ab}\) is finite for each \(a, b\), (18) will then follow from (20) for this considered case.

Since \(\sum_{j=1}^m (r_j + s_j) = a + b\) is odd, there exists a pair \((r_j, s_j)\) in \(\lambda_m \in \mathcal{P}_{ab}\) such that \(r_j + s_j \geq 3\), implying \(c(\lambda_m) \leq m - 1\). In the first phase of the induction, we suppose \(c(\lambda_m) = 0\) and demonstrate that (20) holds. We have when \(c(\lambda_m) = 0\)

\[
|\Delta_i(\lambda_m)| \leq \prod_{j=1}^m \sum_{k \in \mathbb{Z}} |d_k(1)|^{r_j} |d_k(1)|^{s_j} \leq a \sum_{j=1}^m (r_j + s_j - 2) = O\left(a^{r_j/2}\right) = o(1), \tag{21}
\]

where the first inequality follows from \(\min_{1 \leq j \leq m} \max \{r_j, s_j\} \geq 2\) by \(c(\lambda_m) = 0\) and the second inequality results from \(r_j(1) = 1\) and (13). We now make the induction assumption that (20) holds for any \(\lambda_m \in \mathcal{P}_{ab}, c(\lambda) \leq t \in \mathbb{Z}_+\) and all \(a, b \geq 1\) with \(a + b\) odd. We will establish (20) for \(\lambda_m \in \mathcal{P}_{ab}, c(\lambda_m) = t + 1\) and \(a + b\) odd (for which \(m \geq t + 2, a + b \geq 2m + 1\) necessarily) under the previous induction hypothesis.

If \(a = 1\) or \(b = 1\) and \(c(\lambda_m) > 0\), we find \(c(\lambda_m) = 1\) so that using (19) and some algebra

\[
\Delta_i(\lambda_m) = \sum_{k_1 \neq \cdots \neq k_{m-1} \in \mathbb{Z}} \prod_{j=1}^{m-1} \left(r_j(i) - \sum_{s=t}^{m-1} d_s(i) d_{s+t(i)}(i)\right) \tag{22}
\]

where \(r_j = 0, s_j \geq 2\) if \(a = 1\), or \(r_j \geq 2, s_j = 0\) if \(b = 1\), for each \(1 \leq j \leq m - 1\). Applying \(|r_t(i)| \leq 1\) by the Cauchy-Schwartz inequality, (13), and (19):

\[
|\Delta_i(\lambda_m)| \leq a^{a+b-2(m-1)} \{1 + (m-1)a_t^2\} = O(a_t) = o(1). \tag{23}
\]

We may proceed assuming \(a, b > 1\). Then repeating the expansion in (22):

\[
\Delta_i(\lambda_m) = r_t(i) \Delta_i(\lambda_{m0}) - \sum_{j=1}^{m-1} \Delta_i(\lambda_{mj}) \tag{24}
\]

denoting \(\lambda_{mj}, j \geq 1\), as a vector equivalent to \((r^{(1)}, s^{(1)}), \ldots, r^{(m)}, s^{(m)}\) after ordering the pairs \((r_j^{(i)}, s_j^{(i)})\) by the \(\geq\) relation (where \((r_j^{(i)}, s_j^{(i)}) = (r_j + 1, s_j + 1)\) and \((r_i^{(i)}, s_i^{(i)}) = (r_i, s_i)\) for \(i \neq j\)) and writing \(\lambda_{m0} = (r_1, s_1, \ldots, r_{m-1}, s_{m-1})\). Since \(\lambda_{mj} \in \mathcal{P}_{ab}, j \neq 0\) with \(c(\lambda_{mj}) \leq t\), we find \(\max_{1 \leq j \leq m-1} \max_{n \leq s \leq \max t} |\Delta_i(\lambda_{mj})| = o(1)\) by the induction assumption; likewise, \(\lambda_{m0} \in \mathcal{P}_{ab-1}\) with \(c(\lambda_{m0}) = t\) and \(a + b - 2 \geq 1\) is odd so (20) holds for \(\lambda_{m0}\). Because \(|r_t(i)| \leq 1\), we see (20) holds now for \(\lambda_m \in \mathcal{P}_{ab}\) and the proof of Case 1 is complete.
Case 2: a, b odd. As in Case 1, it is enough to show (20) holds. We begin by examining a few preliminary cases.

Consider first $\lambda_m \in \mathcal{P}_{ab}$, $m = 1$. If $a + b > 2$, then $r_1 = a > 2$ or $s_1 = b > 2$ so that (21) follows. When $m = 1$ and $a + b = 2$, then $\Delta_i(\lambda_1) = r_\ell(i)$. Suppose $n$ is large so that $ne/2 > \ell$. The growth rate of $r(k)$ in (21) gives

$$\max_{ne \leq i \leq n} |r_\ell(i)| \leq C\ell^{\alpha} \frac{n}{\ell} M_{n,\ell}$$

where for $M_{n,\ell} = \sup \{L(tn) : \epsilon/2 < t < 2\}$, $M_{n,\ell}/L(n) \rightarrow 1$ by Taqqu (1977, Lemma A1). Hence, (20) is valid for any $\lambda_m \in \mathcal{P}_{ab}$ with $m = 1$ and $a, b$ odd.

Next suppose $a = 1, b > 1$ or $b = 1, a > 1$ and $m > 1$. For any $\lambda_m \in \mathcal{P}_{ab}$ with $m > 1$, there is exactly one $1 \leq j \leq m$ with $r_j = 1$ and $s_j \geq 2, l \neq j$ when $a = 1$ or otherwise one $s_j = 1, r_j \geq 2, l \neq j$ when $b = 1$. If $c(\lambda_m) = 0$ and $m > 1$, then all $r_j \geq 2$ in $\lambda_m$ (if $b = 1$) or all $s_j \geq 2$ (if $a = 1$) and furthermore it must be that $2m < a + b$ (since $2m = a + b$ implies each $r_j + s_j = 2$ which cannot hold here); in which case, (21) holds. If $c(\lambda_m) > 0$ then actually $c(\lambda_m) = 1$ and $m > 1$, so that (22) and (23) follow sequentially. We have now established (20) for $a, b$ odd if either $a = 1$ or $b = 1$.

We may now assume $a, b > 1$ are odd and show (20) holds for any $\lambda_m \in \mathcal{P}_{ab}, m > 1$. As in Case 1, we will use a proof by induction on $c(\lambda_m)$. We start by handling the possibility that $c(\lambda_m) = 0$. When $c(\lambda_m) = 0$ and $m > 1$, there exists some $r_j + s_j \geq 3$ in $\lambda_m$ (implying again $2m < a + b$); if not, then $r_j + s_j = 2$ for all $j$ so that $r_j = 0$ or 2 for each $j$ by $c(\lambda_m) = 0$ and hence $\sum_{j=1}^m r_j = a$ is even, a contradiction. Hence, we find (21) follows when $c(\lambda_m) = 0$ for the same reasons given in Case 1.

We make the induction assumption that (20) is valid for any $\lambda_m \in \mathcal{P}_{ab}$, where $a, b > 1$ are odd, $c(\lambda_m) \leq t \in \mathbb{Z}_+$. We show that the induction hypothesis implies (20) holds for $\lambda_m \in \mathcal{P}_{ab}$ with $c(\lambda_m) = t + 1, a, b > 1$ odd. We can assume $m > 1$ (the $m = 1$ situation being already handled) and decompose $\Delta_i(\lambda_m)$ as in (24). Using the same notation, we have $\lambda_m \in \mathcal{P}_{ab}$ with $c(\lambda_m) \leq t$, so that $\max_{1 \leq j \leq m-1} \max_{ne \leq i \leq n} |\Delta_i(\lambda_m)| = O(1)$ by the induction assumption. The $|\Delta_i(\lambda_m)|$ term cannot be handled under the induction hypothesis because $\lambda_m \in \mathcal{P}_{a-1b-1}$ with $c(\lambda_m) = t$ and $a - 1, b - 1 \geq 2$ even. However, we find $\max_{ne \leq i \leq n} |\Delta_i(\lambda_m)| = O(1)$ by applying (25) and (26) [below]. We have now that (20) holds for $\lambda_m \in \mathcal{P}_{ab}, c(\lambda_m) = t + 1$, which completes the induction argument and the proof of (18) for Case 2.

We briefly establish the following claim:

If $\lambda_m \in \mathcal{P}_{ab}, a, b \geq 1$, then $\max_{ne \leq i \leq n} |\Delta_i(\lambda_m)| = O(1)$.

(26)
We may suppose \( a, b \geq 2 \) because \( \max_{n \leq \ell < n} |\Delta_i(\omega_m)| = o(1) \) for \( a = 1 \) or \( b = 1 \) has been fully established (in the treatments of Case 1 or \( a, b \) odd). If \( m = 1 \), then \( r_i = a, s_i = b \geq 2 \) and \( (21), (26) \) hold. Assume \( (26) \) is true for any \( \lambda_m \in P_{ab} \) with \( a, b \geq 1 \) and \( m \leq t \) for some \( t \geq 1 \in \mathbb{Z}_+ \). We show that this assumption implies \( (26) \) holds for \( \lambda_m \in P_{ab} \) where \( a, b \geq 2 \) and \( m = t + 1 \). If \( c(\lambda_m) = 0 \) [for \( m = t + 1 \)], then \( r_j \geq 2 \) or \( s_j \geq 2 \) for all pairs in \( \lambda_m \) implying the inequality in \( (21) \) is applicable by \( (13) \) and \( (19) \) so that \( |\Delta_i(\omega_m)| \leq a_t^{a+b-2m} = O(1) \). If \( c(\lambda_m) = 1 \), then \( c(A_{m_j}) = 0 \) for \( 1 \leq j \leq t \) so \( \lambda_{m_0} \in P_{a+b-1} \cap \mathbb{Z}_+^2 \) with \( a-1, b-1 \geq 1 \); the induction assumption implies \( \max_{n \leq \ell < n} |\Delta_i(\omega_{m_j})| = O(1) \). Because \( |r_\ell(i)| \leq 1 \), \( (26) \) follows when \( c(\lambda_m) > 0 \) and \( m = t + 1 \), which finishes the induction argument.

Case 3: \( a, b \) even. We first show that \( (20) \) holds for any \( \lambda_m \in P_{ab} \) with \( c(\lambda_m) > 0 \). If \( c(\lambda_m) = 1 \), then \( 1 < m < (a+b)/2 \) and \( (22) \) and \( (23) \) follow subsequently as in Case 1. We make the induction assumption that \( (20) \) is valid for any \( \lambda_m \in P_{ab} \) with \( a, b \) even and \( c(\lambda_m) < t \) for some \( t > 1 \in \mathbb{Z}_+ \) and proceed to show \( (20) \) holds for any \( \lambda_m \) with \( c(\lambda_m) = t + 1 \) and \( a, b \) even. We write \( \lambda_m = \lambda(\omega_m) = (r_1, a, \ldots, r_t, b) \) where \( r_1 \) is either 0 or 2 and \( s_j = 2 - r_j \). Because \( \sum_{j=1}^m (r_j + s_j) = a + b + \sum_{j=1}^m r_j = a \), we have \( 2m = a + b \) and exactly \( a/2 \) of the \( r_j \)’s equal 2 (so exactly \( b/2 \) of the \( s_j \)’s equal 2). Thus, it must be \( \lambda_m = \lambda(a+b)/2 = (r_1, s_1, \ldots, r_{a+b}/2, s_{a+b}/2) \) where \( s_j = 2 - r_j \) and \( r_j := 2 \) if \( j \leq a/2 \), 0 otherwise. There is only one such possible \( \lambda(a+b)/2 \in P_{ab} \). We have that \( K_1(\lambda(a+b)/2) = 1 \) (since \( E(\epsilon^2_j) = 1 \)) and that \( K_2(\lambda(a+b)/2) = \left[ (a-1)(a-3) \cdots (1) \right] \left[ (b-1)(b-3) \cdots (1) \right] = E(Z^a)E(Z^b) \).

The definition of \( \lambda(a+b)/2 \) yields \( \sum_{k \in \mathbb{Z}} (d_k(\lambda(1)))^{r_\ell}(d_k(\lambda(1)))^{s_j} = r_\ell(1) = 1 \) for \( 1 \leq j \leq (a+b)/2, i \geq 1 \) so that we can iteratively use the same algebraic manipulation at in \( (24) \) to write \( \Delta_i(\lambda(a+b)/2) \) as a sum by
parts:
\[
\Delta_l(\lambda_{(a+b)/2}) = 1 - \sum_{i=1}^{(a+b-2)/2} R_{\ell i}(\lambda_{(a+b)/2}),
\]
\[
R_{\ell i}(\lambda_{(a+b)/2}) = \sum_{k_1, \ldots, k_i \in \mathbb{Z}, j=1}^{\ell} \left( \prod_{s=1}^{i} (d_{k_s(i)})^{r_s} \left( \sum_{s=1}^{i} (d_{k_s(i)+1})^{r_{s+1}} (d_{k_s(i)+1})^{r_{s+1}} \right) \right).
\]
Because each \(|R_{\ell i}(\lambda_{(a+b)/2})| \leq (a+b)d^2_l\) from (13) and (19), we find for any \(a, b\) even:
\[
E\left[ (S_i)^{a}(S_j)^{b} \right] = E(Z^a)E(Z^b) + o(1),
\]
using \(K_1(\lambda_{(a+b)/2})\) and \(K_2(\lambda_{(a+b)/2})\). Hence, we have established (18) in Case 3 and completed the proof of Theorem 3 a). □

**Proof of Theorem 3 b).** We again treat only Assumption L processes and essentially follow the proof of Theorem 2.5, HJL (1998). For \(m \geq 1\), define \(d_{m}^2 = N_m^{-1} \sum_{t=1}^{N_m} (M_{mt} - \mu)^2\) for \(N_m = n - m + 1\) and let \(m\) denote \(m_{jn}\) or \(m_{jt}, j = 1, 2\). Using Holder's inequality, (17), and (15)
\[
E|d_{m}^2 - \bar{d}_{m}^2| \leq 4N_m^{-1} \sum_{t=1}^{N_m} \left[ E(\bar{Y}_n - \mu)^2 (E(M_{mt} - \mu)^2 + E(\bar{Y}_n - \mu)^2) \right]^{1/2} = 4n^{-1} \sigma_n (m^{-2} \sigma_m^2 + n^{-2} \sigma_n^2)^{1/2} = o(d_m^2).
\]
From (17), Lemma 2, (17), and (18), it follows that
\[
E(d_{m}^2 - \bar{d}_{m}^2)^2 = d_m^4 O\left[ N_m^{-2} \sum_{t \leq j \leq N_m} \left| \text{Cov}(d_{m}^{-2}(M_{mt} - \mu)^2, d_{m}^{-2}(M_{mt} - \mu)^2) \right|^2 \right] = o(d_m^4).
\]
By (17) again, we have \(E(\bar{d}_{m}^2) = d_m^2 [1 + o(1)]\) so that
\[
d_{m}^2 = \bar{d}_{m}^2 [1 + o(1)]
\]
and then \(\bar{d}_{k}/\bar{d}_{k} \xrightarrow{p} 1\) for \(k = n\) or \(\ell\). Applying this and Theorem 3 a), we find the convergence of \(\tilde{F}_{1n}\) in probability. We are now finished with the proof of Theorem 3 b). □

**5.2 Proofs of Theorems 2 and 4**

We maintain the same notational standard that \(\alpha\) and \(L(\cdot)\) refer to values in (2) and are determined from (3) for each process.

**Lemma 3** Assume the conditions of Theorem 2. Let \(\bar{M}_n = N^{-1} \sum_{t=1}^{N} (M_t - \mu_0)\) and \(\bar{d}_{t}^2 = N^{-1} \sum_{t=1}^{N} (M_t - \mu_0)^2\). Then,
\[
\bar{M}_n/d_n \xrightarrow{d} Y_{\infty}, \quad \bar{d}_{t}^2/d_{t}^2 \xrightarrow{p} 1.
\]
Proof of Lemma 3. Note \( N\tilde{M}_n = n(\tilde{Y}_n - \mu_0) - H_n \) with \( H_n = \ell^{-1} \sum_{j=1}^{\ell} (\ell - j)(Y_j + Y_{n-j+1} - 2\mu_0) \).

Because \( E(H_n) = 0 \) and

\[
\text{Var}[(Nd_n)^{-1}H_n] \leq 4\text{Var}(Y_0)\ell^2(Nd_n)^{-2} \\
\leq C(\ell^2n^{-1})\{n(1-\alpha)L(n)\}^{-1} = o(1),
\]

we have \((Nd_n)^{-1}H_n \xrightarrow{P} 0\). Applying Slutsky's theorem, Theorem 1, and \( n/N \xrightarrow{} 1 \), we have the distributional result for \( \tilde{M}_n \).

The second assertion in Lemma 3 follows from (27) under Assumption L; HJL (1998) [p. 1203] establish (27) for Assumption G processes. \( \square \)

Proof of Theorem 2. Let \( F \) denote the continuous distribution function of \( Y_{\infty} \). From (16) for Assumption L processes [or analogously from (4.1), (4.3) of HJL (1998)], \( F^*(\pm\varepsilon) \xrightarrow{P} F(\pm\varepsilon) > 0 \) for some \( \varepsilon > 0 \) and consequently

\[
P\left( \min_{1 \leq i \leq N} (M_i - \mu_0) < 0 < \max_{1 \leq i \leq N} (M_i - \mu_0) \right) \xrightarrow{P} 1
\]

which implies that a positive \( R_n(\mu_0) \) can be determined with probability approaching 1. Using Lagrange multipliers [cf. Owen (1990)], the probabilities yielding the constrained maximum are given by \( p_i = N^{-1}\{1 + \lambda_{\mu_0}(M_i - \mu_0)\}^{-1} \) where \( \lambda_{\mu_0} \) is the root of

\[
g(\lambda) = N^{-1} \sum_{i=1}^{N} \frac{M_i - \mu_0}{1 + \lambda(M_i - \mu_0)}, \quad \max_{1 \leq i \leq N}(M_i - \mu_0) < \lambda < \frac{-1}{\min_{1 \leq i \leq N}(M_i - \mu_0)}
\]

(i.e. \( g(\lambda_{\mu_0}) = 0 \)). Note (28) shows that, in probability, the function \( g \) is well-defined and strictly decreasing (from \( \infty \) to \( -\infty \)) on the above domain so that \( \lambda_{\mu_0} \) exists and each \( p_i \) is indeed positive.

Let \( Z_n = \max_{1 \leq i \leq N}(M_i - \mu_0) \). We first establish a bound on the order of \( Z_n \). The finiteness of \( E(|Y_1|^{2c}) \) for some \( c > \alpha^{-1} \) ensures that \( \max_{1 \leq i \leq n}|Y_i - \mu_0| \leq n^{1/(2c)} \) eventually [almost surely (a.s.)] by stationarity and the Borel-Cantelli lemma. Then we have \( Z_n = O_P(n^{1/(2c)}\ell^{-1}) \) (a.s.) so that

\[
d^{-2}d_n Z_n = \left\{n^{1/\alpha}n^{-\alpha}L(n)\right\}^{1/2}\left\{\ell^{1-\alpha}L(\ell)\right\}^{-1}\left\{n^{-1/(2c)}\ell Z_n\right\} = o(1) \text{ a.s.}
\]

In a fashion similar to Owen (1990), we next show

\[
|\lambda_{\mu_0}| = O(d_n/d^2).
\]

Assume \( d^2 > 0 \), which occurs in probability by Lemma 3. We can write

\[0 = |g(\lambda_{\mu_0})| \geq \lambda_{\mu_0}/(1 + \lambda_{\mu_0}Z_n) - |\tilde{M}_n|/d^2\]
where the rightmost term is $O_p(d_n/d_f^2)$ by Lemma 3. Hence, \( (d_f^2/d_n)|\lambda_{\mu_0}/(1 + \lambda_{\mu_0} Z_n)| = O_p(1) \). For any $\epsilon > 0$, we may pick $K, N_K > 0$ large enough to ensure that: for $n \geq N_K$

\[
1 - \epsilon/2 < P\left(\frac{d_f^2}{d_n} | \lambda_{\mu_0} | \leq K \{1 + |\lambda_{\mu_0}| Z_n\}/2\right) \\
\leq P\left(\frac{d_f^2}{d_n} | \lambda_{\mu_0} | \leq K + P(K < d_f^2/d_n | \lambda_{\mu_0} |, K/2 \leq K^2 d_f^{-2} d_n Z_n/2) \\
\leq P\left(\frac{d_f^2}{d_n} | \lambda_{\mu_0} | \leq K + P(K^{-1} \leq d_f^{-2} d_n Z_n) \\
\leq P\left(\frac{d_f^2}{d_n} | \lambda_{\mu_0} | \leq K + \epsilon/2 \right)
\]

using (29) for the last inequality above. The order of $|\lambda_{\mu_0}|$ in (30) then follows.

From (29) and (30), we find for $\theta_i = \lambda_{\mu_0}(M_i - \mu_0)$:

\[
\max_{1 \leq i \leq N} |\theta_i| = |\lambda_{\mu_0}| Z_n = O_p(d_f^{-2} d_n) O_p(d_f^2 d_n^{-1}) = o_p(1).
\] (31)

With a little algebra, it holds that

\[
0 = g(\lambda_{\mu_0}) = \lambda_{\mu_0}^2 Z_n d_f^2 + I_n, \quad I_n = N^{-1} \sum_{i=1}^{N} \theta_i^2/(1 + \theta_i)
\]

which can be solved for $\lambda_{\mu_0} = (\lambda_{\mu_0} + I_n)/d_f^2$. Note that

\[
|I_n| \leq |\lambda_{\mu_0}|^2 Z_n d_f^2 \max_{1 \leq i \leq N} |\theta_i|^2 = O_p(d_f^{-4} d_n^2) O_p(d_f^2 d_n^{-1}) O_p(1) = o_p(1),
\] (32)

by (29), (30), and (31). When $|\lambda_{\mu_0}| Z_n < 1$ in (31), a Taylor's expansion gives

\[
\log(1 + \theta_i) = \theta_i - \theta_i^2/2 + \nu_i, \quad |\nu_i| \leq |\lambda_{\mu_0}|^2 Z_n (M_i - \mu_0)^2 (1 - |\lambda_{\mu_0}| Z_n)^{-3},
\]

for each $1 \leq i \leq N$. We now write

\[
-2B_n \log R_n(\mu_0) = 2B_n \sum_{i=1}^{N} \log(1 + \theta_i)
\]

\[
= 2B_n \lambda_{\mu_0} \sum_{i=1}^{N} (M_i - \mu_0) - B_n \lambda_{\mu_0}^2 \sum_{i=1}^{N} (M_i - \mu_0)^2 + 2B_n \sum_{i=1}^{N} \nu_i
\]

\[
= Q_{1n} - Q_{2n} + Q_{3n}
\]

where $Q_{1n} = (d_n^{-1} \lambda_{\mu_0})^2/(d_f^{-2} d_n^2) \xrightarrow{d} Y_{\infty}^2$ and $Q_{2n} = (d_n^{-1} I_n)^2/(d_f^{-2} d_n^2) = o_p(1)$ applying Lemma 3 and (32), and $Q_{3n} = 2B_n \sum \nu_i$ so that (when $|\lambda_{\mu_0}| Z_n < 1$)

\[
|Q_{2n}| \leq 2d_n^{-1} |\lambda_{\mu_0}|^3 Z_n d_f^2/(1 - |\lambda_{\mu_0}| Z_n)^3 = d_n^{-1} d_f^{-2} O_p(d_n^{-4} d_n^{-1}) O_p(d_f^2) O_p(1) = o_p(1).
\]

Theorem 2 then follows from Slutsky's theorem. \qed
Proof of Theorem 4. The convergence of $\tilde{B}_n/B_n$ follows from Theorem 3 and (27). HJL (1998) show (27) under Assumption G.

Again let $F(x)$ denote the distribution of $Y_{\infty}$. By Theorems 1 and 3, $(Y_n - \mu_0)/\tilde{d}_n$ converges in distribution to $Y_{\infty}$ and hence $\|F - F\|_\infty = o(1)$ by the continuity of $F$. Applying Theorem 3, $\|\tilde{F}_n - F\|_\infty = o_P(1)$. By the continuity of $F$, it follows that $\tilde{Q}_{1n} \xrightarrow{d} Q_\beta$.

We now show the convergence of $\tilde{Q}_{jn}$, $j = 2, 3$. By construction, $\tilde{F}_n(x) = \tilde{F}_n(x\tilde{d}_j/d_4)$. Then, we can write

$$\sup_{z \in \mathbb{R}} |\tilde{F}_n(x) - F(x)| \leq \|\tilde{F}_n - F\|_\infty + \|F_n - F\|_\infty + \sup_{z \in \mathbb{R}} |F(z\tilde{d}_j/d_4) - F(x)| = o_P(1)$$

by applying Theorem 3, (27), and the continuity of $F$ so that $\|\tilde{F}_n - F\|_\infty$, $\sup_{x \in \mathbb{R}} |F(x\tilde{d}_j/d_4) - F(x)| = o_P(1)$; and $\|F_n - F\|_\infty = o(1)$ by Theorem 1. We then have $\tilde{Q}_{jn} \xrightarrow{d} Q_\beta$ by the continuity of $F$.

With Slutsky's theorem and Theorem 2,

$$-2\tilde{B}_n \log R_n(\mu_0) - [\tilde{Q}_{jn}]^2 + [Q_\beta]^2 \xrightarrow{d} Y_{\infty}^2.$$ 

The continuity of $F$ and the definition of the quantile $[Q_\beta]^2$ then give

$$P(\mu_0 \in C_{\epsilon,\alpha}(\beta)) = P(-2\tilde{B}_n \log R_n(\mu_0) \leq [\tilde{Q}_{jn}]^2) \xrightarrow{} P(Y_{\infty}^2 \leq [Q_\beta]^2) = \beta.$$ 

This finishes the proof of Theorem 4.

References


Politis, D.N. and Romano, J.P. (1992). A general resampling scheme for triangular arrays of $\alpha$-mixing


Frequency domain empirical likelihood for short- and long-range dependent processes

A paper to be submitted to Biometrika

Daniel J. Nordman and Soumendra N. Lahiri

Abstract

This paper introduces a spectral version of empirical likelihood, using periodogram ordinates to create a nonparametric spectral likelihood. The first-order properties of frequency domain empirical likelihood are studied for linear time processes which could exhibit either long- or short-range forms of dependence. The method results in likelihood ratios which can be used to build nonparametric, asymptotically correct confidence regions for spectral parameters like autocorrelations. Using estimating equations based on spectral mean conditions, maximum empirical likelihood estimators are available for parameter estimation and testing under both types of dependence. We consider the effects of tapering for weakly dependent processes. Our methodology can be applied to many inference problems, such as Whittle estimation and spectral goodness-of-fit testing, and has parallels with the frequency domain bootstrap.

Key Words: Autocorrelation, bootstrap, cumulants, empirical likelihood, estimating equations, long-range dependence, nonparametric estimation, periodogram, spectral distribution, testing hypotheses, Wilks's theorem, Whittle estimation

1 Introduction

Empirical likelihood (EL) allows likelihood-based inference without specifying a parametric distribution for the data. The method, proposed by Owen (1988, 1990) for independent samples, generates a nonparametric likelihood which has been successful in a broad range of applications [cf. Owen (2001)]. Like the fully parametric likelihood, EL yields ratio statistics with limiting chi-square distributions for
estimation and testing purposes [cf. Qin and Lawless (1994)]. The technique has often been paralleled

to the bootstrap in both its mechanics and performance [cf. Efron and Tibshirani (1993)] and desirable

desirable properties of EL have been well documented for independent observations [cf. Hall and La Scala (1990)].

However, most developments with EL in the literature focus on independent data generating pro­

cesses. In this paper, we would like to consider inference on dependent data, in which correlations

among the observations could conceivably remain strong over long periods of time. To explain further,

let \( \{X_t\}, t \in \mathbb{Z} \), be a stationary sequence of random variables with mean \( \mu \) and spectral density \( f \) on

\( \Omega = [-\pi, \pi] \) where

\[
 f(\lambda)/|\lambda|^{-\alpha} \to C_\alpha, \quad \lambda \to 0
\]  

for \( \alpha \in (-1, 1) \) and a constant \( C_\alpha > 0 \) involving \( \alpha \) [cf. Robinson (1995), Lahiri (1999)]. When \( \alpha = 0 \),

we can classify the process \( \{X_t\} \) as short-range dependent (SRD). For \( \alpha > 0 \), the process will be
called long-range dependent (LRD). (The case \( \alpha < 0 \) can result from overdifferencing but typically is
of less practical interest.) This classification resembles the one from Lahiri (1999) and encompasses the
definition of long-range dependence (LRD) where \( f \) is assumed to have a pole at \( \lambda = 0 \) [cf. Beran (1994)].

Recent extensions of EL to dependent data [cf. Kitamura (1997), Monti (1997)] rely exclusively on

a weak or short-range dependence (SRD) structure. The time processes considered have correlations
which decrease rapidly enough over time so that, in essence, observations are nearly independent after
relatively short time separations and the autocovariances are absolutely summable. However, the situ­
ation changes dramatically with strongly dependent data. The rate of decay of the covariance function

\( r(k) = \text{Cov}(X_j, X_{j+k}) \) is characteristically much slower under LRD, namely:

\[
 r(k) = C_\alpha k^{-(1-\alpha)} \quad k \to \infty,
\]

a representation of LRD which is also equivalent to (1) for \( \alpha > 0 \) [cf. Zygmund (1968)]. While the
sum of the covariances converge under SRD, the slow decay in \( r(k) \) causes the same sum to diverge to
infinity under LRD. Strong dependence can then complicate statistical inference in that, for example,
\( \text{Var}(\bar{X}_n) \) decreases to 0 at a slower (and unknown) rate \( n^{-1+\alpha} \) instead of being proportional to the usual
\( n^{-1} \) from SRD. Consequently, the usual scaling factor \( n^{1/2} \) for the centered sample mean \((\bar{X}_n - \mu) \) does
not produce a proper limit distribution, even with Gaussian processes.

It is well-known that the process dependence strength affects the relevance and applicability of many
statistical procedures. An inference method proposed for weak dependence may fail terribly in the
presence of strong dependence, or at least require significant modification. For example, Lahiri (1993)
has shown that the slow growth of \( \text{Var}(\bar{X}_n) \) under LRD causes the moving block bootstrap, which works
so well under SRD (mixing), to produce a degenerate approximation of the sampling distribution of $\tilde{X}_n$.

Strong dependence similarly invalidates the EL statistics proposed by Kitamura (1997) for mixing time series ($\alpha = 0$). These statistics result from applying Owen (1990)’s EL formulation to observational “time blocks” formed with the same blocking techniques used in the bootstrap [cf. Künsch (1989), Liu and Singh (1992)]. Such “blockwise” EL statistics can asymptotically diverge to infinity under LRD because they employ a “blocking” re-calibration [see Kitamura (1997), p. 2089]) rooted in the assumption that $a_n = n^{1/2}$ correctly scales $a_n (\tilde{X}_n - \mu)$ for a normal limit (whereas valid scaling $a_n$ depends intricately on the unknown $\alpha$).

We demonstrate, however, that EL methodology can provide a common tool for inference on both SRD and LRD time processes. We give a new development of EL based on the process spectral distribution $f$ rather than its probability distribution. Using spectral estimating equations and periodogram ordinates, frequency domain empirical likelihood (FDEL) can build a nonparametric spectral likelihood for estimation of parameters, such as autocorrelations or the long-memory constant $\alpha$. We establish that FDEL statistics have limiting chi-square distributions for setting confidence regions. EL tests are possible to assess both parameter conjectures and the validity of (spectral) moment conditions, similar to EL features available in the independent data formulation [Owen (1990), Qin and Lawless (1994)]. In addition, FDEL allows for estimation of the same population quantities targeted by the frequency domain bootstrap of Dahlhaus and Janas (1996) under SRD, while simultaneously having justification for LRD processes as well.

Our results are directly applicable to linear processes with spectral densities satisfying (1), which include several important models for both SRD (eg. ARMA) and LRD. In particular, two such LRD processes are the fractional Gaussian processes of Mandelbrot and Ness (1968) with spectral density and parameter $1/2 < H < 1$:

$$f_{H, a^2}(\lambda) = \frac{4 a^2 \Gamma(2H - 1)}{(2\pi)^{2H+2}} \cos(\pi H - \pi/2) \sin^2(\lambda/2) \sum_{k=-\infty}^{\infty} |\lambda/(2\pi) + k|^{-1-2H}, \quad \lambda \in \Omega$$

(2)

and the fractional autoregressive integrated moving average (FARIMA) processes of Granger and Joyeux (1980) and Hosking (1981) with density and parameters $0 < d < 1$, $\rho = (\rho_1, \ldots, \rho_p)$, $\varrho = (\varrho_1, \ldots, \varrho_q)$:

$$f_{d, \rho, \varrho^2}(\lambda) = \frac{\sigma^2}{2\pi} |1 - \rho_1 e^{i\lambda}|^{-d} \left| \frac{\sum_{j=0}^{p} \rho_j (e^{i\lambda})^j}{\sum_{j=0}^{q} \varrho_j (e^{i\lambda})^j} \right|^2 , \quad \lambda \in \Omega, \, \rho_0 = \varrho_0 = 1$$

(3)

(where the polynomials in the above ratio have unlike roots outside the unit circle). These models fulfill (1) with $\alpha = 2H - 1$ and $\alpha = d$, respectively [cf. Fox and Taqqu (1986)] and are known to arise naturally in astronomy, economics, hydrology, and geophysics [cf. Beran (1994)].
1.1 Related Literature

We point out one important related reference to our work. Monti (1997) has proposed periodogram-based EL confidence regions in the context of quasi-Whittle estimation with SRD linear processes. However, our FDEL methodology differs from that of Monti (1997) in both its form and philosophical underpinnings. In particular, Monti's development of EL treats the collection of periodogram ordinates (at positive frequencies $2\pi/n, 4\pi/n, \ldots$) as asymptotically independent. However, this assumption creates serious complications, like those encountered in the frequency domain bootstrap of Dahlhaus and Janas (1996), because the dependencies among the whole collection of periodogram ordinates are typically not negligible. Monti's justification also depends considerably on the existence and properties of Whittle estimators. Our results include those in Monti (1997) as a special case and further imply the periodogram-based EL theory in Monti (1997) may not be fully accurate.

The rest of the paper is organized as follows. In Section 2, we describe the construction of spectral EL using the periodogram and estimating functions. Examples of useful estimating functions are provided. Our working assumptions are given in Section 3 along with a brief, background comparison of the mechanics of frequency domain EL and bootstrap methods. In Section 4, we state first-order asymptotic properties with FDEL statistics for confidence region estimation and hypothesis testing. In Section 5, we illustrate applications of FDEL to Whittle estimation and goodness-of-fit testing. Section 6 provides extensions of our methods to inference with parameter restrictions and also tapering. Proofs of the main results are given in Section 7, while the Appendix in Section 8 contains supplementary tools to facilitate the proofs.

2 Definition of spectral empirical likelihood

2.1 Estimating functions.

Consider inference on a parameter $\theta \in \Theta \subset \mathbb{R}^p$ associated with the distribution of a time stretch $X_1, \ldots, X_n$. We suppose, following the EL framework of Qin and Lawless (1994, 1995) for independent, identically distributed observations (iid), that information about $\theta$ exists through a system of functions and their corresponding expectations or, in our case, spectral means. Let

$$g_\theta(\lambda) = (g_{1,\theta}(\lambda), \ldots, g_{r,\theta}(\lambda))^\prime : \Pi \times \Theta \rightarrow \mathbb{R}^r$$

(4)
denote a vector of even, estimating functions with \( r \geq p \). (When \( r > p \), the above functions are said to be "overidentifying" for \( \theta \).) We assume that \( g_\theta \) satisfies the spectral moment condition

\[
\int_0^\pi g_\theta(\lambda)f(\lambda)d\lambda = p
\]

for some known \( p \in \mathbb{R}^r \). Note though that \( g_\theta \) is not an "unbiased" estimating function in the sense of Qin and Lawless (1994) and Kitamura (1997); in fact, \( g_\theta(\cdot) \) is defined on \([-\pi, \pi]\) and not a function of the data. However, by combining the periodogram and the estimating functions in (4), we demonstrate the construction of an EL function for subsequent spectral inference.

