UNIVERSITY OF
MANNHEIM

University of Mannheim / Department of Economics

Working Paper Series

Baxter`s Inequality and Sieve Bootstrap for Random Fields

Marco Meyer Carsten Jentsch Jens-Peter Kreiss

Working Paper 15-06

May 2015

BAXTER'S INEQUALITY AND SIEVE BOOTSTRAP FOR RANDOM FIELDS

MARCO MEYER, CARSTEN JENTSCH, AND JENS-PETER KREISS

Abstract. The concept of the autoregressive (AR) sieve bootstrap is investigated for the case of spatial processes in $\mathbb{Z}^2$. This procedure fits AR models of increasing order to the given data and, via resampling of the residuals, generates bootstrap replicates of the sample. The paper explores the range of validity of this resampling procedure and provides a general check criterion which allows to decide whether the AR sieve bootstrap asymptotically works for a specific statistic of interest or not. The criterion may be applied to a large class of stationary spatial processes. As another major contribution of this paper, a weighted Baxter-inequality for spatial processes is provided. This result yields a rate of convergence for the finite predictor coefficients, i.e. the coefficients of finite-order AR model fits, towards the autoregressive coefficients which are inherent to the underlying process under mild conditions.

The developed check criterion is applied to some particularly interesting statistics like sample autocorrelations and standardized sample variograms. A simulation study shows that the procedure performs very well compared to normal approximations as well as block bootstrap methods in finite samples.

1. Introduction

We consider stationary real-valued spatial processes in the plane $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ with zero mean and finite second moments. By imposing only very mild regularity conditions on the processes the framework of this paper remains very general. Particularly, without making any parametric/linearity assumptions on the process $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$, we are interested in fitting spatial autoregressive models of the form

$$X_{\underline{t}} = \sum_{\underline{k}\in\Theta(p)} a_{\underline{k}}(p)\, X_{\underline{t}-\underline{k}} + e_{\underline{t}} \tag{1.1}$$

to data, where $\Theta(p)$ denotes some suitable finite index set and $(e_{\underline{t}})$ is some white noise process. In few words, this paper has two main purposes: Firstly, we will show that models of the form (1.1) are well-suited to describe the behaviour of very general stationary spatial processes since a very large class of these processes possesses an inherent autoregressive structure. This structure can be approximated well by models such as (1.1), which will be shown by proving a generalization of Baxter's

Date: May 6, 2015.

2000 *Mathematics Subject Classification.* Primary 62M10, 62M15; secondary 62G09.

Key words and phrases. Autoregression, bootstrap, random fields.

inequality, cf. Baxter (1962), to the spatial setting. As a second major contribution of this paper, the concept of the autoregressive sieve bootstrap scheme will be transferred to the case of spatial processes. In the following, the aforementioned purposes will be explained in more detail.

By classical results going back to the work of Whittle (1954), general spatial processes in the plane $\mathbb{Z}^2$ always possess half-plane representations with respect to *each half-plane* of $\mathbb{Z}^2$ that might be chosen, as long as mild assumptions are fulfilled. More precisely, there exist one-sided autoregressive (AR) as well as moving-average (MA) representations

$$X_{\underline{t}} = \sum_{\underline{k}\in\Theta} a_{\underline{k}}\, X_{\underline{t}-\underline{k}} + \varepsilon_{\underline{t}}, \quad \text{and} \quad X_{\underline{t}} = \sum_{\underline{k}\in\Theta} b_{\underline{k}}\, \varepsilon_{\underline{t}-\underline{k}} + \varepsilon_{\underline{t}} \tag{1.2}$$

with respect to some (weak) white noise process $(\varepsilon_{\underline{t}})$, where Θ can be *any* half-plane in the sense of Guyon (1995). Throughout this paper we stick to the so-called *lower half-plane* representation corresponding to lexicographical ordering of the plane $\mathbb{Z}^2$ as described by Helson and Lowdenslager (1958), among others. It is important to note that choosing this particular half-plane representation is *not* restrictive at all because any other choice of the half-plane would be fine as well. It has to be understood as a suitable vehicle to establish meaningful theory in this paper and the lower half-plane is just chosen for notational convenience. During the course of this paper, we will also clarify a common misunderstanding in the discussion of spatial and time series autoregressions, that should at least be mentioned briefly at this point: It is often criticized that, for spatial processes, one has to choose a concept of 'past' values for one-sided autoregressions, i.e. choose a direction from which the random variables $X_{\underline{t}}$ are influenced. This choice is of course arbitrary. Hence, one might come to the conclusion that the whole concept of one-sided autoregressions implies a very specific model assumption which is not fulfilled for real-world data. However, the opposite is true since our assumptions do not constrain the class of processes any further than demanding the spectral density to be positive and smooth.

In contrast to our framework, most of the existing literature on autoregressive modeling in the plane is heavily based on the assumption that the underlying spatial process actually fulfills some specific model structure. Autoregressive processes in the plane have been pioneered in Whittle (1954), where unilateral and bilateral autoregressive models are studied. Correlation properties of these processes have been studied in Besag (1972) and for some special cases in Basu and Reinsel (1993). Spatial autoregressive processes with a 'quarter-plane past' form a popular sub-class of unilateral processes in the plane. These processes have been investigated in detail by Tjøstheim (1978), Tjøstheim (1981) and Tjøstheim (1983). However, although the class of spatial AR processes with a quarter-plane past appears to be appealing at first sight due to its simple structure, we still consider half-plane instead of quarter-plane representations in this paper. This is due to the fact that, under very mild assumptions, general spatial processes are *always* assured to have half-plane representations as in (1.2), which is in general not true for quarter-plane representations (at least not with uncorrelated innovations). Hence, imposing a quarter-plane past structure on the process $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ turns out to be very restrictive and is therefore omitted in this paper. The same is true for models considered in Choi and Politis (2007), who discuss the properties of models with several regions of support. Yule-Walker type estimation of spatial AR models has been investigated by Guyon (1982), Basu and Reinsel (1992) and Ha and Newton (1993), who particularly addressed an inaccuracy in Tjøstheim (1981).

The crucial property that spatial processes can always be represented as in (1.2) is also well-known for time series processes $(X_t)_{t\in\mathbb{Z}}$, cf. among others Pourahmadi (2001). Here, the AR representation corresponding to (1.2) reads

$$X_t = \sum_{k=1}^{\infty} a_k\, X_{t-k} + \varepsilon_t. \tag{1.3}$$

To deal with these infinite dimensional autoregressive representations in the time series case, the famous Baxter-inequality (cf. Baxter (1962) for univariate processes and Hannan and Deistler (1988) or Cheng and Pourahmadi (1993) for the multivariate case) plays a fundamental role and allows for meaningful asymptotic theory. When fitting AR models of finite order p to time series, for instance by Yule-Walker estimation, one typically estimates the so-called finite predictor coefficients $a_1(p), \ldots, a_p(p)$, which are simply the coefficients of the L^2-projection of X_t onto the finite past $span\{X_{t-1}, \ldots, X_{t-p}\}$. Baxter's inequality provides a connection between these finite predictor coefficients and the AR coefficients from (1.3) and reads as follows: Under mild smoothness conditions on the spectral density of the process, there exists a constant $C < \infty$ and $p_0 \in \mathbb{N}$ such that

$$\sum_{k=1}^{p} \nu(k)\, |a_k(p) - a_k| \leq C \cdot \sum_{k=p+1}^{\infty} \nu(k)\, |a_k|, \qquad \forall\, p \geq p_0. \tag{1.4}$$

Here, $\nu(\cdot)$ denotes a weight function which is connected to the smoothness condition on the spectral density. Notice that the right-hand side of (1.4) is finite and therefore converges to zero as $p \to \infty$. Hence, the left-hand side also vanishes for $p \to \infty$ which yields convergence for the predictors $a_k(p)$ towards the AR coefficients a_k. In fact, the weights $\nu(k)$ determine the rate of convergence. If this rate is fast enough, then even autoregressive fits of rather small order p are suitable to describe the process (X_t) properly. The goal in this paper is to derive a similar inequality for the AR fits of shape (1.1) in connection with representations (1.2).

The original proof of (1.4) for univariate time series is mainly based on the analytical result of Baxter (1963). One might think that the original proof of Baxter

(1963) transfers straightforwardly from time series to the spatial case, but this is not the case. Heuristically, this is due to the following observations. The proof of Baxter's inequality for time series is heavily based on the fact that by predicting X_t based on $span\{X_{t-1}, \ldots, X_{t-p}\}$, the two sets $\{s : s < t-p\}$ and $\{s : s \geq t\}$ can be separated arbitrarily far apart, for sufficiently large p, by the set $M(p) := \{s : t-p \leq s \leq t-1\}$. Thus, $|\mathrm{Cov}(X_r, X_q)|$ becomes arbitrarily small for sufficiently large p and for $X_r \in \{X_s, s < t-p\}$ and $X_q \in \{X_s, s \geq t\}$. For the spatial case such a separation is no longer possible as no finite subset analogous to $M(p)$ exists that is capable to separate $\mathbb{Z}^2$ in this fashion. As one major contribution of this paper, we come up with a different approach to prove a version of Baxter's inequality that is suitable for spatial processes. This result allows to derive rigorous asymptotic theory for AR fits of increasing order for spatial processes.

For time series, Baxter's inequality is a key ingredient when establishing validity of the AR sieve bootstrap scheme. This procedure was introduced for stationary univariate linear time series by Kreiss (1988), Kreiss (1992) and Bühlmann (1997) who established validity for different statistics including autocovariances and autocorrelations. The main contribution of the AR sieve methodology is to allow the autoregressive order $p = p(n)$ to increase with the sample size n. Thus, the AR sieve bootstrap extends the model-based (parametric) AR bootstrap – first considered by Freedman (1984) – to the much richer (nonparametric) class of AR(∞)-processes.

Paparoditis and Streitberg (1991) established asymptotic validity of the AR sieve bootstrap to infer properties of high order autocorrelations, and Paparoditis (1996) established its validity in a multivariate linear time series context. Furthermore, the AR sieve bootstrap is used for testing for unit roots in Chang and Park (2003) and Paparoditis and Politis (2005), and in econometrics literature for several purposes such as e.g. forecasting in Alonso, Pena and Romo (2002) or in the setup of time series panels in Smeekes and Urbain (2013).

However, while all the aforementioned results were derived under the explicit assumption of an underlying $AR(\infty)$ process, Kreiss, Paparoditis and Politis (2011) extended the range of applicability of the AR sieve significantly. Under very mild conditions and without having to assume any autoregressive structure of the underlying process, they were able to show that the AR sieve remains valid whenever the so-called companion process mimics the proper limiting distribution, which constitutes a simple and general check criterion. Recently, Meyer and Kreiss (2014+) extended the results of Kreiss, Paparoditis and Politis (2011) to the multivariate case. To generalize their concept, as a second main contribution of this paper, we introduce a spatial AR sieve methodology in the spirit of Kreiss, Paparoditis and Politis (2011) and provide rigorous theory.

The proposed AR sieve bootstrap performs favourably compared to block bootstrap techniques, as will be shown in a simulation study in this paper. Block bootstrap and subsampling for random fields were proposed by Hall (1985) and Künsch (1989), whereas Politis and Romano (1993) addressed block resampling schemes for general statistics. Zhu and Lahiri (2007) proved bootstrap consistency for the empirical process of a non-overlapping block bootstrap. Optimal block size and subsample size selection have been addressed in Nordman and Lahiri (2007) and Nordman and Lahiri (2004), respectively.

The remainder of this paper is organised as follows: In section 2 we will introduce the basic notations and definitions and formulate the algorithm of the AR sieve bootstrap procedure precisely. In addition, we will show how the rate of decay of the autocovariances of a spatial process carries over to its cepstral coefficients – the Fourier coefficients of the spectral density – and then to the AR coefficients.

In section 3, we will establish sufficiently fast convergence of the finite-order AR models that are fitted in the course of the sieve bootstrap procedure, to the aforementioned AR coefficients. Here, we will derive a generalisation of Baxter's inequality, cf. Baxter (1962), to the case of random fields. Beyond its application in connection with the AR sieve bootstrap, this result may be of its own interest.

The conditions for AR sieve bootstrap validity are given in section 4, and the result will be a check-criterion which allows to decide whether the procedure is asymptotically consistent or not; with the criterion being solely based on the asymptotics of the particular test statistic one is looking at. This result closely resembles the concept of the so-called companion process introduced by Kreiss, Paparoditis and Politis (2011). We will apply the derived check criterion in section 5 to some particularly interesting statistics, including variogram estimators. It follows a simulation study in section 6 which compares the performance of the AR sieve bootstrap to normal approximations and the block bootstrap. Section 7 contains the proofs of the two central theorems, Baxter's inequality and the result about bootstrap validity, while all other proofs of auxiliary results are deferred to section 8.

2. Preliminaries

Consider a stationary real-valued spatial process $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ with mean zero and finite second moments. In the following we will switch between the two equivalent notations $X_{\underline{t}} = X_{t_1,t_2}$. While the vector index notation $X_{\underline{t}}$ allows for a more compact presentation of the results, the notation X_{t_1,t_2} is sometimes necessary if we want to describe operations on the components of the index vector. For convenience reasons, we will also sometimes use a mixed notation, e.g. in expressions such as $\sum_{t_1 \in A} \sum_{t_2 \in B} X_{\underline{t}}$.

The autocovariance function of $(X_{\underline{t}})$ at lag $\underline{h} = (h_1, h_2)^T$ is denoted by $\gamma(\underline{h}) = E(X_{\underline{t}+\underline{h}} X_{\underline{t}})$. We assume to have a square-shaped data sample $\{X_{\underline{t}} : 1 \leq t_1, t_2 \leq n\}$ consisting of n^2 observations at hand. Define $\Pi := \{\underline{t} \in \mathbb{Z}^2 : \ 1 \leq t_1, t_2 \leq n\}$ and $\Pi_{\underline{h}} := \{\underline{t} \in \mathbb{Z}^2 : \ 1 \leq t_1, t_2, t_1 + h_1, t_2 + h_2 \leq n\}$; i.e. $\Pi_{\underline{h}}$ describes the set of vectors $\underline{t} \in \mathbb{Z}^2$ such that both $\underline{t}$ and $\underline{t} + \underline{h}$ are elements of Π. The empirical autocovariance function can then be stated as

$$\widehat{\gamma}(\underline{h}) := \frac{1}{|\Pi_{\underline{h}}|} \sum_{\underline{t} \in \Pi_{\underline{h}}} (X_{\underline{t}+\underline{h}} - \overline{X})(X_{\underline{t}} - \overline{X}) \tag{2.1}$$

where $\overline{X} = n^{-2} \sum_{\underline{t} \in \Pi} X_{\underline{t}}$ denotes the sample mean.

We now turn to the algorithm of the autoregressive sieve bootstrap for random fields. Our proposal depends on fitting an autoregressive model of finite order $p \in \mathbb{N}$ to the data. Since it is not obvious how such an AR fit would look like in the spatial setting, we first define the following set of vectors in $\mathbb{Z}^2$ which characterises the collection of sites for the p-th order AR fit:

$$\Theta(p) := \{\underline{k} \in \mathbb{Z}^2 : \ (1 \leq k_1 \leq p \text{ and } k_2 = 0) \text{ or } (-p \leq k_1 \leq p \text{ and } 1 \leq k_2 \leq p)\}. \tag{2.2}$$

An autoregressive model with sites given by $\Theta(p)$ could be stated as

$$X_{\underline{t}} = \sum_{\underline{k} \in \Theta(p)} a_{\underline{k}} X_{\underline{t}-\underline{k}} + e_{\underline{t}} \tag{2.3}$$

for some white noise $(e_{\underline{t}})$. Figure 1 illustrates the shape of these types of AR models with an example of order $p = 3$; the index vectors $\underline{t} - \underline{k}$ from (2.3) are marked by the black dots while $\underline{t}$ can be found at the center. The AR model from (2.3) is one-sided in the sense of so-called lexicographical ordering of the plane $\mathbb{Z}^2$; we will discuss this property extensively further along the line in this section, but first formulate the AR sieve bootstrap algorithm.

Let $T_n = T_n(\{X_{\underline{t}} : \ \underline{t} \in \Pi\})$ be an estimator for some unknown parameter θ of the process, based on the given data sample. For an appropriately increasing sequence of real numbers $(c_n)_{n \in \mathbb{N}}$, we assume that the distributions $\mathcal{L}_n = \mathcal{L}(c_n(T_n - \theta))$ converge to a non-degenerated limiting distribution as $n \to \infty$. Our goal is to estimate the distribution $\mathcal{L}_n$ for some finite number $n \in \mathbb{N}$. We propose the following procedure:

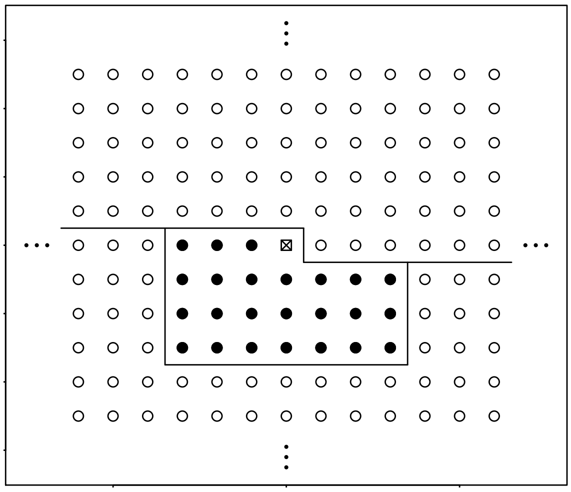

FIGURE 1. Illustration of the shape of an AR(3)-model with respect to $\Theta(3)$, cf. (2.3); locations of sites $\underline{t}-\underline{k}$ marked by the black dots.

The autoregressive sieve bootstrap algorithm for random fields:

(1) Select an order $p = p(n) \in \mathbb{N}$, $p \ll n$ and fit a p-th order autoregressive model of shape (2.3) to the given observations, for example by Yule-Walker estimation. Denote the estimated coefficients by $\{\widehat{a}_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$.

(2) Let $\Pi(n,p) := \{(t_1,t_2) \in \mathbb{Z}^2 : p+1 \le t_1 \le n-p,\ p+1 \le t_2 \le n\}$, i.e. $\Pi(n,p)$ is the set of all vectors $\underline{t} \in \Pi$ such that $(\underline{t}-\underline{k}) \in \Pi$ for all $\underline{k} \in \Theta(p)$. Denote the residuals of the autoregressive fit by $\varepsilon'_{\underline{t}}(p) = X_{\underline{t}} - \sum_{\underline{k}\in\Theta(p)} \widehat{a}_{\underline{k}}(p)\, X_{\underline{t}-\underline{k}}$ for all $\underline{t} \in \Pi(n,p)$, and let $\widehat{F}_n$ be the empirical distribution function of the centered residuals $\widehat{\varepsilon}_{\underline{t}}(p) = \varepsilon'_{\underline{t}}(p) - \overline{\varepsilon}$, where $\overline{\varepsilon} = (n-2p)^{-1}(n-p)^{-1}\sum_{\underline{t}\in\Pi(n,p)} \varepsilon'_{\underline{t}}(p)$. Generate independent random variables $\varepsilon^*_{\underline{j}}$ having identical distribution $\widehat{F}_n$, for example by drawing with replacement from the set of centered residuals. Use these resampled residuals and the parameter estimators to calculate a bootstrap sample $\{X^*_{\underline{t}} : \underline{t} \in \Pi\}$ according to the generating equation

$$X^*_{\underline{t}} = \sum_{\underline{k}\in\Theta(p)} \widehat{a}_{\underline{k}}(p)\, X^*_{\underline{t}-\underline{k}} + \varepsilon^*_{\underline{t}}. \tag{2.4}$$

(3) Let $T^*_{n,(1)} := T_n(\{X^*_{\underline{t}} : \underline{t} \in \Pi\})$ be the same estimator as T_n based on the pseudo sample $\{X^*_{\underline{t}} : \underline{t} \in \Pi\}$ and θ^* the analogue of θ associated with the bootstrap process $(X^*_{\underline{t}})$.

(4) Repeat steps (1)–(3) M times, where M is sufficiently large, in order to obtain independent realisations $T^*_{n,(1)}, \ldots, T^*_{n,(M)}$ of the plug-in estimator.

(5) The estimator for $\mathcal{L}_n$ is then given by the empirical distribution of $\mathcal{L}^*_n = \mathcal{L}^*(c_n(T^*_n - \theta^*))$, based on the observations $T^*_{n,(1)}, \ldots, T^*_{n,(M)}$.

Here, $\mathcal{L}^*$ and E^* denote probability law and expectation, conditional on the given data sample.

In the following, we will investigate under which conditions the underlying process $(X_{\underline{t}})$ possesses one-sided autoregressive representations, since this property is crucial for showing asymptotic validity of the AR sieve bootstrap. For the remainder of this chapter we will be working with spatial processes fulfilling the following assumptions. We use the notation $|\underline{k}|_\infty := \max\{|k_1|, |k_2|\}$ for the maximum vector norm of each $\underline{k} \in \mathbb{Z}^2$. For any arbitrary subset A of some vector space over $\mathbb{R}$ or $\mathbb{C}$, $\overline{sp}(A)$ denotes the closed span of all vectors $a \in A$.

Assumption 1. *Let $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ be a strictly stationary real-valued basic process, i.e. $X_{\underline{t}} \notin \overline{sp}\{X_{\underline{s}},\ \underline{s} \neq \underline{t}\}$, with mean zero and finite second moments. The autocovariance function $\gamma(\cdot)$ of $(X_{\underline{t}})$ fulfils $\sum_{\underline{k}\in\mathbb{Z}^2}(1 + |\underline{k}|_\infty)^r\, |\gamma(\underline{k})| < \infty$ for some $r \in \mathbb{N}_0$ to be specified in the respective results later on. The spectral density of $(X_{\underline{t}})$,*

$$f(\underline{\lambda}) = \frac{1}{4\pi^2} \sum_{\underline{k}\in\mathbb{Z}^2} \gamma(\underline{k})\, e^{-i\langle \underline{k},\underline{\lambda}\rangle}, \quad \underline{\lambda} \in (-\pi,\pi]^2,$$

fulfils the so-called boundedness condition: There exists a constant $c > 0$ such that $f(\underline{\lambda}) \geq c$ uniformly for all frequencies $\underline{\lambda} \in (-\pi,\pi]^2$.

Note that this assumption merely requires the spectral density to be positive and smooth, because the weighted summability condition on the autocovariances just implies that certain partial derivatives of f exist. For $u, v \in \mathbb{N}$ with $u + v \leq r$ we get from differentiating the Fourier series of f:

$$\frac{\partial^{u+v} f}{\partial\lambda_1^u\, \partial\lambda_2^v}(\underline{\lambda}) = \frac{1}{4\pi^2} \sum_{\underline{k}\in\mathbb{Z}^2} (-ik_1)^u (-ik_2)^v\, \gamma(\underline{k})\, e^{-i\langle \underline{k},\underline{\lambda}\rangle}.$$

The derivative of the Fourier series of f on the right-hand side of the latter equation is absolutely summable because $|(-ik_1)^u(-ik_2)^v| \leq (1 + |\underline{k}|_\infty)^r$ and because of Assumption 1. Therefore, the derivative of f itself, given by the left-hand side, exists and is equal to the derivative of the Fourier series.

We will now establish the aforementioned one-sided autoregressive and moving average representations for all processes that fulfil Assumption 1. Here, *one-sided* refers to the lexicographical ordering of the plane $\mathbb{Z}^2$, cf. Guyon (1995). Defining

$$\Theta := \{(k_1,k_2) \in \mathbb{Z}^2 :\ (k_1 \geq 1 \text{ and } k_2 = 0) \text{ or } (k_1 \text{ arbitrary and } k_2 \geq 1)\}$$

one can observe that $\mathbb{Z}^2$ can be partitioned as $\{\underline{0}\} \cup \Theta \cup (-\Theta)$. Θ is commonly referred to as the *upper half-plane* with respect to the origin while $-\Theta$ is the *lower half-plane*, cf. Helson and Lowdenslager (1958). An illustration is given by Figure 2; the upper half-plane Θ is given by the white dots, the lower half-plane by the black dots. Obviously, it holds $\Theta(p) \to \Theta$, as $p \to \infty$.