For illustration, we first provide a few examples of useful estimating functions, some of which target estimation of normalized spectral means: \( \theta = \phi(\pi) = \int_0^\pi \phi f d\lambda / \int_0^\pi f d\lambda \) involving a desirable function \( \phi \).

Dahlhaus and Janas (1996) comment on the importance, and often complete adequacy, of population information expressed in this integral ratio form. For appropriate choices of the estimating vector \( g_\theta \), EL inference is possible for many of the same parameters estimable with the frequency domain bootstrap of Dahlhaus and Janas (1996) (hereafter FDB-DJ).

**Example 1: Autocorrelations.** Consider interest in the autocorrelation function \( p(\cdot) \) at arbitrary lags \( m_1, \ldots, m_k \); that is, \( \theta = (p(m_1), \ldots, p(m_k))' \) where

\[
p(m) = r(m)/r(0) = \int_0^\pi \cos(m\lambda)f(\lambda)d\lambda / \int_0^\pi f(\lambda)d\lambda, \quad m \in \mathbb{Z}.
\]

One can select \( g_\theta(\lambda) = (\cos(m_1\lambda), \ldots, \cos(m_k\lambda))' - \theta \) for autocorrelation inference, in which case the moment condition in (5) is fulfilled with \( p = 0 \in \mathbb{R}^k \). Note here \( r = p = k \).

Suppose \( p(1) = \theta \) and \( p(2) = \theta^2 \), which would be the case if \( \{X_t\} \) satisfies a stationary AR(1) model: \( X_t = \theta X_{t-1} + \epsilon_t \) (for a white noise process \( \{\epsilon_t\} \), \( |\theta| < 1 \)). One could estimate \( \theta \) with overidentifying estimating functions:

\[
\tilde{g}_\theta(\lambda) = (\cos(\lambda) - \theta, \cos(2\lambda) - \theta^2)'
\]

with \( r = 2 > p = 1 \). EL techniques can combine such frequency domain information. Alternatively, if say \( p(m_1) = p(m_2) = 0 \) at some lags (as in a moving average model), the information may be incorporated to estimate \( \theta = p(1) \) with equations \( \tilde{g}_\theta(\lambda) = (\cos(\lambda) - \theta, \cos(m_1\lambda), \cos(m_2\lambda))' \). We can create moment conditions and allow a nonparametric spectral likelihood to suggest parameter estimates as well as quantify our relative certainty in the formulated spectral moments.

**Example 2: Spectral distribution function.** For \( \omega \in [0, \pi] \), denote the spectral distribution function ( sdf ) as \( F(\omega) = \int_0^\omega f(\lambda)d\lambda \). Suppose we wish to estimate \( \theta = (F(\tau_1)/F(\pi), \ldots, F(\tau_k)/F(\pi))' \) for
some \( \tau_1, \ldots, \tau_k \in (0, \pi) \). This normalized parameter often sufficiently characterizes the spectral distribution \( F \) for testing purposes, as in Bartlett's U-Statistic [cf. Dahlhaus (1985b)]. To this end, we can pick \( g_\theta(\lambda) = (I_{(\lambda \leq \tau_1}), \ldots, I_{(\lambda \leq \tau_k)})' - \theta \) where \( I_{(\cdot)} \) denotes the indicator function. Then (5) holds with mean \( p = 0 \in \mathbb{R}^k \).

Example 3: Goodness-of-fit tests. There has been increasing interest lately in frequency domain-based tests to assess model adequacy [cf. Andersen (1993), Paparoditis (2000)]. Consider a test involving a simple null hypothesis \( H_0 : f = f_0 \) against an alternative \( H_1 : f \neq f_0 \) for some candidate density \( f_0 \).

With EL techniques, one immediate approach would be to take \( g(\lambda) = 1/f_0(\lambda) \) so that, under \( H_0 \), the spectral mean of \( g \) is \( \pi \). (Here we treat \( r = 1 \) and the dimension \( p \) of \( \theta \) as 0). A EL ratio test can be produced with \( g \) which resembles a spectral goodness-of-fit test statistic proposed by Milhoj (1981) and shown by Beran (1992) to be useful for long-range dependent Gaussian series.

The more interesting and complicated problem of testing the hypothesis that \( f \) belongs to a given parametric model family (like densities corresponding to ARMA models or fractional Gaussian processes) can also be addressed with frequency domain EL. This testing issue, however, can often involve Whittle estimation [cf. Beran (1992), Paparoditis (2000)]. We return to EL-based, spectral goodness-of-fit tests in Section 5.

Example 4: Whittle-estimation. We denote a parametric collection of spectral densities as

\[ \mathcal{F} = \{f_\theta(\lambda) : \theta \in \Theta \} \]  

and assume the densities are positive on \( \Pi \) and identifiable. (That is, \( \theta \neq \hat{\theta} \in \Theta \) implies the Lebesgue measure of \( \{\lambda : f_\theta(\lambda) \neq f_{\hat{\theta}}(\lambda)\} \) is positive.) Whittle estimation [Whittle (1953)] is a common procedure for fitting the model \( f_\theta \) to the data. The Whittle estimator minimizes the Whittle likelihood involving the periodogram \( I_n \):

\[ W_n(\theta) = (4\pi)^{-1} \int_0^\pi \left\{ \log f_\theta(\lambda) + \frac{I_n(\lambda)}{f_\theta(\lambda)} \right\} d\lambda \]

to estimate the value of \( \theta \) at which the theoretical "distance" measure (the Kullback-Liebler distance for Gaussian processes)

\[ W(\theta) = (4\pi)^{-1} \int_0^\pi \left\{ \log f_\theta(\lambda) + \frac{f(\lambda)}{f_\theta(\lambda)} \right\} d\lambda \]

achieves its minimum [cf. Dzhaparidze (1986)]. The model class may be misspecified (possibly \( f \notin \mathcal{F} \)) but Whittle estimation aims for the density in \( \mathcal{F} \) "closest" to \( f \) as measured by \( W(\theta) \). Dahlhaus
and Wefelmeyer (1996) describe features of Whittle estimation for potentially misspecified parametric models.

To consider a particular parameterization of (6), suppose

$$f_\theta(\lambda) = \sigma^2 k(\lambda, \theta), \quad \theta = (\sigma, \theta')', \quad \Theta \subset (0, \infty) \times \mathbb{R}^{p-1}, \quad \theta = (\theta_1, \ldots, \theta_{p-1})',$$

(8)

and that Kolmogorov's formula holds with

$$(2\pi)^{-1} \int_{-\pi}^{\pi} \log f_\theta(\lambda) d\lambda = \log[\sigma^2/(2\pi)]$$

(as would be the case for many linear models where $\sigma^2$ denotes the innovation variance) [cf. Brockwell and Davis (1991), Section 5.8]. The model class in (8) is commonly considered in the context of Whittle estimation for both SRD and LRD time processes and includes those LRD processes formulated in (2) and (3) [cf. Hannan (1973), Fox and Taqqu (1986), Dahlhaus (1989), and Giraitis and Surgailis (1990)]. Under appropriate regularity conditions, the true minimum argument $\theta_0 = (\sigma_0^2, \theta_0')'$ of $W(\theta)$ is determined by the stationary solution of:

$$\frac{dW(\theta)}{d\theta} = 0$$

$$\frac{f''(\lambda)}{f(\lambda)} \left\{ \frac{f(\lambda)}{f'(\lambda)} \right\} d\lambda = 0$$

$$f_\theta^{-1}(\lambda) \equiv 1/f_\theta(\lambda)$$

$$\int_0^{2\pi} f(\lambda) f_\theta^{-1}(\lambda) d\lambda = 1.$$

The moment conditions in (9) imply that a natural set of estimating functions can be used for EL inference on $\theta$ in densities from (8). Namely, the choice

$$g_\theta^w(\lambda) = (f^{-1}_\theta(\lambda), \partial f^{-1}_\theta(\lambda)/\partial \theta_1, \ldots, \partial f^{-1}_\theta(\lambda)/\partial \theta_{p-1})', \quad \varphi_w = (\pi, 0, \ldots, 0)' \in \mathbb{R}^p$$

(10)

will fulfill (5).

Note that if one wishes to treat $\sigma^2$ as a nuisance parameter, which is often the case in the density formulation of (8) because $\sigma^2$ has no bearing on the essential shape of the density $f_\theta$, estimating functions

$$g_\theta^{w*}(\lambda) = \partial k^{-1}(\lambda, \theta)/\partial \theta, \quad \varphi_{w*} = 0 \in \mathbb{R}^{p-1}$$

(11)

provide structure for inference on $\theta$.

2.2 Construction of spectral empirical likelihood

Denote the periodogram of the sequence $X_1, \ldots, X_n$ by

$$F_n(\lambda) = \frac{1}{2\pi n} \left| \sum_{i=1}^{n} X_i \exp(-it\lambda) \right|^2,$$

$\lambda \in \Pi = [-\pi, \pi].$
where $i = \sqrt{-1}$. For a given $\theta \in \Theta$, the profile frequency domain empirical likelihood function is given by

$$L_n(\theta) = \sup \left\{ \prod_{j=1}^{N} m_j : m_j \geq 0, \sum_{j=1}^{N} m_j = \pi, \sum_{j=1}^{N} m_j g_{\theta}(\lambda_j) I_n(\lambda_j) = p \right\}$$

$$= \pi^N \sup \left\{ \prod_{j=1}^{N} p_j : p_j \geq 0, \sum_{j=1}^{N} p_j = 1, \sum_{j=1}^{N} \pi p_j g_{\theta}(\lambda_j) I_n(\lambda_j) = p \right\}$$  \hspace{1cm} (12)

where $\lambda_j = 2\pi j/n, j \in \{1, \ldots, N\}$, are Fourier frequencies and $N = \lfloor (n - 1)/2 \rfloor$. To determine the above EL function, a weight $m_j$ is assigned to each ordinate $\lambda_j$ (or $I_n(\lambda_j)$) so that an integral over $[0, \pi]$ of $g_{\theta}$ with respect to the sample spectral density $f(\lambda) = \sum_{j=1}^{N-1} I_{\{\lambda_{j-1} < \lambda \leq \lambda_j\}} I_n(\lambda_j) + I_{\{\lambda_N < \lambda\}} I_n(\lambda_N)$ equals $p$. These point masses correspond to a finite measure on $[0, \pi]$ and do not have the usual, immediate interpretation of multinomial probabilities assigned to observations (or blocks of them) as in Owen (1990), Qin and Lawless (1994), and Kitamura (1997). However, (12) shows that EL function can be defined with such probabilities (after a scalar adjustment by $\pi^N$). We can define $L_n(\theta) = -\infty$ when the conditioning set in (12) is empty.

If $p$ is interior to the convex hall of $\{g_{\theta}(\lambda_j) I_n(\lambda_j)\}_{j=1}^{N}$, then $L_n(\theta)$ is actually a positive maximum and optimizing

$$L(p_1, \ldots, p_N, \gamma, t) = \sum_{j=1}^{N} \log(p_j) + \gamma \left( 1 - \sum_{j=1}^{N} p_j \right) - Nt' \left( \sum_{j=1}^{N} p_j [\pi g_{\theta}(\lambda_j) I_n(\lambda_j) - p] \right)$$

with Lagrange multipliers $\gamma$ and $t = (t_1, \ldots, t_r)'$ provides the unique solution for the constrained extrema [as in Owen (1988, 1990)]. Then, the empirical likelihood function in (12) may be reformulated as

$$L_n(\theta) = \prod_{j=1}^{N} \pi p_j(\theta), \quad p_j(\theta) = N^{-1} \left[ 1 + t'_e (\pi I_n(\lambda_j) g_{\theta}(\lambda_j) - p) \right]^{-1}$$  \hspace{1cm} (13)

where $t'_e$ is the stationary point of the function $q(t) = \sum_{j=1}^{N} \log(1 + t'_e (\pi I_n(\lambda_j) g_{\theta}(\lambda_j) - p))$ [see Kitamura (1997)]. (When $\sum_{j=1}^{N} (\pi I_n(\lambda_j) g_{\theta}(\lambda_j) - p) [\pi I_n(\lambda_j) g_{\theta}(\lambda_j) - p]'$ is positive definite in some neighborhood in $\Theta$, $t'_e$ can be locally written as a continuously differentiable function of $\theta$ by the implicit function theorem. Owen (1990) and Qin and Lawless (1994) describe these and other features of the EL.) Without the integral-type linear constraint in (12), $\prod_{j=1}^{N} m_j$ has a maximum when each $m_j = \pi/N$ so that we can form a profile empirical likelihood ratio:

$$R_n(\theta) = L_n(\theta) \left/ \left( \pi N^{-1} \right)^N \right. = \prod_{j=1}^{N} \left[ 1 + t'_e (\pi I_n(\lambda_j) g_{\theta}(\lambda_j) - p) \right]^{-1}.$$  \hspace{1cm} (14)

We shall refer to the maximum of (14), say $\hat{\theta}_n$, as the (spectral) maximum empirical likelihood estimator (MELE).
It may hold also that the spectral density belongs to a parametric class of densities \( F \) as in (6) and is a function of the parameter of interest \( \theta \). That is, we may wish to determine which \( f_\theta \in F \) corresponds to \( f \) (or possibly even test if \( f \in F \)). In this estimation scenario, we can construct a “centered” versions of the EL function and likelihood ratio by approximating \( E(I_n(\lambda_j)) \) with a density evaluation \( f_\theta(\lambda_j) \), for \( f_\theta \in F \). Namely, let:

\[
L_{n,\tau}(\theta) = \sup \left\{ \prod_{j=1}^{N} m_j : m_j \geq 0, \sum_{j=1}^{N} m_j = \pi, \sum_{j=1}^{N} m_j g_\theta(\lambda_j)[I_n(\lambda_j) - f_\theta(\lambda_j)] = 0 \right\}
\]

\[
R_{n,\tau}(\theta) = (N/\pi)^{N} L_{n,\tau}(\theta).
\]

(15)

An exact form for \( L_{n,\tau}(\theta) \) can be deduced as with \( L_n(\theta) \) in (13). We shall denote the maximum of (15) as \( \hat{\theta}_{n,\tau} \). We consider the densities \( f_\theta \) and prospective functions \( g_\theta \) as dependent on the same parameters, which causes no loss of generality.

We emphasize that the profile ratio \( R_{n,\tau}(\theta) \) always requires specification of a candidate spectral density class. However, the EL function in (15) and its associated statistics, such as \( R_{n,\tau}(\hat{\theta}_{n,\tau}) \), can be useful in Whittle-like estimation and goodness-of-fit testing.

3 Preliminaries

To facilitate our discussion, we will now set down some assumptions on the time process under consideration and the potential vector of estimating functions, \( g_\theta \). In the following, let \( \theta_0 \) denote the true (but unknown) value of the parameter \( \theta \) which satisfies the moment condition in (5).

3.1 Assumptions

(A.1) \( \{X_t\} \) is a real-valued, linear process with a moving average representation of the form:

\[
X_t = \mu + \sum_{j=-\infty}^{\infty} b_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}
\]

(16)

where \( \mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\} \), \( \{\varepsilon_t\} \) are iid random variables with \( E(\varepsilon_t) = 0 \), \( E(\varepsilon_t^2) = \alpha > 0 \), \( E(\varepsilon_t^p) < \infty \) and \( \{b_t\} \) is a sequence of constants satisfying \( \sum_{t \in \mathbb{Z}} b_t^2 < \infty \) and

\[
f(\lambda) = \frac{\alpha^2}{2\pi} |b(\lambda)|^2, \quad \lambda \in \Pi
\]

with \( b(\lambda) = \sum_{j \in \mathbb{Z}} b_j e^{ij\lambda} \). It is assumed that, for some \( \alpha \in [0, 1) \), \( C > 0 \), that

\[
f(\lambda) \leq C|\lambda|^{-\alpha}, \quad \lambda \in \Pi
\]
and $f(\lambda)$ is continuous on $(0, \pi]$.

(A.2) For $g_0$ from (4) at $\theta = \theta_0$, each component $g_{j,0}$ is an even, integrable function such that $|g_{j,0}(\lambda)| \leq C|\lambda|^{\beta_j}, \lambda \in \Pi$, where $0 \leq \beta_j < 1$, $\alpha - \beta_j < 1/2$, $j = 1, \ldots, r$.

(A.3) For each $g_{j,0}(\lambda)$, $j = 1, \ldots, r$, one of the following is fulfilled:

Condition 1. $g_{j,0}$ is Lipschitz of order greater than $1/2$ on $[0, \pi]$.

Condition 2. $g_{j,0}$ is continuous on $\Pi$ and

$$\left| \frac{\partial g_{j,0}(\lambda)}{\partial \lambda} \right| \leq C|\lambda|^{\beta_j-1}, \quad 2\alpha - \beta_j < 1.$$  

Condition 3(a). $g_{j,0}$ is of bounded variation on $[0, \pi]$ with finite discontinuities and $\alpha < 1/2$.

(A.4) The $r \times r$ matrix $W_{00} = \int_{\Pi} f^2(\lambda)g_{0e}(\lambda)g_{0e}(\lambda)d\lambda$ is positive definite.

(A.5) On $(0, \pi]$, $f$ is differentiable and

$$\left| \frac{\partial f(\lambda)}{\partial \lambda} \right| \leq C|\lambda|^{-\alpha-1},$$

or each $f(\lambda)g_{j,0}(\lambda)$ is of bounded variation or piecewise Lipschitz of order greater than $1/2$ on $[0, \pi]$, $j = 1, \ldots, r$.

Remark 1: The assumptions on the white noise process can be weakened at the expense of greater complexity. Ergodicity of $\{\varepsilon_t\}$ is required along with absolute summability of the $k$-th order cumulants, for $k \leq 8$, to explicitly define the cumulant spectral density of $\{X_t\}$ [cf. Brillinger (1981), Yajima (1989), Lahiri (1999)]. In addition, we would require that cumulant($\varepsilon_t, \varepsilon_{t+t_1}, \varepsilon_{t+t_2}, \varepsilon_{t+t_3}$) = $C$ if $t_1 = t_2 = t_3 = 0$ and is equal to 0 otherwise, resembling the 4th-order cumulant conditions of Dahlhaus (1985b), Heyde and Gay (1993) and Hosoya (1997, Assumption F) in the context of periodogram-based inference. (Of particular note, Hoyosa's (1997) Assumption F allows the development of quasi-likelihood ratio statistics, in his sense, which is also a desirable feature for EL.) The asymptotic normal limit law for periodogram-based spectral means is needed, which many authors have justified for linear processes in various contexts under both SRD [cf. Brillinger (1981), Dahlhaus (1983,1985a)] and LRD for functions satisfying a “growth rate” condition as in A.2 [cf. Fox and Taqqu (1986), Giraitis and Surgailis (1990),]
Heyde and Gay (1993)]. The standard proofs usually involve stipulations on the innovations \( \{ \varepsilon_t \} \), such as further cumulant summability, Gaussian distributions, or a mixing and/or martingale difference structure.

The bound on \( f \) in Assumption A.1 allows for the process \( \{ X_t \} \) to exhibit both short- and long-range forms of dependence with a possible pole of \( f \) occurring at \( \lambda = 0 \), encompassing the most common definition of LRD [cf. Beran (1994)]. A.1 embodies a slight generalization of (1) which permits the growth rate of \( f \) also possibly depend on a slowly varying function at zero [cf. Fox and Taqqu (1986), Lahiri (1999)].

The behavior of \( g_{q_0} \) in Assumption A.2 helps to reduce the growth rate of the scaled periodogram ordinates, \( g_{q_0}(\lambda_j)/n(\lambda_j) \), at low frequencies under LRD. The condition ensures that the entries of \( W_{q_0} \) are finite in A.4 and that central limit theorems for quadratic forms involving the mean(\( \mu \))-corrected periodogram are possible [cf. Giraitis and Surgailis (1990)]. Additional fractional poles in \( f \) (other than the origin) could be allowed at the expense of greater complexity, making allowances for the vector \( g_{q_0} \) considered in A.2 to dampen any "unstable spots" in \( f \) in the product of the two functions [as in Hosoya (1997)]. However, important processes are permissible under A.1 and, for these, useful estimating equations often satisfy A.2, allowing a fairly broad formulation of EL in the frequency domain.

Assumption A.3 outlines smoothness criteria for the estimating functions so that (for one reason) Riemann integrals of the periodogram sufficiently approximate spectral means in (5), similar to Brillinger (1981, Theorem 5.10.1). The important functions treated in Dahlhaus and Janas (1996) meet the requirements of A.3, including the estimating functions for autocorrelations and normalized spectral distribution given in Examples 1 and 2. The functions in \( f_{q_0}^{-1} \) and \( \partial f_{q_0}^{-1}/\partial \theta \) considered in Examples 3 and 4 satisfy A.3 for many SRD and LRD models so that, for instance, quasi-Whittle estimation and goodness-of-fit testing are available under weak or strong dependence. Hannan (1973) justifies Whittle estimation for ARMA densities where components of \( g_{q_0}^r \) in (10) meet Condition 1. For the fractional Gaussian and FARIMA models for LRD in (2) and (3), \( f_{q_0}^{-1} \) and \( \partial f_{q_0}/\partial \theta \) fulfill Condition 2 [see Conditions A-B, Fox and Taqqu (1986), Dahlhaus (1989), Giraitis and Surgailis (1990)]. Process dependence that is not "too strong" \( [f \in L_2(\Pi)] \) allows more flexibility in choosing estimating functions in Condition 3.

Assumption A.4 reflects requirements of functional linear independence among potential estimating functions. The conditions set in Assumption A.5 will be used only in the context of inference with EL statistics based on \( L_{n,\bar{X}} \) or \( R_{n,\bar{X}} \) and are generally not restrictive.
3.2 Discussion: bootstrap and empirical likelihood mechanics

A few comments will help to explain some of the anticipated properties of FDEL. Comparisons between bootstrap and EL techniques have often been made with iid or weakly dependent observations [cf. Owen (1991), Hall and La Scala (1991), Kitamura (1997)]. Hence, it seems natural to contrast FDEL here with the FDB-DJ, especially since both methods target estimation of similar population quantities for linear processes. There are, of course, other periodogram bootstrap versions [cf. Härdle, Horowitz, and Kreiss (2001)], but the parallels between FDEL techniques and the bootstrap of Dahlhaus and Janas (1996) seem the strongest and most pertinent. We note, though, that FDB-DJ has been proposed and justified only for SRD linear processes.

Define the mean corrected periodogram

\[ I_{nc}(\lambda) = (2\pi n)^{-1} \left| \sum_{t=1}^{n} (X_t - \mu) \exp(it\lambda) \right|^2, \quad \lambda \in \Pi. \]

For a vector function \( \phi \), the FDB-DJ attempts to approximate the distribution of a spectral mean estimator \( \int_0^\pi \phi I_{nc} d\lambda - \int_0^\pi \phi f d\lambda \) with a bootstrap version:

\[
\frac{\pi}{N} \sum_{k=1}^{N} \frac{1}{\phi_k} \sum_{j=1}^{N} (\lambda_j - \hat{\lambda}_j) \exp(it\lambda_j) - \frac{\pi}{N} \sum_{j=1}^{N} (\lambda_j - \hat{\lambda}_j) \exp(it\lambda_j)
\]

where \( \hat{\phi} \) is a kernel estimator of \( f \) and each \( \epsilon_j^* \) is an independent, random selection from the Studentized periodogram ordinates \( \{\epsilon_j = I_{nc}(\lambda_j)/\hat{f}(\lambda_j)\} \). The bootstrap procedure treats \( \{\epsilon_j^*\} \) as a set of iid standard exponential variables (because, for fixed frequencies \( \{\omega_1, \ldots, \omega_k\} \subset (0, \pi) \), variables \( I_{nc}(\omega_j)/f(\omega_j) \) have independent exponential limit distributions under both SRD and LRD [cf. Lahiri (1999)]). However, the dependencies among the entire collection of periodogram ordinates cannot typically be ignored so that the bootstrap does not reliably estimate the asymptotic variance

\[
\lim_{n \to \infty} n \text{Var} \left( \int_0^\pi \phi I_{nc} d\lambda \right) = \pi \int_\Pi \phi f^2 d\lambda + \kappa_{4,\epsilon}/\sigma^4 \left( 1/2 \cdot \int_\Pi \phi f d\lambda \right)^2 \left( 1/2 \cdot \int_\Pi \phi f d\lambda \right)^2
\]

unless the 4-th order cumulant of the white noise process \( \kappa_{4,\epsilon} = 0 \) or \( \int_\Pi \phi f d\lambda = 0 \). That is, the frequency domain bootstrap only captures the first component in the sum from (17). Consequently, the FDB-DJ uses modified functions \( \phi_\theta = \phi - \theta \) to estimate a normalized parameter \( \theta = \int_\Pi \phi f d\lambda/ \int_\Pi f d\lambda \) which guarantees that \( \int_\Pi \phi_\theta f d\lambda = 0 \).

FDEL methods choose estimates of a parameter \( \theta \) by profiling the ordinate evaluations \( \{g_\theta(\lambda_j) I_n(\lambda_j)\} \) to achieve a desired integral equality as specified by the moment relationship in (5). Like the bootstrap, FDEL can experience its own difficulties with variance estimation. The EL ratio can involve variance...
estimates for empirical spectral means, $\sqrt{n} \int_0^\pi \phi_{nc} d\lambda$, of the form

$$\frac{2\pi^2}{n} \sum_{j=1}^N \left( \phi(\lambda_j) I_n(\lambda_j) - 1/2 \cdot \int_{-\pi}^{\pi} \phi f d\lambda \right) \left( \phi(\lambda_j) I_n(\lambda_j) - 1/2 \cdot \int_{-\pi}^{\pi} \phi f d\lambda \right)^*, $$

in its mechanics. Heuristically, $E(\|^2(\lambda)/f^2(\lambda)) \approx 2$ under both SRD and LRD so that the above Riemann integral of the periodogram can consistently estimate (17) when $\int_{-\pi}^{\pi} \phi f d\lambda = 0$. Important estimating equations can however be judiciously selected so that $\rho = 0$ in (5), which eliminates potential variance estimation problems. In fact, if the dependence structure of the underlying process is not too strong ($\alpha < 1/2$), estimating functions for normalized spectral parameters can be constructed in a fashion following the bootstrap of Dahlhaus and Janas (1996) (selecting $g_\theta = \phi_\theta$ as above). Note that unlike its bootstrap counterpart, FDEL requires no kernel density estimates of $f$ and no direct computation of skewness estimates.

To further explain the differences between the FDB-DJ and FDEL, consider their respective approaches to inference on the first-lag autocorrelation $\theta = \rho(1)$. As described above, the bootstrap approximates the distribution of

$$\frac{\hat{\rho}_n(1) - \rho(1)}{\sqrt{\text{Var}(\hat{\rho}_n(1))}}$$

to generate empirical quantiles and set a confidence interval for $\rho(1)$, where $\hat{\rho}_n(1)$ is the Yule-Walker estimate of $\rho(1)$. The FDEL method suggests plausible values of $\rho(1)$ through profiled ratios [see (12) and Example 1]

$$\sum_{j=1}^N p_j \cos(\lambda_j) I_n(\lambda_j) / \sum_{j=1}^N p_j I_n(\lambda_j)$$

involving optimally selected constrained weights $\{p_j\}$. These ratios are then "calibrated" to obtain an interval estimate for $\rho(1)$ with appropriate confidence. The theoretical properties of FDEL are discussed in the next section.

4 Main Results

4.1 A Wilks's theorem

We first establish a Wilks's theorem [Wilks (1938)] for EL ratios which can be used to successfully construct nonparametric confidence regions and simple hypothesis tests as in Owen (1988, 1990, 1991). Before proceeding, we define two differently scaled log-profile EL ratio statistics:

$$\ell_n(\theta) = -4 \log R_n(\theta) \quad \text{and} \quad \ell_{n,\pi}(\theta) = -2 \log R_{n,\pi}(\theta),$$

(18)
using (14) and (15). The difference in the scalar adjustments to log-likelihoods in (18) owes to the assumption that the periodogram ordinates are approximately mean-corrected in the construction of $\ell_{n,\tau}(\theta)$. We note, however, that the spectral EL ratios in (18) do not require complicated scaling adjustments like the time domain blockwise EL ratios in Kitamura (1997) and more closely resemble those found in the iid sample setting of Owen (1990) and Qin and Lawless (1994).

In the remaining discussion, $\chi^2_{\nu}$ denotes a chi-square distribution with $\nu$ degrees of freedom and $\chi^2_{\nu,1 - \gamma}$ represents the $1 - \gamma$ quantile of the same distribution. Let $ch^0A$ denote the interior convex hull of a finite set $A \subset \mathbb{R}^p$.

**Theorem 1** Suppose Assumptions A.1-A.4 hold. If $p = 0 \in \mathbb{R}^p$, then

(i) $\ell_n(\theta_0) \xrightarrow{d} \chi^2_{\nu}$.

(ii) If, in addition, Assumption A.5 holds, $f = f_{\theta_0}$, and $P(0 \in ch^0\{\pi g_{\theta_0}(\lambda_j) f_n(\lambda_j) - f(\lambda_j)\}) \to 1$,

then $\ell_{n,\tau}(\theta_0) \xrightarrow{d} \chi^2_{\tau}$.

(iii) If $\kappa_{4,\ell} = 0$, statement (ii) remains valid if $p \neq 0 \in \mathbb{R}^p$.

**Remark 2**: If $\{X_i\}$ is Gaussian, then the 4th order innovation cumulant $\kappa_{4,\ell} = E(e^{2\xi^4}) - 3\sigma^4_x = 0$.

**Remark 3**: The probabilistic condition in Theorem 1(ii) implies only that the EL ratio $\ell_{n,\tau}$ can be finitely computed at $\theta_0$ and resembles assumptions made in EL for linear models [cf. Owen (1991)] and in the EL formulation of Monti (1997) [see Section 5.1 of this paper].

The formulation of estimating equations which satisfy $p = 0$ in (5) is necessary for $\ell_n(\theta_0)$ to have a chi-square limit law. As mentioned in Section 3.2, this moment restriction is shared also by the the bootstrap of Dahlhaus and Janas (1996). However, useful and important inference is nonetheless possible with FDEL. (Examples 1, 2, and 4 involve estimating equations with $p = 0$.) For Gaussian inference, there is more flexibility with possible estimating functions in $\ell_{n,\tau}$.

From Theorem 1(i), an approximate $100(1 - \gamma)\%$ joint confidence region for the parameter $\theta$ is given by

$$CR(\theta | \ell_n, 1 - \gamma) = \{\theta : \ell_n(\theta) \leq \chi^2_{\nu,1 - \gamma}\},$$

which has asymptotically correct coverage. If the parameter of interest $\theta$ can be incorporated into an estimating equation of the form $g_\theta = \phi - \theta$ for some functional $\phi$ (for example, autocorrelations or normalized sdf when $\alpha < 1/2$), then the confidence region in (19) will always be convex. For these types of parameters, EL confidence regions are also possible for "smooth" functions of $\theta$ [see Hall and
La Scala (1991)).

If \( f \in \mathcal{F} \) in (6), then a confidence region similar to (19) can be set with \( \ell_{n,\sigma}(\theta) \). Monti (1997) suggested such confidence regions for Whittle estimation with SRD linear processes and spectral densities parameterized as in (8) (e.g., ARMA models). The performance of \( \ell_{n,\sigma} \) under model misspecification closely parallels the behavior of the EL ratio for misspecified regression models described in Owen (1991, Section 5.4). If the candidate density class is incorrect, then consistent variance estimators are replaced by upwardly biased ones within the EL ratio (with extra positive terms due to misspecification). To illustrate, consider the case \( r = p = 1 \) and \( \phi = 0 \). If \( f \neq f_{0} \), then

\[
\ell_{n,\sigma}(\theta_{0}) \triangleq a_{\theta_{0}} \cdot \chi_{i}^{2} + o_{p}(1), \quad a_{\theta_{0}} = \frac{\int_{\mathcal{F}} g_{\sigma_{0}}^{2} f^{2} d\lambda}{\int_{\mathcal{F}} g_{\sigma_{0}}^{2} f^{2} d\lambda + \int_{\mathcal{F}} g_{\sigma_{0}}^{2} (f - f_{0})^{2} d\lambda}
\]

because \( \pi^{2} N^{-1} \sum_{j=1}^{N} g_{\sigma_{0}}^{2}(\lambda_{j}) |I_{n}(\lambda_{j}) - f_{0}(\lambda_{j})|^{2} \) is used to estimate \( n \text{Var}(\int_{0}^{T} g_{\sigma_{0}} I_{nc} d\lambda) = \pi \int_{\mathcal{F}} g_{\sigma_{0}}^{2} f^{2} d\lambda \).

Confidence intervals for parameters become conservative to a degree dependent on the misspecification. Owen (1991) provides more details on the effect of misspecified linear model in EL, which are relevant here.

4.2 Maximum empirical likelihood estimation

Maximum empirical likelihood estimates (MELEs) \( \hat{\theta}_{n} \) and \( \hat{\theta}_{n,\sigma} \) have features and uses similar to those from conventional parametric likelihoods and other EL frameworks [Qin and Lawless (1994), Kitamura (1997)]. We present some important first-order asymptotic properties of frequency domain estimates \( \hat{\theta}_{n} \) and \( \hat{\theta}_{n,\sigma} \) and statistics based on them.

Qin and Lawless (1994) showed the local consistency and asymptotic normality of the MELE in a radius \( n^{-1/3} \) "ball" around \( \theta_{0} \) with iid samples and Kitamura (1997) established the consistency and normal limit of a (time domain) blockwise MELE for mixing SRD processes. For SRD and LRD linear time series, we can establish the existence of EL maximums \( \hat{\theta}_{n} \) and \( \hat{\theta}_{n,\sigma} \) in a neighborhood of \( \theta_{0} \):

\[
B_{n} = \{ \theta \in \Theta : ||\theta - \theta_{0}|| \leq n^{-m} \}, \quad m = \max\{1/3, 1/4 + (1 + \alpha - \beta)/2, (1 + \alpha + \delta)/4\} < 1/2
\]

(see Lemma 8). (Because the likelihood function \( \ell_{n}(\theta) \) or \( \ell_{n,\sigma}(\theta) \) could involve weakly or strongly dependent processes, local maximums are guaranteed in potentially larger regions around the true parameter than in the iid setting.) We prove that these point estimates have an asymptotic normal distribution.

In the following, the Euclidean norm is denoted by \( || \cdot || \).