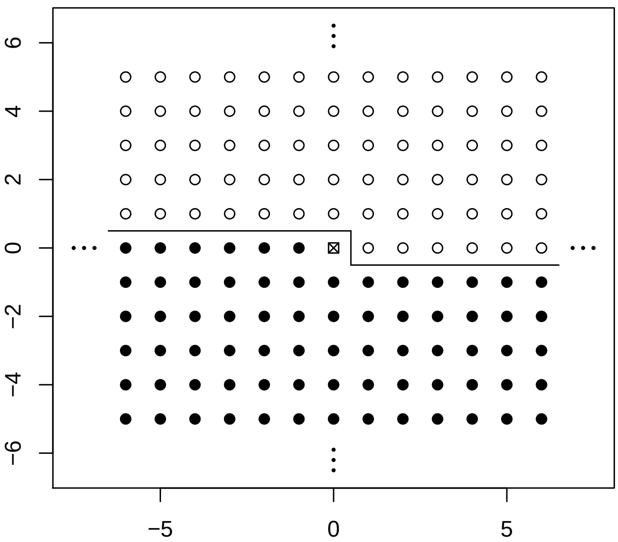

FIGURE 2. Illustration of the upper (white dots) and lower (black dots) half-plane of $\mathbb{Z}^2$.

We now get the following result on one-sided representations for spatial processes:

Lemma 2.1. *Let $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ be a spatial process that fulfils Assumption 1 with some $r \geq 1$. Then there exist uniquely determined autoregressive (AR) coefficients $(a_{\underline{k}})_{\underline{k}\in\Theta}$, uniquely determined moving average (MA) coefficients $(b_{\underline{k}})_{\underline{k}\in\Theta}$ and a uniquely determined uncorrelated white noise process $(\varepsilon_{\underline{t}})$, $\underline{t} \in \mathbb{Z}^2$, such that $(X_{\underline{t}})$ possesses the one-sided AR and MA representations*

$$X_{\underline{t}} = \sum_{\underline{k}\in\Theta} a_{\underline{k}}\, X_{\underline{t}-\underline{k}} + \varepsilon_{\underline{t}}, \quad X_{\underline{t}} = \sum_{\underline{k}\in\Theta} b_{\underline{k}}\, \varepsilon_{\underline{t}-\underline{k}} + \varepsilon_{\underline{t}}, \tag{2.5}$$

respectively, and $\sum_{\underline{k}\in\Theta} a_{\underline{k}}\, X_{\underline{t}-\underline{k}}$ represents the L^2-projection of $X_{\underline{t}}$ onto $\overline{sp}\{X_{\underline{t}-\underline{k}} : \underline{k} \in \Theta\}$. The white noise process $(\varepsilon_{\underline{t}})$ is called the innovation process of $(X_{\underline{t}})$. The coefficients in (2.5) *fulfil the summability conditions*

$$\sum_{\underline{k}\in\Theta} (1 + |\underline{k}|_\infty)^{r-1}\, |a_{\underline{k}}| < \infty, \quad \sum_{\underline{k}\in\Theta} (1 + |\underline{k}|_\infty)^{r-1}\, |b_{\underline{k}}| < \infty. \tag{2.6}$$

It should be noted that the existence of representations (2.5) has already been proven by Whittle (1954). However, we are especially interested in the summability conditions (2.6), which are not available in the literature. Hence, we derive these conditions in the proof of Lemma 2.1, which can be found in section 8.

Remark 2.2. At this point we should clarify a common misunderstanding in the discussion of spatial and time series autoregressions: For time series, the 'past' and the 'future' of a time value $t \in \mathbb{Z}$ are naturally defined, and it is generally accepted that random variables X_t are influenced by its past values $X_{t-1}, X_{t-2}, \ldots$. Since this is not the case for spatial processes, it is often criticized that one has to choose a concept of 'past' values, i.e. choose a direction from which the random variable $X_{\underline{t}}$ is influenced, such as the lower half-plane illustrated by Figure 2. This choice is of course arbitrary, which is why one might come to the conclusion that the whole concept of one-sided autoregressions implies a very specific model assumption which is not fulfilled for real-world data. However, the opposite is true: The AR sieve bootstrap, as an example, only uses the one-sided autoregressions as a vehicle in the proof of bootstrap validity. Under the mild conditions from Assumption 1, which only depend on the spectral density and which do not include any choice of direction whatsoever, the process $(X_{\underline{t}})$ possesses autoregressive representations with respect to *each half-plane* of $\mathbb{Z}^2$ that might be chosen. Therefore, the whole procedure is by no means arbitrary; and the concept of approximating a particular one-sided autogression does not constrain the class of processes any further than demanding the spectral density to be positive and smooth. □

In order to prove the summability conditions from Lemma 2.1 we need the following auxiliary result. The AR and MA coefficients are strongly connected to the so-called *cepstral coefficients* of the process, that is the Fourier coefficients of the logarithm of the spectral density. The following lemma provides a result that carries over the summability condition from the Fourier coefficients of a function f to the Fourier coefficients of its logarithm. The result holds not only for spectral densities but for arbitrary integrable functions, and seems not to be available in the literature so far, at least not in this explicit form.

Lemma 2.3. *Denote for every integrable function $f : (-\pi,\pi]^2 \to \mathbb{R}$ its Fourier coefficients by $\widetilde{f}_{\underline{k}} = (1/4\pi^2) \int_{(-\pi,\pi]^2} f(\underline{\lambda})\, e^{-i\langle \underline{k},\underline{\lambda}\rangle}\, d\underline{\lambda}$ and its formal Fourier series by $\sum_{\underline{k}\in\mathbb{Z}^2} \widetilde{f}_{\underline{k}}\, e^{i\langle \underline{k},\underline{\lambda}\rangle}$. We define the following classes of functions:*

$$\begin{aligned} C_r \;&:=\; \Big\{ f : (-\pi,\pi]^2 \to \mathbb{R},\ \|f\|_r := \sum_{\underline{k}\in\mathbb{Z}^2} (1+|\underline{k}|_\infty)^r\, |\widetilde{f}_{\underline{k}}| < \infty \Big\}, \\ D_{r_1,r_2} \;&:=\; \Big\{ f : (-\pi,\pi]^2 \to \mathbb{R},\ \|f\|_{r_1,r_2} := \sum_{\underline{k}\in\mathbb{Z}^2} (1+|k_1|)^{r_1} (1+|k_2|)^{r_2}\, |\widetilde{f}_{\underline{k}}| < \infty \Big\}. \end{aligned}$$

Assume that $f(\underline{\lambda}) \ge c > 0$ for all $\underline{\lambda} \in (-\pi,\pi]^2$. Then it holds:

(i) If $f \in C_r$ for some $r \ge 2$, it follows $\log f \in C_{r-1}$.
(ii) If $f \in D_{r_1,r_2}$ for some $r_1, r_2 \ge 1$, it follows $\log f \in D_{r_1,r_2}$.

Remark 2.4. In Assumption 1 and Lemma 2.3 (i) we use the weight function $\nu(\underline{k}) = (1+|\underline{k}|_\infty)^r$. This is due to the fact that we will later establish a weighted version of a Baxter-inequality for spatial processes, cf. Theorem 3.2. The proof of this Baxter-inequality requires the weights to be strictly non-decreasing in $|\underline{k}|_\infty$, i.e. $\nu(\underline{k}) \ge \nu(\underline{j})$ whenever $|\underline{k}|_\infty \ge |\underline{j}|_\infty$ or, in other words, whenever $\underline{j} \in \Theta(p)$ and $\underline{k} \in \Theta \setminus \Theta(p)$. Other weights one might think of, like replacing the $|\cdot|_\infty$-norm in $\nu(\underline{k})$

by the Euclidean norm, the 1-norm or letting $\widetilde{\nu}(\underline{k}) = (1 + |k_1|)^{r_1}(1 + |k_2|)^{r_2}$, do not fulfil the property of being strictly non-decreasing in $|\underline{k}|_\infty$ and are, therefore, not suitable in order to establish a weighted Baxter-inequality. However, for Assumption 1 to be fulfilled, it suffices to check whether $\sum_{\underline{k}\in\mathbb{Z}^2}(1+|\underline{k}|)^r\,|\gamma(\underline{k})| < \infty$ for any vector norm $|\underline{k}|$, since all vector norms are equivalent. One could also switch to any other vector norm $|\cdot|_\alpha$, but in this case the projection set $\Theta(p)$ has to be modified such that $|\underline{k}|_\alpha \geq |\underline{j}|_\alpha$ whenever $\underline{j} \in \Theta(p)$ and $\underline{k} \in \Theta \setminus \Theta(p)$. □

Remark 2.5. Classes of functions with weighted absolutely summable Fourier coefficients, such as C_r and D_{r_1,r_2} from Lemma 2.3, are commonly referred to as *Beurling algebras*; C_r represents the special case for the weight function $\nu(\underline{k}) = (1 + |\underline{k}|_\infty)^r$. Remark 2.4 explains why we are looking at these particular weights, although we get the somehow unsatisfactory result that $f \in C_r$ does *not imply* $\log f \in C_r$, but instead $\log f \in C_{r-1}$. While we will only work with assertion (i) from Lemma 2.3 for the remainder of this paper, it is still worthwile to consider the class D_{r_1,r_2} from (ii). Here, we get with analogous arguments as in (i) that $f \in D_{r_1,r_2}$ implies $\log f \in D_{r_1,r_2}$, i.e. the Fourier coefficients of $\log f$ fulfil the same summability condition as the ones of f. This result is strongly connected to the well-known Wiener-Lévy-Theorem (cf. Zygmund (2002), Chapter VI, Theorem 5.2); and, for the special case of $\phi(f) = \log f$, our result even represents a slight generalisation of the latter, with respect to functions in several variables. We will shed some light on this situation:

Originally, Norbert Wiener proved for functions in one variable that if $f \neq 0$ has absolutely summable Fourier coefficients, then the same holds true for $1/f$. This assertion, also known as Wiener's lemma, can be transferred to functions in several variables; and, moreover, weighted summability versions in the spirit of Lemma 2.3 are available, cf. Theorem 6.2 in Gröchenig (2007). For functions in one variable, Paul Lévy generalised Wiener's result, concluding that if f has absolutely summable Fourier coefficients, the same holds true for $\phi(f)$, where ϕ is a smooth functional. This assertion became known as the Wiener-Lévy-Theorem. In contrast to what happens for $\phi(f) = 1/f$, weighted versions in several variables are much harder to come by for general functions ϕ. Typically, one only gets that $\phi(f)$ is the element of a Beurling algebra with weights increasing at a slower rate than the ones of f, cf. Bhatt and Dedania (2003).

Our proof of Lemma 2.3 (ii) shows that a generalisation to functions in several variables for the special case of $\phi(f) = \log f$ is possible. However, the proof relies heavily on the structure of the logarithmic function and cannot be generalised to other functions. □

3. Convergence of finite-order model fits

In this section we will establish results that ensure convergence of the estimated parameters $\{\widehat{a}_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$ from step (1) of the AR sieve bootstrap procedure, cf. section 2, towards the autoregressive coefficients $\{a_{\underline{k}} : \underline{k} \in \Theta\}$ of the underlying process given by Lemma 2.1. We will split up the results in two subsections: The first one will be concerned with convergence of the finite predictor coefficients of the process $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ towards $\{a_{\underline{k}} : \underline{k} \in \Theta\}$. The finite predictors are the L^2-projection coefficients of random variable $X_{\underline{t}}$ to the finite-dimensional space $sp\{X_{\underline{t}-\underline{k}} : \underline{k} \in \Theta(p)\}$. Here, if A is an arbitrary subset of some vector space over $\mathbb{R}$ or $\mathbb{C}$, $sp(A)$ denotes the span of all vectors $a \in A$. In this context we will introduce a Baxter-inequality for spatial processes. Section 3.2 deals with conditions which ensure that the difference between the estimators $\{\widehat{a}_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$ and the finite predictor coefficients vanishes asymptotically in probability. The results from both subsections combined then yield the desired convergence of the finite-order AR model fits.

3.1. **Convergence of finite predictor coefficients.**

The finite predictor coefficients with respect to the set $\Theta(p)$ are the coefficients of the L^2-projection of $X_{\underline{t}}$ onto $sp\{X_{\underline{t}-\underline{k}} : \underline{k} \in \Theta(p)\}$, and will be denoted by $\{a_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$. They can be obtained from solving the minimization problem

$$\{a_{\underline{k}}(p) : \underline{k} \in \Theta(p)\} := \underset{\left\{c_{\underline{k}}(p) :\, \underline{k}\in\Theta(p)\right\}}{\arg\min}\; E\left(X_{\underline{t}} - \sum_{\underline{k}\in\Theta(p)} c_{\underline{k}}(p)\, X_{\underline{t}-\underline{k}}\right)^2 . \tag{3.1}$$

Solving (3.1) leads to the well-known Yule-Walker equations. We now want to introduce the notation which allows us to write the Yule-Walker equations in a convenient form: The number of elements in $\Theta(p)$ is $\bar{p} := 2p(p+1)$. Let $\underline{k}_1, \ldots, \underline{k}_{\bar{p}}$ be an arbitrary enumeration of the vectors $\underline{k} \in \Theta(p)$. Define $\underline{a}(p) := (a_{\underline{k}_1}(p), \ldots, a_{\underline{k}_{\bar{p}}}(p))^T \in \mathbb{R}^{\bar{p}}$ and $\underline{Y}_{\underline{t}} := (X_{\underline{t}-\underline{k}_1}, \ldots, X_{\underline{t}-\underline{k}_{\bar{p}}})^T$. Note that the indices $\underline{k}_j$ appear in the same order in both vectors. Due to the projection property it is easy to see that any solution of (3.1) fulfils

$$E\left((X_{\underline{t}} - \underline{a}(p)^T \underline{Y}_{\underline{t}}) \cdot \underline{Y}_{\underline{t}}^T\, \underline{e}_j\right) = 0, \quad j = 1, \ldots, \bar{p}, \tag{3.2}$$

where $\underline{e}_j$ denotes the j-th unit vector. Using the notation $\Gamma(p) := E(\underline{Y}_{\underline{t}}\, \underline{Y}_{\underline{t}}^T)$ and $\underline{\gamma}(p) := E(X_{\underline{t}} \underline{Y}_{\underline{t}})$, system (3.2) is equivalent to

$$\Gamma(p)\, \underline{a}(p) = \begin{pmatrix} \gamma(\underline{k}_1 - \underline{k}_1) & \cdots & \gamma(\underline{k}_1 - \underline{k}_{\bar{p}}) \\ \vdots & \ddots & \vdots \\ \gamma(\underline{k}_{\bar{p}} - \underline{k}_1) & \cdots & \gamma(\underline{k}_{\bar{p}} - \underline{k}_{\bar{p}}) \end{pmatrix} \cdot \begin{pmatrix} a_{\underline{k}_1}(p) \\ \vdots \\ a_{\underline{k}_{\bar{p}}}(p) \end{pmatrix} = \begin{pmatrix} \gamma(\underline{k}_1) \\ \vdots \\ \gamma(\underline{k}_{\bar{p}}) \end{pmatrix} = \underline{\gamma}(p). \tag{3.3}$$

System (3.3) is called the Yule-Walker equations. Note that the matrix $\Gamma(p)$ is symmetric, regardless of the order of indices in the vectors $\underline{Y}_{\underline{t}}$ and $\underline{a}(p)$. The following result ensures the existence of a unique solution of (3.3). Moreover, we establish a uniform bound for the spectral norms of the inverse matrices $\Gamma(p)^{-1}$, which will turn out to be crucial for proving the Baxter-inequality. The spectral norm of a real-valued quadratic matrix A is defined as the square root of the largest eigenvalue of A^TA, denoted by $\|A\|_{spec} = \sqrt{\sigma_{max}(A^TA)}$. For symmetric positive definite matrices this formula can be simplified to $\|A\|_{spec} = \sigma_{max}(A)$.

Lemma 3.1. *Let $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ be a process that fulfils Assumption 1. Then the matrix $\Gamma(p)$ from the Yule-Walker equations (3.3) is invertible for all $p \in \mathbb{N}$. Furthermore, it holds $\|\Gamma(p)^{-1}\|_{spec} \leq (4\pi^2c)^{-1}$ for all $p \in \mathbb{N}$, where c is the lower bound of the spectral density from Assumption 1, and $\|\cdot\|_{spec}$ denotes the spectral norm.*

The previous lemma justifies calling the unique solution $\{a_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$ of (3.3) *the* finite predictor coefficients of the process for order p. As already mentioned, it is of critical importance for our sieve bootstrap scheme that the $a_{\underline{k}}(p)$ converge towards the autoregressive coefficients $\{a_{\underline{k}} : \underline{k} \in \Theta\}$ of the underlying process from (2.5), as p tends to infinity. In particular, we have to ensure that this convergence is fast enough. Therefore, we introduce the following version of Baxter's inequality for random fields:

Theorem 3.2. (Baxter's Inequality) *Let $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ be a process that fulfils Assumption 1 with some $r \geq 2$ and $c > 0$. Let $\{a_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$ be its finite predictor coefficients as defined above, and $\{a_{\underline{k}} : \underline{k} \in \Theta\}$ be its autoregressive coefficients given by (2.5). Denote by $K := \sum_{\underline{k}\in\mathbb{Z}^2} |\gamma(\underline{k})|$. Then it holds for all $s \in \mathbb{N}_0$ with $s+1 < r$ and for all $p \in \mathbb{N}$:*

$$\sum_{\underline{k}\in\Theta(p)} (1+|\underline{k}|_\infty)^s\, |a_{\underline{k}}(p) - a_{\underline{k}}| \leq \frac{K}{2\sqrt{2}\,\pi^2 c} \cdot \sum_{\underline{k}\in\Theta\setminus\Theta(p)} (1+|\underline{k}|_\infty)^{s+1}\, |a_{\underline{k}}|.$$

Due to Lemma 2.1 the right-hand side converges to zero as $p \to \infty$.

The established convergence of the autoregressive coefficients in Baxter's inequality is closely related to a similar convergence of moving average parameters, which shall be derived in the next step. To do this, we take a look at so-called *z-transforms*, also called *transfer functions*, cf. Brockwell and Davis (1991), section 4.4. Based on the AR and MA representations from (2.5) with the coefficients $(a_{\underline{k}})$ and $(b_{\underline{k}})$, we define the z-transforms

$$A(\underline{z}) = 1 - \sum_{\underline{k}\in\Theta} a_{\underline{k}}\, z_1^{k_1} z_2^{k_2}, \quad B(\underline{z}) = 1 + \sum_{\underline{k}\in\Theta} b_{\underline{k}}\, z_1^{k_1} z_2^{k_2} \quad \forall\, \underline{z} \in S, \tag{3.4}$$

where

$$S := \{\underline{z} \in \mathbb{C}^2 : |z_1| = 1,\, |z_2| \leq 1\}.$$

The series $A(\underline{z})$ and $B(\underline{z})$ converge absolutely on its domain S because of Lemma 2.1. It is worth noting that we have to make the distinction between z_1 and z_2 in S. Since z_2 shows up exclusively with exponents $k_2 \geq 0$ in (3.4), as can be seen from the definition of Θ in section 2, we have $|z_2|^{k_2} \leq 1$ for the entire closed disk $|z_2| \leq 1$, while z_1 shows up with both positive and negative exponents k_1. Hence we get $|z_1|^{k_1} \leq 1$, and thus absolute convergence of the series $A(\underline{z})$ and $B(\underline{z})$, only for the circle $|z_1| = 1$.

In analogy to the definition of $A(\underline{z})$, we now define the z-transform of the finite predictor coefficients $\{a_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$ by

$$A_p(\underline{z}) = 1 - \sum_{\underline{k} \in \Theta(p)} a_{\underline{k}}(p)\, z_1^{k_1} z_2^{k_2} \quad \forall\, \underline{z} \in S_p, \tag{3.5}$$

where

$$S_p := \left\{ \underline{z} \in \mathbb{C}^2 : \ \frac{p}{p+1} \leq |z_1| \leq \frac{p+1}{p},\ 0 \leq |z_2| \leq \frac{p+1}{p} \right\}.$$

Note that $A_p(\underline{z})$ is defined on an extended domain compared to $A(\underline{z})$, but for $p \to \infty$ the domains S_p converge to S.

From the proof of Lemma 2.1 we already have $B(\underline{z}) = 1/A(\underline{z})$ for all $\underline{z} \in S$. In particular, both $A(\underline{z})$ and $B(\underline{z})$ are non-zero on their domain S. The next lemma shows that, for p large enough, the inverse of $A_p(\underline{z})$ has a z-transform similar to the one of $B(\underline{z})$.

Lemma 3.3. *Let $(X_{\underline{t}})_{\underline{t} \in \mathbb{Z}^2}$ be a process that fulfils the conditions of Theorem 3.2 with some $r \geq 2$. Then there exists $\delta > 0$ such that it holds $|A_p(\underline{z})| \geq \delta$ uniformly for all $\underline{z} \in S_p$ and all p large enough. For those p, $B_p(\underline{z}) := 1/A_p(\underline{z})$ can be expressed as a convergent series of the form*

$$B_p(\underline{z}) = 1 + \sum_{\underline{k} \in \Theta} b_{\underline{k}}(p)\, z_1^{k_1} z_2^{k_2} \quad \forall\, \underline{z} \in S_p, \tag{3.6}$$

for suitable coefficients $\{b_{\underline{k}}(p) : \underline{k} \in \Theta\}$.

We conclude this section with a result which transfers the convergence of the autoregressive parameters from Baxter's inequality to the moving average parameters $\{b_{\underline{k}}(p) : \underline{k} \in \Theta\}$ and $\{b_{\underline{k}} : \underline{k} \in \Theta\}$:

Lemma 3.4. *Let $(X_{\underline{t}})_{\underline{t} \in \mathbb{Z}^2}$ be a process that fulfils the conditions of Theorem 3.2 with some $r \geq 2$. For all p large enough such that $A_p(\underline{z}) \neq 0$ for all $\underline{z} \in S_p$, let $\{b_{\underline{k}}(p) : \underline{k} \in \Theta\}$ be the coefficients as defined in (3.6) and let $(a_{\underline{k}})_{\underline{k} \in \Theta}$ and $(b_{\underline{k}})_{\underline{k} \in \Theta}$ be the AR and MA coefficients of $(X_{\underline{t}})$ given by (2.5). Then there exists a constant $C < \infty$ such that it holds for all p large enough, and for all $s \in \mathbb{N}_0$ with $s + 1 < r$:*

$$\sum_{\underline{k} \in \Theta} (1 + |\underline{k}|_\infty)^s\, |b_{\underline{k}}(p) - b_{\underline{k}}| \leq C \cdot \sum_{\underline{k} \in \Theta \setminus \Theta(p)} (1 + |\underline{k}|_\infty)^{s+1}\, |a_{\underline{k}}|.$$

Due to Lemma 2.1, the right-hand side converges to zero as $p \to \infty$.

The proofs for all lemmas in this section can be found in section 8, except for Theorem 3.2, which can be found in section 7.

3.2. **Conditions on the fitted-model order $p(n)$ and convergence of estimated coefficients.**

It is important for the validity of the AR sieve bootstrap scheme that the parameter estimators $\{\widehat{a}_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$ used in step 1 of the procedure converge towards the finite predictor coefficients $\{a_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$ at a sufficient rate. At this point one has to keep in mind that the order p of the autoregressive fits actually depends on the sample size n, which is suppressed in the notation for most parts of this paper due to convenience reasons. In order to use the results from the previous section, we need $p = p(n) \to \infty$ as $n \to \infty$. This implies that the dimension of the Yule-Walker matrices $\Gamma(p)$ given by (3.3) also increases for $n \to \infty$.

Probably the most popular form of fitting an AR model as in step (1) of the sieve bootstrap procedure, is Yule-Walker estimation: One replaces the autocovariances in $\Gamma(p)$ by its empirical versions, cf. (2.1), and solves the linear system. Informally speaking, we then have to make sure that $p(n)$ increases slowly enough such that for n large enough all autocovariances showing up in $\Gamma(p)$ can be estimated sufficiently well, in order to obtain a small difference between $\{\widehat{a}_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$ and $\{a_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$.

The following assumption formalizes this condition. Essentially it contains two assertions: Firstly, the underlying process allows for consistent estimation of the finite predictor coefficients $\{a_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$. Secondly, by restricting the rate of increase of $p = p(n)$, we can achieve sufficiently fast uniform convergence of the estimators $\{\widehat{a}_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$.