**Theorem 2** Assume A.1-A.4 hold and \( \phi = 0 \). Suppose, in a neighborhood of \( \theta_{0} \), \( \partial g_{\sigma}(\lambda)/\partial \theta \), \( \partial^{2} g_{\sigma}(\lambda)/\partial \theta \partial \theta' \) are continuous in \( \theta \); \( ||\partial g_{\sigma}(\lambda)/\partial \theta||, ||\partial^{2} g_{\sigma}(\lambda)/\partial \theta \partial \theta'|| \) are bounded by \( G(\lambda) = C|\lambda|^{-\delta} \) for some \( \delta < 1 \),
\[\delta + \alpha < 1; \text{ the set of discontinuities of } \partial g_\theta / \partial \theta \text{ has measure zero; and } D_{\theta_0} \equiv \int \lambda f(\lambda) \partial g_\theta(\lambda) / \partial \theta d\lambda \text{ has full column rank } p.\]

(i) As \( n \to \infty, \)
\[
\sqrt{n} \left( \hat{\theta}_n - \theta_0 \right) \xrightarrow{d} N \left( 0, \begin{bmatrix} V_{\theta_0} & 0 \\ 0 & U_{\theta_0} \end{bmatrix} \right)
\]
where \( V_{\theta_0} = 4\pi \left( D_{\theta_0}' W_{\theta_0}^{-1} D_{\theta_0} \right)^{-1} \) and \( U_{\theta_0} = \pi W_{\theta_0}^{-1} \left( I_{\cdot \cdot \cdot} - \left( 4\pi \right)^{-1} D_{\theta_0} V_{\theta_0} D_{\theta_0}' W_{\theta_0}^{-1} \right). \)

(ii) In addition, suppose A.5 and \( f = f_\theta \) hold; the set of discontinuities of \( \partial f_\theta / \partial \theta \) has measure zero; \( P(0 \in \text{ch} \{ \pi g_\theta(\lambda_j) [I_n(\lambda_j) - f_\theta(\lambda_j)] \}, \theta \in B_n) \to 1; \) and, in the \( \theta_0 \)-neighborhood, \( \int_0^\pi g_\theta(\lambda) f_\theta(\lambda) d\lambda = \varphi \) and \( \| \partial f_\theta / \partial \theta \|, \| \partial^2 f_\theta(\lambda) / \partial \theta^2 \| \leq C|\lambda|^{-\alpha}, \lambda \in (0, \pi]. \) Then, the distributional result in (i) is valid for \((\hat{\theta}_{n,r} - \theta_0, \hat{\theta}_{n,r}/2)^r. \)

Remark 4: As in Theorem 1(iii), if \( k_{4,\varepsilon} = 0 \) then Theorem 2(ii) holds even if \( p \neq 0 \) in (5). When assuming that \( f \in F \) belongs to a selected density class in (ii), a constant function \( \int_0^\pi g_\theta(\lambda) f_\theta(\lambda) d\lambda = \varphi \) of \( \theta \) represents a natural relationship between the chosen estimating functions and \( F. \) For example, the Whittle estimating equations \( g_\theta^w \) in (10) would often satisfy this condition. The additional conditions on \( f \) in Theorem 2(ii) are satisfied, for example, by many SRD and LRD processes, including those densities in (2) and (3) [cf. Fox and Taqqu (1986)]. The probability assumption in Theorem 2(ii) implies that the EL ratio in (14) can be computed finitely in a neighborhood of \( \theta_0. \)

Combined with a proper variance estimate, Theorem 2 can be used to set approximate confidence intervals for \( \theta. \) The asymptotic variance of \( \sqrt{n}(\hat{\theta}_n - \theta_0) \) is consistently estimated by
\[
\hat{V}_{\theta_0}^{-1} = N^{-1} \left( \sum_{j=1}^N \frac{\partial g_\theta(\lambda_j)}{\partial \theta} I_n(\lambda_j) \right)' \left( \sum_{j=1}^N g_\theta^2(\lambda_j) I_n^2(\lambda_j) \right)^{-1} \left( \sum_{j=1}^N \frac{\partial g_\theta(\lambda_j)}{\partial \theta} I_n(\lambda_j) \right),
\]
A consistent estimator also results by weighting each periodogram ordinate by \( p_j(\hat{\theta}_n) \) from (13) (and removing \( N^{-1} \)) and analogous estimators can be formulated with \( \hat{\theta}_{n,r}. \)

The MELEs \( \hat{\theta}_n \) and \( \hat{\theta}_{n,r} \) allow testing of parameter and moment hypotheses with further log-likelihood ratio statistics, as in the estimating equation framework of Qin and Lawless (1994). To test the parameter assumption \( H_0 : \theta = \theta_0, \) we form an EL ratio \( L_n(\theta_0)/L_n(\hat{\theta}_n) \) [or \( L_{n,r}(\theta_0)/L_{n,r}(\hat{\theta}_{n,r}) \)] as we would with a parametric likelihood. After a log-transformation, the resulting statistic \( \ell_n(\theta_0) - \ell_n(\hat{\theta}_n) \) [or \( \ell_{n,r}(\theta_0) - \ell_{n,r}(\hat{\theta}_{n,r}) \)] has a limiting \( \chi^2 \) distribution for assessing \( H_0. \)
The log-ratio statistics \( \ell_n(\hat{\theta}_n) \) and \( \ell_{n,r}(\hat{\theta}_{n,r}) \) are also useful for testing \( H_0: \) the true parameter \( \theta_0 \) satisfies the spectral mean condition in (5). The practice of testing implications, formed with moment restrictions based on estimating functions, is common in economic applications and much recent research has focused on time-domain tests [cf. Imbens, Spady, and Johnson (1998)]. With \( \ell_n(\hat{\theta}_n) \) or \( \ell_{n,r}(\hat{\theta}_{n,r}) \), spectral moment conditions can also be tested.

Intuitively, the log-EL ratio \( \ell_n(\theta_0) \) admits a decomposition into separate parameter and moment assessment contrasts: \( \ell_n(\theta_0) = \{ \ell_n(\theta_0) - \ell_n(\hat{\theta}_n) \} + \ell_n(\hat{\theta}_n) \). We show these contrast statistics have limiting chi-square distributions for testing the above hypotheses.

**Theorem 3** Under the assumptions of Theorem 2,

(i) \( \ell_n(\theta_0) - \ell_n(\hat{\theta}_n) \) and \( \ell_n(\hat{\theta}_n) \) are asymptotically independent and

\[
\ell_n(\theta_0) - \ell_n(\hat{\theta}_n) \xrightarrow{d} \chi^2_p, \quad \ell_n(\hat{\theta}_n) \xrightarrow{d} \chi^2_{r-p}.
\]

(ii) Additionally, if assumptions in Theorem 2(ii) are satisfied, then \( \ell_{n,r}(\theta_0) - \ell_{n,r}(\hat{\theta}_{n,r}) \xrightarrow{d} \chi^2_p, \quad \ell_{n,r}(\hat{\theta}_{n,r}) \xrightarrow{d} \chi^2_{r-p} \) and both statistics are asymptotically independent.

If \( \kappa_4, \kappa_r = 0 \) and the moment conditions in (5) are set with \( p \neq 0 \), the test statistics in Theorem 3(ii) again retain their distributional properties if the parameter and moment assumptions hold true.

**Example 1** (continued). For an overidentifying number of estimating functions \( (r > p) \), the asymptotic variance \( \text{Var}(\hat{\theta}_n) \) in Theorem 2 cannot decrease if an estimating function is dropped [cf. Qin and Lawles (1994)]. Additional parameter information can sharpen the performance (accuracy) of confidence regions and test statistics. We examine a simple to illustrate how estimating functions may useful even when there is no apparent advantage for variance reduction.

Suppose estimation of \( p(1) = \theta \) is sought with \( \tilde{g}_{\theta}(\lambda) = (\tilde{g}_{1,\theta}(\lambda), \tilde{g}_{2,\theta}(\lambda))' \) as described in Example 1. By this function selection, we have again \( r = 2, p = 1 \) and \( p = 0 \) in (5). If \( \{X_t\} \) represents a stationary AR(1) process (satisfying the moment conditions with \( \tilde{g}_{\theta} \)), it follows that:

\[
w_{11} = -d_1/(4\pi \sigma_1^2), \quad w_{12} = w_{21} = -d_2/(4\pi \sigma_2^2), \quad W_{\theta_0} = \begin{bmatrix} w_{11} & w_{12} \\
 w_{21} & w_{22} \end{bmatrix}, \quad D_{\theta_0} = (d_1, d_2)',
\]

implying \( \text{Var}(\hat{\theta}_n) = 4\pi w_{11}/d_1^2 = (1 - \theta_0^2) \). When the AR(1) model holds, the extra information from \( \tilde{g}_{2,\theta} \) in \( \tilde{g}_{\theta} \) does not help (or hurt) in reducing the limiting variance of \( \sqrt{n} \cdot \hat{\theta}_n \), which is the same as the MELE based solely on \( \tilde{g}_{1,\theta} \). However, \( \tilde{g}_{\theta} \) still permits a nonparametric likelihood assessment of the spectral moment assumption \( p(2) = \theta^2 \) using \( \ell_n(\hat{\theta}_n) \).
5 Applications

5.1 Whittle estimation

Example 4 (continued). Monti (1997), considering periodogram-based EL for Whittle-like estimation with SRD linear processes, suggested confidence regions based on $\ell_{n,r}(\theta)$ for spectral density parameters $\theta$ determining $f_\theta \in \mathcal{F}$, for some given density class. However, Theorem 1 indicates that some suggestions and results of Monti (1997) may not be entirely correct.

- Monti (1997) constructs a likelihood function from standardized ordinates $I_n(\lambda_j)/f_\theta(\lambda_j), j = 1, \ldots, N$, treated as approximately iid random variables (in fact, the development philosophically shares the same starting point as the FDB-DJ). The EL ratio in (4.1) of Monti (1997) essentially corresponds to $\ell_{n,r}(\theta)$ using the estimating functions $g_\theta^w$ from (10), which are intended for inference on parameters $\theta = (\sigma^2, \theta')'$ characterizing $f_\theta \in \mathcal{F}$ from (8). For this choice of functions, the spectral mean $p_w \neq 0$ in (10) due to the first estimating function $f_\theta^{-1}$ intended to additionally prescribe $\sigma^2$.

Theorem 1 implies the choice of $\ell_{n,r}(\theta)$ and $g_\theta^w$ (with $p_w \neq 0$) seems most appropriate when the 4-th order innovation cumulant $\kappa_{4,\epsilon} = 0$. Indeed, Dahlhaus and Janas (1996) describe the inconsistency of the bootstrap Whittle estimate of $\sigma^2$ when $\kappa_{4,\epsilon} \neq 0$ and one might anticipate similar problems for FDEL as well. It appears however that Monti (1997) implicitly assumes the 4-th order innovation cumulant is zero (if not, the statement of the Whittle estimator's [inverse] variance matrix "V" [Monti (1997), p. 404] appears mistaken as there should be additional components depending on $\kappa_{4,\epsilon}$ owing to $f_\theta^{-1}$). Hence, we respectfully question Monti's (1997) claim that the author's EL formulation is generally valid for both Gaussian and non-Gaussian SRD linear processes. The EL ratio statistic $\ell_{n,r}(\theta)$ may arguably be inadequate for setting joint confidence regions for Whittle parameters minimizing (7) when the model class is misspecified.

- Treating $\sigma^2$ as a nuisance parameter and concentrating it out of the Whittle likelihood function [see (9)], Monti (1997) suggests a EL ratio statistic for estimation of the remaining $p - 1$ parameters $\theta$ in (8) via confidence regions. The statistic (6.1) of that paper behaves asymptotically like $1/2 \cdot \ell_n(\theta)$ based on the $p - 1$ estimating functions $g_\theta^{w*}$ from (11). The author then claims this statistic has a limiting chi-square distribution with $p - 1$ degrees of freedom (evaluated at the true $\theta_0$). Note that, for the functions $g_\theta^{w*}$, it holds that $p_{w*} = 0$. However, Theorem 1(i) and (18) imply Monti's (1997) normalization of the EL ratio is incorrect.
For quasi-Whittle estimation in the parameterization from (8), \( t_\theta \) based on the functions \( g_{\theta}^{*\star} \) in (11) appears preferable. This selection results in asymptotically correct confidence regions for \( \theta \) under both SRD and LRD, even for some misspecified situations where the moments in (9) still hold. We do inherently treat \( \sigma^2 \) as a nuisance parameter in this case. Given parameters \( \hat{\theta} \in CR(\theta | \ell_n, 1 - \gamma) \) (using \( g_{\theta}^{*\star} \) in \( \ell_n(\theta) \)), one can sensibly pick an associated, data-based estimate of \( \sigma^2 \): from (13), take

\[
\sigma^2 \ell_n = \sum_{j=1}^{N} p_j(\hat{\theta}) k^{-1}(\lambda_j, \theta) I_n(\lambda_j).
\]

The profile likelihood ratio here admittedly does not simultaneously suggest a range of \( \sigma^2 \) values along with \( \theta \), but possibly this may be expected. As mentioned previously, the FDB-DJ usually has difficulties with estimation of \( \sigma^2 \) and, for Whittle estimation under LRD, estimates of \( \sigma^2 \) are not typically included in the distributional results given for other parameter estimates, eg. \( \theta \). [cf. Fox and Taqqu (1986), Giraitis and Surgailis (1990), Heyde and Gay (1993)]. Dahlhaus (1989) establishes a CLT for Whittle estimates of \( \theta = (\sigma^2, \theta') \) for correct model-specified, LRD Gaussian processes and, for this case, one could use \( g_{\theta}^{*\star} \) and \( \ell_n, \ell_{\theta} \) to make valid confidence regions for \( \theta \) under Theorem 1(iii).

5.2 Goodness-of-fit tests

Example 3 (continued). We return to the simple hypothesis test \( H_0 : f = f_0 \) for some possible density \( f_0 \). To assess the goodness of fit, Milhoj (1981) and Beran (1992) proposed the test statistic

\[
T_n = A_n/(2B_n^2), \quad A_n = \frac{2\pi}{n} \sum_{j=1}^{N} \left( \frac{I_n(\lambda_j)}{f_0(\lambda_j)} \right)^2, \quad B_n = \frac{2\pi}{n} \sum_{j=1}^{N} \frac{I_n(\lambda_j)}{f_0(\lambda_j)},
\]

for mixing linear processes and long-memory Gaussian processes, respectively. The variable \( 2\pi T_n \) represents a sum of squared autocorrelation estimates of the innovations process \( \{e_i\} \) using all possibly estimable lags 1 through \( n - 1 \). Both authors show \( T_n \) has an asymptotic normal distribution and its distribution may also be calculated through the limiting bivariate normal law of \( \sqrt{n}\{(A_n, B_n') - (2\pi, \pi')\} \) under the null hypothesis.

For this test, the involved time processes are linear with \( \kappa_{4, \lambda} = 0 \) by Gaussianity (or at least assumed so [see Milhoj (1981)]). By Theorem 1(iii), we can construct a single statistic \( \ell_{n, \gamma} \) to test \( H_0 \) in the same setting by treating \( \mathcal{F} = \{f_0\} \) in (6) and taking a single estimating function \( f_0^{-1} \) (which satisfies (5) with \( p = \pi \) under \( H_0 \)). We reject \( H_0 \) if \( \ell_{n, \gamma} > \chi_{p, 1 - \gamma}^2. \) By a first order expansion, we find

\[
\ell_{n, \gamma} = n (B_n - \pi)^2 / (\pi A_n) + o_p(1),
\]

and the EL ratio statistic asymptotically incorporates much of the same information in \( 2\pi T_n \) under \( H_0 \). If \( f \neq f_0 \), the power of the test depends on the degree to which \( n > \pi \int_0^\gamma (f/f_0 - 1)^2 d\lambda / \int_0^\gamma (f/f_0 - 1)^2 d\lambda \).

If \( f \neq f_0 \), the power of the test depends on the degree to which \( n > \pi \int_0^\gamma (f/f_0 - 1)^2 d\lambda / \int_0^\gamma (f/f_0 - 1)^2 d\lambda \).
1)dλ)^2 ≥ 1 and \( T_n \) has similar power characteristics under \( H_1 \) [Beran (1992)].

The choice of constant \( f_0 \) on \( \Pi \) coincides with the claim that \( \{X_t\} \) is white noise. A goodness-of-fit test based on the process autocorrelations for the first \( m \) lags can be applied, as considered in corresponding Portmanteau tests [cf. Box and Ljung (1978), Li and McLeod (1986)]. The estimating functions \( g(\lambda) = (\cos(\lambda), \ldots, \cos(m\lambda))' \) satisfy (5) with \( p = 0 \in \mathbb{R}^m \) under this \( H_0 \) and yield a single EL ratio \( \ell_n \) which pools information across \( m \) EL estimated autocorrelation lags [see Example 1].

Example 5. Consider testing a composite hypothesis \( H_0 : f \in \mathcal{F} \), that the spectral density belongs to a specified parametric class. Although the exact test statistics differ in form, Milhoj (1981), Beran (1992), and Paparoditis (2000) have proposed frequency domain tests for \( H_0 \) which share similar characteristics. Each test procedure uses Whittle estimation to select the “best fitting” model in a collection \( \mathcal{F} \) from (6) and then compares this fitted density to the periodogram (the sample density) across all ordinates. Beran (1992) and Paparoditis (2000) formulate their tests for LRD and SRD Gaussian processes, respectively. We briefly show that FDEL techniques can produce similar goodness-of-fit tests, while at the same time expanding our EL theory slightly.

Suppose \( \{X_t\} \) is a Gaussian time series and we wish to test if \( f \in \mathcal{F} \) for some parametric family as in (8). The class \( \mathcal{F} \) could correspond to SRD models as in Paparoditis (2000) or the LRD spectral densities as in Beran (1992), including (2) and (3). In the spirit of their proposed tests, we explore here EL methods to simultaneously incorporate both components of model fitting and model comparison. To this end, consider possible spectral moment conditions

\[
\int_0^\pi g_\theta' f \, d\lambda = \nu_\theta \quad \int_0^\pi (f / f_\theta)^2 \, d\lambda = \pi, \tag{20}
\]

where \( g_\theta = (f_\theta^{-1}, \partial f_\theta^{-1}/\partial \theta')' \) are the Whittle estimating functions from (10) for the parameters \( \theta = (\sigma^2, \theta')' \in \mathbb{R}^p \) in \( f_\theta \). Note that we introduce an overidentifying \( L_2 \) moment restriction on \( f \) (so that \( r = p + 1 \)). We then extend the log-likelihood statistic \( \ell_{n, \mathcal{F}} \) in (18) to include \( I_n^2 \) ordinates: \( \ell_{n, \mathcal{F}}(\theta) = -2 \log R_{n, \mathcal{F}}(\theta) \) for

\[
R_{n, \mathcal{F}}(\theta) = N^N \sup \left\{ \prod_{j=1}^N p_j : p_j \geq 0, \sum_{j=1}^N p_j = 1, n \sum_{j=1}^N p_j y_\theta(\lambda_j) = 0 \right\},
\]

\[
y_\theta(\lambda_j) = \left( \frac{I_{n, \mathcal{F}}^2(\lambda_j) / \{2f_\theta^2(\lambda_j)\}}{f_\theta^2(\lambda_j) / \{2f_\theta^2(\lambda_j)\}} - 1 \right) \frac{g_\theta(\lambda_j)}{g_\theta(\lambda_j) - f_\theta(\lambda_j)} \right). \tag{21}
\]

We use \( f_\theta \) and \( f_\theta^2 \) in (21) to approximate the means of \( I_n \) and \( I_n^2 / 2 \) for each ordinate. To evaluate \( H_0 : f \in \mathcal{F} \), we can test if the moment conditions in (20) hold for some \( \theta \) value. Following the testing
prescription in Theorem 3, we find the argument maximum of \( R_{1,2}(\theta) \), say \( \hat{\theta}_{1,2} \), and form a test statistic \( \ell_{1,2}(\hat{\theta}_{1,2}) \) for \( H_0 \).

We can then extend Theorem 3 to obtain a distributional result for our test statistic.

**Proposition 1** Suppose \( \{X_i\} \) is Gaussian and the assumptions in Theorem 3(ii) hold for \( f_0 \) and \( g_f \) with \( \alpha - \beta < \eta \), for an arbitrarily small \( \eta > 0 \). Under the null hypothesis \( f = f_0 \in \mathcal{F} \),

\[
\ell_{1,2}(\hat{\theta}_{1,2}) \xrightarrow{d} \chi^2_1.
\]

The 1 degree of freedom results from the overidentifying estimating equation using \( f^2 \). The distributional result is valid even with nonzero spectral mean conditions (i.e. \( \rho_w \neq 0 \)) because the process is Gaussian.

The power of the test will not be considered extensively here, but we can make a few comments about model misspecification. Suppose \( f \notin \mathcal{F} \) but \( \theta_0 \) still represents the parameter value which minimizes the asymptotic distance measure \( W(\theta) \) in (7) and \( f_{\theta_0} \) satisfies (9) (i.e. the moment condition \( \int g_{\theta_0} \, d\lambda = \rho_w \) holds). Under technical assumptions like those in Proposition 1, we can establish a consistency property: as \( n \to \infty \),

\[
n^{-1} \ell_{1,2}(\hat{\theta}_{1,2}) \xrightarrow{p} a_0 \left\{ \int_0^\infty \left( \frac{f(\lambda)}{f_{\theta_0}(\lambda)} - 1 \right)^2 \, d\lambda \right\} > 0
\]

where \( a_0 > 0 \) depends on \( f \) and \( f_{\theta_0} \). We are assured that the test statistic can at least determine if \( H_0 : f \in \mathcal{F} \) is true as the sample size increases. However, the exact power will depend on the misspecification as in the goodness-of-fit test proposed by Beran (1992).

### 6 Extensions

#### 6.1 Inference with parameter restrictions

For inference on subsets of parameters or functions of them, we can consider EL estimation subject to a system of parameter constraints on \( \theta \):

\[
\psi(\theta) = 0 \in \mathbb{R}^q
\]

(22)

where \( q < p \) and \( \Psi(\theta) = \partial \psi(\theta) / \partial \theta \) is of full row rank \( q \). By maximizing the EL functions in (12) or (15) under the above restrictions, we find constrained MELEs \( \hat{\theta}^*_n \) or \( \hat{\theta}^*_{n,r} \). Following the fully parametric likelihood framework [cf. Aitchison and Silvey (1958)], we can then use FDEL sequentially to

1. test if the true, unknown parameter value satisfies \( H_0 : \psi(\theta_0) = 0 \) with a log-likelihood ratio statistic \( \ell_n(\hat{\theta}_n) - \ell_n(\hat{\theta}_n) \).
2. if failing to reject $H_0$, make an approximate confidence region for constrained $\theta$ values

$$CT^\psi(\theta | n, 1 - \gamma) = \{\theta : \psi(\theta) = 0, \ell_n(\theta) - \ell_n(\hat{\theta}_n^*) \leq \chi^2_{p-q}\}.$$ 

One could use $\hat{\theta}_{n,r}^*$, $\hat{\theta}_{n,r}$ and $\ell_{n,r}$ above as well. We provide the necessary large sample distributional results in the following theorem.

**Theorem 4** Suppose the conditions in Theorem 2 hold and, in an neighborhood of $\theta_0$, $\psi(\theta)$ is continuously differentiable, $||\partial^2 \psi(\theta)/\partial \theta \partial \theta'||$ is bounded, and $\Psi(\theta_0)$ is rank $q$. Then, under $H_0 : \psi(\theta_0) = 0$,

(i) $\ell_n(\hat{\theta}_n^*) - \ell_n(\hat{\theta}_n) \overset{d}{\to} \chi^2_q$ and $\ell_n(\theta_0) - \ell_n(\hat{\theta}_n^*) \overset{d}{\to} \chi^2_{p-q}$.

(ii) If assumptions in Theorem 2(ii) are satisfied as well,

$$\ell_{n,r}(\hat{\theta}_{n,r}^*) - \ell_n(\hat{\theta}_{n,r}) \overset{d}{\to} \chi^2_q, \ell_n(\theta_0) - \ell_n(\hat{\theta}_{n,r}^*) \overset{d}{\to} \chi^2_{p-q}.$$

The degrees of freedom in the test statistic $\ell_n(\hat{\theta}_n^*) - \ell_n(\hat{\theta}_n)$ correspond naturally to the number of parameter restrictions $q$. Qin and Lawless (1995) first introduced a similar statistic for independent samples and $p = r$ estimating equations (for which $\ell_n(\hat{\theta}_n) = 0$ necessarily) and Kitamura (1997) provided a blockwise version for time domain EL under SRD. Qin and Lawless (1995) describe how to practically implement a constrained maximization of the EL function subject to restrictions as in (22).

### 6.2 Results with tapering

We next consider what happens in our EL framework when possibly tapered observations are used. The results are directly applicable to SRD processes. (Incomplete distributional theory for quadratic forms hinders immediate tapering extensions in FDEL to LRD.)

Let $h : [0, 1] \to \mathbb{R}$ be a function of bounded variation such that $\int_0^1 h^2(x)dx \neq 0$. Then the periodogram of $X_1, \ldots, X_n$ under the “data-taper” function $h(\cdot)$ is

$$I_n^{(T)}(\lambda) = \{2\pi H_2^{(T)}(0)\}^{-1} \left| \sum_{t=0}^{n-1} X_{t+1} h(t/n) \exp(-it\lambda) \right|^2, \lambda \in \Pi,$$

using the window

$$H_2^{(T)}(\lambda) = \sum_{k=0}^{n-1} h^k(t/n) \exp(-it\lambda), \quad k \in \{1, 2, \ldots\}.$$

A data-taper is typically used for handling missing observations (where the function $h(\cdot)$ is set equal to zero over an interval corresponding to a missing data segment) and for reducing leakage, especially where the spectrum contains high peaks [cf. Brillinger (1981), Dalhause and Künsch (1987)].

We simply substitute the tapered periodogram $I_n^{(T)}$ in the definitions of the EL functions $L_n$, $L_{n,r}$
from (12) and (15). We can then obtain MELEs and produce log-profile EL ratios \( \ell_n^T(\theta) \) and \( \ell_{n,r}^T(\theta) \) as before (using the same notation). Tapering in the periodogram however changes the asymptotic variance of spectral mean estimators in (17), a well-known result for SRD time series [cf. Dahlhaus (1983, 1985a)]. Hence, we need to make small adjustments to our developed EL statistics to ensure they have correct asymptotic properties under tapering, while preserving potential small sample-benefits of the taper. To this end, we introduce a scaling factor to account for the taper

\[
A_n^T = \{ H_2^T(0) \}^2 / \{ nH_4^T(0) \}
\]

and show that it provides a simple correction to EL statistics in the following revision of Theorem 1.

We quantify the SRD of the linear process to be tapered [Dahlhaus (1985, Corollary 3.2)] and use estimating functions \( g_0 \) of bounded variation in the theorem.

**Theorem 5** Assume A.1 is satisfied with \( \mu = 0 \) and the filter coefficients are \( b_j = O(j^{-\rho}) \) for \( \rho > 1 \). Suppose \( g_0 \) is componentwise of bounded variation with finite discontinuities and A.4 holds. If \( \rho = 0 \), then

(i) \( A_n^T \ell_n^T(\theta_0) \xrightarrow{d} \chi_1^2 \).

(ii) \( A_n^T \ell_{n,r}^T(\theta_0) \xrightarrow{d} \chi_r^2 \), if \( f = f_{\theta_0} \).

(iii) If \( \kappa = 0 \), statement (ii) remains valid if \( \rho \neq 0 \in \mathbb{R}^r \).

**Remark 5:** We can also reformulate Theorems 2-4 by multiplying \( \ell_{n,r}(\theta) \) and \( \ell_{n,r}(\theta) \) by \( A_n^T \) (and substituting \( V_{\theta_0} \) and \( U_{\theta_0} \) in Theorem 2 with \( V'_{\theta_0} = H \cdot V_{\theta_0} \) and \( U'_{\theta_0} = H \cdot U_{\theta_0} \) for \( H = \int_0^1 h^4 dx / (\int_0^1 h^2 dx)^2 \)).

Dahlhaus and Janas (1996, p. 1953) suggest a tapering correction to improve the frequency domain bootstrap approximation similar to \( A_n^T \) and Theorem 5 shows that adjustments should be made in EL to accommodate tapered periodogram ordinates. In the nontapered case \( h(x) \equiv 1 \), we have \( A_n^T = 1 \) which reduces to the original log-EL ratios in Section 4. Tapering adjustments to EL confidence regions were not used in Monti (1997, Section 7).

We remark finally that the EL results with the tapered periodogram remain valid even if sample mean corrected observations are used \( \{ X_t - \hat{X}_n \}^2_{t=1} \) when \( \mu \neq 0 \).

7 Proofs

We first develop some additional notation and useful functions to help with the proofs. Write \( \lambda_j = 2\pi j/n, j \in \mathbb{Z} \). In the following, \( C \) or \( C(\cdot) \) will denote generic constants that depend on their
arguments (if any) but do not depend on \( n \) or ordinates \( \{\lambda_j\}_{j=1}^{[n/2]} \).

Define the mean corrected discrete Fourier transforms

\[
d_{nc}(\lambda) = \sum_{i=1}^{n} (X_i - \mu)e^{-it\lambda}, \quad \lambda \in \Pi.
\]

Then, \( I_{nc}(\lambda) = (2\pi n)^{-1}|d_{nc}(\lambda)|^2 \), \( |d_{nc}(\lambda)|^2 = d_{nc}(\lambda)d_{nc}(-\lambda) \). It holds that

\[
I_{nc}(\lambda_i) = I_n(\lambda_i), \quad \lambda_i \in \{\lambda_j\}_{j=1}^{[n/2]}, \quad (23)
\]

the mean-corrected and uncorrected periodogram are equal for each ordinate \( \{\lambda_j\}_{j=1}^{N} \).

Let \( H_n(\lambda) = \sum_{t=1}^{n} e^{-it\lambda}, \quad \lambda \in \mathbb{R} \), and write \( K_n(\lambda) = (2\pi n)^{-1}|H_n(\lambda)|^2 \) to denote the Fejer kernel. The function \( K_n \) is nonnegative, even with period \( 2\pi \) on \( \mathbb{R} \) and \( \int_{\Pi} K_n d\lambda = 1 \) [see Brockwell and Davis (1991, p. 71) for these and other basic properties of \( K_n \)].

We adopt the standard that an even function \( g : \Pi \rightarrow \mathbb{R} \) can be periodically extended to \( \mathbb{R} \), with period \( 2\pi \), by \( g(\lambda) = g(-\lambda) \), \( g(\lambda) = g(\lambda + 2\pi) \) for \( \lambda \in \mathbb{R} \). When \( g \) is integrable, define the \( n \)th Cesaro sum of the Fourier series of \( g \) as

\[
c_n g(\lambda) = \int_{\Pi} K_n(\lambda - x)g(x) dx, \quad \lambda \in \Pi
\]

[cf. Edwards (1979), Chapter 5]. For \( g : \mathbb{R} \rightarrow \mathbb{R} \), denote the supremum norm: \( \|g\|_{\infty} = \sup\{\|g(x)\| : x \in \mathbb{R}\} \).

We will make extensive use of the following function from Dahlhaus (1983): Let \( L_{ns} : \mathbb{R} \rightarrow \mathbb{R} \) be the periodic extension (with period \( 2\pi \)) of

\[
L_{ns}(\lambda) = \begin{cases} 
 e^{-s\lambda n} & |\lambda| \leq e^{s}/n \\
 \ln^n(\lambda) & \frac{e^{s}/n}{|\lambda|} < |\lambda| \leq \pi,
\end{cases} \quad \lambda \in \Pi, \quad s = 0, 1.
\]

Then for each \( n \geq 1 \), \( s \in \{0, 1\} \), \( L_{ns}(\cdot) \) is decreasing on \([0, \pi]\) and

\[
|H_n(\lambda)| \leq C L_{n0}(\lambda), \quad \lambda \in \mathbb{R} \quad (24)
\]

Another convenient bound on \( H_n(\lambda) \) is

\[
|H_n(\lambda)| \leq \frac{Cn}{(1 + |\lambda \text{ mod } 2\pi n|)}, \quad \lambda \in \mathbb{R}, \quad \left( L_{n0}(\lambda) \leq \frac{3\pi n}{(1 + |\lambda \text{ mod } 2\pi n|)} \right) \quad (25)
\]

which we will also employ at times. Note \( \lambda \text{ mod } 2\pi \in \Pi \).

We require a few lemmas for the proofs. The first lemma develops bounds for cumulants of discrete Fourier transforms.
Lemma 1 Let \( n \geq 3, 1 \leq i \leq j \leq N \), and \( a_1, \ldots, a_k \in \{\pm \lambda_1, \pm \lambda_j\} \), \( |a_1| \leq |a_2| \leq \cdots \leq |a_k| \) with \( 2 \leq k \leq 8 \). Under Assumption A.1,

\[
\begin{align*}
(i) & \quad \left| \text{cum}(d_{ne}(a_1), d_{ne}(a_2)) \right| \leq C \begin{cases} 
|a_1|^{-\alpha} (|a_2|^{-1} + L_n^1(a_1 + a_2)) & \text{if } \alpha > 0 \\
L_n^1(a_1 + a_2) & \text{if } \alpha = 0
\end{cases} \\
(ii) & \quad \left| \text{cum}(d_{ne}(a_1), \ldots, d_{ne}(a_k)) \right| \leq C \left| a_k |^{\alpha/2-1} |a_{k-1}^{-1/2} + n \ln^{k-1}(n) \right| \prod_{j=1}^{k} |a_j|^{-\alpha/2}.
\end{align*}
\]

The proof of Lemma 1 appears in the Appendix. The next lemma is useful for some evaluations of \( L_n^0 \) and \( L_n^1 \) and describes some basic properties of these functions.

Lemma 2 Let \( 1 \leq i \leq j \leq N \). If \( \tilde{\lambda}_i \in \{\pm \lambda_i\}, \tilde{\lambda}_j \in \{\pm \lambda_j\}, \tilde{\lambda}_i + \tilde{\lambda}_j \neq 0 \), then

\[
\begin{align*}
(i) & \quad L_n^0(\tilde{\lambda}_i + \tilde{\lambda}_j) \leq (2\pi)^{-1} n c_{ij}, \quad c_{ij} = \begin{cases} 
(j - i)^{-1} & \text{sign} \tilde{\lambda}_i \neq \text{sign} \tilde{\lambda}_j \\
(j + i)^{-1} & \text{sign} \tilde{\lambda}_i = \text{sign} \tilde{\lambda}_j, i + j \leq n/4 \\
(n - j - i)^{-1} & \text{sign} \tilde{\lambda}_i = \text{sign} \tilde{\lambda}_j, i + j > n/4.
\end{cases} \\
(ii) & \quad L_n^1(\tilde{\lambda}_i + \tilde{\lambda}_j) \leq \ln(n\pi) L_n^0(\tilde{\lambda}_i + \tilde{\lambda}_j), \\
& \quad L_n^1(\tilde{\lambda}_i + \tilde{\lambda}_j) \leq C n \{c_{ij} \}^d, \text{ for a given } d, \max\{\alpha, \delta, 1/2\} < d < 1.
\end{align*}
\]

(iii) For integers \( j > i \geq 1 \),

\[
\begin{align*}
\frac{2}{(n - j - i)} \leq j \leq \frac{n^2}{(n - j - i)} & \quad \text{or } i = 1; \\
\frac{1}{2} \leq \frac{j}{n} & \quad \text{otherwise.}
\end{align*}
\]

For every integer \( n, r \geq 1 \), there exists \( C > 0 \) independent of \( n \) such that

\[
\begin{align*}
(iv) & \quad \int_\pi L_n^r(\lambda) d\lambda \leq \begin{cases} 
C \ln n & r = 1 \\
C n^{r-1} & r > 1.
\end{cases} \\
(v) & \quad \int_\pi L_n^0(\lambda_1 + \lambda)L_n^0(\lambda_2 + \lambda) d\lambda \leq C L_n^1(\lambda_1 + \lambda_2), \quad \lambda_1, \lambda_2 \in \mathbb{R}.
\end{align*}
\]

proof: Parts (iv) and (v) of Lemma 2 correspond to Lemmas 1-2 of Dahlhaus (1983). Lemma 2 (i) follows from the fact that \( [(\tilde{\lambda}_i + \tilde{\lambda}_j) \mod 2\pi] \geq 2\pi/n \) if \( \tilde{\lambda}_i + \tilde{\lambda}_j \neq 0 \), along with the definition of \( L_n^0 \).