Assumption 2. *For $p = p(n)$, with $p(n) \to \infty$ as $n \to \infty$, assume for the following sequence in n:*

$$p^4 \cdot \sum_{\underline{k} \in \Theta(p)} |\widehat{a}_{\underline{k}}(p) - a_{\underline{k}}(p)| = \mathcal{O}_P(1).$$

In the remainder of this section we will investigate whether the fitted AR models can also be represented as moving averages of possibly infinite order, which will be crucial for asymptotic inference later on. Based on the parameter estimators $\widehat{a}_{\underline{k}}(p)$ we can define the z-transform $\widehat{A}_p(\underline{z})$ analogously to $A_p(\underline{z})$ in (3.5) as

$$\widehat{A}_p(\underline{z}) = 1 - \sum_{\underline{k} \in \Theta(p)} \widehat{a}_{\underline{k}}(p)\, z_1^{k_1} z_2^{k_2} \quad \forall\, \underline{z} \in S_p.$$

The following calculations will make sure that $\widehat{A}_p(\underline{z})$ is bounded away from zero for n large enough. Assumption 2 implies

$$\sup_{\underline{z} \in S_p} \left| \widehat{A}_p(\underline{z}) - A_p(\underline{z}) \right| \quad \leq \quad \sum_{\underline{k} \in \Theta(p)} \left| \widehat{a}_{\underline{k}}(p) - a_{\underline{k}}(p) \right| \left(\frac{p+1}{p} \right)^{|k_1| + k_2}$$

$$
\begin{aligned}
&\leq \left(\frac{p+1}{p}\right)^{2p} \sum_{\underline{k}\in\Theta(p)} |\widehat{a}_{\underline{k}}(p) - a_{\underline{k}}(p)| \\
&= \frac{1}{p^4}\,\mathcal{O}_P(1) = o_P(1), \qquad (3.7)
\end{aligned}
$$

because $((p+1)/p)^{2p}$ is a bounded sequence (convergent with limit e^2), and because the definition of S_p yields

$$
\begin{aligned}
|z_1|^{k_1} &\leq \begin{Bmatrix} \left(\frac{p+1}{p}\right)^{k_1}, & \text{for } k_1 \geq 0 \\ \left(\frac{p}{p+1}\right)^{k_1}, & \text{for } k_1 < 0 \end{Bmatrix} = \left(\frac{p+1}{p}\right)^{|k_1|}, \\
|z_2|^{k_2} &\leq \left(\frac{p+1}{p}\right)^{k_2},
\end{aligned}
$$

for all $\underline{z} \in S_p$. Assumption 2 ensures $p \to \infty$, as $n \to \infty$, which implies that $A_p(\underline{z})$ is bounded away from zero for all n large enough, cf. Lemma 3.3. It follows from (3.7) that $\widehat{A}_p(\underline{z})$ is uniformly bounded away from zero in probability for all $\underline{z} \in S_p$ and for all n large enough. For all those n large enough, the inverse of $\widehat{A}_p(\underline{z})$ possesses the expansion

$$
\widehat{B}_p(\underline{z}) = \frac{1}{\widehat{A}_p(\underline{z})} = 1 + \sum_{\underline{k}\in\Theta} \widehat{b}_{\underline{k}}(p)\, z_1^{k_1} z_2^{k_2} \quad \forall\, \underline{z} \in S_p, \tag{3.8}
$$

in probability, following the same arguments as for (3.6). Hence, the bootstrap process given by (2.4), which can be described by the transfer function $\widehat{A}_p(\underline{z})$, has the moving average representation

$$
X^*_{\underline{t}} = \sum_{\underline{k}\in\Theta} \widehat{b}_{\underline{k}}(p)\, \varepsilon^*_{\underline{t}-\underline{k}} + \varepsilon^*_{\underline{t}} \tag{3.9}
$$

for all n large enough, in probability. The convergence of the parameter estimators $\widehat{a}_{\underline{k}}(p)$ towards $a_{\underline{k}}(p)$ in Assumption 2 carries over to the corresponding moving average parameters, as shows the following lemma.

Lemma 3.5. *Let $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ be a process that fulfils the conditions of Theorem 3.2 and Assumption 2. Then, for all n large enough (and thus p large enough) such that $A_p(\underline{z})$ and $\widehat{A}_p(\underline{z})$ are bounded away from zero (the latter in probability), it holds uniformly for all $\underline{k} \in \Theta$ and for some $C < \infty$:*

$$
\left|\widehat{b}_{\underline{k}}(p) - b_{\underline{k}}(p)\right| \leq C \cdot \left(1 + \frac{1}{p}\right)^{-|k_1|-k_2} \frac{1}{p^4} \quad \textit{in probability.}
$$

The proof can be found in section 8.

4. Asymptotic validity of the bootstrap

In this section we will derive asymptotic validity of the AR sieve bootstrap procedure under appropriate conditions for a class of statistics which will be specified in

Assumption 3. Similar to what happens in the time series case, cf. Kreiss, Paparoditis and Politis (2011), it turns out that the bootstrap procedure asymptotically mimics the behaviour of the so-called companion process, a modification of the underlying process $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$. This yields a check criterion which basically says that the bootstrap procedure works asymptotically for a test statistic T_n, whenever the asymptotic distributions of T_n applied to the underlying and the companion process coincide. We will elaborate this, and start with the definition of the companion process:

Based on representation (2.5) for the underlying process, we define the *companion process* of $(X_{\underline{t}})$ as the stationary spatial process $(\widetilde{X}_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$, generated by

$$\widetilde{X}_{\underline{t}} = \sum_{\underline{k}\in\Theta} a_{\underline{k}}\, \widetilde{X}_{\underline{t}-\underline{k}} + \widetilde{\varepsilon}_{\underline{t}}, \tag{4.1}$$

where the coefficients $a_{\underline{k}}$ are exactly the ones from (2.5) and $(\widetilde{\varepsilon}_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ is an i.i.d. white noise process with identical marginal distribution as $(\varepsilon_{\underline{t}})$, i.e. $\mathcal{L}(\widetilde{\varepsilon}_{\underline{t}}) = \mathcal{L}(\varepsilon_{\underline{t}})$. Therefore, the companion process also possesses the moving average representation

$$\widetilde{X}_{\underline{t}} = \sum_{\underline{k}\in\Theta} b_{\underline{k}}\, \widetilde{\varepsilon}_{\underline{t}-\underline{k}} + \widetilde{\varepsilon}_{\underline{t}}, \tag{4.2}$$

with the exact same coefficients $b_{\underline{k}}$ as in (2.5). The only difference between $(X_{\underline{t}})$ and $(\widetilde{X}_{\underline{t}})$ is the dependence structure of the respective noise processes $(\varepsilon_{\underline{t}})$ and $(\widetilde{\varepsilon}_{\underline{t}})$. While $(\widetilde{\varepsilon}_{\underline{t}})$ is i.i.d., $(\varepsilon_{\underline{t}})$ is strictly stationary but *not necessarily independent*, the random variables $\varepsilon_{\underline{s}}$ and $\varepsilon_{\underline{t}}$ in general are only uncorrelated for $\underline{s} \neq \underline{t}$. Nevertheless, it is easy to see from (4.2) that all second order properties of $(X_{\underline{t}})$ and $(\widetilde{X}_{\underline{t}})$ are identical, i.e. the two processes possess identical autocovariances and spectral densities.

In our main theorem we will establish bootstrap validity for a class of statistics which will be specified in the following Assumption 3. This class is a natural extension of the so-called *functions of generalized means*, introduced by Künsch (1989), to the case of random fields. These statistics will be based on smooth functions g applied to rectangular-shaped subsamples of the available data sample $\{X_{\underline{t}} : \underline{t} \in \Pi\}$, with $\Pi := \{\underline{t} \in \mathbb{Z}^2 : \ 1 \leq t_1, t_2 \leq n\}$. We first specify the necessary notation: For $1 \leq m_1, m_2 \leq n$ let

$$\begin{aligned} S(m_1,m_2) : &= \Big\{\underline{s} = (s_1,s_2)^T \in \mathbb{N}_0^2 : \ 0 \leq s_1 \leq m_1 - 1, \quad 0 \leq s_2 \leq m_2 - 1\Big\} \\ &= \{\underline{s}(1), \ldots, \underline{s}(m_1 m_2)\}, \end{aligned}$$

i.e. $\underline{s}(1), \ldots, \underline{s}(m_1 m_2)$ is any fixed enumeration of the $m_1 m_2$ vectors in $S(m_1, m_2)$. We define the $m_1 m_2$-dimensional random vector

$$\mathbf{Y}_{\underline{t}} := (X_{\underline{t}+\underline{s}(1)}, \ldots, X_{\underline{t}+\underline{s}(m_1 m_2)})^T.$$

Observe that for each $\underline{t}$ with $1 \leq t_1 \leq n - m_1 + 1$ and $1 \leq t_2 \leq n - m_2 + 1$, the components of $\mathbf{Y}_{\underline{t}}$ form a rectangular-shaped subsample of dimension $m_1 \times m_2$ of the original data sample. We can now specify the class of statistics we will be investigating.

Assumption 3. *Let $\bar{n}_1 := n - m_1 + 1$, $\bar{n}_2 := n - m_2 + 1$ for some $1 \leq m_1, m_2 \leq n$, and let $m := m_1 m_2$. Define the statistic T_n as*

$$T_n = f\left(\frac{1}{\bar{n}_1\bar{n}_2}\sum_{t_1=1}^{\bar{n}_1}\sum_{t_2=1}^{\bar{n}_2} g\Big(\mathbf{Y}_{\underline{t}}\Big)\right)$$

where the functions $g : \mathbb{R}^m \to \mathbb{R}^k$ and $f : \mathbb{R}^k \to \mathbb{R}$, with $k \geq 1$, fulfil the following smoothness conditions: f is continuously differentiable in a neighborhood of $\underline{\theta} := E\, g(\mathbf{Y}_{\underline{t}})$ and the gradient of f at $\underline{\theta}$ does not vanish, i.e.

$$\nabla f(\theta) = \Big(\frac{\partial f(\underline{x})}{\partial x_1}, \ldots, \frac{\partial f(\underline{x})}{\partial x_k}\Big)\Big|_{\underline{x}=\underline{\theta}} \neq (0, \ldots, 0).$$

For some $h \geq 1$ all component functions $g_1, \ldots, g_k$ of g are h times continuously differentiable and all h-th-order derivatives satisfy a Lipschitz condition, i.e. for all $i = 1, \ldots, k$ and for all $(h_1, \ldots, h_m) \in \mathbb{N}_0^m$ with $\sum_{u=1}^m h_u = h$ the derivative

$$\frac{\partial^h g_i(\underline{x})}{\partial^{h_1} x_1 \ldots \partial^{h_m} x_m}$$

is Lipschitz.

Remark 4.1. The conditions from the previous assumption should be explained at this point: The class of statistics from Assumption 3 contains, among other things, the sample mean and versions of the sample autocovariance and sample autocorrelation. To obtain the latter two statistics, one typically uses a function g which is not Lipschitz. For example, in the case of sample autocovariances at lag $\underline{h} = (h_1, h_2)^T$, one may choose $m_1 = h_1 + 1$, $m_2 = h_2 + 1$ and $g(x_1, \ldots, x_m) = x_1 x_m$. Then T_n from Assumption 3 translates to taking the empirical mean of observations $X_{\underline{t}+\underline{h}} X_{\underline{t}}$. Now observe that g itself is *not Lipschitz*, but all of its first order partial derivatives are. This is the why we allow for non-Lipschitz functions g in Assumption 3, and merely assume that there exists a number $1 \leq h < \infty$ such that all derivatives of order h (but *not* up to order h) are Lipschitz. □

In order to state the main theorem, we define $\widetilde{T}_n$ and T_n^* as the statistic T_n applied to samples from the companion process $(\widetilde{X}_{\underline{t}})$ and the bootstrap process $(X_{\underline{t}}^*)$, respectively, i.e.

$$\widetilde{T}_n := f\left(\frac{1}{\bar{n}_1\bar{n}_2}\sum_{t_1=1}^{\bar{n}_1}\sum_{t_2=1}^{\bar{n}_2} g\Big(\widetilde{\mathbf{Y}}_{\underline{t}}\Big)\right), \quad T_n^* := f\left(\frac{1}{\bar{n}_1\bar{n}_2}\sum_{t_1=1}^{\bar{n}_1}\sum_{t_2=1}^{\bar{n}_2} g\Big(\mathbf{Y}_{\underline{t}}^*\Big)\right)$$

where

$$\widetilde{\mathbf{Y}}_{\underline{t}} := (\widetilde{X}_{\underline{t}+\underline{s}(1)}, \ldots, \widetilde{X}_{\underline{t}+\underline{s}(m_1 m_2)})^T, \quad \mathbf{Y}_{\underline{t}}^* := (X_{\underline{t}+\underline{s}(1)}^*, \ldots, X_{\underline{t}+\underline{s}(m_1 m_2)}^*)^T.$$

We can prove bootstrap validity under the following assumptions, which ensure convergence of empirical moments and the empirical distribution function to their theoretical counterparts for the innovations:

Assumption 4. *For all continuity points $x \in \mathbb{R}$ of the distribution function F of $\varepsilon_{\underline{0}}$ it holds*

$$F_n(x) \xrightarrow{P} F(x) \quad \text{as } n \to \infty,$$

where $F_n(x)$ is the empirical distribution function

$$F_n(x) = \frac{1}{|\Pi(n,p)|} \sum_{\underline{t} \in \Pi(n,p)} \mathbb{1}\{\varepsilon_{\underline{t}} \leq x\},$$

and where $\Pi(n,p) := \{(t_1,t_2) \in \mathbb{Z}^2 : p+1 \leq t_1 \leq n-p,\ p+1 \leq t_2 \leq n\}$.
Furthermore, it holds $E\Big(\varepsilon_{\underline{t}}^{2(h+2)}\Big) < \infty$, where h is the constant specified in Assumption 3, as well as the following convergence of empirical moments:

$$\frac{1}{|\Pi(n,p)|} \sum_{\underline{t} \in \Pi(n,p)} (\varepsilon_{\underline{t}})^{2w} \xrightarrow{P} E\Big((\varepsilon_{\underline{0}})^{2w}\Big) \quad \forall\, w \leq h+2.$$

Theorem 4.2. *Let $(X_{\underline{t}})_{\underline{t} \in \mathbb{Z}^2}$ be a process fulfilling Assumptions 2 – 4, as well as Assumption 1 with $r = 4$.*
Then, for $\widetilde{T}_n$ and T_n^* as defined above, it holds

$$d_K\left(\mathcal{L}^*\Big(n\,(T_n^* - f(\underline{\theta}^*))\Big), \mathcal{L}\Big(n\,(\widetilde{T}_n - f(\underline{\widetilde{\theta}}))\Big)\right) = o_P(1)$$

as $n \to \infty$, where $\underline{\theta}^* = E^*\Big(g(\mathbf{Y}_{\underline{t}}^*)\Big)$, $\underline{\widetilde{\theta}} = E\Big(g(\widetilde{\mathbf{Y}}_{\underline{t}})\Big)$ and d_K denotes the Kolmogorov distance.

This result shows for all statistics from Assumption 3 that the sieve bootstrap procedure asymptotically approximates the distribution $\widetilde{T}_n$ instead of the one of T_n. Therefore, the bootstrap procedure works asymptotically *if and only if* the limiting distributions of T_n and $\widetilde{T}_n$ coincide. We will give a few examples of the application of this check criterion in the following section. The proof of Theorem 4.2 can be found in Section 7.2.

5. Applications

In this section we will give a few examples of prominent statistics to which the check criterion derived in the previous section can be applied. For a simulation study concerning sample autocorrelations, see section 6.

Example 5.1. (Sample mean) We can use the AR sieve bootstrap procedure for the sample mean, even for processes which are not centered as required per Assumption 1. Let $(Z_{\underline{t}})_{\underline{t} \in \mathbb{Z}^2}$ be a strictly stationary process with mean μ which, other than being non-centered, fulfils the conditions stated in Assumption 1. Since all autocovariances of $(Z_{\underline{t}})$ and the centered process $(X_{\underline{t}}) := (Z_{\underline{t}} - \mu)$ coincide, $(X_{\underline{t}})$ obviously fulfils Assumption 1. Now let $\{Z_{\underline{t}},\, \underline{t} \in \Pi\}$ be a data sample generated by $(Z_{\underline{t}})$. We apply the bootstrap procedure described in section 2 to the data $\{Z_{\underline{t}},\, \underline{t} \in \Pi\}$, which produces bootstrap samples $\{X^*_{\underline{t}},\, \underline{t} \in \Pi\}$, generated by

$$X^*_{\underline{t}} = \sum_{\underline{k} \in \Theta(p)} \widehat{a}_{\underline{k}}(p)\, X^*_{\underline{t}-\underline{k}} + \varepsilon^*_{\underline{t}}.$$

Then, compute $Z^*_{\underline{t}} := \overline{Z} + X^*_{\underline{t}}$ for all $\underline{t} \in \Pi$, where $\overline{Z} := |\Pi|^{-1} \sum_{\underline{t} \in \Pi} Z_{\underline{t}}$ (for the bootstrap data, $\overline{Z}^*$ is analogously defined). We can approximate the distribution of $n(\overline{Z} - \mu)$ by the one of $n(\overline{Z}^* - \overline{Z})$. Asymptotic validity of this approach can be established via Theorem 4.2 in the following way:
The companion process associated with $(X_{\underline{t}})$ is denoted by $(\widetilde{X}_{\underline{t}})$ and we define $\widetilde{Z}_{\underline{t}} := \widetilde{X}_{\underline{t}} + \mu$. The functions f and g in assumption 3 can be chosen appropriately such that T_n is the sample mean of $\{X_{\underline{t}},\, \underline{t} \in \Pi\}$, and $\widetilde{T}_n = \overline{\widetilde{X}}$ is the mean of $\{\widetilde{X}_{\underline{t}},\, \underline{t} \in \Pi\}$. For the linear process $(\widetilde{Z}_{\underline{t}})$, with an obvious notation for $\overline{\widetilde{Z}}$, it is known that

$$n\left(\overline{\widetilde{Z}} - \mu\right) = n\left(\overline{\widetilde{X}}\right) = n\,\widetilde{T}_n \xrightarrow{d} \mathcal{N}\Big(0\,,\, \textstyle\sum_{\underline{h} \in \mathbb{Z}^2} \gamma_{\widetilde{Z}}(\underline{h})\Big),$$

where $\gamma_{\widetilde{Z}}$ denotes the autocoavariance function of $(\widetilde{Z}_{\underline{t}})$. Noting that $\overline{Z}^* = \overline{Z} + \overline{X}^*$, it follows immediately from Theorem 4.2

$$n\left(\overline{Z}^* - \overline{Z}\right) = n\left(\overline{X}^*\right) = n\,T^*_n \xrightarrow{d^*} \mathcal{N}\Big(0\,,\, \textstyle\sum_{\underline{h} \in \mathbb{Z}^2} \gamma_{\widetilde{Z}}(\underline{h})\Big) \quad \text{in prob.} \tag{5.1}$$

For the sample mean $\overline{Z}$ of the actually observed data it holds under suitable regularity conditions that

$$n\left(\overline{Z} - \mu\right) \xrightarrow{d} \mathcal{N}\Big(0\,,\, \textstyle\sum_{\underline{h} \in \mathbb{Z}^2} \gamma_{Z}(\underline{h})\Big). \tag{5.2}$$

Now observe that $(Z_{\underline{t}})$ and $(\widetilde{Z}_{\underline{t}})$ have identical second order properties per definition. In particular, $\gamma_Z(\underline{h}) = \gamma_{\widetilde{Z}}(\underline{h})$ for all lags $\underline{h} \in \mathbb{Z}^2$. Thus, the limiting distributions in (5.1) and (5.2) coincide and it follows

$$d_K\left(\mathcal{L}^*\Big(n\left(\overline{Z}^* - \overline{Z}\right)\Big), \mathcal{L}\Big(n\left(\overline{Z} - \mu\right)\Big)\right) = o_P(1).$$

Therefore, the AR sieve bootstrap proposal is asymptotically valid for the sample mean under the stated conditions. $\square$

In contrast to the preceeding example, the limiting distribution of sample autocovariances does not depend exclusively on second-order properties of the underlying process. This result is well-known, particularly for the time-series case, i.e. $d = 1$. Even if the data are generated by a linear spatial process, that is a process of the form

$$X_{\underline{t}} = \sum_{\underline{\nu} \in \mathbb{Z}^2} \alpha_{\underline{\nu}}\, u_{\underline{t}-\underline{\nu}}, \tag{5.3}$$

with absolutely summable coefficients $(\alpha_{\underline{\nu}})_{\underline{\nu}\in\mathbb{Z}^2}$ and an i.i.d. white noise process $(u_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ with finite fourth moments, the limiting variance depends on the fourth-order cumulants of $(u_{\underline{t}})$. This can be verified with analogous calculations as for the times series case, cf. Brockwell and Davis (1991), Proposition 7.3.4. However, the situation is different if one switches to sample autocorrelations of linear processes, instead of autocovariances. Then, the limiting distribution depends only on the autocorrelations of the underlying process, as shows the following theorem, which is a direct generalisation of the well-known Bartlett formula for time series, cf. Brockwell and Davis (1991), Proposition 7.2.1.:

Lemma 5.2. *Let* $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ *be a linear spatial process as defined in* (5.3)*, i.e. with i.i.d. white noise and finite fourth moments, and with autocorrelation function* ρ*. For the sample autocorrelations* $\widehat{\rho}(\underline{h})=\widehat{\gamma}(\underline{h})/\widehat{\gamma}(\underline{0})$*, with* $\widehat{\gamma}(\cdot)$ *as defined in* (2.1)*, we define the comparative quantity* $\check{\rho}(\underline{h}):=\check{\gamma}(\underline{h})/\check{\gamma}(\underline{0})$ *with*

$$\check{\gamma}(\underline{h}) := \frac{1}{|\Pi|}\sum_{\underline{t}\in\Pi} X_{\underline{t}+\underline{h}}\, X_{\underline{t}},$$

where $\Pi=\{\underline{t}\in\mathbb{Z}^2:\ 1\le t_1,t_2\le n\}$*.* $\check{\rho}(\underline{h})$ *and* $\widehat{\rho}(\underline{h})$ *are asymptotically equivalent. Then it holds*

$$n^2\,\mathrm{Cov}(\check{\rho}(\underline{h}),\check{\rho}(\underline{k})) \longrightarrow V(\underline{h},\underline{k}), \qquad \text{as } n\to\infty,$$

where

$$\begin{aligned} V(\underline{h},\underline{k}) &= \sum_{\underline{r}\in\mathbb{Z}^2}\Big\{2\rho(\underline{r})^2\rho(\underline{k})\rho(\underline{h}) - 2\rho(\underline{r}+\underline{k})\rho(\underline{r})\rho(\underline{h}) - 2\rho(\underline{r}-\underline{h})\rho(\underline{r})\rho(\underline{k}) \\ &\qquad + \rho(\underline{r}-\underline{h}+\underline{k})\rho(\underline{r}) + \rho(\underline{r}+\underline{k})\rho(\underline{r}-\underline{h})\Big\}. \end{aligned}$$

The proof is analogous to the time-series case and can be found in section 8.