Likewise, we have by definition

\[
L_n^1(\tilde{\lambda}_i + \tilde{\lambda}_j) = (2\pi)^{-1} n c_{ij} \tilde{\lambda}_{ij}, \quad \tilde{\lambda}_{ij} = \begin{cases} 
\ln(2\pi(j - i)) & \text{sign} \tilde{\lambda}_i \neq \text{sign} \tilde{\lambda}_j \\
\ln(2\pi(j + i)) & \text{sign} \tilde{\lambda}_i = \text{sign} \tilde{\lambda}_j, i + j \leq n/4 \\
\ln(2\pi(n - j - i)) & \text{sign} \tilde{\lambda}_i = \text{sign} \tilde{\lambda}_j, i + j > n/4.
\end{cases}
\]

There then exists \( C \) so that \( i^{-1+d} \ln(2\pi i) \leq C \) for \( i \geq 1 \) so that \( \{c_{ij} \}^{1-d} \tilde{\lambda}_{ij} \leq C \) for all \( n, i, j \) and Lemma 2(ii) follows. Part (iii) is easy to show so we omit the proof. \( \square \)
Lemma 3 Suppose Assumption A.1 holds. Let $\Pi_\rho = [\rho, \pi]$ for $0 < \rho < \pi$. If $a_1, a_2 \in \Pi$, $|a_1| \leq |a_2|$, and $|a_2| \in \Pi_\rho$, then

$$\text{cum}(d_n(a_1), d_n(a_2)) = (2\pi)^{-1} H_n(a_1 + a_2)f(a_2) + R_{n\rho},$$

where $R_{n\rho} = o(n)$ uniformly for $|a_2| \in \Pi_\rho$.

Proof: We note that $f$ is continuous on $\Pi_{\rho/2}$ so that, given $\epsilon > 0$, there exists $\delta_\epsilon > 0$ such that $|f(a_2 - \lambda) - f(a_2)| < \epsilon$ wherever $|\lambda| < \delta_\epsilon$, $a_2 \in \Pi_\rho$. Then we follow the proof of Theorem 1a of Dahlhaus (1983) to find

$$|R_{n\rho}| \leq \epsilon C L_{n1}(a_1 + a_2) + CL_{n0}(\delta_\epsilon) \int (|f(a_2 - \lambda)| + |f(a_2)|) L_{n0}(a_1 + a_2 - \lambda) d\lambda$$

for $C$ independent of $\epsilon, \rho, a_1, a_2, n$. If $\alpha > 0$, we pick $1 < r = r(\alpha) < 1/\alpha$ and apply Holder's inequality:

$$\int |f(a_2 - \lambda)| L_{n0}(a_1 + a_2 - \lambda) d\lambda \leq C \int |\lambda|^{-\alpha} L_{n0}(a_1 + \lambda) d\lambda$$

$$\leq C \left[ \int |\lambda|^{-\alpha r} d\lambda \right]^{1/r} \left[ \int L_{n0}^{(r-1)}(\lambda) \right]^{(r-1)/r}$$

with Lemma 2(v). If $\alpha = 0$, the same integral is bounded by $C \ln(n)$ by Lemma 2(iv) because $|f|$ is bounded. Using $|f(a_2)| \leq C \rho^{-\alpha}$ and Lemma 2(iv), we have

$$|R_{n\rho}| \leq \epsilon C L_{n1}(a_1 + a_2) + \begin{cases} C(\rho, \epsilon) \ln(n) & \alpha = 0 \\ C(\rho, \epsilon)n^{1/r} & \alpha > 0, \quad (r = r(\alpha) > 1) \end{cases}$$

where $C$ and $C(\rho, \epsilon)$ are independent of $a_1, a_2, n$ (and $C$ does not depend on $\epsilon$). The order of $R_{n\rho}$ then follows.

The next lemma ensures that the $\mathbb{R}^p$ zero vector lies in $\text{ch}^0 \{\pi g_{\theta_0}(\lambda_j) I_n(\lambda_j)\}_{j=1}^N$ as $n$ increases, where $\text{ch}^0 A$ denotes the interior convex hull of a finite set $A \subset \mathbb{R}^p$. This guarantees that the log-likelihood ratio $\ell_n(\theta_0)$ exists asymptotically.

Lemma 4 Let $\{X_t\}, t \in \mathbb{Z}$, be ergodic with spectral density $f$, continuous on $(0, \pi]$ and $f(\lambda) \leq C|\lambda|^{-\alpha}$, $0 \leq \alpha < 1$. Suppose $g = (g_1, \ldots, g_p)'$ is even with finite discontinuities on $[0, \pi]$ and satisfies Assumption A.2. If $\int \Pi f g d\lambda = 0$ and $W = (\int \Pi f^2 g_i g_j d\lambda)_{i,j=1,\ldots,p}$ is positive definite, then

(i) $\inf_{y \in \mathbb{R}^p, \parallel y \parallel = 1} \int f' g' y d\lambda \geq s > 0$, (ii) $P\left(0 \in \text{ch}^0 \{\pi g(\lambda) I_n(\lambda)\}_{j=1}^N\right) \rightarrow 1.$
**proof:** To obtain (i), we modify the proof of Owen (1990). Suppose (i) is not true and there exists \(\{y_m\}_{m=1}^{\infty}, \|y_m\| = 1\) such that \(\int_{g' \geq 0} g'y_m \, d\lambda < 1/m\). By compactness, there exists a convergent subsequence \(\{y_{m_j}\}\) where \(y_{m_j} \to y_0, \|y_0\| = 1\). On the set \(\{\lambda \in \Omega : y_0g(\lambda) > 0\}\), \(g'y_{m_j}I_{\{g' \geq 0\}} \to g'y_0\) pointwise so that we may apply the Monotone Convergence Theorem (for nonnegative functions):

\[
\int_{g' > 0} g'y_0 \, d\lambda = \lim_j \int_{g' > 0} g'y_{m_j}I_{\{g' > 0\}} \, d\lambda \leq \lim_j \int_{g' > 0} g'y_{m_j} \, d\lambda = 0.
\]

Since \(\int_{\Omega} g'y_0 \, d\lambda = 0\), we then have \(g'y_0 = 0\) almost everywhere (a.e.) on \(\Omega\), implying further that \(0 < y_0^*Wy_0 = \int_{\Omega} (g'y_0)^2 \, d\lambda = 0\), a contradiction. Lemma 4(i) now follows.

For Lemma 4(ii), suppose the discontinuities of \(g\) are at \(0 < a_1 < a_2 \ldots < a_k < \pi\) on \([0, \pi]\). Pick \(\epsilon > 0\) so that: \(\epsilon < \min_{1 \leq i \leq s-1}(a_{i+1} - a_i)/2\) (if \(k > 1\)); \(\epsilon < \min_{a_i < 0} a_i/2\); and \(\epsilon < \min_{\pi < a} (\pi - a_i)/2\).

Define a function \(g_\epsilon\) on \([0, \pi]\) where

\[
g_\epsilon(\lambda) = \begin{cases} g(\lambda) & \lambda - a_i \geq \epsilon, 1 \leq i \leq d \\ \omega g(a_i + \epsilon) + (1 - \omega)g(a_i + \epsilon) & \lambda = a_i + \epsilon(1 - 2\omega), a_i \neq 0, \omega \in [0, 1] \\ g(\epsilon)e^{-1}\lambda & \lambda < \epsilon, a_1 = 0. \end{cases}
\]

We then extend \(g_\epsilon\) periodically on \(\mathbb{R}\).

On \(\Omega\), \(g_\epsilon\) is continuous, \(fg_\epsilon\) is integrable, \(g_\epsilon(\lambda) = g_0(\lambda)\) if \(\min_{1 \leq i \leq s} |\lambda - a_i| \geq \epsilon\), and \(\|g_\epsilon\|_\infty \leq \|g_0\|_\infty \leq M\), where we denote \(g_0 \equiv g\) for convenience. Pick \(\delta_\epsilon > 0\) so that \(\|g_\epsilon(x_1) - g_\epsilon(x_2)\| < \epsilon\) if \(|x_1 - x_2| < \delta_\epsilon\) on \(\Omega\) and \(n\) large so that \((2M)^{-1}\int_{|x| > \delta_\epsilon} K_n(x) \, dx \leq \epsilon\) [cf. Brockwell and Davis (1991), p. 71].

For \(m \in \{0, \pi\}\) and \(\|y\| = 1\), let \(h_{my}(\lambda) = (|g_{my}'| + g_{my}')/2\), \(\lambda \in \Omega\). Then for large \(n\),

\[
|c_nh_{m\lambda}(\lambda) - h_{m\lambda}(\lambda)| \leq \int_{\Omega} K_n(\lambda - x)|h_{m\lambda}(x) - h_{m\lambda}(\lambda)| \, dx \
\leq \int_{\Omega} K_n(\lambda - x)||g_\epsilon(x) - g_\epsilon(\lambda)|| \, dx \
\leq \epsilon + 2M \int_{|x| > \delta_\epsilon} K_n(x) \, dx \leq 2\epsilon.
\]

That is, the \(n\)th Cesaro mean of the Fourier series of \(h_{m\lambda}(\lambda)\) converges uniformly in \(\lambda \in \Omega\) and \(\|y\| = 1\).

We can then follow the same arguments as in Hannan (1973) or Rosenblatt (2000, Theorem 2.2.1), using the process ergodicity, to show

\[
A_{n1} = \sup_{\|y\| = 1} \left| \frac{2\pi}{n} \sum_{j=1}^{n/2} \text{Im}(\lambda_j)h_{m\lambda}(\lambda_j) - \int_{\Omega} h_{m\lambda}(\lambda)f(\lambda) \, d\lambda \right| \longrightarrow 0 \text{ a.s. P.} \quad (26)
\]

By construction (\(\|g_\epsilon\|_\infty \leq \|g_0\|_\infty \leq M\)) and \(\lambda^{-\alpha} \leq (\min_{a_i < 0} a_i/2)^{-\alpha}\) on \([0, \pi]\) if \(|\lambda - a_i| \leq \epsilon\) for some
2 \leq i \leq d \text{ or if } |\lambda - a_i| \leq \epsilon, \ a_i \neq 0. \ Then,

\begin{align*}
A_{n2} &= \sup_{\|y\|=1} \left| \int h_{y^*}(\lambda) f(\lambda) d\lambda - \int h_{0y} f(\lambda) d\lambda \right| \\
&\leq C \sum_{i=1}^{d} \int_{|\lambda| - \epsilon} (\|g_{0i}(\lambda)\| + \|g_{i}(\lambda)\|)|\lambda|^{-\alpha} d\lambda \\
&\leq C \left[ \epsilon + I_{(a_i=0)} \epsilon^{-1+\beta} \int_{|\lambda| \leq \epsilon} |\lambda|^{-\alpha+1} d\lambda \right] \\
&\leq C \left[ \epsilon + \epsilon^{1+\beta} \right],
\end{align*}

for \( C \) independent of \( \epsilon \). Since \( \epsilon \) above can be made arbitrarily small, we have \( A_{n2} = o(1) \).

We also write

\begin{align*}
A_{n3} &= \sup_{\|y\|=1} \left| 2\pi n \sum_{j=1}^{N} I_{nc}(\lambda_j) \{ h_{y^*}(\lambda_j) - h_{0y}(\lambda_j) \} \right| \\
&\leq C \frac{d}{n} \sum_{i=1}^{d} \sum_{j: \|\lambda_j\| \leq \epsilon} I_{nc}(\lambda_j) \{ |\lambda_j|^{\alpha} + \|g_i(\lambda_j)\| \}
\end{align*}

and use Lemma 1(i) and (23) [i.e. \( E(I_{nc}(\lambda_j)) \leq C \lambda_j^{-\alpha} \)] to show

\begin{align*}
\lim E(A_{n3}) &\leq C \lim n^{-1} \sum_{j=1}^{[s/2]} \left\{ \sum_{i=1}^{d} I_{(\|\lambda_j\| \leq \epsilon)} + I_{(\|\lambda_j\| \geq \epsilon)} \{ |\lambda_j|^{-\alpha+\beta} + |\lambda_j|^{-\alpha+\epsilon^{-1+\beta}} \} \right\} \\
&\leq C \left[ \epsilon + \int_{|\lambda| \leq \epsilon} |\lambda|^{-\alpha+\beta} + |\lambda|^{-\alpha+\epsilon^{-1+\beta}} d\lambda \right] \\
&\leq C \left[ \epsilon + \epsilon^{1-\alpha+\beta} \right].
\end{align*}

It follows that \( A_{n3} = o_p(1) \).

Let

\begin{align*}
A_{n4} &= \sup_{\|y\|=1} \left| 2\pi n \sum_{j=1-N}^{N} I_{nc}(\lambda_j) h_{0y}(\lambda_j) \right| \\
&\leq \frac{2\pi \|g\|_{\infty}}{n} \left( I_{nc}(0) + I_{nc}(\pi) \right).
\end{align*}

Applying Lemma 9(ii) from the Appendix, we find \( E(A_{n4}) = o(1) \) so that \( A_{n4} = o_p(1) \).

We now have

\begin{align*}
\sup_{\|y\|=1} \left| \frac{4\pi}{n} \sum_{j=1}^{N} I_{n}(\lambda_j) h_{0y}(\lambda_j) - \int_{y' > 0} fg' y' d\lambda \right| &\leq \sum_{i=1}^{4} A_{n4} = o_p(1)
\end{align*}

so that

\begin{align*}
P \left( \inf_{\|y\|=1} \frac{4\pi}{n} \sum_{j=1}^{N} I_{n}(\lambda_j) y' g(\lambda_j) I_{y' > 0} \geq \frac{3}{2} \right) \rightarrow 1 \quad (27)
\end{align*}

by the established Lemma 4(i). To finish the proof of Lemma 4(ii), it suffices to show that the event in (27) implies \( 0 \in ch^a \{ \pi g(\lambda_j) I_{n}(\lambda_j) \}_{j=1}^{N} \subset R^p \). Suppose not; then there exists some \( a \in R^p, \|a\| = 1 \) such that: if \( x \in ch \{ \pi g(\lambda_j) I_{n}(\lambda_j) \}_{j=1}^{N} \) (the convex hull) then \( x' a \geq 0' a = 0 \) by the separating/supporting hyperplane theorem [cf. Kelly and Weiss (1977), p. 142-149]. But when the event in (27) holds, there
exists \( r \in \{1, \ldots, N\} \) such that \( \pi f_n(\lambda_r)(-a)'g(\lambda_r) > 0 \), a contradiction. Hence, part (ii) of Lemma 4 follows. \( \square \)

The next lemma sets up an important distributional result for Riemann integrals based on the periodogram, involving central limit theorems for certain quadratic forms under both LRD and SRD. Write

\[
J_n = \frac{2\pi}{n} \sum_{j=1}^{N} g(\lambda_j)I_n(\lambda_j), \quad E_n = \frac{2\pi}{n} \sum_{j=1}^{N} g(\lambda_j)f(\lambda_j).
\]

**Lemma 5** Suppose Assumptions A.1-A.3 hold with respect to an even function \( g(\cdot) \). Then,

\[
\sqrt{n}(J_n - \int_0^\pi f g d\lambda) \overset{d}{\rightarrow} N(0, V), \quad V = \pi \int_\Pi f^2 g^2 d\lambda + \frac{\kappa_{4, e}}{\sigma_e^2} \left( \int_\Pi f g d\lambda \right)^2.
\]

If A.5 holds additionally,

\[
\sqrt{n}(J_n - E_n) \overset{d}{\rightarrow} N(0, V).
\]

**proof:** By Assumptions A.1-A.2, we have that

\[
\sqrt{n} \left( \int_0^\pi g(\lambda)I_{nc}(\lambda) d\lambda - E \int_0^\pi g(\lambda)I_{nc}(\lambda) d\lambda \right) \overset{d}{\rightarrow} N(0, V)
\]

from Theorem 1 of Giraitis and Surgailis (1990) and the Cramer-Wold device [cf. Billingsley (1986)].

To show (28), it suffices to establish

\[
\sqrt{n} \left| J_n - \int_0^\pi g(\lambda)I_{nc} d\lambda \right| = o_P(1),
\]

(30)

\[
\sqrt{n} \left| E \int_0^\pi g(\lambda)I_{nc} d\lambda - \int_0^\pi f(\lambda)g(\lambda) d\lambda \right| = o(1).
\]

(31)

When each functional component of \( g = (g_1, \ldots, g_p)' \) satisfies one of the Conditions 1-3 set by Assumption A.3, we have (29) follows easily from (28) and Lemma 10 in the Appendix (which shows \( \sqrt{n} |E_n - \int_0^\pi f g d\lambda| = o(1) \)).

WLOG we assume that \( p = 1 \), since we need only establish (30) and (31) componentwise. We begin showing (30) and (31) are valid under Condition 1 of Assumption A.3. By the Lipschitz property, it holds that

\[
\| g(\lambda) - c_n g(\lambda) \|_\infty = \sup_{\lambda \in \Pi} |g(\lambda) - c_n g(\lambda)| = o(n^{-1/2})
\]

(32)

[cf. Theorem 6.5.3, Edwards (1979)]. Let

\[
B_n = \frac{2\pi}{n} \sum_{j=-N}^{[n/2]} c_n g(\lambda_j)I_{nc}(\lambda_j) = \int_\Pi c_n g(\lambda)I_{nc}(\lambda) d\lambda.
\]

(33)
Then,
\[
\sqrt{n} E|2J_n - B_n| \leq \frac{2\pi}{\sqrt{n}} \left[ |c_n g(0)| E(I_{nc}(0)) + \|g\|_\infty E(I_{nc}(\tau)) \right] + \sqrt{n} \|c_n g - g\|_\infty \text{Var}(X_0)
\]
\[
\leq C n^{-1/2}(n^{\alpha-\beta} + 1) + o(1) = o(1)
\]
using (23), Lemma 9(ii)-(iii), \|c_n g\|_\infty \leq \|g\|_\infty, (32), and
\[
\frac{2\pi}{n} \sum_{j=-N}^{[n/2]} I_{nc}(\lambda_j) = \frac{1}{n} \sum_{i=1}^{n} (X_t - \mu)^2 = \int_{\Pi} I_{nc}(\lambda) d\lambda.
\]
(34)

Hence, \(\sqrt{n} |2J_n - B_n| = o_p(1)\). We also find
\[
\sqrt{n} E\left|B_n - \int_{\Pi} g(\lambda) I_{nc}(\lambda) d\lambda\right| \leq \sqrt{n} \|c_n g - g\|_\infty E\left(\int_{\Pi} I_{nc}(\lambda) d\lambda\right)
\]
\[
\leq \sqrt{n} \|c_n g - g\|_\infty \text{Var}(X_0) = o(1)
\]
by (32) and (34) so that \(\sqrt{n} |B_n - \int_{\Pi} g(\lambda) I_{nc} d\lambda| = o_p(1)\). Since \(g\) and \(I_{nc}\) are even functions, we now have (30) under Condition 1.

To show that (31) follows from Condition 1, note
\[
E(I_{nc}(\lambda)) = (2\pi n)^{-1} \text{cum}(d_{nc}(\lambda), d_{nc}(-\lambda)) = \int_{\Pi} K_n(\lambda - y)f(y) dy
\]
(35)
[cf. Lahiri (1999), Lemma 3.1], since E\(d_{nc}(\lambda)\) = 0, so that
\[
E\left(\int_{\Pi} g(\lambda) I_{nc}(\lambda) d\lambda\right) = \int_{\Pi} g(\lambda) \left(\int_{\Pi} K_n(\lambda - y)f(y) dy\right) d\lambda
\]
\[
= \int_{\Pi} f(y) \left(\int_{\Pi} K_n(y - \lambda)g(\lambda) d\lambda\right) dy
\]
\[
= \int_{\Pi} f(y) c_n g(y) dy
\]
(36)
by Fubini's theorem and the evenness of \(K_n(\cdot)\). Then
\[
\sqrt{n} \left| \int_{\Pi} c_n g(\lambda)f(\lambda) d\lambda - \int_{\Pi} g(\lambda)f(\lambda) d\lambda \right| \leq \sqrt{n} \|c_n g - g\|_\infty \int_{\Pi} f d\lambda = o(1),
\]
by (32). Hence, (31) holds under Condition 1 of Assumption A 3 \((c_n g, g, f\) are even functions).

We now consider showing (30) and (31) when \(g\) is real-valued and satisfies Condition 3 of Assumption A.3. First write
\[
\int_{\Pi} gI_{nc} d\lambda = \frac{1}{2\pi} \sum_{u=-n}^{n-1} r_n(u)\tilde{g}(u), \quad \bar{B}_n = \frac{2\pi}{n} \sum_{j=-N}^{n/2} g(\lambda_j) I_{nc}(\lambda_j) = \frac{1}{2\pi} \sum_{u=-n}^{n-1} r_n(u)\tilde{g}^*(u),
\]
where
\[
r_n(u) = \frac{1}{n} \sum_{t=1}^{n-|u|} (X_t - \mu)(X_{t+|u|} - \mu), \quad \tilde{g}(u) = \int_{\Pi} e^{-u\lambda} g(\lambda) d\lambda, \quad \tilde{g}^*(u) = \frac{2\pi}{n} \sum_{t=1}^{n} g(\lambda_t)e^{-u\lambda_t}.
\]
Since \( \alpha < 1/2 \) under Condition 3, we can pick \( 1/(1 - \alpha) < q = q(\alpha) < 2 \) so that 
\[
\sum_{u=-\infty}^{\infty} |r(u)|^q \leq C \sum_{u=1}^{\infty} u^{-q(1-\alpha)} + |r(0)|^q < \infty, \quad r(u) = \text{Cov}(X_0, X_u)
\]
using Lemma 9(i). By this summability result, the stationarity of \( \{X_t\} \), the fact that \( I_{nc} \) is mean-corrected, the bounded variation of \( g \) and the square integrability of the 4th order cumulant partial density \( f_4(\omega_1, \omega_2, \omega_3) = \kappa_4, b(\sum_{j=1}^{3} \omega_j) \prod_{j=1}^{3} b(-\omega_j) \) of \( \{X_t\} \) over \( \Omega^3 \) (because \( |b(\lambda)|^4 \) is finitely integrable over \( \Omega \) by \( \alpha < 1/2 \)), we have 
\[
\sqrt{n} \left( \int g I_{nc} d\lambda - E \int g I_{nc} d\lambda \right) - (\hat{B}_n - E \hat{B}_n) = o_p(1), \tag{37}
\]
applying Theorem 3.2 of Dahlhaus (1985). Note, that by the evenness of \( g \),
\[
\sqrt{n} E|2J_n - \hat{B}_n| \leq n^{-1/2} 2\pi \|g\|_{\infty} E|I_{nc}(0) + I_{nc}(\pi)| \leq C n^{-1/2} (n^{\alpha} + 1) = o(1)
\]
from (23), Lemma 9(ii) and \( \alpha < 1/2 \). Hence, \( \sqrt{n} |2J_n - \hat{B}_n| = o_p(1) \) so that (30) will follow from (37) if we establish 
\[
\sqrt{n} \left| E \int g I_{nc} d\lambda - E \hat{B}_n \right| = o(1), \tag{38}
\]
since \( g, I_{nc} \) are even. Because \( g \) is of bounded variation,
\[
|\hat{g}(u) - g^*(u)| \leq \frac{C}{n - |u|} \quad |u| < n, \ n \geq 1,
\]
by Lemma 3.1 of Dahlhaus (1985). So then 
\[
\sqrt{n} \left| E \int g I_{nc} d\lambda - E \hat{B}_n \right| = \sqrt{n} \left| \sum_{u=-(n-1)}^{n-1} r(u)(\hat{g}(u) - g^*(u)) \right| 
\leq C \sqrt{n} \sum_{u=0}^{n-1} |r(u)|(n - u)^{-1}
\leq C n^{-1/2} \left( |r(0)| + \sum_{u=1}^{[n/2]} u^{-1+\alpha} + \sum_{u=[n/2]+1}^{n-1} n^{\alpha}(n - u)^{-1} \right)
\leq C n^{-1/2} (1 + n^{\alpha} + n^\alpha \ln n) = o(1),
\]
using Lemma 9(i) and \( \alpha < 1/2 \). We have now shown that (38), and consequently (30), holds under Condition 3 of Assumption A.3. Also under Condition 3, we have (31) follows directly from Lemma 4 of Dahlhaus (1983) because \( g \) is of bounded variation and \( f^2 \) is finitely integrable on \( \Omega \).

The proofs that (30), (31) hold under Condition 2 of Assumption A.3 (along with A.1-A.2) are rather involved so we defer these proofs until the Appendix (Lemmas 12 and 11, respectively). The proof of Lemma 5 is now finished. \( \square \)
Lemma 6  With Assumption A.1, suppose \( g \) and \( w \) are real-valued, even Riemann integrable functions on \( \Pi \) such that \(|g(\lambda)|, |w(\lambda)| \leq C|\lambda|^2, 0 \leq \beta < 1, \alpha - \beta < 1/2. \) Then as \( n \to \infty, \)
\[
\frac{2\pi}{n} \sum_{j=1}^{N} g(\lambda_j) w(\lambda_j) I_n^2(\lambda_j) \overset{p}{\to} \int_{\Pi} g w f^2 \, d\lambda, \tag{39}\]
\[
\frac{2\pi}{n} \sum_{j=1}^{N} g(\lambda_j) w(\lambda_j) (I_n(\lambda_j) - f(\lambda_j))^2 \overset{p}{\to} \frac{1}{2} \int_{\Pi} g w f^2 \, d\lambda. \tag{40}\]

Proof: We will first show that (39) holds. Applying the Lebesgue Dominated Convergence Theorem,
\[
\frac{2\pi}{n} \sum_{j=1}^{N} g(\lambda_j) w(\lambda_j) f^2(\lambda_j) \to \int_{0}^{\Pi} g w f^2 \, d\lambda
\]
(writing the function \( \phi_n(\lambda) = \sum_{j=1}^{N} g(\lambda_j) w(\lambda_j) f^2(\lambda_j) I_{\lambda_{j-1} < \lambda \leq \lambda_j} \), we have \( \phi_n \to g w f^2 \) a.e. on \([0, \Pi] \), and \( |\phi_n(\lambda)| \leq C \sum_{j=1}^{N} \lambda_j^{2\beta - 2\alpha} I_{\lambda_{j-1} < \lambda \leq \lambda_j} \). Hence to prove (39), it suffices to establish for
\[
S_n = \frac{2\pi}{n} \sum_{j=1}^{N} g(\lambda_j) w(\lambda_j) I_n^2(\lambda_j),
\]
\[
\left| E(S_n) - \frac{4\pi}{n} \sum_{j=1}^{N} g(\lambda_j) w(\lambda_j) f^2(\lambda_j) \right| = o(1), \quad \text{Var}(S_n) = o(1). \tag{41}\]

By (23) and the product theorem for cumulants [Brillinger (1985), Theorem 2.3.2]
\[
(2\pi n)^2 E(I_n^2(\lambda_j)) = \text{cum}^2(d_{nc}(\lambda_j), d_{nc}(\lambda_j)) + 2 \text{cum}^2(d_{nc}(\lambda_j), d_{nc}(-\lambda_j)) + \text{cum}(d_{nc}(\lambda_j), d_{nc}(\lambda_j), d_{nc}(-\lambda_j), d_{nc}(-\lambda_j)) \tag{42}\]
using \( E(d_{nc}(\lambda)) = 0. \) Then we see
\[
\left| E(S_n) - \frac{2\pi}{n} \sum_{j=1}^{N} g(\lambda_j) w(\lambda_j) f^2(\lambda_j) \right| \leq s_{1n} + s_{2n} + s_{3n},
\]
where we define and bound the terms \( s_{in} \) in the following.

Using Lemma 1(i) and Lemma 2,
\[
s_{1n} = n^{-3} \sum_{j=1}^{N} |g(\lambda_j) w(\lambda_j)| \text{cum}^2(d_{nc}(\lambda_j), d_{nc}(\lambda_j)) \leq C n^{-3} \sum_{j=1}^{N} \lambda_j^{2\beta - 2\alpha} \left( \sum_{j=1}^{N} j^{-1} + \sum_{j=\lfloor n/4 \rfloor + 1}^{N} (n - 2j)^{-1} \right) = o(1).
\]

By Lemma 1(ii),
\[
|\text{cum}(d_{nc}(\lambda_j), d_{nc}(\lambda_j), d_{nc}(-\lambda_j), d_{nc}(-\lambda_j))| \leq C n^{(1/2 + \ln^3(n))^{-2\alpha}}
\]
so that
\[ s_{2n} = n^{-3} \sum_{j=1}^{N} |g(\lambda_j)w(\lambda_j)\operatorname{cum}(\lambda_j, dnc(\lambda_j), dnc(-\lambda_j), dnc(-\lambda_j))| \]
\[ \leq C n^{-1/2} \left( n^{-1} \sum_{j=1}^{N} \lambda_j^{2\alpha - 2} \right) = o(1). \]

Pick 0 < \rho < \pi and write using Lemma 2(i) and Lemma 3
\[ s_{3n} = \left| \frac{4\pi}{n} \sum_{j=1}^{N} g(\lambda_j)w(\lambda_j) \left( \frac{\operatorname{cum}(\lambda_j, dnc(\lambda_j), dnc(-\lambda_j))}{2\pi n} - f(\lambda_j) \right) \left( \frac{\operatorname{cum}(\lambda_j, dnc(-\lambda_j))}{2\pi n} + f(\lambda_j) \right) \right| \]
\[ \leq C n^{-1} \sum_{\lambda_1 \leq \lambda_j < \rho} \lambda_j^{2\alpha - 2} + C n^{-1} |R_n| \left( n^{-1} \sum_{\lambda_1 \leq \lambda_j \leq \lambda_n} \lambda_j^{2\alpha - 2} \right) \]
so that \( \lim s_{3n} \leq C \int_{\rho}^{\pi} \lambda^{2\alpha - 2} d\lambda \) implying \( s_{3n} = o(1) \) because \( \rho \) can be made arbitrarily small. We have now established the first claim in (41). The second claim in (41), \( \operatorname{Var}(S_n) = o(1) \), is more tedious to prove and we show this result in the Appendix (Lemma 13). We now have (39).

The convergence in (40) will follow from (39) if we establish additionally that for
\[ \tilde{S}_n = \frac{2\pi}{n} \sum_{j=1}^{N} g(\lambda_j)w(\lambda_j)f(\lambda_j)I_n(\lambda_j), \]
\[ \left| \mathbb{E}(\tilde{S}_n) - \frac{2\pi}{n} \sum_{j=1}^{N} g(\lambda_j)w(\lambda_j)f^2(\lambda_j) \right| = o(1), \quad \operatorname{Var}(\tilde{S}_n) = o(1) \quad (43) \]
so that \( \tilde{S}_n \to f^\pi gwf^2 d\lambda \). Using (23), (35), Lemma 2(i) and Lemma 3: for a fixed 0 < \rho < \pi
\[ \left| \mathbb{E}(\tilde{S}_n) - \frac{2\pi}{n} \sum_{j=1}^{N} g(\lambda_j)w(\lambda_j)f^2(\lambda_j) \right| \leq \frac{C}{n} \sum_{\lambda_1 \leq \lambda_j \leq \rho} \lambda_j^{2\alpha - 2} + \frac{C|R_n|}{n} \sum_{j=1}^{N} \lambda_j^{2\alpha - 2} \]
\[ = s_{1n} + s_{2n} \]
and \( \lim (\tilde{s}_{1n} + \tilde{s}_{2n}) \leq C \int_{\rho}^{\pi} \lambda^{2\alpha - 2} d\lambda \), which can be made arbitrarily small. Hence, we have shown the first statement in (43).

For the second statement in (43),
\[ \operatorname{Var}(\tilde{S}_n) = (2\pi n^{-1}) \sum_{j=1}^{N} g(\lambda_j)w(\lambda_j)f(\lambda_j)g(\lambda_j)w(\lambda_j)f(\lambda_j)\operatorname{cum}(I_n(\lambda_j), I_n(\lambda_j)). \]
Using (42),
\[ (2\pi n^2) \left| \operatorname{cum}(I_n(\lambda_j), I_n(\lambda_j)) \right| \leq \operatorname{cum}^2(dnc(\lambda_j), dnc(-\lambda_j)) + \operatorname{cum}^2(dnc(-\lambda_j), dnc(-\lambda_j)) \]
\[ + \left| \operatorname{cum}(dnc(\lambda_j), dnc(-\lambda_j), dnc(\lambda_j), dnc(-\lambda_j)) \right|. \quad (44) \]
When \(i = j\), (44) is bounded by \(C n^{2} \lambda_j^{-2\alpha}\) by Lemma 1 so that

\[
(2\pi n^{-1})^{2} \sum_{j=1}^{N} g^{2}(\lambda_{j})w^{2}(\lambda_{j})f^{2}(\lambda_{j})\sum_{j=1}^{n}\lambda_{j}^{\delta-2\alpha}
\leq C n^{-1+\max\{0,2\alpha-2\delta\}} \left( n^{-1} \sum_{j=1}^{N} \lambda_{j}^{2\delta-2\alpha} \right) = o(1).
\tag{45}
\]

When \(i \neq j\),

\[
\left| \sum_{1 \leq i < j \leq N} (\lambda_{i}\lambda_{j})^{2\delta-\alpha} \left[ \sum_{1 \leq i < j \leq N} (\lambda_{i}\lambda_{j})^{2\delta-\alpha} \left( \sum_{1 \leq i < j \leq N} (d_{nc}(\lambda_{i}), d_{nc}(\lambda_{j}), d_{nc}(-\lambda_{i})) \right) \right] \right| \leq C(n^{3/2} + n \ln^{3}(n))(\lambda_{i}\lambda_{j})^{-\alpha},
\]

\[
\leq C n^{-1/2} \left( n^{-1} \sum_{j=1}^{N} \lambda_{j}^{2\delta-2\alpha} \right)^{2} = o(1).
\tag{46}
\]

Pick \(0 < \rho < \pi/2\) and partition the sum

\[
\sum_{1 \leq i < j \leq N} (\lambda_{i}\lambda_{j})^{2\delta-\alpha} \left[ \sum_{1 \leq i < j \leq N} (\lambda_{i}\lambda_{j})^{2\delta-\alpha} \left( \sum_{1 \leq i < j \leq N} (d_{nc}(\lambda_{i}), d_{nc}(\lambda_{j})) \right) \right] = s_{4n} + s_{5n}
\]

into two sums where \(s_{4n}\) represents a sum over \(1 \leq i < j \leq N\) where \(\lambda_{i}\) or \(\lambda_{j} \geq \rho\) and \(s_{5n}\) as a sum over the remaining terms with \(1 \leq i < j \leq N\), \(\lambda_{i} < \rho\). By Lemma 3, (24), Lemma 2(i):

\[
s_{4n} \leq C(\rho)n^{-4} \sum_{1 \leq i < j \leq N} (\lambda_{i})^{2\delta-\alpha} \left[ R_{\rho}\lambda_{i}^{2} + L_{\rho}(\lambda_{i} + \lambda_{j}) + L_{\rho}(\lambda_{i} - \lambda_{j}) \right]
\leq C(\rho)n^{-4} \left[ \sum_{i=1}^{N} (\lambda_{i})^{2\delta-\alpha} \right] + \sum_{1 \leq i < j \leq N} n^{2}(i+j)^{-2} + \sum_{1 \leq i < j \leq N} n^{2}(n-j-i)^{-2}
\leq C(\rho)n^{-1} \left[ \sum_{j=1}^{N} (\lambda_{j})^{2\delta-\alpha} \right] = o(1).
\]

While by Lemma 1(i), Lemma 2(iv) (for some \(\max\{\alpha, 1/2\} < \delta < 1\))

\[
s_{5n} \leq C n^{-2} \sum_{1 \leq i < j \leq N} (\lambda_{i})^{2\delta-2\alpha} \lambda_{i}^{-2\alpha} \left[ (\lambda_{j}^{-2} + n^{2}(j+i)^{-2} + n^{2}(j-i)^{-2}) \right]
\leq C n^{-2} \sum_{1 \leq i < j \leq N} (\lambda_{i})^{2\delta-2\alpha} (j^{-2} + I_{j\geq\delta/2}((i-1))^{-2} + I_{j\geq\delta/2}((i-1))^{-2} + I_{j\geq\delta/2}((i-1))^{-2})(j-i)^{-2\alpha}
\leq C \left( \sum_{1 \leq i < \rho} \lambda_{i}^{2\delta-2\alpha} \right)^{2}.
\]
Thus, for a $C$ independent of $p$, $\lim (s_{4n} + s_{5n}) \leq C \int_0^1 \lambda^{2\beta - 2\alpha} d\lambda$ so that $\lim (s_{4n} + s_{5n}) = 0$. By this, (44), (45) and (46), we have $\text{Var}(\tilde{S}_n) = o(1)$ in (43). The proof of Lemma 6 is now finished. □

Lemma 7 Suppose Assumption A.1 holds and $0 \leq \beta < 1$, $\alpha - \beta < 1/2$. Let $b = \max\{1/3, (1/2)(1/2 + \alpha - \beta)\}$. Then,

$$\max_{1 \leq i \leq N} f_n(\lambda_i)\lambda_i^\beta = o_p(n^b), \quad \max_{1 \leq i \leq N} f(\lambda_i)\lambda_i^\beta = o(n^b).$$

Proof: Note that $f(\lambda_i) \leq C\lambda_i^{-\alpha}$ implies $f(\lambda_i)\lambda_i^\beta \leq Cn^{\max(0, \alpha - \beta)} = o(n^b)$, showing $\max_{1 \leq i \leq N} f(\lambda_i)\lambda_i^\beta = o(n^b)$.