Example 5.3. (Sample autocorrelations/correlogram) Let $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ be a spatial process fulfilling Assumption 1 with corresponding companion process $(\widetilde{X}_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$. We consider the autocorrelation function $\rho(\underline{h})=\gamma(\underline{h})/\gamma(\underline{0})$ at lag $\underline{h}$, together with the usual estimator $T_n:=\widehat{\rho}(\underline{h})=\widehat{\gamma}(\underline{h})/\widehat{\gamma}(\underline{0})$, where $\widehat{\gamma}(\cdot)$ is given by (2.1). For spatial processes, $\rho(\underline{h})$ (and accordingly $\widehat{\rho}(\underline{h})$) are often referred to as the (sample) correlogram, cf. Cressie (1993), Section 2.3.2. Note that the autocorrelations of $(\widetilde{X}_{\underline{t}})$ are given by the function ρ as well. Under suitable assumptions on the dependence structure of the process, such as weak dependence or mixing conditions, it is known that

$$n(\widehat{\rho}(\underline{h})-\rho(\underline{h})) \xrightarrow{d} \mathcal{N}(0,\tau_X^2), \quad n(\widetilde{T}_n-\rho(\underline{h})) \xrightarrow{d} \mathcal{N}(0,\tau_{\widetilde{X}}^2),$$

where the limiting variances τ_X^2 and $\tau_{\widetilde{X}}^2$ in general depend on the fourth order cumulants of $(X_{\underline{t}})$ and $(\widetilde{X}_{\underline{t}})$, respectively. Hence, it follows $\tau_X^2\neq\tau_{\widetilde{X}}^2$ in general, because $(X_{\underline{t}})$ and $(\widetilde{X}_{\underline{t}})$ share second order but not fourth order properties. For T_n^*, denoting the sample autocorrelation applied to the bootstrap sample $\{X^*_{\underline{t}},\ \underline{t} \in \Pi\}$, Theorem 4.2 yields

$$n\Big(T_n^* - f(\underline{\theta}^*)\Big) \xrightarrow{d} \mathcal{N}(0, \tau^2_{\widetilde{X}}).$$

Therefore, $\tau^2_X \neq \tau^2_{\widetilde{X}}$ implies that the AR sieve bootstrap in general is asymptotically not valid for sample autocorrelations.
However, if the data are generated by a linear process $(X_{\underline{t}})$ as given by (5.3), Lemma 5.2 shows that the limiting variance of $n(\check{\rho}(\underline{h}) - \rho(\underline{h}))$ is given by

$$\begin{aligned}\tau^2_X \;=\; \sum_{\underline{r}\in\mathbb{Z}^2} \Big\{ & 2\rho(\underline{r})^2\rho(\underline{h})^2 - 2\rho(\underline{r}+\underline{h})\rho(\underline{r})\rho(\underline{h}) - 2\rho(\underline{r}-\underline{h})\rho(\underline{r})\rho(\underline{h}) \\ & +\rho(\underline{r})^2 + \rho(\underline{r}+\underline{h})\rho(\underline{r}-\underline{h})\Big\}. \end{aligned} \tag{5.4}$$

Since $\check{\rho}(\underline{h})$ and $\widehat{\rho}(\underline{h})$ are asymptotically equivalent, $n(\widehat{\rho}(\underline{h}) - \rho(\underline{h}))$ also has limiting variance τ^2_X. This expression depends only on the autocorrelations of the underlying process, which coincide for $(X_{\underline{t}})$ and $(\widetilde{X}_{\underline{t}})$. Thus, it follows for this case $\tau^2_X = \tau^2_{\widetilde{X}}$, and the bootstrap procedure is asymptotically valid for sample autocorrelations of data generated from linear processes. □

Remark 5.4. When checking for asymptotic validity of the AR sieve bootstrap procedure, it is of critical importance to ensure that the limiting distributions of T_n and $\widetilde{T}_n$ are identical, as has been done in the previous examples. In general, this will be the case whenever the limiting distribution depends only on second order entities such as autocovariances or the spectral density of the underlying process. For data generated by a linear process $X_{\underline{t}} = \sum_{\underline{\nu}\in\mathbb{Z}^2} \alpha_{\underline{\nu}}\, u_{\underline{t}-\underline{\nu}}$, one might be tempted to conclude that $(X_{\underline{t}})$ and its companion process $(\widetilde{X}_{\underline{t}})$ are identical since $(u_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ is already i.i.d.. However, Example 3.2 from Kreiss, Paparoditis and Politis (2011) shows for the special case of time series that this is not the case. To be precise, the companion process $(\widetilde{X}_{\underline{t}})$ is always derived from the AR representation (2.5), where $(\varepsilon_{\underline{t}})$ is the uniquely determined innovation process of $(X_{\underline{t}})$. Even if the process has linear representation $X_{\underline{t}} = \sum_{\underline{\nu}\in\mathbb{Z}^2} \alpha_{\underline{\nu}}\, u_{\underline{t}-\underline{\nu}}$ with i.i.d. noise $(u_{\underline{t}})$, its innovation process might differ from $(u_{\underline{t}})$, and might be only uncorrelated but not i.i.d.. Remark 2.1 of Kreiss, Paparoditis and Politis (2011) gives a specific example of this situation. Therefore, linear processes are in general not identical to their companion processes, which makes a careful inspection of the limiting distributions as in the previous examples a necessity. □

Example 5.5. (Standardized sample variogram) Let $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ be a spatial process fulfilling Assumption 1 with autocovariance function γ. The variogram at lag $\underline{h} \in \mathbb{Z}^2$ is defined as

$$V(\underline{h}) = \mathrm{Var}(X_{\underline{t}} - X_{\underline{t}+\underline{h}}) = E\Big((X_{\underline{t}} - X_{\underline{t}+\underline{h}})^2\Big) = 2\gamma(\underline{0}) - 2\gamma(\underline{h})$$

for centered fields, and $V^{(s)}(\underline{h}) := V(\underline{h})/\gamma(\underline{0})$ is called the standardized variogram. Using the notation from (2.1), two classical estimators for $V(\underline{h})$ are given by

$$\widehat{V}_1(\underline{h}) = 2\widehat{\gamma}(\underline{0}) - 2\widehat{\gamma}(\underline{h}), \quad \widehat{V}_2(\underline{h}) = \frac{1}{|\Pi_{\underline{h}}|} \sum_{\underline{t} \in \Pi_{\underline{h}}} (X_{\underline{t}} - X_{\underline{t}+\underline{h}})^2,$$

which are asymptotically equivalent, cf. Cressie (1993), Section 2.4. In particular, one can easily check that

$$n\Big(\widehat{V}_1(\underline{h}) - \widehat{V}_2(\underline{h})\Big) = o_P(1). \tag{5.5}$$

Versions of both of these estimators are included in the class of functions of generalized means, as given by Assumption 3. Furthermore, both $\widehat{V}_1(\underline{h})$ and $\widehat{V}_2(\underline{h})$ can be used to construct standardized sample variogram estimators via $\widehat{V}_j^{(s)}(\underline{h}) := \widehat{V}_j(\underline{h})/\widehat{\gamma}(\underline{0})$, $j = 1, 2$. It holds

$$\widehat{V}_1^{(s)}(\underline{h}) = 2 - 2\widehat{\rho}(\underline{h}).$$

Now assume the data are generated by a linear process. Then it follows from Example 5.3, with τ_X^2 as defined there,

$$n\Big(\widehat{V}_1^{(s)}(\underline{h}) - V^{(s)}(\underline{h})\Big) = (-2) \cdot n\Big(\widehat{\rho}(\underline{h}) - \rho(\underline{h})\Big) \xrightarrow{d} \mathcal{N}(0, 4\tau_X^2),$$

and $n\Big(\widehat{V}_2^{(s)}(\underline{h}) - V^{(s)}(\underline{h})\Big)$ has the very same limiting distribution due to (5.5). An analogous argumentation as in Example 5.3 therefore yields asymptotic validity of the AR sieve bootstrap procedure for the standardized sample variogram, as long as the data are generated by a linear spatial process. □

Remark 5.6. Our main result Theorem 4.2 provides a check criterion for asymptotic validity of the AR sieve bootstrap for all statistics from Assumption 3. This class of statistics contains, among other things, the statistics from Examples 5.1-5.5. However, we conjecture that analogous results can be proven, in the same spirit as in the proof of Theorem 4.2, for a much wider class of statistics beyond those covered by Assumption 3. If T_n denotes an estimator for some parameter θ, under the condition that $\mathcal{L}\Big(c_n(T_n - \theta)\Big)$ has a non-degenerated limiting distribution for some sequence (c_n), we conjecture that the AR sieve bootstrap procedure is asymptotically valid, as long as the limiting distribution depends on second order properties of the underlying process, only.

For example, according to Section 4.5 in Guyon (1995), one can prove central limit theorems for kernel-based nonparametric spectral density estimators for strictly stationary spatial processes under appropriate mixing conditions. The limiting distribution then depends exclusively on the spectral density of the underlying process, which is a second order quantity, and we conjecture that the AR sieve bootstrap is asymptotically valid in this situation. □

6. A simulation study

In this section, we will present simulation results that compare the performance of the AR sieve bootstrap to classic normal approximations and block bootstrap methods. First, we generated square-shaped samples $\{X_{\underline{t}} = X_{t_1,t_2} : 1 \leq t_1, t_2 \leq n\}$ as defined in section 2, where the sample size is set to be $n = 15$ which corresponds to $15 \times 15 = 225$ observations. The samples are generated by a moving average model given by

$$X_{t_1,t_2} = e_{t_1,t_2} + 0.5 \cdot e_{t_1+1,t_2} - 0.2 \cdot e_{t_1-1,t_2} + 0.3 \cdot e_{t_1,t_2+1} + 0.1 \cdot e_{t_1,t_2-1}, \tag{6.1}$$

where $(e_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ is an i.i.d. white noise process with marginal distribution $\mathcal{N}(0,1)$. The process $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ fulfils the conditions of Assumption 1. Furthermore, each realisation $X_{\underline{t}}$ depends on noise terms from four different directions, two from the lower and two from the upper half-plane, cf. section 2. This means that the process is *not* 'tailor-made' for an AR approximation in the direction of the lower half-plane as performed in the AR sieve algorithm. In fact, the data generating process from (6.1) does not 'favor' any direction of one-sided autoregressive fits; one could as well fit models that are one-sided with respect to the upper, left or right half-plane.

The statistic that we investigated is the sample autocorrelation $\widehat{\rho}(\underline{h})$ as defined in Example 5.3, with $\underline{h} = (1,-1)^T$. For the process from (6.1), the true autocorrelation is given by $\rho(1,-1) = 0.13/1.39$. We approximated the distribution of

$$n\Big(\widehat{\rho}(1,-1) - \rho(1,-1)\Big) \tag{6.2}$$

for $n = 15$ with a normal approximation and with the AR sieve bootstrap, via the empirical distribution of $n(\widehat{\rho}^*(1,-1) - \widehat{\rho}(1,-1))$. To implement the normal approximation, we considered the limiting distribution of (6.2) given by $\mathcal{N}(0,\tau_X^2)$ with τ_X^2 from (5.4), cf. Example 5.3. For the process $(X_{\underline{t}})$ from (6.1) one can easily verify that τ_X^2 is given by

$$\tau_X^2 = \sum_{|\underline{r}|_\infty \leq 2} \Big\{2\rho(\underline{r})^2\rho(1,-1)^2 - \ldots\Big\}, \tag{6.3}$$

since all summands with $|\underline{r}|_\infty := \max\{|r_1|,|r_2|\} > 2$ vanish due to $\rho(\underline{r}) = 0$ for all $|\underline{r}|_\infty > 2$. Hence, we estimated τ_X^2 by replacing ρ with $\widehat{\rho}$ in (6.3). It should be noted that this approach represents a best-case szenario for the normal approximation because we used the additional information that for the present data τ_X^2 has the special form (6.3), i.e. we chose the optimal point of cutting off the infinite sum in (5.4). For real-world data, this information would not be known, and one would have to estimate τ_X^2 based on equation (5.4) by cutting off the infinite sum at some non-optimal point which would generate an additional error in the estimation.

In Figure 3, the display in the top left corner shows the comparison of three different choices for the order p of the AR sieve bootstrap. We simulated the 95%-quantile

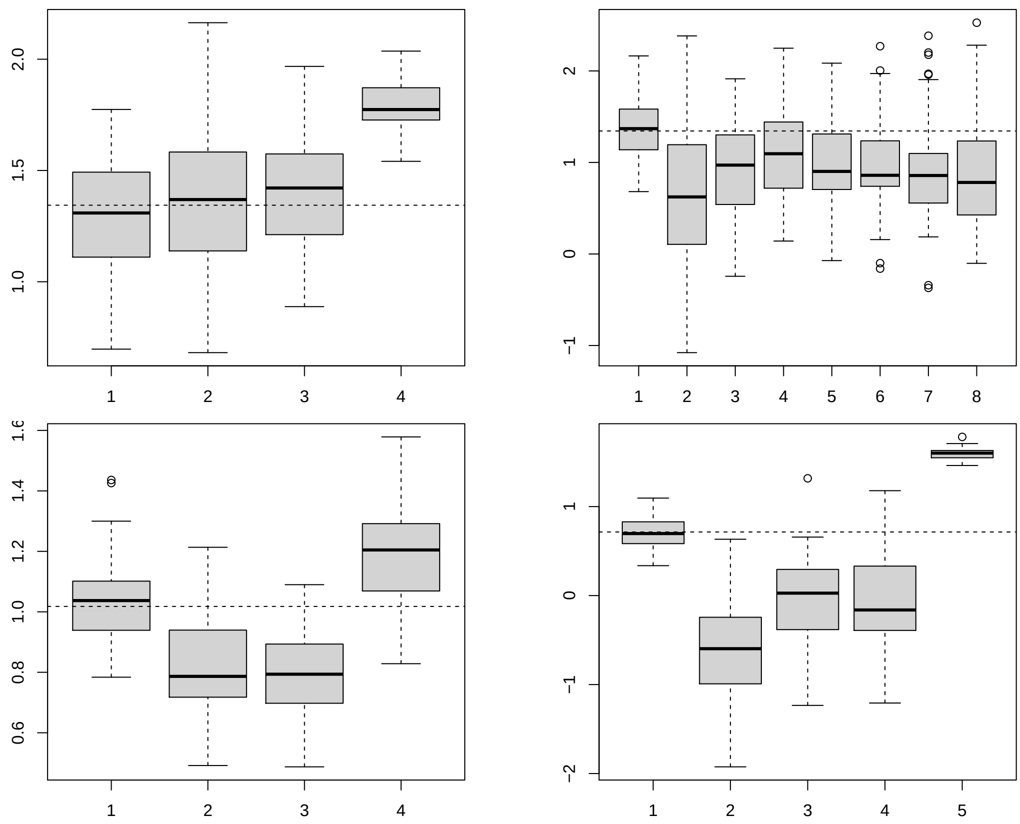

Figure 3. Top left: Approximations of the 95%-quantile of the distribution of $n(\widehat{\rho}(1,-1)-\rho(1,-1))$ for $n = 15$, data generated by model (6.1); boxplots based on $N = 50$ iterations. Boxes 1 to 3: AR sieve bootstrap (based on $M = 500$ repetitions) with $p = 1$, $p = 2$ and $p = 3$, followed by the normal approximation in box 4. Target value given by the horizontal dashed line.
Top right: Same setting as in top left; approximation of the 95%-quantile with the AR sieve bootstrap in box 1, approximations with the block bootstrap and block sizes $l = 2, \ldots, 8$ in boxes $2, \ldots, 8$ (each bootstrap with $M = 500$ repetitions).
Bottom left: Same setting as in top left; approximation of the variance of $n(\widehat{\rho}(1,-1) - \rho(1,-1))$. Box 1: AR sieve bootstrap with $p = 2$. Boxes 2 and 3: Block bootstrap with block sizes $l = 5, 6$. Box 4: Normal approximation.
Bottom right: Approximations of the 95%-quantile of the distribution of $n(\widehat{\rho}(1,-1) - \rho(1,-1))$ for $n = 25$, data generated by model (6.4); boxplots based on $N = 50$ iterations and each bootstrap method based on $M = 300$ repetitions. Box 1: AR sieve bootstrap with $p = 4$. Boxes 2, 3, 4: Block bootstrap with block sizes $l = 8, 9, 10$. Box 5: Normal approximation.

of the distribution of (6.2) for $n = 15$. In each iteration, we generated $M = 500$ bootstrap samples to approximate this quantile, subsequently using the AR sieve bootstrap with orders $p = 1$, $p = 2$ and $p = 3$. We also calculated the normal approximation estimate of the quantile in each iteration as described previously. All of this was carried out for $N = 50$ independent iterations to generate boxplots of the locations of the estimates. The three AR sieve approximations with $p = 1, 2, 3$ are shown in the boxplots $1, 2, 3$ in the top left display of Figure 3, while the normal approximation values are given boxplot 4. The target value, i.e. the 95%-quantile of the distribution of (6.2), is determined from Monte-Carlo simulations with $500,000$ repetitions and illustrated by the horizontal dashed line. One can see that the AR sieve bootstrap works very well compared to the normal approximation, even for small orders p and even though the normal approximation is already improved by additional information, as was explained earlier.

In the aforementioned setting, we also compared the performances of the AR sieve bootstrap and block bootstrap techniques (each based on $M = 500$ repetitions). The target was again the 95%-quantile of the distribution of (6.2) for $n = 15$. The order of the AR sieve bootstrap was fixed to $p = 2$ and we considered block sizes of $l = 2, \ldots, 8$. Here, the block size refers to square-shaped blocks, i.e. a block size of l means drawing blocks of $l \times l$ observations from the original data sample and then sticking the blocks together to form a sample of size $n \times n$. The result can be seen in the top right corner in Figure 3. Boxplot 1 corresponds to the AR sieve bootstrap and the results for the block bootstrap are given in boxes $2, \ldots, 8$ with block length l depicted in box l. Arguably the best result for the block bootstrap is achieved for $l = 4$; however, the AR sieve bootstrap performs considerably stronger than all block bootstrap approaches implemented here.

In order to show that the results obtained so far are not only specific to the 95%-quantile but to the distribution of (6.2) as a whole, we will now look at an approximation of the variance of this distribution instead of a single quantile. The picture in the bootom left corner of Figure 3 shows these approximations of the variance with all parameters as before. The AR sieve bootstrap ($p = 2$) is depicted in box 1, the block bootstrap in boxes 2 and 3 (block lengths $l = 5, 6$) and the normal approximation in box 4. Similar to what happens for the 95%-quantile, the AR sieve bootstrap outperforms the other methods.

To conclude this section, we modified some of the parameters from the simulations performed so far. The data are still generated by a moving average model, but now following the model equation

$$X_{t_1,t_2} = e_{t_1,t_2} + 4 \cdot e_{t_1+1,t_2} - 5 \cdot e_{t_1-1,t_2} + 3 \cdot e_{t_1,t_2+1} - 2 \cdot e_{t_1,t_2-1}, \tag{6.4}$$

where the noise is no longer symmetricly distributed but has an i.i.d. centered exponential distribution. In this model, the dependence of neighbouring random variables is higher than in model (6.1). For example, the true autocorrelation at lag $\underline{h} = (1, -1)^T$ is here given by $\rho(1, -1) = 0.4$ compared to $\rho(1, -1) \approx 0.094$ in model (6.1). We also increase the sample size to $n = 25$ – corresponding to $25 \times 25 = 625$ observations – and choose the order $p = 4$ for the AR sieve bootstrap. The picture in the bottom right corner of Figure 3 shows the results for the approximation of the 95%-quantile of the distribution of (6.2) for $n = 25$; box 1 shows the AR sieve bootstrap, boxes $2, 3, 4$ the block bootstrap with $l = 8, 9, 10$ and box 5 the normal approximation. It can be seen that, for this increased sample size, the normal approximation is close to its limit which, however, differs considerably from the true quantile of the finite sample distribution. This is mainly due to a negative bias for the distribution of (6.2) which can be obtained from the Monte Carlo simulations that were performed to determine the 95%-quantile. The block bootstrap clearly does not show desirable results, this might stem from the increased dependence between neighbouring realisations in the present model compared to the model used previously. However, the AR sieve bootstrap performs very well for this choice of (increased values of) n and p. This emphasizes the fact that convergence of the AR sieve bootstrap can be achieved as long as $p = p(n) \to \infty$ at an appropriate rate.

7. Proofs of the main results

The proof of Theorem 4.2 depends in large parts on some auxiliary results that will be collected in the following lemmas. We will make use of a truncated version $(X^*_{\underline{t},M})$ of the bootstrap process, which is based on the moving average representation of $(X^*_{\underline{t}})$ from (3.9). For arbitrary $M \in \mathbb{N}$ we define

$$X^*_{\underline{t},M} = \sum_{\underline{k} \in \Theta(M)} \widehat{b}_{\underline{k}}(p)\, \varepsilon^*_{\underline{t}-\underline{k}} + \varepsilon^*_{\underline{t}}, \tag{7.1}$$

where the finite collection of sites $\Theta(M)$ is defined in (2.2), whereas the non-truncated version $(X^*_{\underline{t}})$ has the infinite collection of sites Θ. Analogously, a truncated version $(\widetilde{X}_{\underline{t},M})$ of the companion process can be defined by replacing Θ with $\Theta(M)$ in (4.2). As a natural extension of the definition of $\mathbf{Y}^*_{\underline{t}}$ and $\widetilde{\mathbf{Y}}_{\underline{t}}$, we denote by

$$\mathbf{Y}^*_{\underline{t},M} := (X^*_{\underline{t}+\underline{s}(1),M}, \ldots, X^*_{\underline{t}+\underline{s}(m_1 m_2),M})^T, \quad \widetilde{\mathbf{Y}}_{\underline{t},M} := (\widetilde{X}_{\underline{t}+\underline{s}(1),M}, \ldots, \widetilde{X}_{\underline{t}+\underline{s}(m_1 m_2),M})^T.$$

With the notations introduced so far we can state the following auxiliary results:

Lemma 7.1. *Let the Assumptions 1 - 4 be fulfilled with* $r = 4$ *and* h *as specified in Assumption 3. Let* $\underline{c} \in \mathbb{R}^k$ *be arbitrary. Then it holds:*

- $$\sum_{\underline{k} \in \Theta} (1 + |\underline{k}|_\infty)^2 \left|\widehat{b}_{\underline{k}}(p)\right| = \mathcal{O}_P(1), \tag{7.2}$$
- $$E^*\Big(|\varepsilon^*_{\underline{t}}|^{2w}\Big) \xrightarrow{P} E\Big(|\varepsilon_{\underline{t}}|^{2w}\Big) \quad \forall\, w \le h + 2, \tag{7.3}$$

- $$\left(X^*_{\underline{t}_1}, \dots, X^*_{\underline{t}_d}\right)^T \xrightarrow{d^*} \left(\widetilde{X}_{\underline{t}_1}, \dots, \widetilde{X}_{\underline{t}_d}\right)^T \quad \text{in } P\text{-prob.} \tag{7.4}$$
  for all $d \geq 1$ *and all* $\underline{t}_1, \dots, \underline{t}_d \in \mathbb{Z}^2$,
- $$E^*\left(\left|\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M})\right|^{2+2/(h+1)}\right) = \mathcal{O}_P(1), \; E\left(\left|\underline{c}^T g(\widetilde{\mathbf{Y}}_{\underline{t},M})\right|^{2+2/(h+1)}\right) \leq C \tag{7.5}$$
  uniformly for all $t \in \mathbb{Z}^2$,
- $$\mathrm{Cov}^*\left(\underline{c}^T g(\mathbf{Y}^*_{\underline{h},M}), \underline{c}^T g(\mathbf{Y}^*_{\underline{0},M})\right) \xrightarrow{P} \mathrm{Cov}\left(\underline{c}^T g(\widetilde{\mathbf{Y}}_{\underline{h},M}), \underline{c}^T g(\widetilde{\mathbf{Y}}_{\underline{0},M})\right) \tag{7.6}$$
  for all $\underline{h} \in \mathbb{Z}^2$.
- *The series* $\Sigma^{(u,v)} := \sum_{\underline{h} \in \mathbb{Z}^2} \mathrm{Cov}\left(g_u(\widetilde{\mathbf{Y}}_{\underline{h}}), g_v(\widetilde{\mathbf{Y}}_{\underline{0}})\right)$ *converges* (7.7)
  absolutely for all $1 \leq u, v \leq k$.

The following auxiliary result will also be used several times:

Lemma 7.2. *Let the Assumptions 1 - 4 be fulfilled with* $r = 4$. *Let* $W \subset \Theta \cup \{\underline{0}\}$ *be any subset of vectors in the upper half-plane* Θ *or in the origin. We define* $\widetilde{\mathbf{Y}}^{(W)}_{\underline{t}}$ *and* $\mathbf{Y}^{*(W)}_{\underline{t}}$ *to be truncated versions of* $\widetilde{\mathbf{Y}}_{\underline{t}}$ *and* $\mathbf{Y}^*_{\underline{t}}$, *respectively, where*

$$\widetilde{\mathbf{Y}}^{(W)}_{\underline{t}} := (\widetilde{X}^{(W)}_{\underline{t}+\underline{s}(1)}, \dots, \widetilde{X}^{(W)}_{\underline{t}+\underline{s}(m_1 m_2)})^T, \quad \mathbf{Y}^{*(W)}_{\underline{t}} := (X^{*(W)}_{\underline{t}+\underline{s}(1)}, \dots, X^{*(W)}_{\underline{t}+\underline{s}(m_1 m_2)})^T,$$

and

$$\widetilde{X}^{(W)}_{\underline{t}} := \sum_{\underline{k} \in W \setminus \{\underline{0}\}} b_{\underline{k}}\, \widetilde{\varepsilon}_{\underline{t}-\underline{k}} + \widetilde{\varepsilon}_{\underline{t}}\, \mathbb{1}_{\{\underline{0} \in W\}}, \quad X^{*(W)}_{\underline{t}} := \sum_{\underline{k} \in W \setminus \{\underline{0}\}} \widehat{b}_{\underline{k}}(p)\, \varepsilon^*_{\underline{t}-\underline{k}} + \varepsilon^*_{\underline{t}}\, \mathbb{1}_{\{\underline{0} \in W\}}.$$

Then there exists $C < \infty$, *such that it holds for any* $\underline{t} \in \mathbb{Z}^2$ *and any* $v = 1, \dots, k$

$$\begin{aligned} \left\| g_v(\widetilde{\mathbf{Y}}_{\underline{t}}) - g_v(\widetilde{\mathbf{Y}}^{(W)}_{\underline{t}}) \right\|_2 &\leq C \cdot \left(\sum_{\underline{k} \in \Theta \setminus W} |b_{\underline{k}}| + \mathbb{1}_{\{\underline{0} \notin W\}} \right), \\ \left\| g_v(\mathbf{Y}^*_{\underline{t}}) - g_v(\mathbf{Y}^{*(W)}_{\underline{t}}) \right\|_{*2} &\leq \mathcal{O}_P(1) \cdot \left(\sum_{\underline{k} \in \Theta \setminus W} \left|\widehat{b}_{\underline{k}}(p)\right| + \mathbb{1}_{\{\underline{0} \notin W\}} \right), \end{aligned}$$

where $\|z\|_2 := \left(E(z)^2\right)^{1/2}$ *and* $\|z\|_{*2} := \left(E^*(z)^2\right)^{1/2}$ *denote the usual* L^2*-norms.*

The previous lemma explicitly incorporates the two cases $\underline{0} \in W$ and $\underline{0} \notin W$, both of which will be needed in the proofs later on. The proofs of the lemmas from this section can be found in section 8, the proof of Theorem 4.2 in section 7.2.