We next show, for each $\epsilon > 0$,

$$P\left(\max_{1 \leq i \leq N} f_n(\lambda_i)\lambda_i^\beta > \epsilon n^b\right) \leq \frac{n^{1/4}}{\epsilon n^b} \left(\frac{1}{n} \sum_{i=1}^N \lambda_i^{4\beta - 2\alpha} \text{E}(f_n^2(\lambda_i))\right)^{1/4} = o(1).$$

(47)

Then, by (23), (42), (44) and Lemma 1,

$$E(f_n^2(\lambda_i)) = \text{cum}(f_n^2(\lambda_i), f_n^2(\lambda_i)) + [E(f_n^2(\lambda_i))]^2 \leq C\lambda_i^{-4\alpha}.$$

Hence, (47) will follow by showing

$$k_n = n^{-b+1/4} \left(\frac{1}{n} \sum_{i=1}^N \lambda_i^{4\beta - 4\alpha}\right)^{1/4} = o(1).$$

Note that $n^{-1} \sum_{i=1}^N (\lambda_i)^{4\beta - 4\alpha}$ is: $O(1)$ if $4\beta - 4\alpha > -1$; $O(\ln(n))$ if $4\beta - 4\alpha = -1$; and $O(n^{-1-4\beta+4\alpha})$ if $4\beta - 4\alpha < -1$. In the first two cases, $-b + 1/4 < 0$ implies $k_n = o(1)$. In the last case,

$$n^{-b+1/4} (n^{-1-4\beta+4\alpha})^{1/4} = n^{-b+4\alpha-\beta} = o(1),$$

since $b > \alpha - \beta$ so that $k_n = o(1)$. We have now established (47) and Lemma 7. □

Proof of Theorem 1. We will begin with Theorem 1(i) and give a detailed argument. The remaining elements (ii) and (iii) of Theorem 1 will then follow with some minor modifications.

By Lemma 4, $0 \in \text{ch}\{xg_{\theta_0}(\lambda_i)I_n(\lambda_i)\}$ with probability approaching 1 as $n \to \infty$ so that a positive $R_n(\theta_0)$ exists in probability. In view of (14), we can express the extrema $R_n(\theta_0) = \Pi_{i=1}^N (1 + \gamma_i)^{-1}$ with $\gamma_i = t_{\theta_0} g_{\theta_0}(\lambda_i)I_n(\lambda_i), \quad |\gamma_i| < 1,$ where $t_{\theta_0} \in \mathbb{R}^p$ satisfies $Q_n(\theta_0, t_{\theta_0}) = 0$ for

$$Q_n(\theta_0, \epsilon) = \frac{2\pi}{n} \sum_{i=1}^N \frac{g_{\theta_0}(\lambda_i)I_n(\lambda_i)}{1 + \epsilon' g_{\theta_0}(\lambda_i)I_n(\lambda_i)}.$$
As in Owen (1988, 1990), we next show

$$||t_{\theta_0}|| = O_p(n^{-1/2}). \quad (48)$$

Set $||t_{\theta_0}|| = b_0u_0$, $||u_0|| = 1$. Let

$$W_{n\theta_0} = \frac{2\pi}{n} \sum_{i=1}^{N} g_{\theta_0}(\lambda_i)g'_{\theta_0}(\lambda_i)I_n(\lambda_i), \quad I_{n\theta_0} = \frac{2\pi}{n} \sum_{i=1}^{N} g_{\theta_0}(\lambda_i)I_n(\lambda_i).$$

Then,

$$0 = ||Q_n(\theta_0, t_{\theta_0})|| \geq ||u_0Q_n(\theta_0, t_{\theta_0})|| = \frac{2\pi}{n} \left| u_0 \left( \sum_{i=1}^{N} g_{\theta_0}(\lambda_i)I_n(\lambda_i) - b_0 \sum_{i=1}^{N} \frac{g_{\theta_0}(\lambda_i)I_n(\lambda_i)u_0g_{\theta_0}(\lambda_i)I_n(\lambda_i)}{1 + b_0u_0g_{\theta_0}(\lambda_i)I_n(\lambda_i)} \right) \right| \geq \frac{b_0u_0W_{n\theta_0}u_0}{1 + b_0Y_n} - \sum_{j=1}^{p} |e_jI_{n\theta_0}|,$$

where $e_1, \ldots, e_p$ denote the standard basis vectors for $\mathbb{R}^p$ and

$$Y_n = \max_{1 \leq i \leq N} ||g_{\theta_0}(\lambda_i)||I_n(\lambda_i) = o_p(n^{1/2}), \quad (49)$$

by Assumption A.2 and Lemma 7. By Lemma 5 and $p = 0$, we have

$$|e_jI_{n\theta_0}| = O_p(n^{-1/2}). \quad (50)$$

We apply Lemma 6 to find

$$||W_{n\theta_0} - W_{\theta_0}|| = o_p(1) \quad (51)$$

so that $W_{n\theta_0}$ is nonsingular in probability and $u_0W_{n\theta_0}u_0 \geq \sigma_\omega + o_p(1)$ where $\sigma_\omega > 0$ is the smallest eigenvalue of $W_{\theta_0}$. Hence, $(1 + b_0Y_n)^{-1}b_0 = O_p(n^{-1/2})$. It then follows from (49) that $b_0 = ||t_{\theta_0}|| = O_p(n^{-1/2})$, establishing (48).

We note that by (48) and (49),

$$\max_{1 \leq i \leq N} |\gamma_i| \leq ||t_{\theta_0}||Y_n = O_p(n^{-1/2})o_p(n^{1/2}) = o_p(1). \quad (52)$$

With a little algebra, we write

$$0 = Q_n(\theta_0, t_{\theta_0}) = I_{n\theta_0} - W_{n\theta_0}t_{\theta_0} + \frac{2\pi}{n} \sum_{i=1}^{N} \frac{g_{\theta_0}(\lambda_i)I_n(\lambda_i)\gamma_i^2}{1 + \gamma_i},$$

and can solve for $t_{\theta_0} = W_{n\theta_0}^{-1}I_{n\theta_0} + \phi_n$. By Lemma 6, (48), (49), and (52),

$$||\phi_n|| \leq Y_n||t_{\theta_0}||W_{n\theta_0}^{-1} \left\{ \frac{2\pi}{n} \sum_{i=1}^{N} ||g_{\theta_0}(\lambda_i)||^2 I_n(\lambda_i) \right\} \max_{1 \leq i \leq N} (1 + \gamma_i)^{-1} \leq o_p(n^{1/2})O_p(n^{-1})O_p(1)O_p(1)O_p(1) = o_p(n^{-1/2}).$$
When \( ||t_0||Y_n < 1 \) in (52), we apply a Taylor's expansion
\[
\ln(1 + \gamma_i) = \gamma_i - \frac{\gamma_i^2}{2} + \Delta_i, \quad |\Delta_i| \leq ||t_0||^3 Y_n ||g_{\theta_0}(\lambda_i)||^2 t_n^2(\lambda_i)(1 - ||t_0||Y_n)^{-3},
\]
for each \( 1 \leq i \leq N \). Then
\[
\ell_n(\theta_0) = 4 \sum_{i=1}^{N} \ln(1 + \gamma_i) = 2 \left[ 2 \sum_{i=1}^{N} \gamma_i - \sum_{i=1}^{N} \gamma_i^2 \right] + 4 \sum_{i=1}^{N} \Delta_i \tag{53}
\]
\[
= n l_{n,\theta_0}^\prime (\pi W_{n,\theta_0})^{-1} l_{n,\theta_0} - n \phi_n^\prime (\pi W_{n,\theta_0})^{-1} \phi_n.
\]
By Lemmas 5 and 6, \( n l_{n,\theta_0}^\prime (\pi W_{n,\theta_0})^{-1} l_{n,\theta_0} \xrightarrow{d} \chi^2_1 \). We also have
\[
n \phi_n^\prime (\pi W_{n,\theta_0})^{-1} \phi_n = n o_p(1) o_p(n^{-1}) = o_p(1)
\]
and, in probability,
\[
\sum_{i=1}^{N} |\Delta_i| \leq ||t_0||^3 Y_n (1 - ||t_0||Y_n)^{-3} n \left( \frac{1}{n} \sum_{i=1}^{N} ||g_{\theta_0}(\lambda_i)||^2 t_n^2(\lambda_i) \right) \]
\[
= n o_p(n^{-3/2}) o_p(n^{1/2}) o_p(1) = o_p(1)
\]
by (48), (49), (52), and Lemma 6. Applying Slutsky's Theorem, we have Theorem 1(i).

We now establish Theorem 1(ii) and (iii). We have here that \( f_{\theta_0} = f \). By assumption, \( 0 \in \text{ch}^* (\pi g_{\theta_0}(\lambda_i)(I_n(\lambda_i) - f(\lambda_i))) \) \( \frac{N}{n} \) in probability ensuring that a positive \( R_{n,\pi,\theta_0} \) exists in probability. We then repeat the exact same arguments for proving Theorem 1(i) replacing each occurrence of \( I_n(\lambda_i) \) with \( I_n(\lambda_i) - f(\lambda_i) \) instead; we denote the resulting quantities with a tilde:
\[
\tilde{\gamma}_i, \tilde{\Delta}_i, \tilde{\phi}_n, \text{etc.}
\]
All the previous points made follow except for two, which are related and straightforward to remedy:

- By Lemma 6, \( ||2\tilde{W}_{n,\theta_0} - W_{\theta_0}|| = o_p(1) \). Note by Lemma 5, \( \pi W_{\theta_0} \) is the limiting covariance matrix of both \( I_{n,\theta_0} \) and \( \tilde{I}_{n,\theta_0} \) under Theorem 1(ii)-(iii) conditions.

- In (53), we must write
\[
\ell_{n,\pi}(\theta_0) = 2 \sum_{i=1}^{N} \ln(1 + \gamma_i) = n \tilde{I}_{n,\theta_0}^\prime (2\pi \tilde{W}_{n,\theta_0})^{-1} \tilde{I}_{n,\theta_0} - n \tilde{\phi}_n^\prime (2\pi \tilde{W}_{n,\theta_0})^{-1} \tilde{\phi}_n + 2 \sum_{i=1}^{N} \tilde{\Delta}_i,
\]
where \( n\tilde{I}_{n\theta}(2\pi\tilde{W}_{n\theta})^{-1}I_{n\theta} \xrightarrow{d} \chi_2^2 \) by Lemma 5 and

\[
n\hat{\phi}_n(2\pi\tilde{W}_{n\theta})^{-1}\hat{\phi}_n = o_p(1), \quad \sum_{i=1}^N \hat{\delta}_i = o_p(1).
\]

The proof of Theorem 1 is now complete. \( \Box \)

We require some additional notation before proceeding with the proof that the spectral MELEs \( \hat{\theta}_n \) and \( \hat{\theta}_{n,\pi} \) exist in probability (Lemma 8). Define the functions on \( \Theta \times \mathbb{R}^r \):

\[
Q_{1n}(\theta, t) = \frac{2\pi}{n} \sum_{i=1}^N \frac{g_\theta(\lambda_i)I_n(\lambda_i)}{1 + t'g_\theta(\lambda_i)I_n(\lambda_i)},
\]

\[
Q_{2n}(\theta, t) = \frac{2\pi}{n} \sum_{i=1}^N \frac{I_n(\lambda_i)\left(\theta g_\theta(\lambda_i)/\theta\right)'t}{1 + t'g_\theta(\lambda_i)I_n(\lambda_i)}.
\]

We will also use the functions above to prove Theorems 2 and 3.

**Lemma 8** Under the assumptions of Theorem 2:

(i) As \( n \to \infty \), the probability that \( \ell_n(\theta) \) attains a maximum \( \hat{\theta}_n \) in the ball \( ||\theta - \theta_0|| < n^{-m} \) converges to 1, for \( m = \max\{1/3, 1/4 + (\alpha - \beta)/2, (1 + \alpha + \delta)/4\} < 1/2 \).

(ii) If Theorem 2(ii) assumptions are satisfied, then (i) above holds with respect to \( \ell_{n,\pi}(\theta) \) and \( \hat{\theta}_{n,\pi} \).

**proof:** Let

\[
B_n = \{ \theta \in \Theta : ||\theta - \theta_0|| \leq n^{-m} \}, \quad \partial B_n = \{ \theta \in \Theta : ||\theta - \theta_0|| = n^{-m} \}.
\]

We give first a detailed proof of Lemma 8(i) and after discuss modifications necessary to show Lemma 8(ii).

Many aspects of the following argument are adapted from Qin and Lawless (1994). We show first that, in probability, \( R_n(\theta) \) exists finitely for \( \theta \in B_n \) and can be written as in (14). First,

\[
\sup_{\theta \in B_n} \frac{4\pi}{n} \sum_{i=1}^N I_n(\lambda_i) \frac{1}{2} \left[ |g'_{\theta}(\lambda_i)t| - |g'_{\theta}(\lambda_i)|_t \right] \leq \sup_{\theta \in B_n} \frac{4\pi}{n} \sum_{i=1}^N ||g_{\theta}(\lambda_i) - g_{\theta}(\lambda_i)||_t \leq C n^{-1-m} \sum_{i=1}^N \lambda_i^{-\delta} I_n(\lambda_i) = O_p(n^{-m}),
\]
by Lemma 14(i). By this, Lemma 4 and (27):

\[ P\left( \inf_{\theta \in \mathcal{B}_n} \frac{4\pi}{n} \sum_{i=1}^{N} I_n(\lambda_i) g_{\theta}(\lambda_i) t_{2i}(\lambda_i) |t_{2i}(\lambda_i)| > 0 \right) \geq \frac{e}{2} \rightarrow 1. \]

Consequently, \( P(0 \in \text{ch} \{ \pi g_{\theta}(\lambda_i) I_n(\lambda_i) \}_{i=1}^{N}, \theta \in \mathcal{B}_n) \rightarrow 1. \) As in Owen (1990), we can then write \( R_n(\theta) = \prod_{i=1}^{N} [1 + \gamma_{i\theta}]^{-1} \) where \( \gamma_{i\theta} = t_{2i} g_{\theta}(\lambda_i) I_n(\lambda_i) \) and \( t_{2i} \) is determined by

\[ Q_{1n}(\theta, t_{2i}) = 0 \quad (54) \]

for each \( \theta \in \mathcal{B}_n \) (with probability approaching 1).

Define

\[ I_{n\theta} = \frac{2\pi}{n} \sum_{i=1}^{N} g_{\theta}(\lambda_i) I_n(\lambda_i), \quad W_{n\theta} = \frac{2\pi}{n} \sum_{i=1}^{N} g_{\theta}(\lambda_i) g_{\theta}(\lambda_i) I_n^2(\lambda_i). \]

We need to establish a few properties of \( I_{n\theta} \) and \( W_{n\theta} \) before proceeding further. By a 1st-order Taylor expansion of \( g_{\theta} \) around \( \theta_0 \),

\[ \sup_{\theta \in \mathcal{B}_n} \| I_{n\theta} \| \leq \| I_{n\theta_0} \| + C n^{-m} \left( \frac{1}{n} \sum_{i=1}^{N} \lambda_i t_{n}\right) = O_p(n^{-m}), \quad (55) \]

where \( \| I_{n\theta_0} \| = O_p(n^{-1/2}) \) by Lemma 5. We also have by Assumption A.2 and the bound on \( \partial g_{\theta}/\partial \theta \),

\[ \sup_{\theta \in \mathcal{B}_n} \| W_{n\theta} - W_{n\theta_0} \| \leq C n^{-m-1} \sum_{i=1}^{N} \left( \| g_{\theta}(\lambda_i) \| + \| g_{\theta}(\lambda_i) \| \right) \| g_{\theta}(\lambda_i) - g_{\theta_0}(\lambda_i) \| I_{n\theta_0}^2(\lambda_i) \]

\[ \leq C n^{-m-1} \sum_{i=1}^{N} \lambda_i t_{n}^2 + C n^{-2m-1} \sum_{i=1}^{N} \lambda_i^{-2d} I_{n\theta_0}^2(\lambda_i) \]

\[ = O_p(n^{-(1-2m)}) = o_p(1), \quad (56) \]

by Lemma 14(ii). It follows from (51) that

\[ \sup_{\theta \in \mathcal{B}_n} \| W_{n\theta} - W_{\theta_0} \| = o_p(1). \quad (57) \]

This also implies

\[ P\left( \sup_{\theta \in \mathcal{B}_n} \| W_{n\theta} - W_{\theta_0} \| \leq \frac{1}{2\| W_{\theta_0} \|^2} \right) \leq P(W_{n\theta} \text{ nonsingular, } \theta \in \mathcal{B}_n) \rightarrow 1. \quad (58) \]

Let \( t_{2i} = b_{\theta u_{\theta}}, \| u_{\theta} \| = 1 \). We now wish to show

\[ \sup_{\theta \in \mathcal{B}_n} \| t_{2i} \| = O_p(n^{-m}). \quad (59) \]

In view of (54), we can follow Owen (1990) and write

\[ 0 = \| Q_{1n}(\theta, t_{2i}) \| \geq |u_{\theta}^T Q_{1n}(\theta, t_{2i})| \]

\[ \geq \frac{b_{\theta u_i}}{1 + b_{\theta Z_n}} \sum_{j=1}^{p} |t_{2j}^i I_{n\theta_0}| - C \| \theta - \theta_0 \|^2 \sum_{i=1}^{N} \lambda_i^{-2d} I_{n\theta_0}(\lambda_i) \].
where \( \{e_j\}_{j=1}^p \) is the standard basis of \( IR^p \) and

\[
Z_n = \sup_{\theta \in B_n} \max_{1 \leq i \leq N} \|g_\theta(\lambda_i)\| I_n(\lambda_i).
\]

The two negative terms above are \( O_p(n^{-m}) \) by Lemma 5 and Lemma 14(i) so that, denoting \( \sigma_w \) as the smallest eigenvalue of \( W_\theta \), it follows from (57) that \( u_\theta^T W_\theta u_\theta \geq \sigma_w + o_p(1) \), where the error is uniform in \( \theta \in B_n \). We have then \( \sup_{\theta \in B_n} |b_\theta(1 + b_\theta Z_n)^{-1}| = O_p(n^{-m}) \). Note that by Lemma 7 and Lemma 14(ii)

\[
Z_n \leq \max_{1 \leq i \leq N} (\|g_\theta(\lambda_i)\| + C \|\theta - \theta_0\| |\lambda_i^{-2} I_n(\lambda_i)^1/2 = o_p(n^m)
\]

so that (59) now follows.

We can obtain an expression for \( t_\theta \) by rewriting (54),

\[
0 = Q_n(\theta, t_\theta) = I_n + W_n t_\theta + \frac{2\pi}{n} \sum_{i=1}^N g_\theta(\lambda_i) I_n(\lambda_i) \frac{\gamma_{i\theta}^2}{1 + \gamma_{i\theta}},
\]

and solving for \( t_\theta = W_n^{-1} I_n + \phi_n\theta \), where

\[
\|\phi_n\theta\| \leq C Z_n \|t_\theta\|^2 \|W_n^{-1}\| \left( \frac{2\pi}{n} \sum_{i=1}^N \|g_\theta(\lambda_i)\|^2 I_n(\lambda_i) \right) \max_{1 \leq i \leq N} (1 + \gamma_{i\theta})^{-1} = o_p(n^{-m}),
\]

uniformly in \( \theta \in B_n \) by (57), (59), (60), Lemma 6, and

\[
\sup_{\theta \in B_n} \max_{1 \leq i \leq N} \|\gamma_{i\theta}\| \leq \sup_{\theta \in B_n} \|t_\theta\| Z_n = o_p(1).
\]

As in (53), we can now write for each \( \theta \in B_n \),

\[
\ell_n(\theta) = n I_n(\pi W_n)^{-1} I_n - n \phi_n'(\pi W_n)^{-1} \phi_n + 4 \sum_{i=1}^N \Delta_{i\theta}
\]

\[
\sup_{\theta \in B_n} \|\phi_n\theta(\pi W_n)^{-1} \phi_n\theta = o_p(n^{1-2m})
\]

and, when \( \sup_{\theta \in B_n} \|t_\theta\| Z_n < 1 \) in (61),

\[
4 \sum_{i=1}^N \|\Delta_{i\theta}\| \leq C n Z_n \|t_\theta\|^2 \left( \frac{1}{n} \sum_{i=1}^N \|g_\theta(\lambda_i)\|^2 I_n(\lambda_i) \right) (1 - \sup_{\theta \in B_n} \|t_\theta\| Z_n)^{-3} = o_p(n^{1-2m})
\]

uniformly in \( \theta \in B_n \). Hence, for \( \theta \in B_n \),

\[
\ell_n(\theta) = n I_n(\pi W_n)^{-1} I_n + o_p(n^{1-2m})
\]

(62)
We may expand
\[ I_{n\theta} = I_{n\theta_0} + D_{n\theta_0}(\theta - \theta_0) + E_{n\theta} \]
\[ = I_{n\theta_0} + D_{n\theta_0}(\theta - \theta_0) + o_p(n^{-m}) \]  
(53)
uniformly in \( \theta \in B_n \) by (55) and (57).

We may expand
\[ N_{\theta \theta} = N_{\theta \theta_0} + E_{\theta \theta} \]
\[ = N_{\theta \theta_0} + E_{\theta \theta} + o_p(n^{-m}) \]
(63)
uniformly in \( \theta \in B_n \), using \( \|E_{\theta \theta}\| \leq C \|\theta - \theta_0\|^2 n^{-1} \sum_{i=1}^{N} \lambda_i^{-\delta} I_n(\lambda_i) \). In addition,
\[ D_{n\theta_0} \xrightarrow{p} \frac{1}{2} \int \frac{\partial g_{\theta_0}(\lambda)}{\partial \theta} f(\lambda) d\lambda \equiv \frac{D_{\theta_0}}{2} \]  
(64)
since \( 2\pi/n \sum_{i=1}^{N} f(\lambda_i) \partial g_{\theta_0}(\lambda_i)/\partial \theta \rightarrow D_{\theta_0}/2 \) by the Lebesgue Dominated Convergence Theorem \((\partial g_{\theta_0}/\partial \theta \text{ is Riemann integrable and bounded by } C|\lambda|^{-\delta}) \) and a straightforward modification of the arguments for (43) [using \( \delta + \alpha < 1 \) instead of \( 2\alpha - 2\beta < 1 \) there] yields
\[ \|E(D_{n\theta_0}) - \int_0^\pi \frac{\partial g_{\theta_0}(\lambda)}{\partial \theta} f(\lambda) d\lambda\| = o(1), \quad \text{Var}(D_{n\theta_0}) = o(1). \]

From (55), (64), and \( \sup_{\theta \in B_n} |E_{\theta \theta}| = o_p(n^{-m}) \), we find
\[ \sup_{\theta \in B_n} n \left\| I_{n\theta}^* (\pi W_{\theta_0})^{-1} I_{n\theta} - \left( I_{n\theta_0} + \frac{D_{\theta_0}}{2} (\theta - \theta_0) \right)^* \left( \pi W_{\theta_0} \right)^{-1} \left( I_{n\theta_0} + \frac{D_{\theta_0}}{2} (\theta - \theta_0) \right) \right\| \]
\[ = o_p(n^{1-2m}). \]  
(65)

Because \( I_{n\theta_0} = O_p(n^{-1/2}) \) from Lemma 5, it holds now in probability (or with arbitrarily large probability as \( n \rightarrow \infty \)) that uniformly for \( \theta \in \partial B_n \),
\[ \ell_n(\theta) = n \left( O_p(n^{-1/2}) + n^{-m} \frac{D_{\theta_0} u_0}{2} \right) \left( \pi W_{\theta_0} \right)^{-1} \left( O_p(n^{-1/2}) + n^{-m} \frac{D_{\theta_0} u_0}{2} \right) + o_p(n^{1-2m}) \]
\[ \geq (\sigma_* - \epsilon) n^{1-2m}, \]
where \( \sigma_* - \epsilon > 0 \) and \( \sigma_* \) is the smallest eigenvalue of \( D'_{\theta_0} (\pi W_{\theta_0})^{-1} D_{\theta_0} \).

From Theorem 1, \( \ell_n(\theta_0) = O_p(1) \). Then (54) and (58) imply that in fact \( t_\theta \) is a continuously differentiable function of \( \theta \) on \( B_n \) and so \( \ell_n(\theta) \) is as well (\( R_n(\theta) \) admits the representation from (14)).

Hence, as in Qin and Lawless 1994, \( \ell_n(\theta) \) attains a minimum \( \hat{\theta}_n \) (or equivalently \( L_n(\theta) \) finds a maximum) in \( B_n \setminus \partial B_n \) with probability approaching 1 and at \( \hat{\theta}_n \):
\[ 0 = Q_{1n}(\hat{\theta}_n, t_{\hat{\theta}_n}), \]
\[ 0 = \frac{\partial \ell_n(\theta)}{\partial \theta} \bigg|_{\theta = \hat{\theta}_n} = \frac{2n}{\pi} Q_{2n}(\hat{\theta}_n, t_{\hat{\theta}_n}) \]  
(66)
using (54).
We sketch the proof of Lemma 8(ii), where we assume \( f = f_\theta \) and \( p = 0 \in \mathbb{R}^r \). (If \( \kappa_{\epsilon,\epsilon} = 0 \), the same proof establishes a version of Lemma 8(ii) that is valid for any \( p \in \mathbb{R}^r \).) We can proceed in the same manner as above, beginning from (54), with \( R_{n,\theta}(\theta) = \prod_{i=1}^N [1 + \tilde{\gamma}_w]^{-1} \) where now \( \tilde{\gamma}_w = \epsilon^{-1}_w g_\theta(\lambda_i)(I_n(\lambda_i) - f_\theta(\lambda_i)) \) for \( \theta \) satisfying \( \tilde{\gamma}_1(\theta) = 0 \). Let

\[
\tilde{I}_{n\theta} = \frac{2\pi n}{n} \sum_{i=1}^N g_\theta(\lambda_i)(I_n(\lambda_i) - f_\theta(\lambda_i)), \quad \tilde{W}_{n\theta} = \frac{2\pi n}{n} \sum_{i=1}^N g_\theta(\lambda_i)g_\theta(\lambda_i)(I_n(\lambda_i) - f_\theta(\lambda_i))^2.
\]

Note that by Lemma 14(ii),

\[
\sup_{\theta \in B_n} \left\| \frac{1}{n} \sum_{i=1}^N (g_\theta(\lambda_i)f_\theta(\lambda_i) - g_\theta(\lambda_i)f_\theta(\lambda_i)) \right\| \leq C n^{-m-1} \sum_{i=1}^N \lambda_i^{-\alpha}(\lambda_i^2 + \lambda_i^{-\delta}) = O(n^{-m}) \quad (67)
\]

and also \( ||\tilde{I}_{n\theta}|| = O_p(n^{-1/2}) \) from Lemma 5 under Lemma 8(ii) conditions so that by (55),

\[
\sup_{\theta \in B_n} ||\tilde{I}_{n\theta}|| = O_p(n^{-m}).
\]

Also, by Lemma 14,

\[
\sup_{\theta \in B_n} \left\| \frac{1}{n} \sum_{i=1}^N \left( g_\theta g_\theta f_\theta(\lambda_i) - g_\theta g_\theta f_\theta(\lambda_i) \right) I_n(\lambda_i) + \frac{1}{n} \sum_{i=1}^N \left( g_\theta g_\theta f_\theta^2(\lambda_i) - g_\theta g_\theta f_\theta^2(\lambda_i) \right) \right\| \leq C n^{-m-1} \sum_{i=1}^N \lambda_i^{-\alpha}(\lambda_i^{2\delta} + \lambda_i^{2\delta} n^{-m} \lambda_i^{-2\delta}) = O_p(n^{-1-2m}) = o_p(1),
\]

so that we have

\[
\sup_{\theta \in B_n} ||\tilde{W}_{n\theta} - \tilde{W}_{n\theta}|| = o_p(1)
\]

using (56). However, by Lemma 6, in this case: \( \|2\tilde{W}_{n\theta} - \tilde{W}_{n\theta}\| = o_p(1) \), where \( \pi W_{n\theta} \) is the limiting covariance matrix of \( \tilde{I}_{n\theta} \) by Lemma 5. Then, (58) holds replacing \( W_{n\theta} \) with \( 2\tilde{W}_{n\theta} \) in the probability statement.

For the \( t_\theta \) defining \( R_{n,\theta}(\theta) \), we find (59) still holds after a straightforward modification of the previous argument to account for \( I_n(\lambda_i) - f_\theta(\lambda_i) \); namely, we use (67) and also the fact that

\[
\hat{Z}_n = \sup \max_{1 \leq i \leq N} ||g_\theta(\lambda_i)||I_n(\lambda_i) - f(\lambda_i)|| = o_p(n^{m})
\]

because

\[
\max_{1 \leq i \leq N} ||g_\theta(\lambda_i)||f_\theta(\lambda_i) \leq C \max_{1 \leq i \leq N} (\lambda_i^\delta + n^{-m} \lambda_i^{-\delta}) \lambda_i^{-\alpha} \leq C \max_{1 \leq i \leq N} \lambda_i^\delta \lambda_i^{-\alpha} + C n^{-m} \left( \sum_{i=1}^N \lambda_i^{-2\delta - 2\alpha} \right)^{1/2} = o(n^m)
\]

by Lemma 7 and Lemma 14(iii).
We can then write $t_\theta = \bar{W}_{n\theta}^{-1} \bar{I}_{\theta} + \tilde{\phi}_n \theta$ for $\theta \in B_n$ with $||\tilde{\phi}_n|| = o_p(n^{-m})$. Since the counterpart of (61) is still valid, namely

$$\sup_{\theta \in B_n} \max_{1 \leq i \leq N} |\tilde{\gamma}_n(\theta)| \leq \sup_{\theta \in B_n} ||Z_n|| = o_p(1),$$

we expand $\ell_{n,\gamma}(\theta)$ for each $\theta \in B_n$ analogous to (62) [using a multiple of 2]

$$\ell_{n,\gamma}(\theta) = \sum_{i=1}^{N} (1 + \tilde{\gamma}_n(\theta)) = n \bar{I}_{\theta}(2\pi \bar{W}_{n\theta})^{-1} \bar{I}_{\theta} + o_p(n^{1-2m})$$

Then following the same essential steps in (63), we make a Taylor expansion around $\theta_0$ to write

$$\bar{I}_{\theta} = \bar{I}_{\theta_0} + D_{n\theta_0}(\theta - \theta_0) - D_{n\theta_0}^* (\theta - \theta_0) + \tilde{E}_{n\theta},$$

uniformly for $\theta \in B_n$, where

$$D_{n\theta_0}^* = \frac{2\pi}{n} \sum_{i=1}^{N} \left( \frac{\partial g_{n\theta_0} (\lambda_i)}{\partial \theta} f_{\theta_0} (\lambda_i) + g_{n\theta_0} (\lambda_i) \left( \frac{\partial f_{\theta_0} (\lambda_i)}{\partial \theta} \right) \right)$$

and $||\tilde{E}_{n\theta}|| \leq C ||\theta - \theta_0||^2 n^{-1} \sum_{i=1}^{N} (\lambda_i^0 + 1 - \lambda_i^0 - \lambda_i) = o_p(n^{-m})$.

By the Lebesgue Dominated Convergence Theorem, as $n \to \infty$,

$$D_{n\theta_0}^* \to \frac{1}{2} \int_{\Omega} \left( \frac{\partial g_{\theta_0} (\lambda)}{\partial \theta} f_{\theta} + g_{\theta_0} \frac{\partial f_{\theta}}{\partial \theta} \right) d\lambda = \frac{1}{2} \frac{\partial}{\partial \theta} \left[ \int_{\Omega} f_{\theta} g_{\theta} d\lambda \right] \bigg|_{\theta = \theta_0} = 0 \in \mathbb{R}^r$$

since the conditions of Theorem 2 imply we may take partial derivatives outside the integral and $\int_{\Omega} f_{\theta} g_{\theta} d\lambda = p$ is constant in the $B_n$-neighborhood of $\theta_0$.