7.1. **Proof of Theorem 3.2.** In order to write the Yule-Walker equations (3.3) in compact form we denoted $\bar{p} = 2p(p+1)$ and introduced the arbitrary but fixed enumeration $\underline{k}_1, \dots, \underline{k}_{\bar{p}}$ of the vectors $\underline{k} \in \Theta(p)$. Now we extend this enumeration to the infinite but countable set Θ, by choosing an arbitrary enumeration $\underline{k}_{\bar{p}+1}, \underline{k}_{\bar{p}+2}, \dots$ of the vectors $\underline{k} \in \Theta \setminus \Theta(p)$ such that

$$\Theta = \{\underline{k}_1, \dots, \underline{k}_{\bar{p}}\} \cup \{\underline{k}_{\bar{p}+1}, \underline{k}_{\bar{p}+2}, \dots\}.$$

While the finite predictor coefficients $(a_{\underline{k}}(p))_{\underline{k}\in\Theta(p)}$ are given by (3.3), Lemma 2.1 shows that the autoregressive coefficients $(a_{\underline{k}})_{\underline{k}\in\Theta}$ determine the L^2-projection of $X_{\underline{t}}$ onto $\overline{sp}\{X_{\underline{t}-\underline{k}} : \underline{k} \in \Theta\}$. Therefore, $X_{\underline{t}} - \sum_{\underline{k}\in\Theta} a_{\underline{k}}\, X_{\underline{t}-\underline{k}}$ is orthogonal to each $X_{\underline{t}-\underline{s}}$, $\underline{s} \in \Theta$. Equivalently, with the introduced enumeration of Θ this means

$$\operatorname{Cov}\Big(X_{\underline{t}} - \sum_{j=1}^{\infty} a_{\underline{k}_j}\, X_{\underline{t}-\underline{k}_j}, X_{\underline{t}-\underline{k}_m}\Big) = \gamma(\underline{k}_m) - \sum_{j=1}^{\infty} a_{\underline{k}_j}\, \gamma(\underline{k}_m - \underline{k}_j) = 0 \quad \forall\, m \in \mathbb{N}.$$

From this system of equations we consider only those ones with $m = 1, \ldots, \bar{p}$, which is equivalent to

$$\Gamma(p) \cdot \begin{pmatrix} a_{\underline{k}_1} \\ \vdots \\ a_{\underline{k}_{\bar{p}}} \end{pmatrix} + \sum_{j=\bar{p}+1}^{\infty} a_{\underline{k}_j} \begin{pmatrix} \gamma(\underline{k}_1 - \underline{k}_j) \\ \vdots \\ \gamma(\underline{k}_{\bar{p}} - \underline{k}_j) \end{pmatrix} = \begin{pmatrix} \gamma(\underline{k}_1) \\ \vdots \\ \gamma(\underline{k}_{\bar{p}}) \end{pmatrix}.$$

Since the right-hand sides of this system and (3.3) coincide, we can infer

$$\begin{pmatrix} a_{\underline{k}_1}(p) - a_{\underline{k}_1} \\ \vdots \\ a_{\underline{k}_{\bar{p}}}(p) - a_{\underline{k}_{\bar{p}}} \end{pmatrix} = \Gamma(p)^{-1} \cdot \sum_{j=\bar{p}+1}^{\infty} a_{\underline{k}_j} \begin{pmatrix} \gamma(\underline{k}_1 - \underline{k}_j) \\ \vdots \\ \gamma(\underline{k}_{\bar{p}} - \underline{k}_j) \end{pmatrix}. \tag{7.8}$$

In the following we will denote the (n, r)-th entry of $\Gamma(p)^{-1}$ by $(\Gamma(p)^{-1})^{(n,r)}$. We are interested in a weighted sum of the absolute values of the entries on the left-hand side of (7.8). For $s \in \mathbb{N}_0$ such that $s + 1 < r$ we get

$$\begin{aligned} &\sum_{n=1}^{\bar{p}} (1 + |\underline{k}_n|_\infty)^s\, |a_{\underline{k}_n}(p) - a_{\underline{k}_n}| \\ = \;& \sum_{n=1}^{\bar{p}} (1 + |\underline{k}_n|_\infty)^s \Big| \sum_{j=\bar{p}+1}^{\infty} a_{\underline{k}_j} \sum_{r=1}^{\bar{p}} (\Gamma(p)^{-1})^{(n,r)}\, \gamma(\underline{k}_r - \underline{k}_j) \Big| \\ \leq \;& \sum_{j=\bar{p}+1}^{\infty} |a_{\underline{k}_j}| \sum_{r=1}^{\bar{p}} |\gamma(\underline{k}_r - \underline{k}_j)| \max_{r=1,\ldots,\bar{p}} \sum_{n=1}^{\bar{p}} (1 + |\underline{k}_n|_\infty)^s \left|(\Gamma(p)^{-1})^{(n,r)}\right| \end{aligned} \tag{7.9}$$

We denote the max-column-sum norm of an arbitrary $n \times n$-matrix B by $\|B\|_1 = \max_{j=1,\ldots,n} \sum_{i=1}^{n} |B^{(i,j)}|$. It is well-known that $\|\cdot\|_1$ is submultiplicative which allows us to derive

$$\begin{aligned} &\max_{r=1,\ldots,\bar{p}} \sum_{n=1}^{\bar{p}} (1 + |\underline{k}_n|_\infty)^s \left|(\Gamma(p)^{-1})^{(n,r)}\right| \\ = \;& \Big\| \operatorname{diag}\Big[(1 + |\underline{k}_1|_\infty)^s, \ldots, (1 + |\underline{k}_{\bar{p}}|_\infty)^s\Big] \cdot \Gamma(p)^{-1} \Big\|_1 \\ \leq \;& \max_{n=1,\ldots,\bar{p}} (1 + |\underline{k}_n|_\infty)^s \cdot \left\|\Gamma(p)^{-1}\right\|_1. \end{aligned}$$

Hence, (7.9) can be bounded from above by

$$\left\|\Gamma(p)^{-1}\right\|_1 \cdot \sum_{j=\bar{p}+1}^{\infty} \max_{n=1,\ldots,\bar{p}} (1 + |\underline{k}_n|_\infty)^s\, |a_{\underline{k}_j}| \sum_{r=1}^{\bar{p}} |\gamma(\underline{k}_r - \underline{k}_j)|$$

$$\leq \left\|\Gamma(p)^{-1}\right\|_1 \sum_{\underline{k}\in\mathbb{Z}^2} |\gamma(\underline{k})| \cdot \sum_{j=\bar{p}+1}^{\infty} \max_{n=1,\ldots,\bar{p}} (1+|\underline{k}_n|_\infty)^s \, |a_{\underline{k}_j}|.$$

Since our numeration was chosen such that $\Theta(p) = \{\underline{k}_1, \ldots, \underline{k}_{\bar{p}}\}$ and $\Theta \setminus \Theta(p) = \{\underline{k}_{\bar{p}+1}, \underline{k}_{\bar{p}+2}, \ldots\}$, the inequality derived so far reads

$$\begin{aligned} &\sum_{\underline{k}\in\Theta(p)} (1+|\underline{k}|_\infty)^s \, |a_{\underline{k}}(p) - a_{\underline{k}}| \\ \leq\;& \left\|\Gamma(p)^{-1}\right\|_1 \sum_{\underline{k}\in\mathbb{Z}^2} |\gamma(\underline{k})| \cdot \sum_{\underline{k}\in\Theta\setminus\Theta(p)} \max_{\underline{v}\in\Theta(p)} (1+|\underline{v}|_\infty)^s \, |a_{\underline{k}}|. \end{aligned} \tag{7.10}$$

Per definition of $\Theta(p)$ we have

$$\max_{\underline{v}\in\Theta(p)} (1+|\underline{v}|_\infty)^s = (1+p)^s \leq (1+|\underline{k}|_\infty)^s \quad \forall\, \underline{k} \in \Theta\setminus\Theta(p), \tag{7.11}$$

as $|\underline{k}|_\infty \geq p+1$ for all $\underline{k} \in \Theta \setminus \Theta(p)$; this is why we need a weight function strictly nondecreasing in $|\underline{k}|_\infty$. Furthermore, it holds $\|A\|_1 \leq \sqrt{n}\,\|A\|_{\text{spec}}$ for all $n \times n$-matrices A, i.e. $\|\Gamma(p)^{-1}\|_1 \leq \sqrt{2p(p+1)}\,\|\Gamma(p)^{-1}\|_{\text{spec}}$ and

$$\sqrt{2p(p+1)} \leq \sqrt{2}\,(p+1) < \sqrt{2}\,(1+|\underline{k}|_\infty) \quad \forall\, \underline{k} \in \Theta\setminus\Theta(p).$$

Therefore, and due to (7.11) and Lemma 3.1, (7.10) can be bounded by

$$\begin{aligned} &\sqrt{2p(p+1)}\left\|\Gamma(p)^{-1}\right\|_{\text{spec}} \sum_{\underline{k}\in\mathbb{Z}^2} |\gamma(\underline{k})| \cdot \sum_{\underline{k}\in\Theta\setminus\Theta(p)} (1+|\underline{k}|_\infty)^s \, |a_{\underline{k}}| \\ \leq\;& \frac{1}{2\sqrt{2}\,\pi^2 c} \sum_{\underline{k}\in\mathbb{Z}^2} |\gamma(\underline{k})| \cdot \sum_{\underline{k}\in\Theta\setminus\Theta(p)} (1+|\underline{k}|_\infty)^{s+1} \, |a_{\underline{k}}|, \end{aligned}$$

which completes the proof. □

7.2. **Proof of Theorem 4.2.** The basic structure of this proof resembles the proof of Theorem 3.3 in Bühlmann (1997). At first, we will neglect the outer function f in T_n^* and show for the bootstrap quantities

$$(\bar{n}_1\bar{n}_2)^{-1/2} \sum_{t_1=1}^{\bar{n}_1} \sum_{t_2=1}^{\bar{n}_2} \Big(g(\mathbf{Y}^*_{\underline{t}}) - E^*\Big(g(\mathbf{Y}^*_{\underline{t}})\Big)\Big) \xrightarrow{d^*} \mathcal{N}(\underline{0}, \Sigma) \text{ in prob.,} \tag{7.12}$$

where the entries of Σ are given by $\Sigma^{(u,v)} := \sum_{\underline{h}\in\mathbb{Z}^2} \operatorname{Cov}\Big(g_u(\widetilde{\mathbf{Y}}_{\underline{h}}), g_v(\widetilde{\mathbf{Y}}_{\underline{0}})\Big)$, for $u, v = 1, \ldots, k$. Note that (7.7) guarantees that this expression is well-defined. Since the companion process $(\widetilde{X}_{\underline{t}})$, just as the bootstrap process, is a linear spatial process (recall that the innovations $(\widetilde{\varepsilon}_{\underline{t}})$ are i.i.d.), one can follow the exact same arguments as in proving (7.12) to derive

$$(\bar{n}_1\bar{n}_2)^{-1/2} \sum_{t_1=1}^{\bar{n}_1} \sum_{t_2=1}^{\bar{n}_2} \Big(g(\widetilde{\mathbf{Y}}_{\underline{t}}) - E\Big(g(\widetilde{\mathbf{Y}}_{\underline{t}})\Big)\Big) \xrightarrow{d} \mathcal{N}(\underline{0}, \Sigma) \tag{7.13}$$

with the very same limiting distribution as above. We will therefore restrict ourselves to providing a thorough reasoning of (7.12) and omit the proof of the CLT for the companion process. In the end we will incorporate the function f by applying the delta method to both CLT's which will complete the proof of Theorem 4.2 since $(\bar{n}_1\bar{n}_2)^{1/2}$ and n are asymptotically equivalent.

The strategy for proving (7.12) is the following: Firstly, one can observe that (7.12) follows if we can show

$$\frac{1}{n}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\Big(g(\mathbf{Y}^*_{\underline{t}}) - E^*\Big(g(\mathbf{Y}^*_{\underline{t}})\Big)\Big) \xrightarrow{d^*} \mathcal{N}(\underline{0},\Sigma) \text{ in prob.,} \tag{7.14}$$

since the expressions in both assertions are asymptotically equivalent per definition of $\bar{n}_1, \bar{n}_2$. We will invoke the Cramér-Wold device and, in the first step, consider the truncated quantity $(\mathbf{Y}^*_{\underline{t},M})$ based on the truncated process $(X^*_{\underline{t},M})$ introduced in (7.1). For arbitrary $M \in \mathbb{N}$ and for all $\underline{c} \in \mathbb{R}^k$ we will show

$$\frac{1}{n}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M}) - E^*\Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M})\Big)\Big) \xrightarrow{d^*} \mathcal{N}\Big(0, \underline{c}^T\Sigma_M\,\underline{c}\Big) \text{ in prob.,} \tag{7.15}$$

where

$$\Sigma_M^{(u,v)} := \sum\nolimits_{h_1=-2M-m_1+1}^{2M+m_1-1}\sum\nolimits_{h_2=-M-m_2+1}^{M+m_2-1}\mathrm{Cov}\Big(g_u(\widetilde{\mathbf{Y}}_{\underline{h},M}), g_v(\widetilde{\mathbf{Y}}_{\underline{0},M})\Big).$$

In order to establish the limiting variance in (7.15), we first recall that $(X^*_{\underline{t},M})$ is strictly stationary as can be seen from (7.1) and the (conditional) i.i.d. property of $(\varepsilon^*_{\underline{t}})$. Therefore, $(\mathbf{Y}^*_{\underline{t},M})$, and consequently $(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M}))$, are also strictly stationary processes in $\underline{t}$. Hence, standard calculations yield

$$\begin{aligned}
&\mathrm{Var}^*\Big(\frac{1}{n}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M}) - E^*\Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M})\Big)\Big)\Big)\\
=\;& \frac{1}{n^2}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{v_1=1}^{n}\sum_{v_2=1}^{n}\mathrm{Cov}^*\Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M}), \underline{c}^T g(\mathbf{Y}^*_{\underline{v},M})\Big)\\
=\;& \sum_{h_1=-(n-1)}^{n-1}\sum_{h_2=-(n-1)}^{n-1}\Big(1-\frac{|h_1|}{n}\Big)\Big(1-\frac{|h_2|}{n}\Big)\mathrm{Cov}^*\Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{h},M}), \underline{c}^T g(\mathbf{Y}^*_{\underline{0},M})\Big)
\end{aligned}$$

A close inspection of the definition of $\mathbf{Y}^*_{\underline{t},M}$, which depends only on a finite number of random variables $\varepsilon^*_{\underline{t}+\underline{j}}$, together with the i.i.d. property of $(\varepsilon^*_{\underline{t}})$, shows that $g(\mathbf{Y}^*_{\underline{h},M})$ and $g(\mathbf{Y}^*_{\underline{0},M})$ are independent whenever $|h_1| \geq 2M+m_1$ or $|h_2| \geq M+m_2$. Therefore, for all $n \geq \min\{2M+m_1, M+m_2\}$ the last right-hand side equals

$$\begin{aligned}
&\sum_{h_1=-2M-m_1+1}^{2M+m_1-1}\sum_{h_2=-M-m_2+1}^{M+m_2-1}\Big(1-\frac{|h_1|}{n}\Big)\Big(1-\frac{|h_2|}{n}\Big)\mathrm{Cov}^*\Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{h},M}), \underline{c}^T g(\mathbf{Y}^*_{\underline{0},M})\Big)\\
=\;& \underline{c}^T\Sigma_M\,\underline{c} + o_P(1),
\end{aligned}$$

as can be seen from (7.6). This establishes the correct asymptotic variance in (7.15), and by abbreviating

$$v_n^* := \mathrm{Var}^*\Big(\frac{1}{n}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M}) - E^*\Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M})\Big)\Big)\Big)$$

assertion (7.15) follows from Slutsky's Theorem if we can show

$$\frac{1}{n\sqrt{v_n^*}}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M}) - E^*\Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M})\Big)\Big) \xrightarrow{d^*} \mathcal{N}(0,1) \text{ in prob.} \tag{7.16}$$

The strategy is to use a blocking technique. We define sequences of integers $a(n), b(n) \in \mathbb{N}$ with $a(n) \to \infty$, $b(n) \to \infty$ and $b(n)/a(n) \to 0$ as $n \to \infty$. Also, we assume that $a(n)$ and $b(n)$ increase slowly enough such that

$$N(n) := \frac{n}{a(n)+b(n)} \longrightarrow \infty, \quad \text{as } n \to \infty,$$

and assume without loss of generality $N(n) \in \mathbb{N}$ for all n. The idea is to split up the n^2 summands in (7.16) into dominating, square-shaped blocks A_{j_1,j_2} of size $a(n) \times a(n)$, and negligible remainder terms B_{j_1,j_2} and C_{j_1,j_2}. In the following, we will often abbreviate $a = a(n)$, $b = b(n)$ and $N = N(n)$ in order to simplify the notation. (7.16) can be decomposed as

$$\frac{1}{n\sqrt{v_n^*}}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}\Big(A_{j_1,j_2} + B_{j_1,j_2} + C_{j_1,j_2}\Big) \xrightarrow{d^*} \mathcal{N}(0,1) \text{ in prob.}, \tag{7.17}$$

where

$$\begin{aligned}
A_{j_1,j_2} &:= \sum_{t_1=(j_1-1)(a+b)+1}^{j_1a+(j_1-1)b}\ \sum_{t_2=(j_2-1)(a+b)+1}^{j_2a+(j_2-1)b}\Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M}) - E^*\Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M})\Big)\Big),\\
B_{j_1,j_2} &:= \sum_{t_1=j_1a+(j_1-1)b+1}^{j_1(a+b)}\ \sum_{t_2=(j_2-1)(a+b)+1}^{j_2a+(j_2-1)b}\Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M}) - E^*\Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M})\Big)\Big),\\
C_{j_1,j_2} &:= \sum_{t_1=(j_1-1)(a+b)+1}^{j_1(a+b)}\ \sum_{t_2=j_2a+(j_2-1)b+1}^{j_2(a+b)}\Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M}) - E^*\Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M})\Big)\Big).
\end{aligned}$$

We will now establish moment bounds for these three expressions. For the constant h from Assumption 3, define $\delta := 2/(3h+3)$. Writing $A_{1,1} := \sum_{t_1}\eta_{t_1}$ with

$$\eta_{t_1} := \sum_{t_2=1}^{a}\Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M}) - E^*\Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M})\Big)\Big), \quad t_1 = 1,\ldots,a\ ,$$

we get from Theorem 1 in Yokoyama (1980) that $E^*(|\eta_{t_1}|^{2+2\delta}) = a(n)^{1+\delta}\,\mathcal{O}_P(1)$ uniformly in t_1, since η_{t_1} consists of $a(n)$ summands which are centered, $(M+m_2)$-dependent in t_2 and fulfil the required moment assumption $E^*(|\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M})|^{2+3\delta}) = \mathcal{O}_P(1)$ (uniformly) due to (7.5). In other words, we get

$$E^*\Big(\Big|\frac{\eta_{t_1}}{a(n)^{1/2}}\Big|^{2+2\delta}\Big) = \mathcal{O}_P(1),$$

i.e. the $(2M+m_1)$-dependent sequence $(\eta_{t_1}/a(n)^{1/2})_{t_1}$ itself fulfils the conditions of Theorem 1 in Yokoyama (1980) which yields

$$E^*\Big(|A_{1,1}|^{2+\delta}\Big) = a(n)^{1+(\delta/2)} \cdot E^*\Big(\Big|\sum_{t_1=1}^{a} \frac{\eta_{t_1}}{a(n)^{1/2}}\Big|^{2+\delta}\Big) = \Big(a(n)\cdot a(n)\Big)^{1+(\delta/2)}\,\mathcal{O}_P(1).$$

With analogous calculations we get the bounds

$$\begin{aligned} E^*\Big(|B_{1,1}|^{2+\delta}\Big) &= \Big(a(n)\cdot b(n)\Big)^{1+(\delta/2)}\,\mathcal{O}_P(1), \\ E^*\Big(|C_{1,1}|^{2+\delta}\Big) &= \Big((a(n)+b(n))\cdot b(n)\Big)^{1+(\delta/2)}\,\mathcal{O}_P(1). \end{aligned}$$

Note that we can not apply Yokoyama's Theorem directly to $A_{1,1}$, but have to take the intermediate step with η_{t_1} as performed above, because the a^2 many summands in $A_{1,1}$ can *not* be transformed into an M-dependent sequence, regardless of the ordering of the summands.

In the preceeding part of this proof we have shown $v^* = \underline{c}^T\Sigma_M\,\underline{c} + o_P(1)$, and we can assume $\underline{c}^T\Sigma_M\,\underline{c} \neq 0$ (in the case $\underline{c}^T\Sigma_M\,\underline{c} = 0$ the desired assertion (7.15) would follow trivially). Hence, we have $1/v^* = \mathcal{O}_P(1)$. We will now use this assertion, as well as the established moment bounds, to show that B_{j_1,j_2} and C_{j_1,j_2} in (7.17) are asymptotically negligible. The blocks B_{j_1,j_2} are identically distributed and, for n large enough such that $a(n), b(n) > 2M$, independent. Standard calculations yield for any $\varepsilon > 0$:

$$\begin{aligned} P^*\Big\{\Big|\frac{1}{n\sqrt{v_n^*}}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N} B_{j_1,j_2}\Big| > \varepsilon\Big\} &\le \frac{1}{n^2\varepsilon^2 v^*}\mathrm{Var}^*\Big(\sum_{j_1=1}^{N}\sum_{j_2=1}^{N} B_{j_1,j_2}\Big) \\ &\le \frac{N^2}{n^2\varepsilon^2 v^*}\Big(E^*\Big(|B_{1,1}|^{2+\delta}\Big)\Big)^{2/(2+\delta)} \\ &\le \frac{a(n)\cdot b(n)}{(a(n)+b(n))^2}\,\mathcal{O}_P(1) = o_P(1). \qquad (7.18) \end{aligned}$$

The same assertion can be shown for the C_{j_1,j_2}-blocks. Therefore, Slutsky's Theorem implies (7.17) if we can show

$$\frac{\tau_N}{n\sqrt{v_n^*}} \cdot \frac{1}{\tau_N}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N} A_{j_1,j_2} \xrightarrow{d^*} \mathcal{N}(0,1) \text{ in prob.}, \qquad (7.19)$$

where

$$\tau_N := \Big(\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}\mathrm{Var}^*(A_{j_1,j_2})\Big)^{1/2}.$$

Observe that, for n large enough such that $b(n) > 2M$, the A_{j_1,j_2}-blocks are independent random variables, and in the following we will only consider those n. Per definition, we can decompose

$$\begin{aligned}\frac{\tau_N^2}{n^2} &= v_n^* + \operatorname{Var}^*\Big(\frac{1}{n}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}(B_{j_1,j_2}+C_{j_1,j_2})\Big)\\ &\quad -2\operatorname{Cov}^*\Big(\frac{1}{n}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}(A_{j_1,j_2}+B_{j_1,j_2}+C_{j_1,j_2}),\frac{1}{n}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}(B_{j_1,j_2}+C_{j_1,j_2})\Big)\\ &= v_n^* + o_P(1),\end{aligned}\tag{7.20}$$

which follows from $n^{-2}\operatorname{Var}^*(\sum_{j_1}\sum_{j_2}(B_{j_1,j_2}+C_{j_1,j_2})) = o_P(1)$, cf. the calculations leading up to (7.18). Therefore, the first factor on the left-hand side of (7.19) converges to 1 in probability.

We will apply Lindeberg's central limit theorem to the second factor in (7.19), recalling that the A_{j_1,j_2}-blocks are i.i.d. random variables for n large enough. Using $N = (n/a(n))\cdot\mathcal{O}(1)$ and the moment condition for $A_{1,1}$ established earlier, as well as the fact that (7.20) implies $n/\tau_N = \mathcal{O}_P(1)$, we can check the Lyapunov condition:

$$\begin{aligned}&\frac{1}{\tau_N^{2+\delta}}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}E^*\Big(|A_{j_1,j_2}|^{2+\delta}\Big) = \frac{N^2}{\tau_N^{2+\delta}}\,a(n)^{2+\delta}\,\mathcal{O}_P(1)\\ =\;& \Big(\frac{n}{\tau_N}\Big)^2\Big(\frac{a(n)}{n}\cdot\frac{n}{\tau_N}\Big)^{\delta}\mathcal{O}_P(1) = \Big(\frac{n}{\tau_N}\Big)^{2+\delta}\Big(\frac{1}{N(n)}\Big)^{\delta}\mathcal{O}_P(1) = o_P(1).\end{aligned}$$

This yields (7.19) and, consequently, (7.15).

We will invoke Proposition 6.3.9 of Brockwell and Davis (1991) to show that (7.15) implies (7.14). The next step is to prove $\Sigma_M^{(u,v)} \to \Sigma^{(u,v)}$, as $M\to\infty$. It holds

$$\begin{aligned}\Sigma_M^{(u,v)} &= \sum_{h_1=-2M-m_1+1}^{2M+m_1-1}\sum_{h_2=-M-m_2+1}^{M+m_2-1}\operatorname{Cov}\Big(g_u(\widetilde{\mathbf{Y}}_{\underline{h}}),g_v(\widetilde{\mathbf{Y}}_{\underline{0}})\Big)\\ &+\sum_{h_1=-2M-m_1+1}^{2M+m_1-1}\sum_{h_2=-M-m_2+1}^{M+m_2-1}\Big(\operatorname{Cov}\Big(g_u(\widetilde{\mathbf{Y}}_{\underline{h},M}),g_v(\widetilde{\mathbf{Y}}_{\underline{0},M})\Big)-\operatorname{Cov}\Big(g_u(\widetilde{\mathbf{Y}}_{\underline{h}}),g_v(\widetilde{\mathbf{Y}}_{\underline{0}})\Big)\Big).\end{aligned}$$

The first summand on the right-hand side converges to $\Sigma^{(u,v)}$, as $M\to\infty$, due to (7.7). As for the second summand, we have

$$\begin{aligned}&\sum_{h_1=-2M-m_1+1}^{2M+m_1-1}\sum_{h_2=-M-m_2+1}^{M+m_2-1}\Big|\operatorname{Cov}\Big(g_u(\widetilde{\mathbf{Y}}_{\underline{h},M}),g_v(\widetilde{\mathbf{Y}}_{\underline{0},M})\Big)-\operatorname{Cov}\Big(g_u(\widetilde{\mathbf{Y}}_{\underline{h}}),g_v(\widetilde{\mathbf{Y}}_{\underline{0}})\Big)\Big|\\ \le&\sum_{h_1}\sum_{h_2}\Big|\operatorname{Cov}\Big(g_u(\widetilde{\mathbf{Y}}_{\underline{h},M})-g_u(\widetilde{\mathbf{Y}}_{\underline{h}}),g_v(\widetilde{\mathbf{Y}}_{\underline{0},M})\Big)+\operatorname{Cov}\Big(g_u(\widetilde{\mathbf{Y}}_{\underline{h}}),g_v(\widetilde{\mathbf{Y}}_{\underline{0},M})-g_v(\widetilde{\mathbf{Y}}_{\underline{0}})\Big)\Big|\\ \le&\sum_{h_1}\sum_{h_2}\Big(\Big\|g_u(\widetilde{\mathbf{Y}}_{\underline{h},M})-g_u(\widetilde{\mathbf{Y}}_{\underline{h}})\Big\|_2\Big\|g_v(\widetilde{\mathbf{Y}}_{\underline{0},M})\Big\|_2+\Big\|g_u(\widetilde{\mathbf{Y}}_{\underline{h}})\Big\|_2\Big\|g_v(\widetilde{\mathbf{Y}}_{\underline{0},M})-g_v(\widetilde{\mathbf{Y}}_{\underline{0}})\Big\|_2\Big)\\ \le&\, C\cdot\sum_{h_1=-2M-m_1+1}^{2M+m_1-1}\sum_{h_2=-M-m_2+1}^{M+m_2-1}\sum_{\underline{k}\in\Theta\setminus\Theta(M)}|b_{\underline{k}}|,\end{aligned}\tag{7.21}$$

for some generic constant $C < \infty$ (which, from now on, may change from line to line). The latter inequality can be derived from Lemma 7.2 (with $W = \Theta(M) \cup \{\underline{0}\}$), using the fact that

$$\left\| g_u(\widetilde{\mathbf{Y}}_{\underline{h}}) \right\|_2 \leq \left\| g_u(\widetilde{\mathbf{Y}}_{\underline{h}}) - g_u(\widetilde{\mathbf{Y}}_{\underline{h}}^{(\emptyset)}) \right\|_2 + \left\| g_u(\underline{0}) \right\|_2 \leq 1 + \sum_{\underline{k} \in \Theta} |b_{\underline{k}}| + \left| g_u(\underline{0}) \right| < \infty$$

follows also from Lemma 7.2 (with $W = \emptyset$), the same being true for $\|g_v(\widetilde{\mathbf{Y}}_{\underline{0},M})\|_2$. For (7.21) it holds

$$\begin{aligned}
& C \cdot \sum_{h_1=-2M-m_1+1}^{2M+m_1-1} \sum_{h_2=-M-m_2+1}^{M+m_2-1} \sum_{\underline{k} \in \Theta \setminus \Theta(M)} |b_{\underline{k}}| \\
\leq\ & C \cdot \sum_{\underline{k} \in \Theta \setminus \Theta(M)} M^2 \, |b_{\underline{k}}| \\
\leq\ & C \cdot \left(\sum_{k_1=M+1}^{\infty} \sum_{k_2=0}^{M} |k_1|^2 \, |b_{\underline{k}}| + \sum_{k_1=-\infty}^{-M-1} \sum_{k_2=0}^{M} |k_1|^2 \, |b_{\underline{k}}| + \sum_{k_1=-\infty}^{\infty} \sum_{k_2=M+1}^{\infty} |k_2|^2 \, |b_{\underline{k}}| \right) \\
\leq\ & C \cdot \sum_{\underline{k} \in \Theta \setminus \Theta(M)} (1 + |\underline{k}|_\infty)^2 \, |b_{\underline{k}}|,
\end{aligned}$$

which converges to zero, as $M \to \infty$, due to (2.6) and the assumption $r = 4$. Hence, we have shown that $\Sigma_M^{(u,v)} \to \Sigma^{(u,v)}$, as $M \to \infty$.

Now we apply Proposition 6.3.9 of Brockwell and Davis (1991) to the bootstrap quantities from (7.15), i.e. we show that it holds

$$\lim_{M \to \infty} \limsup_{n \to \infty} P^* \left\{ \left| \frac{1}{n} \sum_{t_1,t_2=1}^{n} \underline{c}^T \Big(g(\mathbf{Y}_{\underline{t}}^*) - g(\mathbf{Y}_{\underline{t},M}^*) - E^* \Big(g(\mathbf{Y}_{\underline{t}}^*) - g(\mathbf{Y}_{\underline{t},M}^*) \Big) \Big) \right| > \delta \right\} = 0$$
$$\text{in } P\text{-probability}, \qquad (7.22)$$

for any $\delta > 0$. Then (7.14) follows from said Proposition 6.3.9, using the Cramér-Wold device. For condition (7.22) to hold, it is sufficient to show

$$\operatorname{Var}^* \left(\frac{1}{n} \sum_{t_1=1}^{n} \sum_{t_2=1}^{n} \Big(\underline{c}^T g(\mathbf{Y}_{\underline{t}}^*) - \underline{c}^T g(\mathbf{Y}_{\underline{t},M}^*) \Big) \right) \leq \frac{1}{M^2} \, \mathcal{O}_P(1). \qquad (7.23)$$

We abbreviate $Z_{\underline{t},M}^* := \underline{c}^T g(\mathbf{Y}_{\underline{t}}^*) - \underline{c}^T g(\mathbf{Y}_{\underline{t},M}^*)$ and observe that $(Z_{\underline{t},M}^*)$ is a stationary spatial process. Standard calculations yield

$$\begin{aligned}
\operatorname{Var}^* \left(\frac{1}{n} \sum_{t_1=1}^{n} \sum_{t_2=1}^{n} Z_{\underline{t},M}^* \right) &= \sum_{h_1=-(n-1)}^{n-1} \sum_{h_2=-(n-1)}^{n-1} \frac{(n - |h_1|)(n - |h_2|)}{n^2} \operatorname{Cov}^*(Z_{\underline{0},M}^*, Z_{\underline{h},M}^*) \\
&\leq \sum_{h_1=-\infty}^{\infty} \sum_{h_2=-\infty}^{\infty} \left| \operatorname{Cov}^*(Z_{\underline{0},M}^*, Z_{\underline{h},M}^*) \right| \\
&\leq 2 \sum_{h_1=-\infty}^{\infty} \sum_{h_2=0}^{\infty} \left| \operatorname{Cov}^*(Z_{\underline{0},M}^*, Z_{\underline{h},M}^*) \right|,
\end{aligned}$$

noting that $\operatorname{Cov}^*(Z^*_{\underline{0},M}, Z^*_{\underline{h},M}) = \operatorname{Cov}^*(Z^*_{\underline{0},M}, Z^*_{-\underline{h},M})$. Hence, to obtain (7.23), it suffices to show

$$\sum_{h_1=0}^{\infty}\sum_{h_2=0}^{\infty}\left|\operatorname{Cov}^*(Z^*_{\underline{0},M}, Z^*_{\underline{h},M})\right| \leq \frac{1}{M^2}\,\mathcal{O}_P(1), \tag{7.24}$$

since the remaining part $\sum_{h_1=-\infty}^{-1}\sum_{h_2=0}^{\infty}|\operatorname{Cov}^*(Z^*_{\underline{0},M}, Z^*_{\underline{h},M})|$ can be treated analogously. In order to show (7.24) we will make use of three different truncated versions of $Z^*_{\underline{t},M}$, which we will denote by $Z^{*[1]}_{\underline{t},M},\ldots,Z^{*[3]}_{\underline{t},M}$. The truncation points will depend on h_1 and h_2, the indices showing up in (7.24), which will be suppressed in the notation. Each of the truncated versions is generated in a natural way from truncated versions of $X^*_{\underline{t}}$ and $X^*_{\underline{t},M}$. To be precise, we set

$$Z^{*[j]}_{\underline{t},M} := \underline{c}^T g(\mathbf{Y}^{*[j]}_{\underline{t}}) - \underline{c}^T g(\mathbf{Y}^{*[j]}_{\underline{t},M}), \quad j = 1,2,3\,,$$

where

$$\mathbf{Y}^{*[j]}_{\underline{t}} := (X^{*[j]}_{\underline{t}+\underline{s}(1)},\ldots,X^{*[j]}_{\underline{t}+\underline{s}(m_1m_2)})^T, \quad \mathbf{Y}^{*[j]}_{\underline{t},M} := (X^{*[j]}_{\underline{t}+\underline{s}(1),M},\ldots,X^{*[j]}_{\underline{t}+\underline{s}(m_1m_2),M})^T$$

and, setting $\widehat{b}_{(0,0)}(p) := 1$ and $\widehat{b}_{(k_1,0)}(p) := 0$ for $k_1 < 0$,

$$\begin{aligned}
X^{*[1]}_{\underline{t}} &:= \sum_{k_1=-\infty}^{\infty}\sum_{k_2=0}^{h_2-m_2}\widehat{b}_{\underline{k}}(p)\,\varepsilon^*_{\underline{t}-\underline{k}}\cdot\mathbb{1}_{\{h-m_2\geq 0\}} + \sum_{k_1=-\infty}^{\lfloor h_1/2\rfloor}\sum_{k_2=(h_2-m_2+1)\vee 0}^{\infty}\widehat{b}_{\underline{k}}(p)\,\varepsilon^*_{\underline{t}-\underline{k}},\\
X^{*[2]}_{\underline{t}} &:= \sum_{k_1=-\lfloor h_1/2\rfloor+m_1}^{\infty}\sum_{k_2=0}^{\infty}\widehat{b}_{\underline{k}}(p)\,\varepsilon^*_{\underline{t}-\underline{k}},\\
X^{*[3]}_{\underline{t}} &:= \sum_{k_1=-\infty}^{\infty}\sum_{k_2=0}^{h_2-m_2}\widehat{b}_{\underline{k}}(p)\,\varepsilon^*_{\underline{t}-\underline{k}}\cdot\mathbb{1}_{\{h-m_2\geq 0\}}.
\end{aligned}$$

The versions $X^{*[j]}_{\underline{t},M}$, $j = 1,2,3$, can be obtained from the corresponding definitions of $X^{*[j]}_{\underline{t}}$, by replacing each $\widehat{b}_{\underline{k}}(p)$ with $\widehat{b}_{\underline{k}}(p)\cdot\mathbb{1}_{\{\underline{k}\in\Theta(M)\}}$.
With these definitions we can now split up the expression in (7.24) as

$$\begin{aligned}
&\sum_{h_1=0}^{\infty}\sum_{h_2=0}^{\infty}\left|\operatorname{Cov}^*(Z^*_{\underline{0},M}, Z^*_{\underline{h},M})\right|\\
\leq\; &\sum_{h_1=0}^{\infty}\sum_{h_2=0}^{\infty}\left|\operatorname{Cov}^*(Z^*_{\underline{0},M}, Z^*_{\underline{h},M} - Z^{*[1]}_{\underline{h},M})\right| + \sum_{h_1=0}^{\infty}\sum_{h_2=0}^{\infty}\left|\operatorname{Cov}^*(Z^{*[2]}_{\underline{0},M}, Z^{*[1]}_{\underline{h},M})\right|\\
&+ \sum_{h_1=0}^{\infty}\sum_{h_2=0}^{\infty}\left|\operatorname{Cov}^*(Z^*_{\underline{0},M} - Z^{*[2]}_{\underline{0},M}, Z^{*[3]}_{\underline{h},M})\right|\\
&+ \sum_{h_1=0}^{\infty}\sum_{h_2=0}^{\infty}\left|\operatorname{Cov}^*(Z^*_{\underline{0},M} - Z^{*[2]}_{\underline{0},M}, Z^{*[1]}_{\underline{h},M} - Z^{*[3]}_{\underline{h},M})\right|\\
=:\; & I + II + III + IV.
\end{aligned}$$

A close inspection of the definition of the different truncated versions $Z^{*[j]}_{\underline{0},M}$ and $Z^{*[j]}_{\underline{h},M}$ shows that $Z^{*[2]}_{\underline{0},M}$ and $Z^{*[1]}_{\underline{h},M}$ are independent random variables because they depend on disjoint sets of variables $\varepsilon^*_{\underline{t}}$ (this is why the truncated versions are defined as they are), and the $(\varepsilon^*_{\underline{t}})$ are i.i.d.. With the same argument, $Z^{*[3]}_{\underline{h},M}$ is independent of both $Z^*_{\underline{0},M}$ and $Z^{*[2]}_{\underline{0},M}$. Thus, the expressions II and III are identical zero. We can therefore prove (7.24) by showing

$$I \leq \frac{1}{M^2}\,\mathcal{O}_P(1), \quad IV \leq \frac{1}{M^2}\,\mathcal{O}_P(1). \tag{7.25}$$

Using the notation $\|\cdot\|_{*2}$ introduced in Lemma 7.2, we have for I

$$\sum_{h_1=0}^{\infty}\sum_{h_2=0}^{\infty}\left|\mathrm{Cov}^*(Z^*_{\underline{0},M}, Z^*_{\underline{h},M} - Z^{*[1]}_{\underline{h},M})\right| \leq \left\|Z^*_{\underline{0},M}\right\|_{*2}\sum_{h_1=0}^{\infty}\sum_{h_2=0}^{\infty}\left\|Z^*_{\underline{h},M} - Z^{*[1]}_{\underline{h},M}\right\|_{*2},$$

and

$$\begin{aligned}
&\sum_{h_1=0}^{\infty}\sum_{h_2=0}^{\infty}\left\|Z^*_{\underline{h},M} - Z^{*[1]}_{\underline{h},M}\right\|_{*2}\\
\leq\; &\sum_{h_1=0}^{\infty}\sum_{h_2=0}^{\infty}\left(\left\|\underline{c}^T g(\mathbf{Y}^*_{\underline{h}}) - \underline{c}^T g(\mathbf{Y}^{*[1]}_{\underline{h}})\right\|_{*2} + \left\|\underline{c}^T g(\mathbf{Y}^*_{\underline{h},M}) - \underline{c}^T g(\mathbf{Y}^{*[1]}_{\underline{h},M})\right\|_{*2}\right)\\
\leq\; &\sum_{h_1=0}^{\infty}\sum_{h_2=0}^{\infty}\sum_{u=1}^{k}|c_u|\left(\left\|g_u(\mathbf{Y}^*_{\underline{h}}) - g_u(\mathbf{Y}^{*[1]}_{\underline{h}})\right\|_{*2} + \left\|g_u(\mathbf{Y}^*_{\underline{h},M}) - g_u(\mathbf{Y}^{*[1]}_{\underline{h},M})\right\|_{*2}\right)\\
\leq\; &\sum_{h_1=0}^{\infty}\sum_{h_2=0}^{\infty}\sum_{u=1}^{k}|c_u|\left(\mathcal{O}_P(1)\cdot\sum_{k_1=\lfloor h_1/2\rfloor+1}^{\infty}\;\sum_{k_2=(h_2-m_2+1)\vee 0}^{\infty}\left|\widehat{b}_{\underline{k}}(p)\right|\right)
\end{aligned} \tag{7.26}$$

where the $\mathcal{O}_P(1)$-expression on the right-hand side does not depend on $\underline{h}$, u or M. The latter inequality follows directly from Lemma 7.2. The last right-hand side can be bounded by

$$\begin{aligned}
&\mathcal{O}_P(1)\cdot\sum_{h_1=0}^{\infty}\sum_{h_2=0}^{\infty}\sum_{k_1=\lfloor h_1/2\rfloor+1}^{\infty}\;\sum_{k_2=(h_2-m_2+1)\vee 0}^{\infty}\left|\widehat{b}_{\underline{k}}(p)\right|\\
\leq\; &\mathcal{O}_P(1)\cdot m_2\sum_{h_1=0}^{\infty}\sum_{h_2=m_2-1}^{\infty}\sum_{k_1=\lfloor h_1/2\rfloor+1}^{\infty}\;\sum_{k_2=h_2-m_2+1}^{\infty}\left|\widehat{b}_{\underline{k}}(p)\right|\\
\leq\; &\mathcal{O}_P(1)\cdot m_2\sum_{h_1=0}^{\infty}\sum_{k_1=\lfloor h_1/2\rfloor+1}^{\infty}\sum_{k_2=0}^{\infty}(k_2+1)\left|\widehat{b}_{\underline{k}}(p)\right|\\
\leq\; &\mathcal{O}_P(1)\cdot m_2\sum_{k_1=1}^{\infty}\sum_{k_2=0}^{\infty}2k_1\,(k_2+1)\left|\widehat{b}_{\underline{k}}(p)\right|\\
\leq\; &\mathcal{O}_P(1)\cdot 2m_2\sum_{\underline{k}\in\Theta}(1+|\underline{k}|_\infty)^2\left|\widehat{b}_{\underline{k}}(p)\right| = \mathcal{O}_P(1),
\end{aligned}$$

using (7.2). Thus, we have

$$I \leq \left\|Z^*_{\underline{0},M}\right\|_{*2}\cdot\mathcal{O}_P(1).$$

We also get from Lemma 7.2

$$\begin{aligned} M^2 \cdot \left\| Z^*_{\underline{0},M} \right\|_{*2} &\leq M^2 \cdot \sum_{u=1}^{k} |c_u| \left\| g_u(\mathbf{Y}^*_{\underline{0}}) - g_u(\mathbf{Y}^*_{\underline{0},M}) \right\|_{*2} \\ &\leq \mathcal{O}_P(1) \cdot M^2 \sum_{\underline{k} \in \Theta \setminus \Theta(M)} \left| \widehat{b}_{\underline{k}}(p) \right| \\ &\leq \mathcal{O}_P(1) \cdot \sum_{\underline{k} \in \Theta} (1 + |\underline{k}|_\infty)^2 \left| \widehat{b}_{\underline{k}}(p) \right| = \mathcal{O}_P(1), \end{aligned} \tag{7.27}$$

uniformly for all $M \in \mathbb{N}$, due to (7.2) and $M < |\underline{k}|_\infty$ for all $\underline{k} \in \Theta \setminus \Theta(M)$. It follows

$$I \leq \frac{1}{M^2} \mathcal{O}_P(1).$$

Turning to expression IV, we can decompose

$$\begin{aligned} IV &\leq \sum_{h_1=0}^{M} \sum_{h_2=0}^{M} \left| \mathrm{Cov}^*(Z^*_{\underline{0},M} - Z^{*[2]}_{\underline{0},M}, Z^{*[1]}_{\underline{h},M} - Z^{*[3]}_{\underline{h},M}) \right| \\ &\quad + \sum_{h_1=M+1}^{\infty} \sum_{h_2=M+1}^{\infty} \left| \dots \right| + \sum_{h_1=0}^{M} \sum_{h_2=M+1}^{\infty} \left| \dots \right| + \sum_{h_1=M+1}^{\infty} \sum_{h_2=0}^{M} \left| \dots \right| \\ &=: A + B + C + D. \end{aligned}$$

With the same techniques as in (7.27) we get

$$\begin{aligned} M^2 \cdot A &\leq M^2 \cdot \sum_{h_1=0}^{M} \sum_{h_2=0}^{M} \left(\left\| Z^*_{\underline{0},M} \right\|_{*2} + \left\| Z^{*[2]}_{\underline{0},M} \right\|_{*2} \right) \left(\left\| Z^{*[1]}_{\underline{h},M} \right\|_{*2} + \left\| Z^{*[3]}_{\underline{h},M} \right\|_{*2} \right) \\ &\leq \mathcal{O}_P(1) \cdot M^2 \sum_{h_1=0}^{M} \sum_{h_2=0}^{M} \left(2 \sum_{\underline{k} \in \Theta \setminus \Theta(M)} \left| \widehat{b}_{\underline{k}}(p) \right| \right)^2 \\ &\leq \mathcal{O}_P(1) \cdot \left(\sum_{\underline{k} \in \Theta \setminus \Theta(M)} (M+1)^2 \left| \widehat{b}_{\underline{k}}(p) \right| \right)^2 \\ &\leq \mathcal{O}_P(1) \cdot \left(\sum_{\underline{k} \in \Theta} (1 + |\underline{k}|_\infty)^2 \left| \widehat{b}_{\underline{k}}(p) \right| \right)^2 = \mathcal{O}_P(1) \end{aligned} \tag{7.28}$$

uniformly for all $M \in \mathbb{N}$. Since we are interested in an asymptotic result for $M \to \infty$ in (7.22), we can, from now on, consider only those M large enough such that $-\lfloor (M+1)/2 \rfloor + m_1 - 1 < 0$ and $M - m_2 + 2 \geq 0$. With the same calculation as for $\|Z^*_{\underline{h},M} - Z^{*[1]}_{\underline{h},M}\|_{*2}$ in (7.26), we can derive

$$\begin{aligned} \left\| Z^*_{\underline{0},M} - Z^{*[2]}_{\underline{0},M} \right\|_{*2} &\leq \mathcal{O}_P(1) \cdot \sum_{k_1=-\infty}^{-\lfloor h_1/2 \rfloor + m_1 - 1} \sum_{k_2=0}^{\infty} \left| \widehat{b}_{\underline{k}}(p) \right|, \\ \left\| Z^{*[1]}_{\underline{h},M} - Z^{*[3]}_{\underline{h},M} \right\|_{*2} &\leq \mathcal{O}_P(1) \cdot \sum_{k_1=-\infty}^{\lfloor h_1/2 \rfloor} \sum_{k_2=(h_2-m_2+1) \vee 0}^{\infty} \left| \widehat{b}_{\underline{k}}(p) \right| \end{aligned}$$

$$\leq \mathcal{O}_P(1) \cdot \sum_{k_1=-\infty}^{\infty} \sum_{k_2=(h_2-m_2+1)\vee 0}^{\infty} \left|\widehat{b}_{\underline{k}}(p)\right|.$$

With these bounds and for all M large enough as defined before, we get for B:

$$\begin{aligned}
& M^2 \cdot B \\
\leq\ & M^2 \cdot \sum_{h_1=M+1}^{\infty} \sum_{h_2=M+1}^{\infty} \left\| Z^*_{\underline{0},M} - Z^{*[2]}_{\underline{0},M} \right\|_{*2} \left\| Z^{*[1]}_{\underline{h},M} - Z^{*[3]}_{\underline{h},M} \right\|_{*2} \\
\leq\ & \mathcal{O}_P(1) \cdot M^2 \sum_{h_1=M+1}^{\infty} \sum_{k_1=-\infty}^{-\lfloor h_1/2 \rfloor + m_1 - 1} \sum_{k_2=0}^{\infty} \left|\widehat{b}_{\underline{k}}(p)\right| \cdot \sum_{h_2=M+1}^{\infty} \sum_{k_1=-\infty}^{\infty} \sum_{k_2=h_2-m_2+1}^{\infty} \left|\widehat{b}_{\underline{k}}(p)\right| \\
=\ & \mathcal{O}_P(1) \cdot B_1 \cdot B_2,
\end{aligned}$$

where

$$\begin{aligned}
B_1 &:= M \sum_{h_1=M+1}^{\infty} \sum_{k_1=-\infty}^{-\lfloor h_1/2 \rfloor + m_1 - 1} \sum_{k_2=0}^{\infty} \left|\widehat{b}_{\underline{k}}(p)\right| \\
&= M \sum_{k_1=-\infty}^{-\lfloor (M+1)/2 \rfloor + m_1 - 1} \sum_{k_2=0}^{\infty} 2 \left| k_1 - \Big(- \lfloor (M+1)/2 \rfloor + m_1 \Big) \right| \left|\widehat{b}_{\underline{k}}(p)\right| \\
&\leq 2 \sum_{k_1=-\infty}^{-\lfloor (M+1)/2 \rfloor + m_1 - 1} \sum_{k_2=0}^{\infty} M\, |k_1| \left|\widehat{b}_{\underline{k}}(p)\right| \\
&\leq 8 m_1 \sum_{\underline{k} \in \Theta} (1 + |\underline{k}|_\infty)^2 \left|\widehat{b}_{\underline{k}}(p)\right| = \mathcal{O}_P(1), \qquad (7.29)
\end{aligned}$$

and

$$\begin{aligned}
B_2 &:= M \sum_{h_2=M+1}^{\infty} \sum_{k_1=-\infty}^{\infty} \sum_{k_2=h_2-m_2+1}^{\infty} \left|\widehat{b}_{\underline{k}}(p)\right| \\
&= M \sum_{k_1=-\infty}^{\infty} \sum_{k_2=M-m_2+2}^{\infty} \left| k_2 - (M - m_2 + 1) \right| \left|\widehat{b}_{\underline{k}}(p)\right| \\
&\leq \sum_{k_1=-\infty}^{\infty} \sum_{k_2=M-m_2+2}^{\infty} M\, |k_2| \left|\widehat{b}_{\underline{k}}(p)\right| \\
&\leq (m_2+1) \sum_{\underline{k} \in \Theta} (1 + |\underline{k}|_\infty)^2 \left|\widehat{b}_{\underline{k}}(p)\right| = \mathcal{O}_P(1), \qquad (7.30)
\end{aligned}$$

uniformly for all M large enough. The latter inequalities in (7.29) and (7.30) hold because

$$M\,|k_2| \leq (k_2 + m_2 - 2)\, k_2 \leq (m_2+1)(k_2+1)\, k_2 \leq (m_2+1)\,(1 + |\underline{k}|_\infty)^2$$

for all $k_2 \geq M - m_2 + 2$, and, with similar calculations,

$$M\,|k_1| \leq 2m_1\,(2|k_1|+1)\,|k_1| \leq 4m_1\,(1+|k_1|)\,|k_1| \leq 4m_1\,(1+|\underline{k}|_\infty)^2$$

for all $k_1 \leq -\lfloor (M+1)/2 \rfloor + m_1 - 1$. Altogether, this yields $B \leq (1/M^2)\,\mathcal{O}_P(1)$, and, with exactly the same arguments as for A and B, we can also show $C \leq$

$(1/M^2)\,\mathcal{O}_P(1)$ and $D \leq (1/M^2)\,\mathcal{O}_P(1)$. This implies (7.25) and, therefore, (7.24) and (7.23). Hence, the proof of (7.12) is completed.