Hence, we may write uniformly for $\theta \in B_n$

$$\bar{I}_{\theta} = \bar{I}_{\theta_0} + D_{n\theta_0}(\theta - \theta_0) + o_p(n^{-m}).$$

Then, Lemma 5 implies $\bar{I}_{n\theta_0}$ is asymptotically normal with mean 0 and covariance matrix $\pi W_{\theta_0}$ under the assumption that $p = 0$ or $\kappa_{m,r} = 0$ and $p \neq 0$. After substituting $\bar{I}_{n\theta_0}$, $\ell_{n,\gamma}(\theta)$ for $I_{n\theta_0}$, $\ell_n(\theta)$, the remaining arguments from (64) and (66) can now be applied to show that $\ell_{n,\gamma}(\theta)$ attains a minimum $\hat{\theta}_{n,\gamma}$ on $B_n \setminus \partial B_n$ with probability approaching 1 and $\hat{\theta}_{n,\gamma}$ satisfies

$$0 = \frac{\partial \ell_{n,\gamma}(\theta)}{\partial \theta} \bigg|_{\theta = \hat{\theta}_{n,\gamma}} = \left. \frac{n}{\pi} \hat{Q}_{2n}(\hat{\theta}_{n,\gamma}, t_{\hat{\theta}_{n,\gamma}}) \right|_{\theta = \hat{\theta}_{n,\gamma}},$$

in place of (66). The proof of Lemma 8 is now finished. □
Proof of Theorem 2. We follow the essential steps in the proof of Theorem 1 of Qin and Lawless (1994). We first establish the asymptotic normality of the MELE $\hat{\theta}_n$ maximizing $R_n(\theta)$. Taking derivatives, we have

$$
\frac{\partial Q_{1n}(\theta_n,0)}{\partial \theta} = \frac{2\pi}{n} \sum_{i=1}^{N} \frac{\partial g_{\theta_n}(\lambda_i)}{\partial \lambda_i} I_n(\lambda_i), \quad \frac{\partial Q_{2n}(\theta_n,0)}{\partial \theta} = 0,
$$

$$
\frac{\partial Q_{1n}(\theta_n,0)}{\partial t} = -\frac{2\pi}{n} \sum_{i=1}^{N} g_{\theta_n}(\lambda_i) g'_{\theta_n}(\lambda_i) I''_n(\lambda_i), \quad \frac{\partial Q_{2n}(\theta_n,0)}{\partial t} = \left[ \frac{\partial Q_{1n}(\theta_n,0)}{\partial \theta} \right]'.
$$

Expanding $Q_{1n}(\theta_n, t_{\theta_n})$, $Q_{2n}(\theta_n, t_{\theta_n})$ at $(\theta_n, 0)$, we have by (66)

$$
0 = Q_{1n}(\theta_n, t_{\theta_n}) = Q_{1n}(\theta_n, 0) + \frac{\partial Q_{1n}(\theta_n, 0)}{\partial \theta} (\theta_n - \theta_0) + \frac{\partial Q_{1n}(\theta_n, 0)}{\partial t} (t_{\theta_n} - 0) + E_{1n},
$$

$$
0 = Q_{2n}(\theta_n, t_{\theta_n}) = Q_{2n}(\theta_n, 0) + \frac{\partial Q_{2n}(\theta_n, 0)}{\partial \theta} (\theta_n - \theta_0) + \frac{\partial Q_{2n}(\theta_n, 0)}{\partial t} (t_{\theta_n} - 0) + E_{2n}.
$$

For $d_n = ||t_{\theta_n}|| + ||\hat{\theta}_n - \theta_0||$, one may verify:

$$
||E_{1n}||, ||E_{2n}|| = O_p \left( \delta_n n^{-1} \sum_{i=1}^{N} \left( n^{-m} \lambda_i^{-2d} + n^{-m} \lambda_i^{-d} \right) (I_n^2(\lambda_i) + I_n(\lambda_i)) \right) = o_p(\delta_n),
$$

by Lemma 14. Then,

$$
\Sigma_n^{\top} \left( \begin{array}{c} t_{\theta_n} \\ \hat{\theta}_n - \theta_0 \end{array} \right) = \left[ \begin{array}{c} Q_{1n}(\theta_n, 0) + o_p(\delta_n) \\ \frac{\partial Q_{1n}(\theta_n, 0)}{\partial \theta} a_p(\delta_n) \end{array} \right], \quad \Sigma_n = \left[ \begin{array}{cc} \frac{\partial Q_{1n}(\theta_n, 0)}{\partial \theta} & \frac{\partial Q_{2n}(\theta_n, 0)}{\partial \theta} \\ \frac{\partial Q_{2n}(\theta_n, 0)}{\partial \theta} & 0 \end{array} \right].
$$

By Lemma 6 and (64),

$$
\Sigma_n \xrightarrow{p} \frac{1}{2} \begin{bmatrix} -2W_{\theta_0} & D_{\theta_0} \\ D_{\theta_0}^{\top} & 0 \end{bmatrix}, \quad \Sigma_n^{-1} = \begin{bmatrix} A_{11n} & A_{12n} \\ A_{21n} & A_{22n} \end{bmatrix} \xrightarrow{p} \frac{1}{2\pi} \begin{bmatrix} -2U_{\theta_0} & W_{\theta_0}^{-1}D_{\theta_0}V_{\theta_0} \\ V_{\theta_0}D_{\theta_0}^{\top}W_{\theta_0}^{-1} & 2V_{\theta_0} \end{bmatrix}.
$$

Then $\sqrt{n}Q_{1n}(\theta_n, 0) \xrightarrow{d} N(0, \pi W_{\theta_0})$ by Lemma 5 and so it holds that $\delta_n = O_p(n^{-1/2})$. We also have that

$$
\sqrt{n}(\hat{\theta}_n - \theta_0) = -\sqrt{n} A_{21n} Q_{1n}(\theta_n, 0) + o_p(1) \xrightarrow{d} N(0, V_{\theta_0}),
$$

$$
\sqrt{n}(t_{\theta_n} - 0) = -\sqrt{n} A_{11n} Q_{1n}(\theta_n, 0) + o_p(1) \xrightarrow{d} N(0, U_{\theta_0}).
$$

Note finally that $V_{\theta_0}D_{\theta_0}^{\top}W_{\theta_0}^{-1}W_{\theta_0}U_{\theta_0} = 0$, implying $t_{\theta_n}$ and $\hat{\theta}_n - \theta_0$ are asymptotically uncorrelated.

We now show the asymptotic normality of $\sqrt{n}(\hat{\theta}_{n,x} - \theta_0, t_{\theta_n,x}/2)$. Taking derivatives,

$$
\frac{\partial \tilde{Q}_{1n}(\theta_n,0)}{\partial \theta} = \frac{2\pi}{n} \sum_{i=1}^{N} \frac{\partial g_{\theta_n}(\lambda_i)}{\partial \lambda_i} I_n(\lambda_i) - D^*_n \theta_0, \quad \frac{\partial \tilde{Q}_{2n}(\theta_n,0)}{\partial \theta} = 0,
$$

$$
\frac{\partial \tilde{Q}_{1n}(\theta_n,0)}{\partial t} = -\frac{2\pi}{n} \sum_{i=1}^{N} g_{\theta_n}(\lambda_i) g'_{\theta_n}(\lambda_i) (I_n(\lambda_i) - f_\theta(\lambda_i)), \quad \frac{\partial \tilde{Q}_{2n}(\theta_n,0)}{\partial t} = \left[ \frac{\partial \tilde{Q}_{1n}(\theta_n,0)}{\partial \theta} \right]'.
$$
where $D_{n\theta_0}$ is from (69). Analogous to above, we can use (71) and a Taylor expansion to write

$$
\Sigma_n^{-1} = \left[ \begin{array}{cc}
-\bar{Q}_1n(\theta_0,0) + \bar{E}_n & -\bar{Q}_1n(\theta_0,0) + o_p(\delta_n) \\
\frac{\partial}{\partial t} Q_2n(\theta_0,0) & 0
\end{array} \right],
$$

where $\delta_n = ||t_{\theta_0}|| + ||\hat{\theta}_n - \theta_0||$ and

$$
||\bar{E}_1\||, ||\bar{E}_2\|| = o_p\left(\sum_{i=1}^{N} \left[ I_n(\lambda_i) + I_n^2(\lambda_i) + \lambda^{-2a}_i \right] (n^{-2m} \lambda_i^{-2a} + n^{-m} \lambda_i^{-d} + n^{-m} \lambda_i^{-2d}) \right) = o_p(\delta_n)
$$

by Lemma 14. Then by Lemma 6, (64), and (69),

$$
\Sigma_n \rightarrow_p \left[ \frac{1}{2} \begin{array}{cc}
-W_{\theta_0} & D_{\theta_0} \\
D_{\theta_0} & 0
\end{array} \right], \quad \Sigma_n^{-1} \rightarrow_p \left[ \begin{array}{cc}
\bar{A}_{11n} & \bar{A}_{12n} \\
\bar{A}_{21n} & \bar{A}_{22n}
\end{array} \right] \cdot \frac{1}{2\pi} \begin{array}{cc}
-tU_{\theta_0} & W_{\theta_0}^{-1} D_{\theta_0} V_{\theta_0} \\
V_{\theta_0} D_{\theta_0} W_{\theta_0}^{-1} & V_{\theta_0}
\end{array}.
$$

Again, $\sqrt{n} \bar{Q}_1n(\theta_0,0) \rightarrow^d N(0, \pi W_{\theta_0})$ by Lemma 5 and so that $\delta_n = o_p(n^{-1/2})$;

$$
\sqrt{n}(\delta_n - \theta_0) = -\sqrt{n} \bar{A}_{21n} \bar{Q}_1n(\theta_0,0) + o_p(1) \rightarrow^d N(0, U_{\theta_0});
$$

and as before $t_{\delta_n}$ and $\hat{\theta}_n - \theta_0$ are asymptotically uncorrelated.

The proof of Theorem 2 is now finished. \( \Box \)

We need some additional notation for the next proof. For a matrix $X$ with $X'X$ nonsingular, let $P_X = X(X'X)^{-1}X'$ denote the projection matrix of $X$.

**Proof of Theorem 3.** Consider $\ell_n(\theta_0)$ first. By (50), (51), and (53),

$$
\ell_n(\theta) = nQ_1n(\theta,0)(\pi W_{\theta_0})^{-1}Q_1n(\theta,0) + o_p(1)
$$

$$
\ell_n(\theta) = \left[ \sqrt{n}(\pi W_{\theta_0})^{-1/2} Q_1n(\theta,0) \right] \left[ \sqrt{n}(\pi W_{\theta_0})^{-1/2} Q_1n(\theta,0) \right] + o_p(1).
$$

Note as well that if $f = f_{\theta_0}$

$$
\ell_n,f(\theta_0) = nQ_1n(\theta_0,0)(\pi W_{\theta_0})^{-1}Q_1n(\theta_0,0) + o_p(1)
$$

$$
\ell_n,f(\theta_0) = \left[ \sqrt{n}(\pi W_{\theta_0})^{-1/2} Q_1n(\theta_0,0) \right] \left[ \sqrt{n}(\pi W_{\theta_0})^{-1/2} Q_1n(\theta_0,0) \right] + o_p(1),
$$

by the two annotated comments in the last paragraph from the proof of Theorem 1.
Note by (57), (62), (64), and (65), we can write
\[ \ell_n(\hat{\theta}_n) = n \left( Q_{1n}(\theta_0, 0) + \frac{1}{2} \left( \frac{D_{\theta \theta}}{\sqrt{n}} \right) \right) \left( \pi W_{\theta \theta} \right)^{-1} \left( Q_{1n}(\theta_0, 0) + \frac{1}{2} \left( \frac{D_{\theta \theta}}{\sqrt{n}} \right) \right) + o_p(1) \]  
(75)

using (73) and the convergence of \( A_{21n} \) (in probability) for the second equality above and writing \( I_{xx} \) as the \( r \times r \) identity matrix. Likewise by (57), (68), (70), (74), and the convergence of \( \hat{A}_{21n} \):

\[ \ell_{n,r}(\hat{\theta}_{n,r}) = \left[ \sqrt{n}(\pi W_{\theta \theta})^{-1/2} Q_{1n}(\theta_0, 0) \right] \left( I_{xx} - P_{w_{\theta \theta}^{-1/2} d_{w \theta}} \right) \left[ \sqrt{n}(\pi W_{\theta \theta})^{-1/2} Q_{1n}(\theta_0, 0) \right] + o_p(1). \]

By Lemma 5,
\[ \sqrt{n}(\pi W_{\theta \theta})^{-1/2} Q_{1n}(\theta_0, 0) \rightarrow N(0, I_{xx}), \ \sqrt{n}(\pi W_{\theta \theta})^{-1/2} Q_{1n}(\theta_0, 0) \rightarrow N(0, I_{xx}). \]

Theorem 3 now follows from: \( P_{w_{\theta \theta}^{-1/2} d_{w \theta}} \) and \( I_{xx} - P_{w_{\theta \theta}^{-1/2} d_{w \theta}} \) are orthogonal idempotent matrices;

\[ \text{rank} \left( P_{w_{\theta \theta}^{-1/2} d_{w \theta}} \right) = \text{rank} \left( W_{\theta \theta}^{-1/2} D_{\theta \theta} \right) = \text{rank} \left( D_{\theta \theta} \right) = p; \]

\[ \text{rank} \left( I_{xx} - P_{w_{\theta \theta}^{-1/2} d_{w \theta}} \right) = r - \text{trace} \left[ P_{w_{\theta \theta}^{-1/2} d_{w \theta}} \right] = r - \text{rank} \left[ P_{w_{\theta \theta}^{-1/2} d_{w \theta}} \right] = r - p. \]

**Proof of Theorem 4.** We only consider establishing Theorem 4(i), as Theorem 4(ii) follows from minor modification (similar to those found in the proof of Lemma 8, Theorem 2(ii), and Theorem 3(iii)). The subsequent arguments draw heavily from the proofs of Lemma 8, Theorem 2, and Theorem 3. To ease the notation, we write the functions \( \psi(\theta), \Psi(\theta) \) as \( \psi_\theta, \Psi_\theta \) in the following.

We first establish the existence of \( \hat{\theta}_n^* \). Let

\[ Q_{1n}(\theta, t, \nu) = Q_{1n}(\theta, t), \quad Q_{2n}(\theta, t, \nu) = Q_{2n}(\theta, t) + \Psi_\theta \nu, \quad Q_{3n}(\theta, t, \nu) = \psi_\theta \]

and define \( U_n = \left\{ (\theta, t, \nu) \in \mathbb{R}^p \times \mathbb{R}^r \times \mathbb{R}^q : \theta \in B_n, ||t|| + ||\nu|| \leq n^{-(1+2m)/4} \right\} \) with the same \( m \) from Lemma 8. We start by showing that the system of equations:
\[ Q_{1n}^*(\theta, t, \nu) = 0, \quad i = 1, 2, 3 \]
(76)

has a solution \((\theta_n^*, t_n^*, \nu_n^*) \in U_n\). On a set of arbitrarily large probability as \( n \rightarrow \infty \), \( t_\theta \) is a continuously differentiable function of \( \theta \in B_n \), where \( Q_{1n}^*(\theta, t_\theta) = 0 \). Differentiating with respect to \( \theta \), we find
\[ 0 = \frac{\partial Q_{1n}(\theta, t_\theta)}{\partial \theta} \]
\[ = \frac{2\pi}{n} \sum_{i=1}^{N} \frac{I_n(\lambda_i)}{1 + \epsilon_\theta g_\theta(\lambda_i) I_n(\lambda_i)} \left[ \frac{\partial g_\theta(\lambda_i)}{\partial \theta} \right] - \frac{g_\theta(\lambda_i) I_n(\lambda_i)}{(1 + \epsilon_\theta g_\theta(\lambda_i) I_n(\lambda_i))^2} \left[ \frac{\partial g_\theta(\lambda_i)}{\partial \theta} + \epsilon_\theta \frac{g_\theta(\lambda_i)}{\partial \theta} \right] \]
so that

$$\frac{\partial \ell_n}{\partial \theta} = \frac{1}{2} W^{-1}_n D_{e_{0}} + o_p(1) = \frac{1}{2} W^{-1}_n D_{\theta_0} + o_p(1)$$  \hspace{1cm} (77)$$

uniformly for \( \theta \in B_n \), using (57)-(61), (64) and Lemma 14. We can write

$$\pi \frac{\partial \ell_n(\theta)}{\partial \theta} = \frac{2 \pi}{n} \sum_{i=1}^{n} \frac{\partial \ln((1 + t^* g(\lambda_i) I_n(\lambda_i))}{\partial \theta} = \frac{2 \pi}{n} \sum_{i=1}^{n} I_n(\lambda_i) \left( \frac{\partial g(\lambda_i)}{\partial \theta} \right) t^*_i = F_n(\theta) t^*_i.$$  

By (61), (64), and a Taylor's expansion,

$$\sup_{\theta \in B_n} \left\| F_n(\theta) - \frac{1}{2} D_{\theta_0} \right\| = o_p(1).$$  \hspace{1cm} (78)$$

Using Lemma 14, we can also show

$$\sup_{\theta \in B_n} \left\| \frac{\partial F_n(\theta)}{\partial \theta} \right\| = O_p(1).$$  \hspace{1cm} (79)$$

Then, uniformly for \( \theta \in B_n \),

$$\frac{\pi}{2n} \frac{\partial \ell_n(\theta)}{\partial \theta} = \frac{\pi}{2n} \frac{\partial \ell_n(\theta_0)}{\partial \theta} + \left[ \frac{\partial F_n(\theta)}{\partial \theta} t^*_i + F_n(\theta) \frac{\partial t^*_i}{\partial \theta} \right] \theta^* \text{ between } \theta, \theta_0$$

$$= \frac{\pi}{2n} \frac{\partial \ell_n(\theta_0)}{\partial \theta} + \pi V^{-1}_n(\theta - \theta_0) - R_n(\theta),$$  \hspace{1cm} (80)$$

where \( \sup_{\theta \in B_n} \left\| R_n(\theta) \right\| = o_p(n^{-m}) \), \( R_n(\theta) \) continuous by (59), (77), (78), (79) [see the proof of Lemma 8(i)].

For \( \theta \in B_n \), write

$$\psi_\theta - \Psi_{\theta_0} (\theta - \theta_0) = \| \theta - \theta_0 \|^2 k(\theta),$$

where \( k(\theta) \) is continuous and bounded. Note that \( \Psi_{\theta_0} V_{\theta_0} \Psi_{\theta_0}' \) is invertible so that, since

$$\| \Psi_{\theta_0} V_{\theta_0} \Psi_{\theta_0}' - \Psi_{\theta_0} V_{\theta_0} \Psi_{\theta_0} \| \leq C \| \theta - \theta_0 \| \leq C n^{-m},$$

\( \Psi_{\theta_0} V_{\theta_0} \Psi_{\theta_0}' \) is invertible for \( \theta \in B_n \), large \( n \). By (80), we then write: \( \theta \in B_n \),

$$\frac{\pi}{2n} \frac{\partial \ell_n(\theta)}{\partial \theta} + \pi \Psi_{\theta_0} (\theta - \theta_0)^{-1} \left( \| \theta - \theta_0 \|^2 k(\theta) - \frac{1}{\pi} \Psi_{\theta_0} V_{\theta_0} \left[ \frac{2 \pi}{n} \frac{\partial \ell_n(\theta)}{\partial \theta} - \pi V^{-1}_n(\theta - \theta_0) \right] \right)$$

$$= \pi V^{-1}_n(\theta - \theta_0) - \tilde{R}_n(\theta) \quad \tilde{R}_n(\theta) \text{ continuous, } \sup_{\theta \in B_n} \left\| \tilde{R}_n(\theta) \right\| = o_p(n^{-m}).$$

$$= -\eta(\theta).$$

We show now that \( \eta(\theta) \) has a root inside \( B_n \setminus B_n \). Note \( \eta(\theta) \) is continuous on \( B_n \) and

$$\frac{(\theta - \theta_0)^{\gamma}}{\| \theta - \theta_0 \|^\gamma} \eta(\theta) = \frac{-\pi}{\| \theta - \theta_0 \|^\gamma} (\theta - \theta_0)^{\gamma} V^{-1}_n(\theta - \theta_0) + \frac{\theta - \theta_0)^{\gamma}}{\| \theta - \theta_0 \|^\gamma} \tilde{R}_n(\theta)$$

$$\leq \left( -m + \frac{\| \tilde{R}_n(\theta) \|}{\| \theta - \theta_0 \|^\gamma} \right) \| \theta - \theta_0 \|$$
where \( \mu_0 \) is the minimum eigenvalue of \( \pi V_0^{-1} \); for large \( n \), on a set of arbitrarily large probability, 
\[
\sup_{\theta \in \partial B_n} ||\hat{R}_n(\theta)|| n^m < \mu_0,
\]
implies that there exists \( \hat{\theta}_n^* \in {\partial B_n} \) such that
\[
\eta(\hat{\theta}_n^*) = 0,
\]
by Lemma 2 of Aitchison and Silvey (1958). [The lemma states that if \( g : \mathbb{R}^s \to \mathbb{R}^s \) is continuous, such that \( \theta' g(\theta) < 0 \) for every \( ||\theta|| = 1 \), then there exists a point \( \hat{\theta} \) such that \( ||\hat{\theta}|| < 1 \) and \( g(\hat{\theta}) = 0 \).

Then,
\[
0 = -\frac{1}{n} \Psi_{\theta_0} V_{\theta_0} \eta(\hat{\theta}_n^*) = ||\hat{\theta}_n^* - \theta_0||^2 k(\hat{\theta}_n^*) + \Psi_{\theta_0}(\hat{\theta}_n^* - \theta_0) = \psi(\hat{\theta}_n^*). \tag{81}
\]
It also follows that
\[
\frac{\pi}{2n} \frac{\partial \ell_n(\hat{\theta}_n^*)}{\partial \theta} = \Psi_{\theta_0}(\Psi_{\theta_0} V_{\theta_0} \Psi_{\hat{\theta}_n}^{-1}) \frac{\partial \ell_n(\hat{\theta}_n^*)}{\partial \theta}, \tag{82}
\]
so that \( \hat{\theta}_n^* \) and
\[
\nu_n^* = -\pi(\Psi_{\theta_0} V_{\theta_0} \Psi_{\hat{\theta}_n}^{-1}) \frac{\partial \ell_n(\hat{\theta}_n^*)}{\partial \theta}
\]
satisfy (76).

We next prove that any solution of (76) in \( U_n \), say \( (\hat{\theta}, \hat{\tau}, \hat{\nu}) \), minimizes \( \ell_n(\theta) \) on \( B_n \), subject to the condition \( \psi_\theta = 0 \). If \( \theta \in B_n, \psi_\theta = 0 \), then we make a Taylor’s expansion around \( \hat{\theta} \):
\[
\frac{\pi}{2n} \left[ \ell_n(\theta) - \ell_n(\hat{\theta}) \right] = \frac{\pi}{2n} \frac{\partial \ell_n(\hat{\theta})}{\partial \theta} (\theta - \hat{\theta}) + \frac{\pi}{4n} (\theta - \hat{\theta})' \frac{\partial^2 \ell_n(\hat{\theta})}{\partial \theta \partial \theta'} (\theta - \hat{\theta}), \quad \theta^* \text{ between } \theta, \hat{\theta}.
\]

Since \( \hat{\theta} \) satisfies (76), it follows from some algebra that \( \hat{\theta} \) also satisfies (81), substituting \( \hat{\theta} \) for \( \hat{\theta}_n^* \). Note that
\[
0 = \psi_\theta - \psi_{\hat{\theta}} = \Psi_{\theta_0}(\theta - \hat{\theta}) + o(||\theta - \hat{\theta}||^2).
\]
Using \( \hat{\theta} \) in (81), we find
\[
\frac{\pi}{2n} \frac{\partial \ell_n(\hat{\theta})}{\partial \theta} (\theta - \hat{\theta}) = \frac{\pi}{2n} \frac{\partial \ell_n(\hat{\theta})}{\partial \theta} V_{\theta_0} \Psi_{\theta_0}' \left[ (\Psi_{\theta_0} V_{\theta_0} \Psi_{\hat{\theta}}^{-1}) \Psi_{\theta_0}(\theta - \hat{\theta}) = o_p(||\theta - \hat{\theta}||). \tag{80}
\]

It follows from arguments in (80) that
\[
\frac{\pi}{2n} \frac{\partial^2 \ell_n(\theta^*)}{\partial \theta \partial \theta'} = \left. \frac{\partial [\psi_{\theta} F_{\theta}^{-1}(\theta)]}{\partial \theta} \right|_{\theta = \theta^*} = \pi V_{\theta_0}^{-1} + o_p(1).
\]
We then have
\[
\ell_n(\theta) - \ell_n(\hat{\theta}) \geq (\mu_0 + o_p(1)) n||\theta - \hat{\theta}||^2,
\]
where the \( o_p(1) \) term is uniform for \( \theta \in B_n, \psi_\theta = 0 \). We have therefore established that there exists a consistent MELE of \( \theta_0, \hat{\theta}_n^* = \hat{\theta}_n^* \), satisfying the condition \( \psi(\hat{\theta}_n^*) = 0 \). Correspondingly, we will denote
\[
\nu_n^* = \nu_n^* \text{ from (82).}
We now show

\[
\sqrt{n} \left( \hat{\theta}_n^* - \theta_0 \right) \xrightarrow{d} N \left( 0, \pi \begin{bmatrix} P_{\theta_0} & 0 \\ 0 & R_{\theta_0} \end{bmatrix} \right), \quad P_{\theta_0} = \frac{1}{\pi} V_{\theta_0} \left( I_{m,m} - \frac{1}{\pi} \Psi_{\theta_0}' R_{\theta_0} \Psi_{\theta_0} V_{\theta_0} \right),
\]

\[
R_{\theta_0} = \pi \left( \Psi_{\theta_0} V_{\theta_0} \Psi_{\theta_0}' \right)^{-1}.
\]

Expanding \( Q_{in}^{*}(\theta, t, \nu) \) at \((\theta_0, 0, 0)\) and using that \((\hat{\theta}_n^*, \hat{t}_n^*, \nu_n^*)\) satisfies (76), we have [see (72)]:

\[
\begin{bmatrix}
-Q_{1n}(\theta_0, 0) + o_p(\delta_n^*) \\
o_p(\delta_n^*) \\
o_p(\nu_n^*)
\end{bmatrix} = \Sigma_n \begin{bmatrix}
\hat{\theta}_n^* - \theta_0 \\
\hat{t}_n^* \\
\hat{\nu}_n^*
\end{bmatrix},
\]

\[
\Sigma_n = \begin{bmatrix}
\frac{\partial Q_{1n}(\theta_0, 0)}{\partial \theta} & \frac{\partial Q_{1n}(\theta_0, 0)}{\partial t} & \frac{\partial Q_{1n}(\theta_0, 0)}{\partial \nu} \\
\frac{\partial Q_{2n}(\theta_0, 0)}{\partial \theta} & 0 & \Psi_{\theta_0}' \\
0 & \Psi_{\theta_0} & 0
\end{bmatrix},
\]

where \( \delta_n^* = ||\hat{\theta}_n^* - \theta_0|| + ||\hat{t}_n^*|| + ||\hat{\nu}_n^*||. \) Then,

\[
\Sigma_n \xrightarrow{P} \begin{bmatrix}
-W_{\theta_0} & 0 & 0 \\
\frac{\partial Q_{2n}}{\partial t} & 0 & \Psi_{\theta_0}' \\
0 & \Psi_{\theta_0} & 0
\end{bmatrix} = \begin{bmatrix}
C_{11} & C_{12} \\
C_{21} & C_{22}
\end{bmatrix} = \tilde{C},
\]

\[
C_{12} = \begin{bmatrix}
\frac{\partial Q_{1n}}{\partial t} \\
0
\end{bmatrix}, \quad C_{21} = \begin{bmatrix}
0 \\
\Psi_{\theta_0}'
\end{bmatrix}.
\]

Note that \( \text{det}(\tilde{C}) = \text{det}(C_{11}) \text{det}(Q_{cz}^{-1}) = \text{det}(-W_{\theta_0}) \text{det}(\pi V_{\theta_0}^{-1}) \text{det}(-R_{\theta_0}^{-1}) \neq 0, \) for \( Q_{cz} = C_{22} - C_{21} C_{11}^{-1} C_{12}, \) and

\[
\tilde{C}^{-1} = \begin{bmatrix}
-W_{\theta_0}^{-1} + W_{\theta_0}^{-1} C_{12} Q_{cz}^{-1} C_{21} W_{\theta_0}^{-1} \\
Q_{cz}^{-1} C_{21} W_{\theta_0}^{-1}
\end{bmatrix}, \quad Q_{cz}^{-1} = \begin{bmatrix}
P_{\theta_0} & \frac{1}{\pi} V_{\theta_0} \Psi_{\theta_0}' R_{\theta_0} \\
\frac{1}{\pi} R_{\theta_0} \Psi_{\theta_0} V_{\theta_0} & -R_{\theta_0}
\end{bmatrix}.
\]

Since, by Lemma 5, \( \sqrt{n} Q_{1n}(\theta_0, 0) \xrightarrow{d} N(0, \pi W_{\theta_0}), \) it follows that \( \delta_n^* = O_p(n^{-1/2}). \) Then,

\[
\sqrt{n} \left( \hat{\delta}_n^* - \theta_0 \right) = \sqrt{n} Q_{cz}^{-1} C_{21} W_{\theta_0}^{-1} Q_{1n}(\theta_0, 0) + o_p(1)
\]

\[
\xrightarrow{d} N \left( 0, \pi \begin{bmatrix} P_{\theta_0} & 0 \\ 0 & R_{\theta_0} \end{bmatrix} \right).
\]

As in (75), we can then expand

\[
\ell_n(\hat{\theta}_n^*) = n \left[ Q_{1n}(\theta_0, 0) + \frac{D_{\theta_0}}{2} (\hat{\theta}_n^* - \theta_0) \right] (\pi W_{\theta_0})^{-1} Q_{1n}(\theta_0, 0) + o_p(1)
\]

\[
= n Q_{1n}(\theta_0, 0) \left( I_{m,m} - \frac{1}{4} D_{\theta_0} R_{\theta_0} D_{\theta_0}' W_{\theta_0}^{-1} \right) \left( \pi W_{\theta_0} \right)^{-1} \left( I_{m,m} - \frac{1}{4} D_{\theta_0} R_{\theta_0} D_{\theta_0}' W_{\theta_0}^{-1} \right) Q_{1n}(\theta_0, 0) + o_p(1)
\]

\[
= \left[ \sqrt{n}(\pi W_{\theta_0})^{-1/2} Q_{1n}(\theta_0, 0) \right] \left[ I_{m,m} - (P_{w_{\theta_0}^{-1/2} D_{\theta_0}} - P_{H_{\theta_0}}) \right] \left[ \sqrt{n}(\pi W_{\theta_0})^{-1/2} Q_{1n}(\theta_0, 0) \right] + o_p(1),
\]

where \( H_{\theta_0} = W_{\theta_0}^{-1/2} d_{\theta_0} (D_{\theta_0}' R_{\theta_0}^{-1} D_{\theta_0})^{-1} \Psi_{\theta_0}. \) Then,

\[
\ell_n(\hat{\theta}_n^*) = \ell_n(\theta_0) + \ell_n(\delta_n^*) = \left[ \sqrt{n}(\pi W_{\theta_0})^{-1/2} Q_{1n}(\theta_0, 0) \right] P_{H_{\theta_0}} \left[ \sqrt{n}(\pi W_{\theta_0})^{-1/2} Q_{1n}(\theta_0, 0) \right] + o_p(1),
\]

\[
\ell_n(\theta_0) = \ell_n(\delta_n^*)
\]

\[
= \left[ \sqrt{n}(\pi W_{\theta_0})^{-1/2} Q_{1n}(\theta_0, 0) \right] (P_{w_{\theta_0}^{-1/2} D_{\theta_0}} - P_{H_{\theta_0}}) \left[ \sqrt{n}(\pi W_{\theta_0})^{-1/2} Q_{1n}(\theta_0, 0) \right] + o_p(1).
\]
Note now that \( \sqrt{n}(\pi W_{\theta_0})^{-1/2} Q_{1n}(\theta_0, 0) \xrightarrow{d} N(0, I_{n-1}) \) by Lemma 5, \( P_{\theta_0} \) and \( P_{\theta_0}^{-1/2} \sigma_{\theta_0} - P_{\theta_0} \) are idempotent matrices with

\[
\begin{align*}
\text{rank}(P_{\theta_0}) &= \text{rank}(H_{\theta_0}) = \text{rank}(\Psi_{\theta_0}) = q; \\
\text{rank}(P_{\theta_0}^{-1/2} \sigma_{\theta_0} - P_{\theta_0}) &= p - \text{trace}[P_{\theta_0}] = p - \text{rank}(P_{\theta_0}) = p - q.
\end{align*}
\]

(We use \( \text{rank}(H_{\theta_0}) \leq \text{rank}(\Psi_{\theta_0}), \text{rank}(\Psi_{\theta_0}) = \text{rank}(D_{\theta_0}^{-1/2} H_{\theta_0}) \leq \text{rank}(H_{\theta_0}) \) above.) Theorem 4 now follows. \( \Box \)

**Proof of Theorem 5.** By assumption, \( f \) is continuous on \( \Pi \); by Corollary 3.2 of Dahlhaus (1985a),

\[
\sqrt{n} \left( \int_0^\lambda I_n^{(T)} g_{\theta_0} d\lambda - \mathbb{E} \int_0^\lambda I_n^{(T)} g_{\theta_0} d\lambda \right) \xrightarrow{d} N(0, HV),
\]

where \( V \) is the covariance matrix from Lemma 5 and \( H = \int_0^1 h^2 dx / \int_0^1 h^2 dx^2 \); and by Theorem 1 of Dahlhaus (1983): for \( a_1, \ldots, a_k \in \mathbb{R}, 2 \leq k \leq 8 \)

\[
\text{cum}(d_n(a_1), \ldots, d_n(a_k)) = (2\pi)^{k-1} H_k^{(T)} \left( \sum_{j=1}^k \kappa_{k,j} \sum_{j=1}^k a_j \prod_{j=1}^k b(-a_j) + R_n \right)
\]

\[
|R_n| = o(n) \text{ uniformly in } a_1, \ldots, a_k.
\]

Since \( |H_k^{(T)}(\lambda)| \leq CL_{n0}(\lambda) \) [see Dahlhaus (1983)], we find Lemmas 1 and 2 still hold with respect to the tapered discrete Fourier transform:

\[
d_{n,c}^{(T)}(\lambda) = \sum_{t=0}^{n-1} h(t/n) X_{t+1} e^{-i\lambda t}, \quad \lambda \in \Pi,
\]

as does Lemma 9 (with \( \alpha = \beta = 0 \)). Lemma 4 also remains valid after substituting the tapered periodogram \( I_k^{(T)} \). (The same essential arguments can be used to justify the tapered version of Lemma 4 after replacing \( H_n, K_n \) by \( H_1^{(T)}, \Phi_1^{(T)}, \) where

\[
\Phi_1^{(T)}(\lambda) = \begin{cases} 
|H_1^{(T)}(\lambda)|^2 / (2\pi H_2^{(T)}(0)) & H_2^{(T)}(0) \neq 0 \\
0 & H_2^{(T)}(0) = 0.
\end{cases}
\]

Note that \( \Phi_1^{(T)}(\lambda) \) is an approximate identity in that:

\[
\sup_n \int_{\Pi} |\Phi_1^{(T)}| d\lambda < \infty, \quad \lim_{n \to \infty} \int_{\Pi} \Phi_1^{(T)} d\lambda = 1, \quad \lim_{n \to \infty} \int_{\Pi \setminus \{|\lambda| < \rho\}} \Phi_1^{(T)} d\lambda = 0 \text{ for } \rho > 0.
\]

Also, (26) holds in probability with the tapered periodogram again by the same arguments in Hannan (1973), since \( r_n^{(T)}(u) \xrightarrow{d} r(u) = \text{Cov}(X_j, X_{j+u}) \), where

\[
r_n^{(T)}(u) = \frac{1}{2\pi H_2^{(T)}(0)} \sum_{t=0}^{n-1} A(t + |u|) \mathbb{A} \left( \frac{t}{n} \right) X_{t+1+|u|} X_{t+1}, \quad u \in \mathbb{Z}.
\]
From (84) and the proof of Lemma 5 under Condition 3 of Assumption A.3, we find that Lemma 5 holds replacing \( I_n, V \) with \( I_n^{(r)}, HV \).

We can also establish that Lemma 6 still holds. To do so, it suffices to show that, for

\[
S_n^{(r)} = \frac{2\pi}{n} \sum_{j=1}^{N} g(\lambda_j)w(\lambda_j)I_n^{(r)}(\lambda_j)^2, \quad \bar{S}_n^{(r)} = \frac{2\pi}{n} \sum_{j=1}^{N} g(\lambda_j)w(\lambda_j)J(\lambda_j)I_n^{(r)}(\lambda_j),
\]

it holds that

\[
E(S_n^{(r)}) = \frac{2\pi}{n} \sum_{j=1}^{N} g(\lambda_j)w(\lambda_j)J(\lambda_j) = o(1),
\]

because the remaining elements of the proof of Lemma 6 are still valid upon using \( I_n^{(r)} \) [see (41) and (43)]. However, these convergence results follow easily from the product theorem for cumulants [replacing \( (2\pi n)^2E(I_n^2(A_i)) \) with \( E(\{2\pi H_2^{(r)}(0)J_n^{(r)}(\lambda_j)^2\}) \) in (42)], (85), and \( |H_2^{(r)}(\lambda)| \leq C L_{a0}(\lambda) \).

Because Lemma 7 still holds and \( A_n^{(r)} \to H \) as \( n \to \infty \), we can now follow the proof of Theorem 1 with minor and straightforward modifications. The proof of Theorem 5 is now finished. \( \square \)

**Proof of Proposition 1.** We give a sketch of the proof, which essentially repeats many details from the proofs of Lemma 8, Theorem 2(ii), and Theorem 3(ii). Since \( \{X_t\} \) is Gaussian, it follows that \( E(\xi_0^k) < \infty, k \geq 1 \) and we can extend Lemma 1 to include cumulants of order \( k > 8 \), while Lemmas 2 and 3 also remain valid.

Using results in Fox and Taqqu (1987), Beran (1992) established the limiting bivariate normal distribution of \( (\int_{a}^{b} I_n c f^{-1} d\lambda, \int_{a}^{b} I_n c f^{-1} d\lambda)' \) under Proposition 1 conditions. Modifying the proof of Theorem 1 in Beran (1992) (since \( g_2^x \) has the same growth rate as \( f_2^{-1} \) near the origin) and using the Cramer-Wold device, we find: for any \( a \in \mathbb{R}^p, b \in \mathbb{R}^p \)

\[
\int_{a}^{b} b f_2^{-1} d\lambda + I_n c a g_2^w d\lambda \rightarrow N\left(0, \begin{pmatrix} b \end{pmatrix} W_{s_0}^{-1} \begin{pmatrix} b \end{pmatrix} \right),
\]

\[
W_{s_0} = \begin{bmatrix}
10\pi & 4\pi & 0 \\
4\pi & 2\pi & 0 \\
0 & 0 & \infty
\end{bmatrix}, \quad W_{s_0}^{**} = \left( \int_{\Pi} f_2^{-1} \frac{\partial f_2^{-1}}{\partial \theta_i} \frac{\partial f_2^{-1}}{\partial \theta_j} d\lambda \right)_{i,j=1,\ldots,p-1}
\]

with positive definite \( W_{s_0}, W_{s_0}^{**} \); or equivalently,

\[
\left( \int_{a}^{b} I_n c f^{-2} d\lambda, \int_{a}^{b} I_n c g_2^w d\lambda \right) \rightarrow N\left(0, \begin{pmatrix} 2\pi \end{pmatrix} W_{s_0}^{**} \begin{pmatrix} 2\pi \end{pmatrix} \right).
\]
Because $f_n^{-1}$ satisfies Assumption A.3, we can show $f_n^{-2}$ satisfies A.3 as well so that
\[ \left| \frac{2\pi}{n} \sum_{j=1}^{N} \frac{\partial^2 f_n(\lambda_j)}{\partial^2 f_n^2(\lambda_j)} - \frac{2\pi}{n} \sum_{j=1}^{N} 3f_n(\lambda_j) \frac{\partial^2 f_n^2(\lambda_j)}{\partial^2 f_n^2(\lambda_j)} \right| = o_p(n^{-1/2}) \]
using the same arguments in Lemma 5.

By the cumulant product theorem [cf. Brillinger (1981)], we can show that
\[ \left| \frac{2\pi}{n} \sum_{j=1}^{N} \frac{\partial^2 f_n(\lambda_j)}{\partial^2 f_n^2(\lambda_j)} g_{\theta_n}(\lambda_j) - \frac{2\pi}{n} \sum_{j=1}^{N} 3f_n(\lambda_j) g_{\theta_n^2}(\lambda_j) \right| = o_p(1), \]
by proceeding with moment arguments as in (41). From this and Lemma 6, we find
\[ \frac{2\pi}{n} \sum_{j=1}^{N} g_{\theta_n}(\lambda_j) g_{\theta_n^2}(\lambda_j) \xrightarrow{p} \frac{1}{2} W_{\theta_n}, \]
an extension of (40). We can extend Lemma 7 to show additionally that
\[ \max_{1 \leq i \leq N} (f_n(x_j) x_j^2) = o(n^b), \quad \max_{1 \leq i \leq N} (f(x_j) x_j^2) = o_p(n^b). \]
Also, we find $\|\partial f_n^{-2}(\lambda)/\partial \theta_n\|, \|\partial^2 f_n^{-2}(\lambda)/\partial \theta_n^2\|$ $\leq C \lambda^{-4-\beta}$ in a neighborhood of $\theta_0$ and that Lemma 1.4 is valid along with extensions to Lemma 1.4(ii)-(iii):
\[ (ii)' \quad n^{-m-1} \sum_{i=1}^{N} \lambda_i^{2d-\alpha} I_n^2(\lambda_i) = o_p(n^{-(1-2m)}), \]
\[ n^{-2m-1} \sum_{i=1}^{N} \lambda_i^{2d-3\beta} \left( I_n^1(\lambda_i) + I_n^2(\lambda_i) \right) = o_p(n^{-(1-2m)}), \]
\[ (iii)' \quad \sum_{i=1}^{N} \lambda_i^{2d-3\alpha} = O(n^m), \quad \sum_{i=1}^{N} \lambda_i^{2d-2d-4\alpha} = o(n^m). \]

We now have all the tools needed to prove Proposition 1 and we can proceed with the same steps and analogous arguments used to prove Lemma 8, Theorem 2(ii), and Theorem 3(ii). The modifications required are straightforward.

8 Appendix

Proof of Lemma 1. To show Lemma 1(i), note that for $\lambda \in \Pi, |a_1 - \lambda| \leq |a_1|/2$, we have
\[ |H_n(\lambda)| \leq C|a_1|^{-1}, \quad |H_n(a_1 + a_2 - \lambda)| \leq C|a_2|^{-1}, \]
by (24) since $1/2 \leq |(a_1 + a_2 - \lambda) \mod 2\pi|$ (consider cases $|a_2| \leq \pi/2$ or $> \pi/2$); if $|a_1 - \lambda| > |a_1|/2$ then
\[ f(a_1 - \lambda) \leq C|a_1|^{-\alpha}. \]
Integrating separately over $|a_1 - \lambda| \leq |a_1|/2$ and $|a_1 - \lambda| > |a_1|/2$:

$$|\text{cum}(d_{nc}(a_1), d_{nc}(a_2))|$$

$$\leq \int_{|a_1 - \lambda| \leq |a_1|/2} |H_n(a_1 + a_2 - \lambda)||H_n(\lambda)|f(a_1 - \lambda)\,d\lambda$$

$$\leq C|a_2|^{-1}|a_1|^{-\alpha} d\lambda + C|a_1|^{-\alpha} \int L_{n0}(a_1 + a_2 - \lambda)L_{n0}(\lambda)\,d\lambda$$

$$\leq C|a_1|^{-\alpha}(|a_2|^{-1} + L_{n1}(a_1 + a_2)),$$

using Lemma 2(v).

For Lemma 1(ii), note the joint cumulant may be expressed as [cf. Yajima (1989), Lahiri (1999)]:

$$\text{cum}(d_{nc}(a_1), \ldots, d_{nc}(a_k)) = \kappa_{\epsilon,k}(2\pi)^{-k+1} \int_{\Pi^{k-1}} f(z_1, \ldots, z_{k-1}|a_1, \ldots, a_{k-1})\,dz_1, \ldots, dz_{k-1};$$

$$f_n(z_1, \ldots, z_{k-1}|a_1, \ldots, a_{k-1}) = H_n\left(\sum_{j=1}^k a_j - \sum_{j=1}^{k-1} z_j\right)\left(b\left(\sum_{j=1}^{k-1} (a_j - z_j)\right)\prod_{j=1}^{k-1} \{H_n(z_j)b(z_j - a_j)\}\right),$$

where $\kappa_{\epsilon,k}$ denotes the $k$th innovation cumulant, $2 \leq k \leq 8$. Let $B = \{(z_1, \ldots, z_{k-1}) \in \Pi^{k-1} \mid |z_j - a_j| \leq |a_j|/2k, j = 1, \ldots, k\}$. On $B$, $|H(z_j)| \leq C|a_j|^{-1}$ and $|a_k|/2 \leq |a_k - \sum_{j=1}^{k-1} (z_j - a_j)| \leq 3|a_k|/2$ implies

$$\left|H\left(\sum_{j=1}^k a_j - \sum_{j=1}^{k-1} z_j\right)\right| \leq C \max\{|a_k|^{-1}, 2/\pi\} \leq C|a_k|^{-1}$$

by (24). Then applying Holder's inequality

$$\int_{|z_{k-1} - a_k| \leq |a_k|/2k} \left|\sum_{j=1}^{k-1} (a_j - z_j)\right| |b(z_{k-1} - a_{k-1})|\,dz_{k-1}$$

$$\leq \left[\int_{|\lambda - a_{k-1}| \leq |a_{k-1}|/2k} |b(\lambda)|^2 \lambda \cdot \int_{|\lambda - a_{k-1}| \leq |a_{k-1}|/2k} |\lambda - a_{k-1}|^{-\alpha} \,d\lambda\right]^{1/2} \leq C|a_{k-1}|^{(1-\alpha)/2};$$

and for $1 \leq j < k - 1,$

$$\int_{|z_j - a_j| \leq |a_j|/2k} |b(z_j - a_j)|\,dz_j \leq C|a_j|^{1-\alpha/2}.$$

Hence,

$$\int_B |f_n(z_1, \ldots, z_{k-1}|a_1, \ldots, a_{k-1})|\,dz_1, \ldots, dz_{k-1} \leq C(|a_k|^{-1}|a_{k-1}|^{-1/2}) \prod_{j=1}^{k-1} |a_j|^{-\alpha/2}.$$
\[
\int_{\Pi} \left| H \left( \sum_{j=1}^{k} a_j - \sum_{j=1}^{k-1} z_j \right) b \left( \sum_{j=1}^{k-1} (a_j - z_j) \right) \right| \, dz_j = \int_{\Pi} |H(a_k - \lambda)||b(-\lambda)| \, d\lambda
\]

(86)

\[
\leq \int_{|\lambda| \leq |a_k|^{1/2}} C |a_k|^{-1} |\lambda|^{-\alpha/2} \, d\lambda + \int_{|\lambda| > |a_k|^{1/2}} C |a_k|^{-\alpha/2} L_n \Theta(a_k - \lambda) \, d\lambda \leq C |a_k|^{-\alpha/2} \ln(n),
\]

by Lemma 2(iv). For \( j' \in \{1, \ldots, k-1\} \setminus \{j\} \), we follow the same in (86) to get

\[
\int_{\Pi} |H(z_j)||b(x_j - a_j)| \, dx_j = \int_{\Pi} |H(a_j - \lambda)||b(-\lambda)| \, d\lambda \leq C |a_j|^{-\alpha/2} \ln(n).
\]

Since \( \Pi^{k-1} \setminus B = \bigcup_{j=1}^{k-1} B_j \), we have

\[
\int_{\Pi^{k-1} \setminus B} \left| f_n(z_1, \ldots, z_{k-1}[a_1, \ldots, a_{k-1}]) \right| \, dz_1, \ldots, dz_{k-1} \leq \sum_{j=1}^{k-1} \int_{B_j} \left| f_n(z_1, \ldots, z_{k-1}[a_1, \ldots, a_{k-1}]) \right| \, dz_1, \ldots, dz_{k-1} \leq C n \ln^{k-1}(n) \prod_{j=1}^{k} |a_j|^{-\alpha/2}.
\]

The proof of Lemma 1 is now complete. \( \square \)

The following lemma combines some items which are useful in many proofs of the paper.

Lemma 9 Let \( r(k) = \text{Cov}(X_0, X_k), k \in \mathbb{Z} \). Under Assumption A.1,

(i) \(|r(k)| \leq C |k|^{-(1-\alpha)} \) for all \( k \geq 1 \),

(ii) \((2\pi n)^{-1} E\{I_n(0)\} \leq C n^\alpha, \quad (2\pi n)^{-1} E\{I_n(\pi)\} \leq C \).

(iii) If \( g : \Pi \rightarrow \mathbb{R} \) is an even, integrable function with \(|g(\lambda)| \leq C |\lambda|^\beta, 0 \leq \beta < 1 \) then \(|c_n g(0)| \leq C n^{-\beta} \).

proof: To prove Lemma 9(i), fix \( k \geq 1 \). Note \(|r(k)| \leq C \int_{\Pi} |\cos(kz)| z^{-\alpha} \, dz \). Then,

\[
\int_0^{\pi/(2k)} z^{-\alpha} |\cos(kz)| \, dz = \int_0^{\pi/(2k)} z^{-\alpha} |\cos(kz)| \, dz = \int_0^{\pi/(2k)} \frac{z^{1-\alpha} \cos(kz)}{1-\alpha} \, dz + \int_0^{\pi/(2k)} \frac{k z^{1-\alpha} \sin(kz)}{1-\alpha} \, dz \leq C k^{-1+\alpha}.
\]

For \( i = 1, \ldots, k-1 \) and \( k > 1 \), let \( b_i = \pi(2i+1)/(2k) \), \( a_i = \pi(2i-1)/(2k) \) and note

\[
\int_{a_i}^{b_i} z^{-\alpha} |\cos(kz)| \, dz = \int_{a_i}^{b_i} z^{-\alpha} \cos(kz) \, dz = \int_{a_i}^{b_i} \frac{z^{-\alpha} \sin(kz)}{1-\alpha} \, dz + \alpha \int_{a_i}^{b_i} z^{-\alpha-1} \sin(kz) \, dz \leq C k^{-1} \left( (a_i^{-\alpha} - b_i^{-\alpha}) + a_i^{-1-\alpha} (b_i - a_i) \right) \leq C k^{-1+\alpha} \left( (2i-1)^{-\alpha} - (2i+1)^{-\alpha} \right) + C k^{-1+\alpha} (2i-1)^{-1-\alpha} \leq C k^{-1+\alpha} (2i-1)^{-1-\alpha}.
\]
by the mean value theorem. The constant $C$ is not dependent on $i = 1, \ldots, k - 1$. Hence,

$$\sum_{i=1}^{k-1} \int_{a_i}^{a_{i+1}} z^{-\alpha} |\cos(kz)| \, dx \leq C k^{-1+\alpha} \sum_{j=1}^{\infty} j^{-1-\alpha} \leq C k^{-1+\alpha}.$$ 

Finally,

$$\int_{\pi/(2k)}^{\pi} z^{-\alpha} |\cos(kz)| \, dx \leq (\pi/2)^{-\alpha} \int_{\pi/(2k)}^{\pi} |\cos(kz)| \, dx = (\pi/2)^{-\alpha} k^{-1} \leq C k^{-1+\alpha}.$$ 

We have now established part (i) of Lemma 9.

To show Lemma 9(ii), note

$$E\{I_n(0)\} = (2\pi n)^{-1} \int_0^{\pi} K_n(\lambda) f(\lambda) \, d\lambda$$

$$\leq C n^{-1} \left[ \int_{|\lambda|<1/n} n|\lambda|^{-\alpha} \, d\lambda + \int_{|\lambda|>1/n} n^{-\alpha} K_n(\lambda) \, d\lambda \right]$$

$$\leq C n^{-\alpha},$$

where we used above: $K_n$ is nonnegative, $\int \mathbb{1} K_n(\lambda) \, d\lambda = 1$, and $|K_n(\lambda)| \leq C n$.

And

$$E\{I_n(\pi)\} = (2\pi n)^{-1} \int_0^{\pi} K_n(\lambda - \pi) f(\lambda) \, d\lambda$$

$$\leq n^{-1} C \left[ \int_{|\lambda|<\pi/2} n|\lambda|^{-\alpha} \, d\lambda + \int_{|\lambda|>\pi/2} (\pi/2)^{-\alpha} K_n(\lambda - \pi) \, d\lambda \right]$$

$$\leq C.$$ 

We use $K_n(\lambda) \leq C n^{-1} L_{n0}^2(\lambda)$ and the growth rate of $g$ to show

$$|c_n g(0)| \leq \int \mathbb{1} K_n(\lambda) |g(\lambda)| \, d\lambda \leq C n|\lambda|^{\beta} + \int_{|\lambda|>1/n} C n^{-1} |\lambda|^{-2+\beta} \, d\lambda \leq C n^{-\beta},$$

which establishes Lemma 9(iii). □

**Lemma 10** Suppose Assumptions A.1-A.3 and A.5 hold with respect to a real-valued, even $g \equiv g_0$.

Then for $E_{n_f} = (2\pi n) \sum_{j=1}^{N} g(\lambda_j) f(\lambda_j)$,

$$\sqrt{n} \left| E_{n_f} - \int_0^\pi g f \, d\lambda \right| = o(1).$$

**proof:** If $fg$ is of bounded variation, $|E_{n_f} - \int_0^\pi g f \, d\lambda| = O(n^{-1})$ [cf. Brillinger (1981)].

If $fg$ is piecewise Lipschitz of order $\gamma > 1/2$ on $[0, \pi]$, then we may write $0 = a_0 < a_1 < \ldots < a_d < a_{d+1} = \pi$ where $x, y \in (a_i, a_{i+1})$ implies $|g(x) - g(y)| \leq C|x - y|^\gamma$. Say $n$ is large enough that
\( \pi/2 < \min(a_{i+1} - a_i) \) and hence there exists \( j_i \in \{1, \ldots, N\} \), \( \lambda_{j_i} \leq a_i < \lambda_{j_i+1} \) for each \( i = 1, \ldots, d \).

Then,
\[
\left| E_{nf} - \int_0^\pi gf \, d\lambda \right| = \left| E_{nf} - \sum_{j=1}^{N-1} \int_{\lambda_j}^{\lambda_{j+1}} gf \, d\lambda - \int_{\lambda_N}^\pi gf \, d\lambda - \int_0^{\lambda_1} gf \, d\lambda \right|
\leq C \left\{ n^{-\gamma} + \pi - \lambda_N + n^{-1+\alpha-\beta} + \sum_{i=1}^d \left( n^{-1} \lambda_i^{1-\alpha} + \int_{\lambda_i}^{\lambda_{i+1}} \lambda_i^{\beta-\alpha} \, d\lambda \right) \right\}
\leq C \left\{ n^{-\gamma} + n^{-1} n^{\max\{0,\alpha-\beta\}} \right\} = o(n^{-1/2})
\]

using \( \alpha - \beta < 1/2 \) and \( \pi - \lambda_N \leq \pi/n \).

If \( \partial f/\partial x \) exists on \((0, \pi]\),
\[
\left| E_{nf} - \int_0^\pi gf \, d\lambda \right| \leq \left| \sum_{j=1}^{N-1} \left\{ \frac{2\pi}{n} g(\lambda_j) f(\lambda_j) - f(c_j) \int_{\lambda_j}^{\lambda_{j+1}} g \, d\lambda \right\} \right| + C \left( n^{-1} + n^{-1+\alpha-\beta} \right)
\leq u_n + \tilde{u}_n + o(n^{-1/2}),
\]

where \( c_j \in [\lambda_j, \lambda_{j+1}] \) using the continuity of \( f \) and the mean value theorem;
\[
u_n = \frac{2\pi}{n} \sum_{j=1}^{N-1} |f(\lambda_j) - f(c_j)||g(\lambda_j)| \leq C n^{-1} \sum_{j=1}^{N-1} |\lambda_j - c_j| \lambda_j^{1-\alpha} \lambda_j^\beta
\leq C n^{-1+\max\{0,\alpha-\beta\}} \sum_{j=1}^{n} j^{-1}
(87)
\]

while: if \( g \) satisfies Condition 1 of Assumption A.3 for some \( \gamma > 1/2 \):
\[
\tilde{u}_n = \sum_{j=1}^{N-1} f(c_j) \int_{\lambda_j}^{\lambda_{j+1}} g(\lambda) \, d\lambda \leq C n^{-1} \left( \frac{1}{n} \sum_{j=1}^{N-1} \lambda_j^{-\alpha} \right) = o(n^{-1/2});
\]

if \( g \) meets Condition 2, then as in (87)
\[
\tilde{u}_n \leq C n^{-2} \sum_{j=1}^{N-1} \lambda_j^{-\alpha-1+\beta} \leq C n^{-1+\max\{0,\alpha-\beta\}} \sum_{j=1}^{n} j^{-1} = o(n^{-1/2});
\]

under Condition 3, \( g \) is bounded variation so that \( g(x) = h_1(x) - h_2(x) + a \) on \([0, \pi]\) for \( h_i(x) \) nonnegative and nondecreasing [cf. Royden (1988)] and \( |f(c_j)| \leq C n^a \) with \( \alpha < 1/2 \) so that
\[
\tilde{u}_n \leq C n^a \sum_{i=1}^{N-2} \sum_{j=1}^{2\pi/n} (h_i(\lambda_{j+1}) - h_i(\lambda_j)) \leq C n^{a-1} \sum_{i=1}^{2} \left( h_i(\pi) - h_i(0) \right) = o(n^{-1}).
\]

The proof of Lemma 10 is now finished. \( \square \)
Lemma 11 Suppose Assumptions A.1-A.2 hold for real-valued, even \( g = g_0 \) which satisfies Condition 2 of Assumption A.3. Then,

\[
\sqrt{n} \left| \int_0^x g_{\text{nc}} \, d\lambda - \int_0^x g \, d\lambda \right| = o(1).
\]

**proof:** Note \( g, f \) and \( I_{\text{nc}} \) are even functions. As in (36) and using \( \int_\Pi K_n \, d\lambda = 1 \),

\[
\sqrt{n} \left| \int_\Pi g_{\text{nc}} \, d\lambda - \int_\Pi g \, d\lambda \right| = \sqrt{n} \left| \int_\Pi K_n(\lambda - y) f(y) (g(\lambda) - g(y)) \, dy \, d\lambda \right|
\]

\[
\leq \sqrt{n} \int_\Pi K_n(\lambda - y) f(y) |g(\lambda) - g(y)| \, dy \, d\lambda \tag{88}
\]

\[
\leq C n^{3/2} \int_{(0, \pi)^2} \frac{|f(y)||g(\lambda) - g(y)|}{(1 + |(\lambda - y) \mod 2\pi n|^2)} + \frac{|f(y)||g(\lambda) - g(y)|}{(1 + |(\lambda + y) \mod 2\pi n|^2)} \, dy \, d\lambda
\]

where the second to last inequality follows from (25) and the last from \(|(\lambda + y) \mod 2\pi| \geq |\lambda - y|\), \( \lambda, y \in (0, \pi] \).

We now modify the argument in Giraitis and Surgailis (1990), p. 99. If \( 0 < y < \lambda \leq \pi \), \( |g(\lambda) - g(y)| \leq C \lambda^{-1} |\lambda - y| \). Since \( g \) is continuous, we pick some \( 0 < \rho < 1/2 \) and write

\[
f(y)||g(\lambda) - g(y)| \leq C f(y)||g(\lambda) - g(y)||^{1-\rho} \leq C \lambda^{-1+\rho'} |\lambda - y|^{1-\rho}, \quad \rho' = \rho + \beta(1 - \rho) - \alpha.
\]

Likewise, if \( 0 < \lambda < y \leq \pi \), \( f(y)||g(\lambda) - g(y)| \leq C \lambda^{-1+\rho'} |\lambda - y|^{1-\rho}. \) Note \( \rho' < 1 \) but we require that \( \rho' > 0 \) or equivalently, that \( \rho' > (\alpha - \beta)/(1 - \beta) \). If \( \alpha - \beta > 0 \), we need \( 1/2 > (\alpha - \beta)/(1 - \beta) \) (which holds by Condition 2) to find a \( \rho < 1/2 \) and \( 0 < \rho' < 1 \). We will also make use of the fact that there exists \( C > 0 \) whereby

\[
\int_0^\infty \frac{y^{-1+\rho'}}{(1 + |\lambda - y||n|^2)} \, dy \leq C |\lambda|^{-1+\rho'}, \quad \lambda \in \mathbb{R} \quad (0 < \rho' < 1).
\]

Then,

\[
n^{3/2} \int_{(0, \pi)^2} f(y)||g(\lambda) - g(y)|| \, dy \, d\lambda \leq n^{-1/2-\rho' + \rho} \int_{(0, \pi)^2} \frac{y^{-1+\rho'}|\lambda - y|^{1-\rho}}{(1 + |\lambda - y||n|^2)} \, dy \, d\lambda
\]

\[
\leq n^{-1/2-\rho' + \rho} \int_0^{\pi} \left( \int_0^\infty \frac{y^{-1+\rho'}}{(1 + |\lambda - y|^{1+\rho})} \, dy \right) \, d\lambda
\]

\[
\leq C n^{-1/2-\rho' + \rho} \int_0^{\pi} \lambda^{-1+\rho'} \, d\lambda
\]

\[
\leq C n^{-1/2+\rho} = o(1).
\]

The proof of Lemma 11 is now finished. ☐
Lemma 12 Suppose Assumptions A.1, A.2 hold for a real-valued, even \( g = g_0 \) which satisfies Condition 2 of A.3. Then,

\[
\sqrt{n} \left| \int_0^\pi c_n g(\lambda) I_{nc}(\lambda) \, d\lambda - \int_0^\pi g(\lambda) I_{nc}(\lambda) \, d\lambda \right| = o_p(1), \tag{89}
\]

\[
\sqrt{n} \frac{2\pi}{n} \sum_{j=1}^N g(\lambda_j) I_{nc}(\lambda_j) - \int_0^\pi c_n g(\lambda) I_{nc}(\lambda) \, d\lambda \right| = o_p(1). \tag{90}
\]

Proof: We begin showing (89), starting with

\[
2E \left| \int_0^\pi [c_n g(\lambda) - g(\lambda)] I_{nc}(\lambda) \, d\lambda \right| = E \left| \int I_{nc}(\lambda) \left[ \int K_n(\lambda - y) (g(y) - g(\lambda)) \, dy \right] \, d\lambda \right|
\]

\[
\leq E \int I_{nc}(\lambda) \left[ \int K_n(\lambda - y) |g(y) - g(\lambda)| \, dy \right] \, d\lambda
\]

\[
= \int \left[ \int K_n(\lambda - w) f(w) \, dw \right] \left[ \int K_n(\lambda - y) |g(y) - g(\lambda)| \, dy \right] \, d\lambda \equiv t_n,
\]

where we apply the evenness of \( c_n, g, I_{nc} \), the properties of \( K_n \), and (35). It suffices to show \( \sqrt{n} t_n = o(1) \).

With Fubini's theorem,

\[
\sqrt{n} t_n = \sqrt{n} \int \int K_n(\lambda - w) K_n(\lambda - y) f(w) g(y) - g(\lambda) \, dw \, dy \, d\lambda
\]

\[
\leq \sqrt{n} \int \int K_n(\lambda - w) K_n(\lambda - y) f(w) \{|g(y) - g(\lambda)| + |g(\lambda) - g(\lambda)|\} \, dw \, dy \, d\lambda
\]

\[
\leq \sqrt{n} \int \int K_n(\lambda - w) f(w) |g(y) - g(\lambda)| \, dw \, d\lambda
\]

\[
+ \sqrt{n} \int \int K_n(\lambda - w) K_n(y - \lambda) f(w) |g(y) - g(\lambda)| \, d\lambda \, dw \, dy
\]

\[
= \sqrt{n} (t_{1n} + t_{2n})
\]

where we used above \( K_n \) is even, \( \int_\Omega K_n(\lambda - y) \, dy = 1 \). We have already shown that \( \sqrt{n} t_{1n} = o(1) \), following the arguments beginning from (88). We need now only focus on \( t_{2n} \). Applying (2), Lemma 1(v), and (25) sequentially, we get:

\[
\int \int K_n(\lambda - w) K_n(\lambda - y) \, d\lambda \leq C n^{-2} \int \int L_{20}^2(\lambda - w) L_{20}^2(y - \lambda) \, d\lambda
\]

\[
\leq C n^{-1} L_{20}^2(y - w) \leq \frac{C n}{(1 + |(y - w) \mod 2\pi n|^2)}.
\]
Hence,
\[ \sqrt{n} t_{2n} \leq Cn^{3/2} \int_{\mathbb{T}} \frac{f(w)|g(y) - g(w)|}{(1 + |(y - w) \mod 2\pi|n)^2} \, dw \, dy \]
\[ \leq Cn^{3/2} \int_{(0,\pi]^2} \frac{f(w)|g(y) - g(w)|}{(1 + |(y - w) \mod 2\pi|n)^2} + \frac{f(w)|g(y) - g(w)|}{(1 + |(y + w) \mod 2\pi|n)^2} \, dw \, dy \]
\[ = o(1), \]

which we established in Lemma 11 starting with (88). This shows (89).

To establish the validity of (90), note that by (23) and (33),
\[ 2\sqrt{n} \left| \frac{2\pi}{n} \sum_{j=1}^{N} g(\lambda_j)I_n(\lambda_j) - \int_0^\pi c_n g(\lambda)I_{ne}(\lambda) \, d\lambda \right| \leq \tilde{\tau}_n + \tilde{\tau}_n; \]
\[ \tau_n = \frac{4\pi}{\sqrt{n}} \sum_{j=1}^{N} I_n(\lambda_j)|c_n g(\lambda_j) - g(\lambda_j)|, \quad \tilde{\tau}_n = \frac{2\pi}{\sqrt{n}} \left( |c_n g(0)|I_{ne}(0) + |c_n g(\pi)|I_{ne}(\pi) \right). \]

We wish to show that \( \tau_n, \tilde{\tau}_n = o_p(1) \). We easily have
\[ E(\tilde{\tau}_n) \leq C n^{-1/2}[n^{a-\beta} + 1] = o(1) \quad \Rightarrow \quad \tilde{\tau}_n = o_p(1), \]
using Lemma 9(ii) and (iii) and \( \|c_n g\|_\infty \leq \|g\|_\infty < \infty \). To show that \( \tau_n = o_p(1) \), it suffices to show:
1 \( \leq j \leq N, \quad n \geq 3 \]
\[ |c_n g(\lambda_j) - g(\lambda_j)| \leq C \lambda_j^d \left( j^{-1} \ln(n) + a_{jn} \right), \quad a_{jn} = \begin{cases} 0 & j \leq n/4 \\ (n - 2j)^{-1} & j > n/4, \end{cases} \]

where \( C \) is independent of \( j, n \); this is because, from (91), we have
\[ \tau_n \leq C \frac{\ln(n)}{n^{1/2}} \left( \max_{1 \leq j \leq N} I_n(\lambda_j)\lambda_j^d \right) \left( \sum_{j=1}^{n} j^{-1} \right) \leq \frac{C \ln^2(n)}{n^{1/2}} \left( \max_{1 \leq j \leq N} I_n(\lambda_j)\lambda_j^d \right) = o_p(1), \]
by Lemma 7. (Note as well that \( n - 2j \geq n - 2N \geq 1 \) for \( j > n/4 \).)

We now prove (91). First we decompose the difference: \( 1 \leq j \leq N, \)
\[ |c_n g(\lambda_j) - g(\lambda_j)| = \left| \int_{\mathbb{T}} K_n(y)(g(\lambda_j + y) - g(\lambda_j)) \, dy \right| \leq \left| \int_0^{\pi} d_{jn}(y) \, dy \right| + \left| \int_0^{\pi} d_{jn}^*(y) \, dy \right|, \]
\[ d_{jn}(y) = k_n(y)[g(\lambda_j + y) - g(\lambda_j)], \quad d_{jn}^*(y) = k_n(y)[g(\lambda_j - y) - g(\lambda_j)]; \]
we then handle each absolute integral. We begin with \( d_{jn} \) and note
\[ |g(\lambda_j + y) - g(\lambda_j)| \leq C \lambda_j^{-1+\beta} |y|, \quad 0 < y < \pi - \lambda_j \]
so that

\[ \left| \int_0^{1/n} d_j(y) \, dy \right| \leq C \lambda_j^{-1+\beta} \int_0^{1/n} L_{no}(y) \, dy \leq C \lambda_j^{-1+\beta} n^{-1} \leq C j^{-1} \lambda_j^3 \]

\[ \left| \int_{1/n}^{\pi-\lambda_j} d_j(y) \, dy \right| \leq C \lambda_j^{-1+\beta} \int_{1/n}^{\pi-\lambda_j} \frac{ny}{(1+n^2y^2)} \, dy \]

\[ = C \lambda_j^{-1+\beta} \frac{1}{n} \int \left[ \frac{-u}{1+u} + \ln(u+1) \right]_{1}^{(\pi-\lambda_j)n} \]

\[ \leq C n^{-1} \ln(n) \lambda_j^{-1+\beta} \]

\[ \leq C j^{-1} \ln(n) \lambda_j^3, \]

by (24) and (25). (Note \(n(\pi-\lambda_j) \geq 1\).) If \(\lambda_j \leq \pi/2\), then by (25)

\[ \left| \int_{\pi-\lambda_j}^{\pi} d_j(y) \, dy \right| \leq C n^{-1} L_0^2(\pi/2) \int_{\pi-\lambda_j}^{\pi} \|g\|_\infty \, dy \leq C n^{-1} \lambda_j, \]

using above \(\pi/2 < y \leq \pi\). If \(\lambda_j > \pi/2\), then by the symmetric, periodic extension of \(g\):

\[ \left| \int_{\pi-\lambda_j}^{2\pi-2\lambda_j} d_j(y) \, dy \right| = \left| \int_{\pi-\lambda_j}^{\pi} K_n(y) [g(\lambda_j + y - 2\pi) - g(\lambda_j)] \, dy \right| \]

\[ = \left| \int_{\pi-\lambda_j}^{\pi} K_n(2\pi - t - \lambda_j) [g(t) - g(\lambda_j)] \, dt \right| \quad (t = 2\pi - \lambda_j - y) \]

\[ \leq C \lambda_j^3 \int_{\pi-\lambda_j}^{\pi} \frac{(t-\lambda_j)\ln n}{(1+n(2\pi-2\lambda_j)-u)^2} \, du \quad (u = (t-\lambda_j)n) \]

\[ \leq C \lambda_j^3 \frac{1}{n} \int_{1}^{(\pi-\lambda_j)n} \frac{u}{(1+n(2\pi-2\lambda_j)-u)} \, du \leq \frac{C n^{-1} \ln(n) \lambda_j^3}{(1+n(2\pi-2\lambda_j)-u)} \]

\[ \leq C n^{-1} \lambda_j^{-1+\beta} \lambda_j^3, \]

using (25) for the first inequality; using a substitution \(t = 2\pi - \lambda_j - y\) and (25),

\[ \left| \int_{2\pi-2\lambda_j}^{\pi} d_j(y) \, dy \right| \leq \left| \int_{\pi-\lambda_j}^{\lambda_j} K_n(2\pi - t - \lambda_j)2\|g\|_\infty \, dy \right| \]

\[ \leq C \int_{\pi-\lambda_j}^{\lambda_j} \frac{n}{(1+(2\pi-\lambda_j-t)n)^2} \, dt \]

\[ \leq C \left[ (1+(2\pi-\lambda_j)n + z)^{-1} \left( 1+ \binom{2\lambda_j}{z} n^{n} \right)^{-1} \right]_{0}^{(\lambda_j - t)n} \quad (z = (\lambda_j - t)n) \]

\[ \leq C (n-2j)^{-1+\beta} \lambda_j^3 \quad (2\lambda_j/\pi \geq 1), \]

where \([n/4]+1 \leq j \leq N\) so that \(1 \leq n-2j \leq N\). We have now shown that the bound in (91) applies to \(\int_0^\pi d_j \, dy\).