Using analogous arguments as for the bootstrap quantities in (7.12), one can show (7.13) for the non-bootstrap quantities. Since

$$\left\|\underline{\theta}^* - \underline{\widetilde{\theta}}\right\| := \sum_{v=1}^{k} \left|E^*\Big(g_v(\mathbf{Y}^*_{\underline{t}})\Big) - E\Big(g_v(\widetilde{\mathbf{Y}}_{\underline{t}})\Big)\right| = o_P(1) \tag{7.31}$$

follows with the same arguments as in the proof of (7.6) (by simply replacing covariances with expectations), we can incorporate the outer function f from the definition of $\widetilde{T}_n$ and T_n^*, cf. Assumption 3, with the delta method. It follows from (7.12), (7.13) and (7.31) that $(\bar{n}_1\bar{n}_2)^{1/2}(T_n^* - f(\underline{\theta}^*))$ and $(\bar{n}_1\bar{n}_2)^{1/2}(\widetilde{T}_n - f(\underline{\widetilde{\theta}}))$ have identical limiting (normal) distributions. Therefore, since $(\bar{n}_1\bar{n}_2)^{1/2}$ is asymptotically equivalent to n, we have

$$\sup_{x\in\mathbb{R}} \left|P^*\Big\{n(T_n^* - f(\underline{\theta}^*)) \leq x\Big\} - P\Big\{n(\widetilde{T}_n - f(\underline{\widetilde{\theta}})) \leq x\Big\}\right| = o_P(1),$$

which completes the proof. □

8. Proofs of the auxiliary results

Proof of Lemma 2.1:
Finding one-sided AR- and MA-representations as in (2.5) is closely related to finding a spectral factorization $f(\underline{\lambda}) = |B'(\underline{\lambda})|^2$ of the spectral density, where B' is a complex-valued function with one-sided Fourier series in the sense of the half-plane Θ, i.e.

$$B'(\underline{\lambda}) = \sum_{\underline{k}\in\Theta\cup\{\underline{0}\}} \widetilde{b}_{\underline{k}}\, e^{-i\langle\underline{k},\underline{\lambda}\rangle}. \tag{8.1}$$

Under Assumption 1 the spectral density f is equal to its absolutely convergent Fourier series $f(\underline{\lambda}) = \sum_{\underline{k}\in\mathbb{Z}^2}(\gamma(\underline{k})/4\pi^2)\, e^{i\langle\underline{k},\underline{\lambda}\rangle}$. Lemma 2.3 shows that $\log f \in C_{r-1}$ and, in particular, $\log f$ is equal to its absolutely convergent Fourier series,

$$\log f(\underline{\lambda}) = \sum_{\underline{k}\in\mathbb{Z}^2} d_{\underline{k}}\, e^{i\langle\underline{k},\underline{\lambda}\rangle},$$

say. Whittle (1954) showed that the spectral factorization of f can be obtained from the Fourier series of $\log f$ by letting

$$B_0(\underline{z}) := \exp\Big(L(\underline{z})\Big), \quad L(\underline{z}) := \frac{d_{\underline{0}}}{2} + \sum_{\underline{k}\in\Theta} d_{\underline{k}}\, z_1^{k_1} z_2^{k_2} \quad \forall\, \underline{z} \in S, \tag{8.2}$$

where $S = \{\underline{z} \in \mathbb{C}^2,\ |z_1| = 1,\ |z_2| \leq 1\}$. Identifying $B'(\underline{\lambda}) := B_0(e^{-i\lambda_1}, e^{-i\lambda_2})$ gives the spectral factorization $f(\underline{\lambda}) = |B'(\underline{\lambda})|^2$, since one can easily verify

$$B_0(e^{-i\lambda_1}, e^{-i\lambda_2}) \cdot \overline{B_0(e^{-i\lambda_1}, e^{-i\lambda_2})} = f(\underline{\lambda})$$

(note that $\log f(-\underline{\lambda}) = \log f(\underline{\lambda})$ implies $d_{\underline{k}} = d_{-\underline{k}} \in \mathbb{R}$ for all $\underline{k} \in \mathbb{Z}^2$). Through straightforward multiplication and grouping of like summands one can easily verify that each power $L(\underline{z})^j$, $j \in \mathbb{N}_0$, has a series representation with respect to the upper half-plane $\Theta \cup \{\underline{0}\}$, only, i.e.

$$L(\underline{z})^j := \frac{d_{\underline{0}}(j)}{2} + \sum_{\underline{k}\in\Theta} d_{\underline{k}}(j)\, z_1^{k_1} z_2^{k_2} \quad \forall\, \underline{z} \in S.$$

Using this, and expanding $B_0(\underline{z})$ in (8.2) via $\exp(L(\underline{z})) = \sum_{j=0}^{\infty} L(\underline{z})^j/j!$ yields

$$B_0(\underline{z}) = \sum_{\underline{k}\in\Theta\cup\{\underline{0}\}} \widetilde{b}_{\underline{k}}\, z_1^{k_1} z_2^{k_2}$$

for suitable coefficients $\widetilde{b}_{\underline{k}}$. $B'(\underline{\lambda}) = B_0(e^{-i\lambda_1}, e^{-i\lambda_2})$ then gives the desired form (8.1) of the Fourier series of $B'(\underline{\lambda})$. The series $B_0(\underline{z})$, and therefore also the series in (8.1), converge absolutely because the Fourier series of $\log f$ converges absolutely and the power series of the exponential function converges absolutely in $\mathbb{C}$. Furthermore, it holds $|z_1^{k_1}| \le 1$, $|z_2^{k_2}| \le 1$ for all $\underline{z} \in S$ and all $\underline{k} \in \Theta$ (note that Θ contains only vectors $\underline{k}$ with $k_2 \ge 0$). From (8.2) it is obvious that $B_0(\underline{z}) \neq 0$ for all $\underline{z}$ with $|z_1| = 1$, $|z_2| \le 1$ and we can define $A_0(\underline{z}) := 1/B_0(\underline{z})$ on this region which, analogously to B_0, has a one-sided series representation with absolutely summable coefficients $(\widetilde{a}_{\underline{k}})_{\underline{k}\in\Theta\cup\{\underline{0}\}}$, i.e.

$$A_0(\underline{z}) = \frac{1}{B_0(\underline{z})} = \exp\left(\frac{-d_{\underline{0}}}{2} + \sum_{\underline{k}\in\Theta}(-d_{\underline{k}})\, z_1^{k_1} z_2^{k_2}\right) = \sum_{\underline{k}\in\Theta\cup\{\underline{0}\}} \widetilde{a}_{\underline{k}}\, z_1^{k_1} z_2^{k_2}. \tag{8.3}$$

From the definitions of A_0 and B_0 it follows immediately $\widetilde{a}_{\underline{0}} = \exp(-d_{\underline{0}}/2) \neq 0$ and $\widetilde{b}_{\underline{0}} = \exp(d_{\underline{0}}/2) \neq 0$ and we get the standardized versions

$$A(\underline{z}) := \frac{A_0(\underline{z})}{\widetilde{a}_{\underline{0}}} = 1 - \sum_{\underline{k}\in\Theta} a_{\underline{k}}\, z_1^{k_1} z_2^{k_2}, \quad B(\underline{z}) := \frac{B_0(\underline{z})}{\widetilde{b}_{\underline{0}}} = 1 + \sum_{\underline{k}\in\Theta} b_{\underline{k}}\, z_1^{k_1} z_2^{k_2}, \tag{8.4}$$

where $a_{\underline{k}} := -\widetilde{a}_{\underline{k}}/\widetilde{a}_{\underline{0}}$ and $b_{\underline{k}} := \widetilde{b}_{\underline{k}}/\widetilde{b}_{\underline{0}}$ for all $\underline{k} \in \Theta$. (8.4) yields exactly the z-transforms defined in (3.4). We now consider the functions A' and L' on $(-\pi, \pi]^2$ which are, just as B_0 and B', defined via $A'(\lambda_1, \lambda_2) := A_0(e^{-i\lambda_1}, e^{-i\lambda_2})$, $L'(\lambda_1, \lambda_2) := L(e^{-i\lambda_1}, e^{-i\lambda_2})$. Per definition, it holds $B'(\underline{\lambda}) = \exp(L'(\underline{\lambda}))$ and $A'(\underline{\lambda}) = \exp(-L'(\underline{\lambda}))$. Using the submultiplicative C_{r-1}-norm defined in Lemma 2.3, as well as the fact that $\log f \in C_{r-1}$ implies $L' \in C_{r-1}$, we can infer

$$\|B'\|_{r-1} = \left\| \sum_{j=0}^{\infty} \frac{1}{j!}(L')^j \right\|_{r-1} \le \sum_{j=0}^{\infty} \frac{1}{j!}\|L'\|_{r-1}^j = \exp(\|L'\|_{r-1}) < \infty.$$

Analogously, the same argument delivers $A' \in C_{r-1}$ which yields (2.6).
We can now define the process $(\varepsilon_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ via

$$\varepsilon_{\underline{t}} := X_{\underline{t}} - \sum_{\underline{k}\in\Theta} a_{\underline{k}}\, X_{\underline{t}-\underline{k}}$$

which is obviously an L^2-convergent series with spectral density

$$f_\varepsilon(\underline{\lambda}) = |\widetilde{a}_{\underline{0}}^{-1}\, A'(\underline{\lambda})|^2 \cdot f(\underline{\lambda}) = \widetilde{a}_{\underline{0}}^{-2}, \quad \forall\, \underline{\lambda} \in (-\pi,\pi]^2,$$

since $f(\underline{\lambda}) = |B'(\underline{\lambda})|^2 = 1/|A'(\underline{\lambda})|^2$. Hence, $(\varepsilon_{\underline{t}})$ is uncorrelated white noise. Furthermore, the backshift operators of $\varepsilon_{\underline{t}}$ and $X_{\underline{t}}$ coincide and $1/(\widetilde{a}_{\underline{0}}^{-1}\, A'(\underline{\lambda})) = \widetilde{b}_{\underline{0}}^{-1}\, B'(\underline{\lambda})$ implies

$$X_{\underline{t}} = \sum_{\underline{k}\in\Theta} b_{\underline{k}}\, \varepsilon_{\underline{t}-\underline{k}} + \varepsilon_{\underline{t}}.$$

It remains to show that $(\varepsilon_{\underline{t}})$ is the innovation process, i.e. that $\sum_{\underline{k}\in\Theta} a_{\underline{k}}\, X_{\underline{t}-\underline{k}}$ is the L^2-projection of $X_{\underline{t}}$ onto $H_{\underline{t}}(X) := \overline{sp}\{X_{\underline{t}-\underline{k}} : \underline{k} \in \Theta\}$. Let $\underline{j} \in \Theta$ be arbitrary. We show that $X_{\underline{t}} - \sum_{\underline{k}\in\Theta} a_{\underline{k}}\, X_{\underline{t}-\underline{k}}$ is orthogonal to $X_{\underline{t}-\underline{j}}$ via

$$\mathrm{Cov}\Big(X_{\underline{t}} - \sum_{\underline{k}\in\Theta} a_{\underline{k}}\, X_{\underline{t}-\underline{k}}\,,\, X_{\underline{t}-\underline{j}}\Big) = \mathrm{Cov}\Big(\varepsilon_{\underline{t}}\,,\, \varepsilon_{\underline{t}-\underline{j}} + \sum_{\underline{k}\in\Theta} b_{\underline{k}}\, \varepsilon_{\underline{t}-\underline{j}-\underline{k}}\Big) = 0,$$

since $(\varepsilon_{\underline{t}})$ is white noise. It remains to show that the coefficients in (2.5) are uniquely determined. Assume there was a different sequence of coefficients $(a'_{\underline{k}})_{\underline{k}\in\Theta}$ such that

$$X_{\underline{t}} = \sum_{\underline{k}\in\Theta} a'_{\underline{k}}\, X_{\underline{t}-\underline{k}} + \varepsilon_{\underline{t}}.$$

Then we get from (2.5) $\sum_{\underline{k}\in\Theta}(a_{\underline{k}} - a'_{\underline{k}})\, X_{\underline{t}-\underline{k}} = 0$. If there exists $\underline{s} \in \Theta$ such that $a_{\underline{s}} \neq a'_{\underline{s}}$ it follows

$$X_{\underline{t}-\underline{s}} = -\sum_{\underline{k}\in\Theta,\, \underline{k}\neq\underline{s}} \frac{a_{\underline{k}} - a'_{\underline{k}}}{a_{\underline{s}} - a'_{\underline{s}}}\, X_{\underline{t}-\underline{k}} \in \overline{sp}\{X_{\underline{j}},\, \underline{j} \neq \underline{t} - \underline{s}\},$$

which contradicts the basic process condition from Assumption 1. Therefore, the coefficients $(a_{\underline{k}})$ are unique. Analogously, assume the MA representation in (2.5) was also fulfilled with coefficients $(b'_{\underline{k}})$, $b_{\underline{s}} \neq b'_{\underline{s}}$. Then we get

$$\varepsilon_{\underline{t}-\underline{s}} = -\sum_{\underline{k}\in\Theta,\, \underline{k}\neq\underline{s}} \frac{b_{\underline{k}} - b'_{\underline{k}}}{b_{\underline{s}} - b'_{\underline{s}}}\, \varepsilon_{\underline{t}-\underline{k}}.$$

Since $f_\varepsilon(\underline{\lambda}) > 0$ implies $\mathrm{Var}(\varepsilon_{\underline{t}}) > 0$ this yields a contradiction via

$$0 < \mathrm{Cov}(\varepsilon_{\underline{t}-\underline{s}}, \varepsilon_{\underline{t}-\underline{s}}) = \mathrm{Cov}\Big(\varepsilon_{\underline{t}-\underline{s}}\,,\, -\sum_{\underline{k}\in\Theta,\, \underline{k}\neq\underline{s}} \frac{b_{\underline{k}} - b'_{\underline{k}}}{b_{\underline{s}} - b'_{\underline{s}}}\, \varepsilon_{\underline{t}-\underline{k}}\Big) = 0,$$

as $(\varepsilon_{\underline{t}})$ is white noise. Hence, the coefficients $(b_{\underline{k}})$ are uniquely determined. □

Proof of Lemma 2.3:
For any $r \geq 0$ and arbitrary functions $g, h \in C_r$ the Fourier series of gh is given by $\sum_{\underline{k}\in\mathbb{Z}^2}(\sum_{\underline{j}\in\mathbb{Z}^2} \widetilde{g}_{\underline{j}} \widetilde{h}_{\underline{k}-\underline{j}})\, e^{i\langle\underline{k},\underline{\lambda}\rangle}$, and from

$$\begin{aligned}(1 + |\underline{k}|_\infty)^r &\leq (1 + |\underline{j}|_\infty + |\underline{k} - \underline{j}|_\infty + |\underline{j}|_\infty \cdot |\underline{k} - \underline{j}|_\infty)^r \\ &= (1 + |\underline{j}|_\infty)^r \cdot (1 + |\underline{k} - \underline{j}|_\infty)^r\end{aligned}$$

one can easily see that $\|\cdot\|_r$ is submultiplicative, i.e.

$$\|gh\|_r \leq \|g\|_r \cdot \|h\|_r. \tag{8.5}$$

Since $f \in C_r$ for $r \geq 2$, its formal Fourier series $\sum_{\underline{k}\in\mathbb{Z}^2} \widetilde{f}_{\underline{k}}\, e^{i\langle \underline{k},\underline{\lambda}\rangle}$ converges absolutely and is therefore equal to $f(\underline{\lambda})$ everywhere on $(-\pi,\pi]^2$. Also, f is twice continuously differentiable and the derivatives are equal to their respective Fourier series

$$\frac{\partial f}{\partial \lambda_j}(\underline{\lambda}) = \sum_{\underline{k}\in\mathbb{Z}^2} ik_j\, \widetilde{f}_{\underline{k}}\, e^{i\langle \underline{k},\underline{\lambda}\rangle}, \qquad \frac{\partial^2 f}{\partial \lambda_1 \partial \lambda_2}(\underline{\lambda}) = \sum_{\underline{k}\in\mathbb{Z}^2} (-k_1 k_2)\, \widetilde{f}_{\underline{k}}\, e^{i\langle \underline{k},\underline{\lambda}\rangle}$$

as well, since these series are obviously absolutely convergent. To be more precise, one has from $|k_1|\cdot|k_2| \leq |\underline{k}|_\infty^2$ that

$$\left\|\frac{\partial^2 f}{\partial \lambda_1 \partial \lambda_2}\right\|_{r-2} = \sum_{\underline{k}\in\mathbb{Z}^2} (1+|\underline{k}|_\infty)^{r-2}\, |k_1|\, |k_2|\, |\widetilde{f}_{\underline{k}}| \leq \sum_{\underline{k}\in\mathbb{Z}^2} (1+|\underline{k}|_\infty)^{r}\, |\widetilde{f}_{\underline{k}}| < \infty,$$

i.e. the second order derivative is in C_{r-2}. The same holds true for the first order derivatives, and Theorem 6.2 in Gröchenig (2007) (the weight function $(1+|\underline{k}|_\infty)^r$ obviously fulfils the required GRS-condition) implies $\|1/f\|_r < \infty$ and in particular $(1/f) \in C_{r-2}$. Since $f(\underline{\lambda}) \geq c > 0$, $\log f$ is also twice continuously differentiable and it holds

$$\frac{\partial^2 \log f}{\partial \lambda_1 \partial \lambda_2}(\underline{\lambda}) = \frac{1}{f^2(\underline{\lambda})} \cdot \left(\frac{\partial^2 f}{\partial \lambda_1 \partial \lambda_2}(\underline{\lambda}) \cdot f(\underline{\lambda}) - \frac{\partial f}{\partial \lambda_1}(\underline{\lambda}) \cdot \frac{\partial f}{\partial \lambda_2}(\underline{\lambda}) \right).$$

With this representation, (8.5) and the results established so far we can now infer

$$\left\|\frac{\partial^2 \log f}{\partial \lambda_1 \partial \lambda_2}\right\|_{r-2} \leq \left\|\frac{\partial^2 f}{\partial \lambda_1 \partial \lambda_2}\right\|_{r-2} \cdot \left\|\frac{1}{f}\right\|_{r-2} + \left\|\frac{\partial f}{\partial \lambda_1}\right\|_{r-2} \cdot \left\|\frac{\partial f}{\partial \lambda_2}\right\|_{r-2} \cdot \left\|\frac{1}{f}\right\|_{r-2}^2 < \infty.$$

The same holds true for the first order derivatives $\partial \log f/\partial\lambda_j$. Now let $(d_{\underline{k}})$ be the Fourier coefficients of $\log f$. As seen above, $\partial \log f/\partial \lambda_j$ and $\partial^2 \log f/\partial\lambda_1\partial\lambda_2$ have Fourier coefficients $ik_j\, d_{\underline{k}}$ and $-k_1k_2\, d_{\underline{k}}$, respectively. Hence, it holds

$$\sum_{\underline{k}\in\mathbb{Z}^2} (1+|\underline{k}|_\infty)^{r-2}\, |k_1|\, |k_2|\, |d_{\underline{k}}| < \infty, \ \sum_{\underline{k}\in\mathbb{Z}^2} (1+|\underline{k}|_\infty)^{r-2}\, |k_j|\, |d_{\underline{k}}| < \infty, \ j = 1,2. \tag{8.6}$$

For $k_1, k_2 \in \mathbb{Z}\setminus\{0\}$ it holds $1+|\underline{k}|_\infty \leq 2\,|\underline{k}|_\infty \leq 2\,|k_1|\,|k_2|$. Analogously, $1+|\underline{k}|_\infty$ can be bounded from above by $2\,|k_1|$ if $k_1 \neq 0$, $k_2 = 0$ and by $2\,|k_2|$ if $k_2 \neq 0$, $k_1 = 0$. Therefore, we obtain

$$\begin{aligned} \|\log f\|_{r-1} \leq |d_{\underline{0}}| \quad &+ \sum_{k_1\neq 0} (1+|(k_1,0)'|_\infty)^{r-2}\, 2\,|k_1|\,|d_{\underline{k}}| \\ &+ \sum_{k_2\neq 0} (1+|(0,k_2)'|_\infty)^{r-2}\, 2\,|k_2|\,|d_{\underline{k}}| \\ &+ \sum_{k_1,k_2\neq 0} (1+|\underline{k}|_\infty)^{r-2}\, 2\,|k_1|\,|k_2|\,|d_{\underline{k}}| \end{aligned}$$

which is finite due to (8.6). This completes the proof of assertion (i). Assertion (ii) can be proven with analogous arguments for all $r_1, r_2 \geq 1$. $\square$