Now consider \(d_j^n\). Note that

\[ |g(\lambda_j - y) - g(\lambda_j)| \leq C(\lambda_j - y)^{-1+\beta} \lambda_j^3, \quad 0 < y < \lambda_j \]
and hence by (24)
\[
\int_0^{1/n} d^r_{n}(y) \, dy \leq C \int_0^{1/n} (\lambda_j - y)^{-1+\beta} \, dy \leq C n^{-1} \lambda_j^{-1+\beta} \leq C j^{-1} \lambda_j^\beta,
\]
\[
\int_{1/n}^{\lambda_j/2} d^r_{n}(y) \, dy \leq C \frac{\int_{1/n}^{\lambda_j/2} y^{-1} (\lambda_j - y)^{-1+\beta} \, dy}{n} \leq C \frac{\ln(n) \lambda_j^{-1+\beta}}{n} \leq C \ln(n) j^{-1} \lambda_j^\beta,
\]
\[
\int_{\lambda_j/2}^{\lambda_j} d^r_{n}(y) \, dy \leq \frac{C L_{n0}^2(\lambda_j/2)}{n} \int_{\lambda_j/2}^{\lambda_j} y (\lambda_j - y)^{-1+\beta} \, dy
\]
\[
\leq \frac{C}{n \lambda_j} \int_{\lambda_j/2}^{\lambda_j} (\lambda_j - y)^{-1+\beta} \, dy \leq C j^{-1} \lambda_j^\beta \quad \text{(for } \beta > 0),
\]
\[
\int_{\lambda_j/2}^{\lambda_j} d^r_{n}(y) \, dy \leq C n^{-1} \|g\|_{\infty} (\lambda_j/2) L_{n0}^1(\lambda_j/2) \leq C j^{-1} \lambda_j^\beta \quad \text{(for } \beta = 0).
\]
If \( \lambda_j \leq \pi/2 \), then with (24):
\[
\int_{\lambda_j}^{2\lambda_j} d^r_{n}(y) \, dy \leq \frac{C L_{n0}^2(\lambda_j)}{n} \int_{\lambda_j}^{2\lambda_j} |g(y - \lambda_j) - g(\lambda_j)| \, dy \quad \left( \leq C j^{-1} \lambda_j^\beta, \text{ if } \beta = 0 \right)
\]
\[
\leq \frac{C}{n \lambda_j^2} \int_{\lambda_j}^{2\lambda_j} (2\lambda_j - y) (\lambda_j - y)^{-1+\beta} \, dy
\]
\[
= \frac{C}{n \lambda_j^2 \beta} \left[ (2\lambda_j - y) (y - \lambda_j)^\beta + (y - \lambda_j)^\beta + (1 + \beta)^{-1} \right]_{\lambda_j}^{2\lambda_j}
\]
\[
\leq \frac{C \lambda_j^\beta}{n \lambda_j} \leq C j^{-1} \lambda_j^\beta;
\]
\[
\int_{2\lambda_j}^{\pi} d^r_{n}(y) \, dy \leq \frac{C}{n} \int_{2\lambda_j}^{\pi} y^{-2} |g(y - \lambda_j) - g(\lambda_j)| \, dy
\]
\[
\leq \frac{C \lambda_j^{-1+\beta}}{n} \int_{2\lambda_j}^{\pi} y^{-2} (y - 2\lambda_j) \, dy
\]
\[
\leq C j^{-1} \lambda_j^\beta \left[ -y^{-1} (y - 2\lambda_j) + \ln(y) \right]_{2\lambda_j}^{\pi} \leq C \ln(n) j^{-1} \lambda_j^\beta,
\]
using \( |g(y - \lambda_j) - g(\lambda_j)| \leq C \lambda_j^{-1+\beta} (y - 2\lambda_j) \) for \( 2\lambda_j < y < \pi \). If \( \lambda_j > \pi/2 \) (so that \( \pi - \lambda_j \leq \lambda_j \)), then
\[
\int_{\lambda_j}^{\pi} d^r_{n}(y) \, dy \leq C n^{-1} L_{n0}^2(\lambda_j) \|g\|_{\infty} (\pi - \lambda_j) \leq C j^{-1} \lambda_j^\beta, \quad \text{(for } \beta = 0)
\]
\[
\int_{\lambda_j}^{\pi} d^r_{n}(y) \, dy \leq \frac{C L_{n0}^2(\lambda_j)}{n} \int_{2\lambda_j}^{\pi} (2\lambda_j - y) (\lambda_j - y)^{-1+\beta} \, dy \quad \text{(for } \beta > 0)
\]
\[
\leq \frac{C}{n \lambda_j^2 \beta} \left[ (2\lambda_j - y) (y - \lambda_j)^\beta + (y - \lambda_j)^\beta + (1 + \beta)^{-1} \right]_{\lambda_j}^{\pi}
\]
\[
\leq \frac{C \lambda_j^\beta}{n \lambda_j} \leq C j^{-1} \lambda_j^\beta,
\]
using $|g(y - \lambda_j) - g(\lambda_j)| \leq C(2\lambda_j - y)(y - \lambda_j)^{-1+\beta}$ for $2\lambda_j < y < \pi, \beta > 0$. We have now that $\int_0^\pi d\eta_n dy |$ conforms to the bound in (91) as well. Hence, we are finished with the proof of Lemma 12. \hfill \Box

**Lemma 13** With Assumption A.1, suppose $g, w$ are real-valued, even Riemann integrable functions on $\Pi$ such that $|g(\lambda)|, |w(\lambda)| \leq C|\lambda|^\beta, 0 \leq \beta < 1, \alpha - \beta < 1/2$. Then,

$$v_n = \text{Var}\left(\frac{2\pi}{n} \sum_{j=1}^N g(\lambda_j)w(\lambda_j) f_n^2(\lambda_j)\right) = o(1).$$

**proof:** Expand $v_n$ as

$$v_n = \left(\frac{2\pi}{n}\right)^2 \sum_{j=1}^N g^2(\lambda_j)w^2(\lambda_j)cum\left(f_n^2(\lambda_j), f_n^2(\lambda_j)\right) + \left(\frac{2\pi}{n}\right)^2 \sum_{1 \leq i < j \leq N} 2g(\lambda_i)w(\lambda_i)g(\lambda_j)w(\lambda_j)cum\left(f_n^2(\lambda_i), f_n^2(\lambda_j)\right)$$

$$|v_n| \leq C n^{-6} \sum_{j=1}^N \sum_{d}^4 cum\left(|d_n(\lambda_j)|^4, |d_n(\lambda_j)|^4\right) + C n^{-6} \sum_{1 \leq i < j \leq N} (\lambda_i, \lambda_j)^{2g} |cum\left(|d_n(\lambda_i)|^4, |d_n(\lambda_j)|^4\right)| \equiv v_{1n} + v_{2n},$$

using (23) above and $|d_n(\lambda)|^2 = d_n(\lambda)d_n(-\lambda)$. Define $\mathcal{P}$ to be the set of all indecomposable partitions of the labels in the two row table

$$\begin{array}{cccc}
a_{11} & a_{12} & a_{13} & a_{14} \\
a_{21} & a_{22} & a_{23} & a_{24}
\end{array}$$

[cf. Brillinger (1981), Section 2.3]. We write the elements of a partition $P = (P_1, \ldots, P_r), 1 \leq r \leq 7$. For $1 \leq i < j \leq N$, define $a_s^{ij}, s = 1, 2, 1 \leq t \leq 4$ so that

$$a_{s1}^{ij} = a_{s2}^{ij} = -a_{s3}^{ij} = -a_{s4}^{ij} = \lambda_i, \quad a_{s1}^{ij} = a_{s2}^{ij} = -a_{s3}^{ij} = -a_{s4}^{ij} = \lambda_j.$$  

By the product theorem for cumulants [Brillinger (1981), Theorem 2.3.2],

$$cum\left(|d_n(\lambda_i)|^4, |d_n(\lambda_j)|^4\right) = \sum_{P=(P_1, \ldots, P_r) \in \mathcal{P}} \text{cum}_{ij}(P), \quad cum_{ij}(P) = \prod_{u=1}^r \text{cum}(d_{nc}(a_u^{ij}), a_{st} \in P_u).$$

Because $E(d_{nc}(\lambda)) = 0$, we may WLOG only consider

$$\mathcal{P}' \equiv \{ P = (P_1, \ldots, P_r) \in \mathcal{P} : 1 < |P_1| \leq \cdots \leq |P_r|, 1 \leq r \leq 6\},$$

where $|A|$ denotes the size of a finite set $A$; that is, consider only partitions in $\mathcal{P}$ where each set in the partition has 2 or more elements $a_{st}$. Then,

$$v_{1n} \leq C \sum_{P \in \mathcal{P}'} G_{in}(P), \quad i = 1, 2;$$
\[ G_{1n}(P) = n^{-6} \sum_{j=1}^{N} (\lambda_j)^{d} |\text{cum}_{ijn}(P)|, \quad G_{2n}(P) = n^{-6} \sum_{1 \leq i < j \leq N} (\lambda_i \lambda_j)^{2\beta} |\text{cum}_{ijn}(P)|. \]

To show \( v_{in} = o(1) \), it suffices to show
\[ G_{in}(P) = o(1), \quad P \in P', \quad i = 1, 2. \] (92)

We start with \( v_{1n} \). By Lemma 1, we have \(|\text{cum}_{ijn}(P)| \leq C n^4 \lambda_j^{-1+\alpha}, P \in P' \) so that
\[ G_{1n}(P) \leq C n^{-3} \sum_{j=1}^{N} (\lambda_j)^{d-4\alpha} \leq C n^{\max(0,2\alpha-2\beta)-1} \left( \frac{1}{n} \sum_{j=1}^{N} \lambda_j^{2\beta-2\alpha} \right) = o(1)O(1) = o(1), \]
since \( \alpha - \beta < 1/2 \). Hence, (92) is established for \( G_{1n} \) and \( v_{1n} = o(1) \).

We now consider \( G_{2n}(P) \) and several cases of \( P = (P_1, \ldots, P_r) \in P \) to show (92) holds under each:

**case 1:** \( r = 1 \)

**case 4:** \( r = 2, |P_2| = 4 \)

**case 2:** \( r = 2, |P_2| = 6 \)

**case 5:** \( r = 3, |P_3| = |P_2| = 3 \)

**case 3:** \( r = 2, |P_2| = 5 \)

**case 6:** \( r = 3, |P_3| = 4, |P_2| = 2 \)

**case 7:** \( r = 4, |P_1| = |P_2| = |P_3| = |P_4| = 2 \) for which we have the following subcases:

7.1: There exists \( k \neq k' \) where \( \sum_{a_{it} \in P_k} a_{ij}^{it} = 0 = \sum_{a_{it} \in P_{k'}} a_{ij}^{it} \).

7.2: There exists exactly one \( k \) where \( \sum_{a_{it} \in P_k} a_{it}^{ij} = 0 \).

7.3: For each \( m, \sum_{a_{it} \in P_k} a_{it}^{ij} \neq 0; \) and \( |\sum_{a_{it} \in P_k} a_{it}^{ij}| = 2\lambda_i, |\sum_{a_{it} \in P_{k'}} a_{it}^{ij}| = 2\lambda_j \) for some \( k, k' \).

7.4: For each \( m, |\sum_{a_{it} \in P_k} a_{it}^{ij}| \notin \{0, 2\lambda_i, 2\lambda_j\} \).

We handle each case separately and assume throughout that \( j > i \) and \( P = (P_1, \ldots, P_r) \).

**case 1:** Here \( |P_1| = 8 \). By Lemma 1(ii) and (24), \(|\text{cum}_{ijn}(P)| \leq C \{ n^{3/2} + n \ln^7(n) \} (\lambda_i \lambda_j)^{-2\alpha} \) so
\[ G_{2n}(P) \leq C n^{-3/2} \left( \frac{1}{n} \sum_{j=1}^{N} \lambda_j^{2\beta-2\alpha} \right)^2 = o(1). \]

**case 2:** If \( \sum_{a_{it} \in P_k} a_{it}^{ij} \neq 0, \) then the same bound in case 1 applies and \( G_{2n}(P) = o(1) \). If \( \sum_{a_{it} \in P_k} a_{it}^{ij} = 0 \), we can bound \(|\text{cum}(d_n(a_{it}^{ij}) : a_{it} \in P_k)| \leq C n \lambda_i^{-\alpha} \) or \( C n \lambda_j^{-\alpha} \) by Lemma 1(i) and (24) and obtain \(|\text{cum}_{ijn}(P)| \leq C \{ n^{3/2} + n^2 \ln^5(n) \} (\lambda_i \lambda_j)^{-2\alpha} \) with Lemma 1; hence,
\[ G_{2n}(P) \leq C n^{-3/2} \left( \frac{1}{n} \sum_{j=1}^{N} \lambda_j^{2\beta-2\alpha} \right)^2 = o(1). \]

**case 3:** Apply Lemma 1(ii) and (24) twice, \(|\text{cum}_{ijn}(P)| \leq C \{ n^{3/2} + n \ln^4(n) \}^2 (\lambda_i \lambda_j)^{-2\alpha} \) and
\[ G_{2n}(P) \leq C n^{-1} \left( \frac{1}{n} \sum_{j=1}^{N} \lambda_j^{2\beta-2\alpha} \right)^2 = o(1). \]
case 4: The bound and result in case 3 applies.

case 5: If $\alpha > 0$, then by considering all possible $\{a_{st}^{ij} : a_{st} \in P_k\}$ and using Lemma 1,

$$|\text{cum}_{ijn}(P)| \leq C n (\lambda_i^{-\alpha} n^{3/2 - \alpha} + n \ln^2(n)) \{n(3-\alpha)/2 + n \ln^2(n)\} \lambda_i^{-\alpha} \lambda_j^{-3\alpha/2};$$

$$G_{2n}(P) \leq C n^{-1/2} \ln^2(n) \left( \frac{1}{n} \sum_{j=1}^{N} \lambda_j^{2\beta - 2\alpha} \right) \left( \frac{1}{n} \sum_{j=1}^{N} \lambda_j^{\beta - \alpha} \right) + C n^{-\alpha} \left( \frac{1}{n} \sum_{j=1}^{N} \lambda_j^{2\beta - 2\alpha} \right)^2 = o(1).$$

If $\alpha = 0$, $|\text{cum}(d_{nc}(a_{st}^{ij}) : a_{st} \in P_k)| \leq C (\lambda_i^{-3/2} + n \ln^2(n))$ for $k = 2, 3$, and $|\text{cum}(d_{nc}(a_{st}^{ij}) : a_{st} \in P_k)| \leq C n$ so that

$$G_{2n}(P) = C n^{-1} \ln^4(n) \left( \frac{1}{n} \sum_{j=1}^{N} \lambda_j^2 \right)^2 + C n^{-2} \sum_{i,j=1}^{N} i^{-3} = o(1).$$

case 6: Note $\sum_{a_{st} \in P_k} I_{\{|a_{st}^{ij}| = \lambda_k\}} \geq 1$ for $k = i$ or $j$. If $\{|a_{st}^{ij}| : a_{st} \in P_k\} = \{\lambda_j\}$ for $k = 1$ or 2 (and there can be at most one such $k$), then by Lemma 1(i)

$$\prod_{k=1}^{2} |\text{cum}(d_{nc}(a_{st}^{ij}) : a_{st} \in P_k)| \leq C n^2 (\lambda_i \lambda_j)^{-\alpha},$$

$$|\text{cum}(d_{nc}(a_{st}^{ij}) : a_{st} \in P_k)| \leq C \{\lambda_j^{\alpha/2 - 1} \lambda_i^{-1/2} + n \ln^3(n)\} \lambda_j^{-3\alpha/2} \lambda_i^{-\alpha/2};$$

if not, then $\sum_{a_{st} \in P_k} I_{\{|a_{st}^{ij}| = \lambda_k\}} \geq 2$ so that, for $k = 1, 2$, $|\text{cum}(d_{nc}(a_{st}^{ij}) : a_{st} \in P_k)| \leq C n \lambda_i^{-\alpha}$ and $|\text{cum}(d_{nc}(a_{st}^{ij}) : a_{st} \in P_k)| \leq C \{\lambda_j^{\alpha/2 - 1} \lambda_i^{-1/2} + n \ln^3(n)\} (\lambda_i, \lambda_j)^{-\alpha}$. Hence,

$$|\text{cum}_{ijn}(P)| \leq C n^2 \lambda_i^{\alpha/2 - 1} \lambda_j^{-1/2} + n \ln^3(n) \lambda_i^{-3\alpha/2} \lambda_j^{-\alpha},$$

$$G_{2n}(P) \leq C n^{-1/2} \left( \frac{1}{n} \sum_{j=1}^{N} \lambda_j^{2\beta - 2\alpha} \right)^2 + n^2 \ln^3(n) \left( \frac{1}{n} \sum_{j=1}^{N} \lambda_j^{2\beta - 2\alpha} \right) \left( \frac{1}{n} \sum_{j=1}^{N} \lambda_j^{\beta - \alpha} \right) = o(1).$$

Before proceeding to the treatment of case 7, we define a few sets to ease our upcoming summation notation. For $0 < \rho < \pi/2$, let

$$A^0 = \{(j, i) : 1 \leq i < j \leq N, \lambda_j \geq \rho\},$$

$$A^\rho = \{(j, i) : 1 \leq i < j \leq N, \lambda_j < \rho\}$$

$$A_{\rho_1} = \{(j, i) : 1 \leq i < j \leq N, \lambda_j < \rho, j > i^2/(i - 1) \text{ or } i = 1\},$$

$$A_{\rho_2} = \{(j, i) : 2 \leq i < j \leq N, \lambda_j < \rho, j \leq i^2/(i - 1)\},$$

$$A^{n/4} = \{(j, i) : 1 \leq i < j < N, i + j > n/4\},$$

$$A_{n/4} = \{(j, i) : 1 \leq i < j < N, i + j \leq n/4\},$$

We will use $A_{\rho_1}, A_{\rho_2}, A^{n/4}, A_{n/4}$ only to handle subcase 7.1.
subcase 7.1: WLOG say $k = 1$, $k' = 2$. We have $|\sum_{a_{ij}\in F_k} a_{ij}^t| = |\sum_{a_{ij}\in F_k} a_{ij}^t| = \sum_{j} \lambda_{ij}$, and, by Lemma 2(i), $\prod_{i=1}^m |\text{cum}(d_{\text{rev}}(a_{ij}^t)) : a_{ij} \in F_k| \leq C n^2 (\lambda_i \lambda_j)^{\alpha}$. If $|\sum_{a_{ij}\in F_m} a_{ij}^t| = \sum_{j} \lambda_{ij} - \lambda_i$, $m = 3, 4$, then by Lemma 3, (24), and Lemma 2(i) (for the sum involving $\lambda_i$ or $\lambda_j \geq \rho$) and Lemma 1(i), (24), and Lemma 2 (iii) (for the sum with $\lambda_j < \rho$)

\[
G_2n(P) \leq \frac{C(\rho)}{n^4} \sum_{\lambda_i, \lambda_j, \lambda_k < \rho} \lambda_i^{\beta - \alpha} \left[ \frac{n}{j-i} + |R_{n\rho}| \right]^2 + \frac{C}{n^4} \sum_{\lambda_i, \lambda_j} \lambda_i^{\beta - 3\alpha} \lambda_j^{\beta - \alpha} \left[ \frac{1}{j-i} + \frac{n}{(j-i)^2} \right]^2
\]

\[
eq g_1n(\rho) + g_2n(\rho),
\]

\[
g_1(\rho) \leq C(n^{-2} \left( \sum_{1 \leq i < j \leq N} \lambda_i^{\beta - \alpha} (j-i)^{-2} + \frac{(n^{-1} R_{n\rho})^2}{(j-i)^2} \sum_{1 \leq i < j \leq N} \lambda_i^{\beta - \alpha} \right) \leq C(\rho) \left( n^{-1 + \max(0, \alpha - \beta)} \sum_{j=1}^N j^{-2} + (n^{-1} R_{n\rho})^2 \left( \frac{1}{n} \sum_{j=1}^N \lambda_j^{\beta - \alpha} \right) \right) = o(1).
\]

Using Lemma 2(iii) on the sums over $\lambda_i A$ and $\lambda_j B$:

\[
g_2n(\rho) \leq C(\rho) \left( \frac{1}{n} \sum_{\lambda \leq \lambda_i < \rho} \lambda_i^{\beta - 2\alpha} \right)^2 + Cn^{-2} \sum_{\lambda_i A} \left( \lambda_i \lambda_j \right)^{\beta - 2\alpha} \left( j - i \right)^{\beta - \alpha} + Cn^{-2} \sum_{\lambda_i A} \left( \lambda_i \lambda_j \right)^{\beta - 2\alpha} \left( j - i \right)^{\beta - \alpha}
\]

\[
\leq C(n^{-1} \sum_{\lambda \leq \lambda_i < \rho} \lambda_i^{\beta - 2\alpha})^2.
\]

Then,

\[
\frac{1}{\rho} n \int_0^\rho \lambda^{\beta - 2\alpha} \frac{1}{(j-i)^2} = C \rho^{1+2\beta - 2\alpha} \rho (1 + 2\beta - 2\alpha)
\]

for a $C$ that does not depend on $0 < \rho < \pi/2$. Hence, $G_2n(P) = o(1)$.

If $|\sum_{a_{ij}\in F_m} a_{ij}^t| = \lambda_i + \lambda_j, m = 3, 4$, with the same steps as above:

\[
G_2n(P) \leq \frac{C(\rho)}{n^4} \sum_{\lambda_i, \lambda_j, \lambda_k < \rho} \lambda_i^{\beta - \alpha} |R_{n\rho}| + |R_{n\rho}|^2 + \frac{C}{n^4} \sum_{\lambda_i, \lambda_j} \lambda_i^{\beta - 3\alpha} \lambda_j^{\beta - \alpha} \left[ \frac{1}{j-i} + \frac{n}{(j-i)^2} \right]^2
\]

\[
eq g_1n(\rho) + g_2n(\rho),
\]

Using Lemma 2(i),

\[
g_2n(\rho) \leq C_{\rho} n^{-2} \left( \sum_{\lambda_i A} \lambda_i^{\beta - \alpha} (n-j-i)^2 + \sum_{\lambda_i A} \lambda_i^{\beta - \alpha} \left( \frac{1}{n} \sum_{j=1}^N \lambda_j^{\beta - \alpha} \right) \right) + C_{\rho} n^{-1 + \max(0, \alpha - \beta)} \sum_{j=1}^N j^{-2} + (n^{-1} R_{n\rho})^2 \left( \frac{1}{n} \sum_{j=1}^N \lambda_j^{\beta - \alpha} \right) = o(1).
\]

Note $g_4n(\rho) \leq g_2n(\rho)$ so that $\lim (g_2n(\rho) + g_4n(\rho)) \leq C \rho^{1+2\beta - 2\alpha}$ for $0 < \rho < \pi/2$ and again $G_2n(P) = o(1)$.

subcase 7.2: WLOG suppose $k = 1$. There exists $m, m'$ where $|\sum_{a_{ij}\in F_m} a_{ij}^t| = \lambda_j + \lambda_i, |\sum_{a_{ij}\in F_m'} a_{ij}^t| = \lambda_j + \lambda_i$.
\[ \lambda_j - \lambda_i. \] Then by applying Lemma 3, (24), Lemma 1(i), and Lemma 2(ii) for a selected \( 0 < \rho < \pi/2: \]

\[
\begin{align*}
G_{2n}(P) &\leq \frac{C(\rho)}{n^4} \sum_{A_{\sigma}} \lambda_{\sigma}^{d-\alpha} \left[ L_{n0}(\lambda_i + \lambda_j) + |R_{n0}| \right] \left[ \frac{n}{j-i} + |R_{n0}| \right] \\
&\quad + \frac{C}{n^4} \sum_{A_{\sigma}} (\lambda_i \lambda_j)^{2d-\alpha} \lambda_i^{2-\alpha} \left[ \lambda_j^{-1} + L_{n1}(\lambda_i + \lambda_j) \right] \left[ \lambda_j^{-1} + L_{n1}(\lambda_i - \lambda_j) \right] \\
&\equiv g_{n}(\rho) + g_{2n}(\rho).
\end{align*}
\]

By Lemma 2(ii)

\[
\begin{align*}
g_{n}(\rho) &\leq C n^{-4} \sum_{A_{\sigma}} \lambda_{\sigma}^{d-3\alpha} \lambda_j^{d-\alpha} \left[ \lambda_j^{-1} + \frac{n}{(j-i)^2} \right] = C g_{2n}(\rho), \\
g_{2n}(\rho) &\leq \frac{C(\rho)}{n^2} \left( \sum_{A_{\sigma}/\tau} \lambda_{\sigma}^{d-\alpha} \left[ \frac{1}{n-j-i} + n^{-1}|R_{n0}| \right] \right) \\
&\quad + \sum_{A_{\sigma}/\tau} \frac{\lambda_{\sigma}^{d-\alpha}}{i-j} \left[ \frac{1}{n-j-i} + n^{-1}|R_{n0}| \right] \\
&\quad + (n^{-1}|R_{n0}|) \left( \frac{1}{n} \sum_{j=1}^{N} \lambda_j^{d-\alpha} \right) \\
&\leq C(\rho) \left( n^{-1+\max(0,\alpha-\delta)} \sum_{j=1}^{n} j^{-2} + (n^{-1}|R_{n0}|) \left( \frac{1}{n} \sum_{j=1}^{N} \lambda_j^{d-\alpha} \right) \right) = o(1).
\end{align*}
\]

Hence, \( \lim (g_{n}(\rho) + g_{2n}(\rho)) \leq C \rho^{1+2\delta-2\alpha} \) so that \( G_{2n}(P) = o(1) \) follows.

subcase 7.3: There exists some \( m \) such that \( |\sum_{a_{st} \in F_k} a_{st}^{ij}| = \lambda_j - \lambda_i \) or \( \lambda_j + \lambda_i \). Note \( |\text{cum}(d_{nc}(a_{st})^{ij}) : a_{st} \in F_k) = C n \lambda_j^{-\alpha}, \text{cum}(d_{nc}(a_{st})^{ij}) : a_{st} \in F_k)| \leq C n \lambda_j^{-\alpha} \) by Lemma 1(i).

Suppose first that \( |\sum_{a_{st} \in F_k} a_{st}^{ij}| = \lambda_j - \lambda_i \). Again pick \( 0 < \rho < \pi/2 \). Then by Lemma 3, (24), Lemma 1(i), and Lemma 2(ii): 

\[
\begin{align*}
G_{2n}(P) &\leq \frac{C(\rho)}{n^5} \sum_{A_{\sigma}} \lambda_{\sigma}^{d-\alpha} \left[ n + |R_{n0}| \right] \left[ \frac{n}{j-i} + |R_{n0}| \right]^2 \\
&\quad + \frac{C}{n^4} \sum_{A_{\sigma}} (\lambda_i \lambda_j)^{2d-\alpha} \lambda_i^{2-\alpha} \left[ \lambda_j^{-1} + \frac{n}{(j-i)^2} \right]^2 \\
&\equiv g_{n}(\rho) + g_{2n}(\rho).
\end{align*}
\]

Then, we have that \( g_{n} \leq C(\rho)(1 + n^{-1}|R_{n0}|)g_{2n}(\rho) \) and \( g_{2n}(\rho) = C g_{2n}(\rho) \) implying further that \( \lim (g_{n}(\rho) + g_{2n}(\rho)) \leq C \rho^{1+2\delta-2\alpha} \). Since \( \rho \) is arbitrary, \( G_{2n}(P) = o(1) \) follows.

If \( |\sum_{a_{st} \in F_k} a_{st}^{ij}| = \lambda_j + \lambda_i \), then repeating essentially the same steps as above 

\[
\begin{align*}
G_{2n}(P) &\leq \frac{C(\rho)}{n^5} \sum_{A_{\sigma}} \lambda_{\sigma}^{d-\alpha} \left[ n + |R_{n0}| \right] \left[ L_{n0}(\lambda_i + \lambda_j) + |R_{n0}| \right]^2 \\
&\quad + \frac{C}{n^4} \sum_{A_{\sigma}} (\lambda_i \lambda_j)^{2d-\alpha} \lambda_i^{2-\alpha} \left[ \lambda_j^{-1} + \frac{n}{(j-i)^2} \right]^2 \\
&\equiv g_{n}(\rho) + g_{2n}(\rho).
\end{align*}
\]
Then, $g_{3n}(\rho) \leq C_\rho (1 + n^{-1}|R_{n\rho}|)g_{3n}(\rho)$ and $g_{10n}(\rho) = C g_{4n}(\rho)$ so that we again have $\lim (g_{3n}(\rho) + g_{10n}(\rho)) \leq C \rho^{1+2\delta-2\alpha}$ for any $0 < \rho < \pi/2$ and $G_{2n}(P) = o(1)$ follows.

**subcase 7.4:** There are 3 possible partitions $P$: for each $m$, $|\sum_{a_i \in P^m} a_{4i}^2| = \lambda_j + \lambda_i$; for each $m$, $|\sum_{a_i \in P^m} a_{4i}^2| = \lambda_j - \lambda_i$; or there exist $m, m'$ where $|\sum_{a_i \in P^m} a_{4i}^2| = \lambda_j + \lambda_i, |\sum_{a_i \in P^{m'}} a_{4i}^2| = \lambda_j - \lambda_i$.

Fix $0 < \rho < \pi/2$.

With the first possibility, we use Lemma 1(i), (24), Lemma 3, and Lemma 2(ii) to find

$$G_{2n}(P) \leq \frac{C(\rho)}{n^\delta \lambda_j} \sum_{A^2} \lambda_i^2 \left[ \frac{n}{j-i} + |R_{n\rho}| \right] \left[ \frac{n}{j-i} + |R_{n\rho}| \right]^{\frac{d}{4} - 2\delta} \lambda_i^{-4\alpha} \lambda_j^{-1}$$

$s1n(p) + g_{10n}(\rho)$.

Then, we can write $g_{11n}(\rho) \leq C(\rho) [1 + n^{-1}|R_{n\rho}|]^{3}\sum_{n} g_{3n}(\rho)$ since $\{\lambda_j/(2\pi)\}^{-\alpha} \geq 1$. Note that $n^{-2} \lambda_i^{-\alpha}[\lambda_j^{-1} + n(j + i)^{-d}]^2 \lambda_j^{-\alpha} [j^{-1} + j^{-d}] \leq 2 \lambda_j^{-\alpha}$, using $-d + \alpha < 0$. Hence, we have $g_{12n}(\rho) \leq C g_{4n}(\rho)$.

Consider now the next possible partition, $|\sum_{a_i \in P^m} a_{4i}^2| = \lambda_j - \lambda_i$ for each $m$. We again use Lemma 1(i), (24), Lemma 3, and Lemma 2(ii) to find

$$G_{2n}(P) \leq \frac{C(\rho)}{n^\delta \lambda_j} \sum_{A^2} \lambda_i^2 \left[ \frac{n}{j-i} + |R_{n\rho}| \right] \left[ \frac{n}{j-i} + |R_{n\rho}| \right]^{\frac{d}{4} - 2\delta} \lambda_i^{-4\alpha} \lambda_j^{-1}$$

$s1n(p) + g_{14n}(\rho)$.

By Lemma 2(iii), $n^{-2} \lambda_i^{-\alpha}[\lambda_j^{-1} + n(j - i)^{-d}]^2 \lambda_j^{-\alpha} [j^{-1} + j^{-d}] \leq \lambda_j^{-\alpha} + 2 \lambda_j^{-\alpha}$ so that $g_{14n}(\rho) \leq C g_{2n}(\rho)$. Also, $g_{13n}(\rho) \leq C_\rho [1 + n^{-1}|R_{n\rho}|]^{3} g_{1n}(\rho)$.

In the last possibility, the sets in the partition $P$ correspond to $\{\lambda_i, -\lambda_j\}, \{\lambda_i, -\lambda_j\}, \{-\lambda_i, \lambda_j\}, \{-\lambda_i, \lambda_j\}$. Again we use Lemma 1(i), (24), Lemma 3, and Lemma 2(ii) to write

$$G_{2n}(P) \leq \frac{C(\rho)}{n^\delta \lambda_j} \sum_{A^2} \lambda_i^2 \left[ \frac{n}{j-i} + |R_{n\rho}| \right] \left[ \frac{n}{j-i} + |R_{n\rho}| \right]^{\frac{d}{4} - 2\delta} \lambda_i^{-4\alpha} \lambda_j^{-1}$$

$s1n(p) + g_{16n}(\rho)$.

Then, $g_{15n}(\rho) \leq C_\rho [1 + n^{-1}|R_{n\rho}|]^{3} g_{5n}(\rho)$ using again $\{\lambda_i/(2\pi)\}^{-\alpha} \geq 1$. It follows from the preceding use of Lemma 2(iii) that

$$n^{-2} \lambda_i^{-\alpha}[\lambda_j^{-1} + n(j + i)^{-d}] [\lambda_j^{-1} + n(j - i)^{-d}] \leq 3 \lambda_j^{-\alpha},$$

so that we have $g_{16n}(\rho) \leq C g_{5n}(\rho)$.

Now for any $0 < \rho < \pi/2$,

$$\lim \left( \sum_{k=11}^{18} g_{kn}(\rho) \right) \leq C \rho^{1+2\delta-2\alpha}.$$
so that for any of the 3 possible partitions $P$ allowed under the considered subcase 7.4, we have $G_{2n}(p) = o(1)$.

We have now established (92) and the proof of Lemma 13 is complete. □

We use the following lemma in the proofs of Lemma 8 and Theorems 2-3.

**Lemma 14** Let $0 < \delta < 1$, $0 \leq \beta < 1$ such that $\alpha + \delta < 1$ and $\alpha - \beta < 1/2$. Let $m = \max\{1/3, 1/4 + (\alpha - \beta)/2, (\alpha + \delta + 1)/4\}$. Under Assumption A.1,

(i) $n^{-1} \sum_{i=1}^{N} \lambda_{i}^{-2d} I_{n}(\lambda_{i}) = O_{p}(1)$,

(ii) $n^{-m-1} \sum_{i=1}^{N} \lambda_{i}^{d-\delta} \{I_{n}^{2}(\lambda_{i}) + I_{n}(\lambda_{i})\} = O_{p}(n^{-(1-2m)})$,

(iii) $n^{-2m-1} \sum_{i=1}^{N} \lambda_{i}^{d-\delta-2\alpha} = O(n^{3m})$, $\sum_{i=1}^{N} \lambda_{i}^{-(d-\delta-2\alpha)} = o(n^{4m})$.

**proof:** By (23), Lemma 1, (24), and (42), we have $E(I_{n}(\lambda_{i})) \leq C\lambda_{i}^{-\alpha}$, $E(I_{n}^{2}(\lambda_{i})) \leq C\lambda_{i}^{-2\alpha}$ for $1 \leq i \leq N$. Lemma 14(i) then follows from $n^{-1} \sum_{i=1}^{N} \lambda_{i}^{-d-\alpha} = O(1)$; and to show Lemma 14(ii), it suffices to establish Lemma 14(iii).

Let $s_{1n} = \sum_{i=1}^{N} \lambda_{i}^{d-\delta-2\alpha}$, $s_{2n} = \sum_{i=1}^{N} \lambda_{i}^{d-\delta-2\alpha}$. If $\beta - \delta - 2\alpha > -1$, $s_{1n}^{*} = O(n)$ and $3m \geq 1$; if $\beta - \delta - 2\alpha = -1$, then $s_{1n}^{*} = O(n \ln(n))$ and $m > 1/3$; if $\beta - \delta - 2\alpha < -1$, then $s_{1n}^{*} = O(n^{2\alpha + \delta - \beta})$ and $m \geq \alpha - \beta$, $2m \geq \alpha + \delta$. Hence, we have $s_{1n}^{*} = O(n^{3m})$.

If $2\delta + 2\alpha \leq 1$, then $s_{2n}^{*} = O(n \ln(n))$ and $4m > 1$; if $2\delta + 2\alpha > 1$, then $s_{2n}^{*} = O(n^{2\delta + 2\alpha})$ and $2m > \alpha + \delta$. Hence, $s_{2n}^{*} = o(n^{4m})$.

The proof of Lemma 14 is now complete. □

**References**