Proof of Lemma 3.1:
As a preliminary consideration we recall for the vectors $\underline{k}_1, \ldots, \underline{k}_{\bar{p}}$ from (3.3) and arbitrary $r, s \in \{1, \ldots, \bar{p}\}$

$$\int_{(-\pi,\pi]^2} \exp(i\langle \underline{k}_r - \underline{k}_s, \underline{\lambda}\rangle)\, d\underline{\lambda} = \begin{cases} 4\pi^2 & , \ r = s \\ 0 & , \ r \neq s \end{cases}, \tag{8.7}$$

because $\underline{k}_r = \underline{k}_s$ if and only if $r = s$. Let $\underline{d} \in \mathbb{R}^{\bar{p}}$ be arbitrary with $\underline{d} \neq \underline{0}$ and denote by $\underline{w}(\underline{\lambda}) := (\exp(i\langle \underline{k}_1, \underline{\lambda}\rangle), \ldots, \exp(i\langle \underline{k}_{\bar{p}}, \underline{\lambda}\rangle))'$. Observe that $|\underline{d}'\underline{w}(\underline{\lambda})|^2 = \sum_{r,s=1}^{\bar{p}} d_r\, d_s\, \exp(i\langle \underline{k}_r - \underline{k}_s, \underline{\lambda}\rangle)$. Using (8.7) as well as $\gamma(\underline{h}) = \int_{(-\pi,\pi]^2} f(\underline{\lambda})\, e^{i\langle \underline{h}, \underline{\lambda}\rangle}\, d\underline{\lambda}$ and $f(\underline{\lambda}) \geq c > 0$, cf. Assumption 1, we can derive

$$\begin{aligned} \underline{d}'\, \Gamma(p)\, \underline{d} &= \int_{(-\pi,\pi]^2} f(\lambda)\, |\underline{d}'\underline{w}(\underline{\lambda})|^2\, d\underline{\lambda} \\ &\geq c \cdot \int_{(-\pi,\pi]^2} |\underline{d}'\underline{w}(\underline{\lambda})|^2\, d\underline{\lambda} \\ &= c \cdot \sum_{r,s=1}^{\bar{p}} d_r\, d_s \int_{(-\pi,\pi]^2} \exp(i\langle \underline{k}_r - \underline{k}_s, \underline{\lambda}\rangle)\, d\underline{\lambda} \\ &= 4\pi^2 c \cdot \underline{d}'\, \underline{d}. \end{aligned}$$

On the one hand this shows that $\Gamma(p)$ is positive definite and therefore invertible for each $p \in \mathbb{N}$. On the other hand it follows

$$\frac{\underline{d}'\, \Gamma(p)\, \underline{d}}{\underline{d}'\, \underline{d}} \geq 4\pi^2 c,$$

which implies for the smallest eigenvalue $\sigma_{\min}(\Gamma(p)) \geq 4\pi^2 c$, cf. Lütkepohl (1996), 5.2.2 (2). This yields for the largest eigenvalue of the inverse matrix $\sigma_{\max}(\Gamma(p)^{-1}) \leq (4\pi^2 c)^{-1}$ for all $p \in \mathbb{N}$. The spectral norm of the symmetric matrix $\Gamma(p)^{-1}$ is given by its largest eigenvalue, i.e. $\|\Gamma(p)^{-1}\|_{\text{spec}} \leq (4\pi^2 c)^{-1}$ for all $p \in \mathbb{N}$, which yields the desired assertion. $\square$

Proof of Lemma 3.3:
Let $p \in \mathbb{N}$ be arbitrary. For any $\underline{z} = (z_1, z_2) \in S_p$ we define $\underline{\widetilde{z}} = (\widetilde{z}_1, \widetilde{z}_2)$ as the unique vector in S that minimizes the distance to $\underline{z}$ componentwise. To be more precise, let

$$\widetilde{z}_1 := \arg\min_{|u_1|=1} |u_1 - z_1|, \quad \widetilde{z}_2 := \arg\min_{|u_2|\leq 1} |u_2 - z_2|.$$

In the first step we derive an expression $D(p)$ such that

$$\sup_{\underline{z} \in S_p} \left| A_p(\underline{z}) - A(\underline{\widetilde{z}}) \right| \leq D(p). \tag{8.8}$$

For any $\underline{z} \in S_p$ we have

$$\begin{aligned}\left|A_p(\underline{z}) - A(\underline{\widetilde{z}})\right| &\leq \sum_{\underline{k}\in\Theta(p)} \left|a_{\underline{k}}(p)\, z_1^{k_1}\, z_2^{k_2} - a_{\underline{k}}\, \widetilde{z}_1^{k_1}\, \widetilde{z}_2^{k_2}\right| + \sum_{\underline{k}\in\Theta\backslash\Theta(p)} |a_{\underline{k}}|\, |\widetilde{z}_1|^{k_1}\, |\widetilde{z}_2|^{k_2} \\ &\leq \sum_{\underline{k}\in\Theta(p)} |a_{\underline{k}}(p)| \left|z_1^{k_1}\, z_2^{k_2} - \widetilde{z}_1^{k_1}\, \widetilde{z}_2^{k_2}\right| + \sum_{\underline{k}\in\Theta(p)} |a_{\underline{k}}(p) - a_{\underline{k}}|\, |\widetilde{z}_1|^{k_1}\, |\widetilde{z}_2|^{k_2} \\ &\quad + \sum_{\underline{k}\in\Theta\backslash\Theta(p)} |a_{\underline{k}}|\, |\widetilde{z}_1|^{k_1}\, |\widetilde{z}_2|^{k_2},\end{aligned}$$

which implies

$$\begin{aligned}\sup_{\underline{z}\in S_p}\left|A_p(\underline{z}) - A(\underline{\widetilde{z}})\right| &\leq \sum_{\underline{k}\in\Theta(p)} |a_{\underline{k}}(p)| \sup_{\underline{z}\in S_p}\left|z_1^{k_1}\, z_2^{k_2} - \widetilde{z}_1^{k_1}\, \widetilde{z}_2^{k_2}\right| \\ &\quad + \sum_{\underline{k}\in\Theta(p)} |a_{\underline{k}}(p) - a_{\underline{k}}| + \sum_{\underline{k}\in\Theta\backslash\Theta(p)} |a_{\underline{k}}|, \qquad (8.9)\end{aligned}$$

since $|\widetilde{z}_1|^{k_1} = 1$ and $|\widetilde{z}_2|^{k_2} \leq 1$ for any $\underline{k} \in \Theta(p)$. In order to get a bound for the remaining supremum on the right-hand side, consider the following: For arbitrary $\underline{z} \in S_p$, if $|z_2| \leq 1$, it follows per definition $\widetilde{z}_2 = z_2$, and thus $|z_2^{k_2} - \widetilde{z}_2^{k_2}| = 0$. However, if $|z_2| > 1$, we can write $z_2 = r\, e^{i\varphi}$ for some $-\pi < \varphi \leq \pi$ and some $1 < r \leq (p+1)/p$. For this z_2, it holds $\widetilde{z}_2 = e^{i\varphi}$ and therefore

$$\left|z_2^{k_2} - \widetilde{z}_2^{k_2}\right| = \left|r^{k_2}\, e^{ik_2\varphi} - e^{ik_2\varphi}\right| = r^{k_2} - 1 \leq \left(\frac{p+1}{p}\right)^{k_2} - 1.$$

Similarly, one can show that

$$\left|z_1^{k_1} - \widetilde{z}_1^{k_1}\right| \leq \left(\frac{p+1}{p}\right)^{|k_1|} - 1,$$

for any $\underline{z} \in S_p$, which yields

$$\begin{aligned}\sup_{\underline{z}\in S_p}\left|z_1^{k_1}\, z_2^{k_2} - \widetilde{z}_1^{k_1}\, \widetilde{z}_2^{k_2}\right| &\leq \sup_{\underline{z}\in S_p}\left(\left|z_1^{k_1}\right|\left|z_2^{k_2} - \widetilde{z}_2^{k_2}\right| + \left|\widetilde{z}_2^{k_2}\right|\left|z_1^{k_1} - \widetilde{z}_1^{k_1}\right|\right) \\ &\leq \left(\frac{p+1}{p}\right)^{|k_1|} \cdot \left(\left(\frac{p+1}{p}\right)^{k_2} - 1 + \left(\frac{p+1}{p}\right)^{|k_1|} - 1\right).\end{aligned}$$

Inserting this inequality into (8.9) yields (8.8) with

$$\begin{aligned}D(p) &:= \sum_{\underline{k}\in\Theta(p)} |a_{\underline{k}}(p)| \left(\frac{p+1}{p}\right)^{|k_1|} \cdot \left(\left(\frac{p+1}{p}\right)^{k_2} + \left(\frac{p+1}{p}\right)^{|k_1|} - 2\right) \\ &\quad + \sum_{\underline{k}\in\Theta(p)} |a_{\underline{k}}(p) - a_{\underline{k}}| + \sum_{\underline{k}\in\Theta\backslash\Theta(p)} |a_{\underline{k}}|.\end{aligned}$$

In the next step we show $D(p) \to 0$, as $p \to \infty$. In order to handle the first summand in the definition of $D(p)$, consider that it holds for any $p \in \mathbb{N}$ and any $\underline{k} \in \Theta(p)$

$$\left|\left(\frac{p+1}{p}\right)^{|k_1|} \cdot \left(\left(\frac{p+1}{p}\right)^{k_2} + \left(\frac{p+1}{p}\right)^{|k_1|} - 2\right)\right| \leq 4\left(\frac{p+1}{p}\right)^{2p} \leq 4\, e^2. \qquad (8.10)$$

This, together with $|a_{\underline{k}}(p)| \leq |a_{\underline{k}}| + |a_{\underline{k}}(p) - a_{\underline{k}}|$, yields

$$\begin{aligned} D(p) \;\leq\; & \sum_{\underline{k}\in\Theta(p)} |a_{\underline{k}}| \left(\frac{p+1}{p}\right)^{|k_1|} \cdot \left(\left(\frac{p+1}{p}\right)^{k_2} + \left(\frac{p+1}{p}\right)^{|k_1|} - 2 \right) \\ & + (4\,e^2+1) \sum_{\underline{k}\in\Theta(p)} |a_{\underline{k}}(p) - a_{\underline{k}}| + \sum_{\underline{k}\in\Theta\setminus\Theta(p)} |a_{\underline{k}}|. \end{aligned} \tag{8.11}$$

For the latter two summands on the right-hand side we immediately get from Theorem 3.2 for some constant $C < \infty$

$$(4\,e^2+1) \sum_{\underline{k}\in\Theta(p)} |a_{\underline{k}}(p) - a_{\underline{k}}| + \sum_{\underline{k}\in\Theta\setminus\Theta(p)} |a_{\underline{k}}| \leq \Big(C(4\,e^2+1)+1\Big) \sum_{\underline{k}\in\Theta\setminus\Theta(p)} (1+|\underline{k}|_\infty)|a_{\underline{k}}|,$$

which converges to zero, as $p \to \infty$, due to summability condition (2.6) and since $\Theta(p) \to \Theta$. For the first summand on the right-hand side of inequality (8.11) we can apply Lebesgue's dominated convergence theorem because (8.10) provides a dominating and summable sequence via

$$\begin{aligned} & \sum_{\underline{k}\in\Theta} \mathbb{1}_{\{\underline{k}\in\Theta(p)\}} |a_{\underline{k}}| \left| \left(\frac{p+1}{p}\right)^{|k_1|} \cdot \left(\left(\frac{p+1}{p}\right)^{k_2} + \left(\frac{p+1}{p}\right)^{|k_1|} - 2 \right) \right| \\ \leq \;& 4\,e^2 \sum_{\underline{k}\in\Theta} |a_{\underline{k}}| < \infty. \end{aligned}$$

Hence,

$$\begin{aligned} & \lim_{p\to\infty} \sum_{\underline{k}\in\Theta} \mathbb{1}_{\{\underline{k}\in\Theta(p)\}} |a_{\underline{k}}| \left(\frac{p+1}{p}\right)^{|k_1|} \left(\left(\frac{p+1}{p}\right)^{k_2} + \left(\frac{p+1}{p}\right)^{|k_1|} - 2 \right) \\ = \;& \sum_{\underline{k}\in\Theta} |a_{\underline{k}}| \lim_{p\to\infty} \left(\frac{p+1}{p}\right)^{|k_1|} \left(\left(\frac{p+1}{p}\right)^{k_2} + \left(\frac{p+1}{p}\right)^{|k_1|} - 2 \right) \\ = \;& 0. \end{aligned}$$

Therefore, we have $D(p) \to 0$, as $p \to \infty$. From the representation of $A(\underline{\tilde{z}})$ as an exponential of a bounded function, cf. (8.4) and (8.3), we have $|A(\underline{\tilde{z}})| \geq 2\delta$ uniformly for all $\underline{\tilde{z}} \in S$ and for some $\delta > 0$. Then, choosing p large enough such that $D(p) \leq \delta$, (8.8) implies

$$|A_p(\underline{z})| \geq \delta \quad \forall \underline{z} \in S_p, \tag{8.12}$$

which is the first assertion of Lemma 3.3.

Now let p be large enough such that (8.12) holds, but fixed. We will derive the series representation of $B_p(\underline{z}) = 1/A_p(\underline{z})$ as in Lemma 3.3. Using the equivalent notations $a_{\underline{k}}(p)$ and $a_{(k_1,k_2)}(p)$ for the coefficients, we can write

$$A_p(\underline{z}) = 1 + \sum_{\underline{k}\in\Theta(p)} a_{\underline{k}}(p)\, z_1^{k_1} z_2^{k_2} = \sum_{k_2=0}^{p} \alpha(z_1,k_2)\, z_2^{k_2},$$

where

$$\alpha(z_1, k_2) := \sum_{k_1=-p}^{p} a_{(k_1,k_2)}(p)\, z_1^{k_1},$$

with $a_{(0,0)}(p) := 1$ and $a_{(k_1,0)}(p) := 0$ for all $k_1 < 0$. Let z_1 be fixed with $p/(p+1) \leq |z_1| \leq (p+1)/p$. Then, since $A_p(z_1, z_2)$ is a polynomial in z_2 (for each fixed z_1) with a finite number of complex roots, which is bounded away from zero on $|z_2| \leq (p+1)/p$, it actually has no complex roots on a slightly larger open disk $|z_2| < (p+1)/p + \varepsilon$, and $B_p(z_1, z_2) = 1/A_p(z_1, z_2)$ is an analytic function (in z_2) on this open disk $|z_2| < (p+1)/p + \varepsilon$. Thus, $B_p(z_1, z_2)$ can be represented as a power series (in z_2)

$$B_p(z_1, z_2) = \Big(\sum_{k_2=0}^{p} \alpha(z_1, k_2)\, z_2^{k_2} \Big)^{-1} = \sum_{k_2=0}^{\infty} \beta(z_1, k_2)\, z_2^{k_2}, \tag{8.13}$$

which converges absolutely on $|z_2| \leq (p+1)/p$. The coefficients $\beta(z_1, k_2)$ can be determined recursively from

$$1 = \sum_{k_2=0}^{p} \alpha(z_1, k_2)\, z_2^{k_2} \cdot \sum_{k_2=0}^{\infty} \beta(z_1, k_2)\, z_2^{k_2} = \sum_{k_2=0}^{\infty} \Big(\sum_{l_2=0}^{p\wedge k_2} \alpha(z_1, l_2)\, \beta(z_1, k_2 - l_2) \Big) z_2^{k_2}.$$

For example, if $p \geq 2$, one can derive

$$\beta(z_1, 0) = \frac{1}{\alpha(z_1, 0)}, \quad \beta(z_1, 1) = \frac{-\alpha(z_1, 1)}{\alpha(z_1, 0)^2}, \quad \beta(z_1, 2) = \frac{\alpha(z_1, 1)^2 - \alpha(z_1, 0)\alpha(z_1, 2)}{\alpha(z_1, 0)^3},$$

and so on. In general, one can obtain that

$$\beta(z_1, k_2) = \frac{\eta(z_1, k_2)}{\alpha(z_1, 0)^{k_2+1}}, \tag{8.14}$$

where $\eta(z_1, k_2)$ is some finite linear combination of certain k_2-fold products of the coefficients $\alpha(z_1, 0), \ldots, \alpha(z_1, p)$. Hence, it is easy to see that each $\eta(z_1, k_2)$ can be expressed as

$$\eta(z_1, k_2) = \sum_{k_1=-pk_2}^{pk_2} c_{(k_1,k_2)}(p)\, z_1^{k_1}, \tag{8.15}$$

defined on $p/(p+1) \leq |z_1| \leq (p+1)/p$, for suitable coefficients $c_{(k_1,k_2)}(p)$.

We will now develop Laurent series expansions (in z_1) for each $\beta(z_1, k_2)$. At first, observe that in (3.4) the z-transform $A(z_1, z_2)$ was defined on the domain S, i.e. on $|z_1| = 1$, $|z_2| \leq 1$, because the series converges absolutely on S due to

$$|A(z_1, z_2)| \leq 1 + \sum_{\underline{k}\in\Theta} |a_{(k_1,k_2)}| \left|z_1^{k_1}\right| \left|z_2^{k_2}\right| \leq 1 + \sum_{\underline{k}\in\Theta} |a_{(k_1,k_2)}| < \infty.$$

Note that in the series expansion in (3.4), only exponents $k_2 \geq 0$ but both positive and negative exponents k_1 show up. However, for $z_2 = 0$ fixed, the series reduces to

$$A(z_1, 0) = 1 + \sum_{k_1=1}^{\infty} a_{(k_1,0)}\, z_1^{k_1}$$

with only positive exponents k_1. Therefore the series expansion $A(z_1, 0)$ actually converges absolutely not only on the unit circle $|z_1| = 1$, but on the entire disk $|z_1| \leq 1$. Analogously, $A_p(z_1, 0)$ reduces to a polynomial

$$A_p(z_1, 0) = 1 + \sum_{k_1=1}^{p} a_{(k_1,0)}(p)\, z_1^{k_1},$$

which is defined not only on the ring $p/(p+1) \leq |z_1| \leq (p+1)/p$ but on the closed disk $|z_1| \leq (p+1)/p$. Again, from the representation of $A(z_1, 0)$ as an exponential of a bounded function, cf. (8.4) and (8.3), we get that $|A(z_1, 0)|$ is uniformly bounded away from zero on $|z_1| \leq 1$. Also, with the very same technique as for showing (8.8), we can derive

$$\sup_{|z_1|\leq 1} \Big| A_p(z_1, 0) - A(z_1, 0) \Big| \leq D(p),$$

and therefore, for the fixed p large enough chosen above, we have

$$|A_p(z_1, 0)| \geq \delta \quad \forall\, |z_1| \leq (p+1)/p.$$

Hence $1/A_p(z_1, 0)$ can be expanded as an absolutely convergent power series on $|z_1| \leq (p+1)/p$. Since we also have per definition $\alpha(z_1, 0) = A_p(z_1, 0)$ and $\beta(z_1, 0) = 1/\alpha(z_1, 0)$, it holds

$$\beta(z_1, 0) = \frac{1}{A_p(z_1, 0)} = 1 + \sum_{k_1=1}^{\infty} b_{(k_1,0)}(p)\, z_1^{k_1},$$

for suitable coefficients $b_{(k_1,0)}(p)$. It follows immediately that for each $k_2 \geq 1$

$$\frac{1}{\alpha(z_1, 0)^{k_2+1}} = \beta(z_1, 0)^{k_2+1} = 1 + \sum_{k_1=1}^{\infty} \widetilde{b}_{(k_1,k_2)}(p)\, z_1^{k_1},$$

for suitable coefficients $\widetilde{b}_{(k_1,k_2)}(p)$, the series absolutely convergent on $|z_1| \leq (p+1)/p$. This expansion, together with (8.15) and (8.14), shows that for all $k_2 \geq 1$

$$\beta(z_1, k_2) = \sum_{k_1=-pk_2}^{\infty} b_{(k_1,k_2)}(p)\, z_1^{k_1},$$

absolutely convergent on $p/(p+1) \leq |z_1| \leq (p+1)/p$, for suitable coefficients $b_{(k_1,k_2)}(p)$. Inserting this into (8.13), and setting $b_{(k_1,k_2)}(p) := 0$ for all $k_1 < -k_2 p$, yields

$$\begin{aligned} B_p(z_1, z_2) &= \sum_{k_2=0}^{\infty} \beta(z_1, k_2)\, z_2^{k_2} \\ &= 1 + \sum_{k_1=1}^{\infty} b_{(k_1,0)}(p)\, z_1^{k_1} + \sum_{k_2=1}^{\infty} \sum_{k_1=-pk_2}^{\infty} b_{(k_1,k_2)}(p)\, z_1^{k_1}\, z_2^{k_2} \\ &= 1 + \sum_{\underline{k}\in\Theta} b_{\underline{k}}(p)\, z_1^{k_1}\, z_2^{k_2}, \end{aligned}$$

which completes the proof. □

Proof of Lemma 3.4:
We will use the space C_r of functions on $(-\pi,\pi]^2$ with finite norm $\|\cdot\|_r$ as defined in Lemma 2.3. In the proof of Lemma 2.1 we introduced the functions $A', B' \in C_{r-1}$ which possess the representations

$$A'(\underline{\lambda}) = \widetilde{a}_{\underline{0}}\Big(1 - \sum_{\underline{k}\in\Theta} a_{\underline{k}}\, e^{-i\langle\underline{k},\underline{\lambda}\rangle}\Big), \quad B'(\underline{\lambda}) = \widetilde{b}_{\underline{0}}\Big(1 + \sum_{\underline{k}\in\Theta} b_{\underline{k}}\, e^{-i\langle\underline{k},\underline{\lambda}\rangle}\Big),$$

respectively, as can be seen from (8.4). Hence, the Fourier coefficients of A', B' are given by the autoregressive and moving average parameters $a_{\underline{k}}$ and $b_{\underline{k}}$, up to the constant non-zero factors $\widetilde{a}_{\underline{0}}, \widetilde{b}_{\underline{0}}$. In order to simplify the notation in the remainder of this proof we define for all $p \geq p_0$ the functions $A'_p, B'_p : (-\pi,\pi]^2 \to \mathbb{R}$ via $A'_p(\underline{\lambda}) := \widetilde{a}_{\underline{0}}\, A_p(e^{-i\lambda_1}, e^{-i\lambda_2})$, $B'_p(\underline{\lambda}) := \widetilde{b}_{\underline{0}}\, B_p(e^{-i\lambda_1}, e^{-i\lambda_2})$ and obtain

$$A'_p(\underline{\lambda}) = \widetilde{a}_{\underline{0}}\Big(1 - \sum_{\underline{k}\in\Theta(p)} a_{\underline{k}}(p)\, e^{-i\langle\underline{k},\underline{\lambda}\rangle}\Big), \quad B'_p(\underline{\lambda}) = \widetilde{b}_{\underline{0}}\Big(1 + \sum_{\underline{k}\in\Theta} b_{\underline{k}}(p)\, e^{-i\langle\underline{k},\underline{\lambda}\rangle}\Big),$$

cf. (3.5) and (3.6) for the definitions of $A_p(\underline{z})$ and $B_p(\underline{z})$. Since $\widetilde{a}_{\underline{0}} = 1/\widetilde{b}_{\underline{0}}$, we can conclude from (8.3) and (3.6) that

$$A'_p(\underline{\lambda}) = 1/B'_p(\underline{\lambda}), \quad A'(\underline{\lambda}) = 1/B'(\underline{\lambda}) \quad \forall\, \underline{\lambda} \in (-\pi,\pi]^2. \tag{8.16}$$

We now have established the necessary notation to prove the assertion of Lemma 3.4. For all $s \in \mathbb{N}_0$ with $s+1 < r$ we derive, using (8.16) and the submultiplicativity of $\|\cdot\|_s$ established in (8.5),

$$\begin{aligned}
&\sum_{k\in\Theta} (1+|\underline{k}|_\infty)^s\, |b_{\underline{k}}(p) - b_{\underline{k}}| \\
=\;& \Big\| (1/\widetilde{b}_{\underline{0}}) \cdot (B'_p - B') \Big\|_s && (8.17)\\
=\;& (1/\widetilde{b}_{\underline{0}}) \cdot \Big\| B'_p \cdot \big[A' - A'_p\big] \cdot B' \Big\|_s \\
\leq\;& (1/\widetilde{b}_{\underline{0}}) \cdot \Big(\|B'_p - B'\|_s + \|B'\|_s\Big) \cdot \|A' - A'_p\|_s \cdot \|B'\|_s. && (8.18)
\end{aligned}$$

From Baxter's inequality, cf. Theorem 3.2, we can infer

$$\begin{aligned}
\|A' - A'_p\|_s \;&=\; \widetilde{a}_{\underline{0}} \cdot \Big(\sum_{k\in\Theta(p)} (1+|\underline{k}|_\infty)^s\, |a_{\underline{k}}(p) - a_{\underline{k}}| + \sum_{k\in\Theta\setminus\Theta(p)} (1+|\underline{k}|_\infty)^s\, |a_{\underline{k}}| \Big) \\
&\leq\; \widetilde{a}_{\underline{0}}\Big(\frac{M}{2\sqrt{2}\,\pi^2 c} + 1\Big) \cdot \sum_{k\in\Theta\setminus\Theta(p)} (1+|\underline{k}|_\infty)^{s+1}\, |a_{\underline{k}}| \qquad \forall\, p \geq p_0. && (8.19)
\end{aligned}$$

Because the right-hand side converges to zero as $p \to \infty$, one can always find $p \in \mathbb{N}$ such that $\|A' - A'_p\|_s$ becomes arbitrarily small. In particular, for some arbitrary $\delta \in (0,1)$, choose $p_1 \geq p_0$ such that

$$\|A' - A'_p\|_s \cdot \|B'\|_s \leq \delta$$

for all $p \geq p_1$. Taking the difference of (8.18) and (8.17) we get

$$\begin{aligned} \|B'_p - B'\|_s &\leq \frac{\|B'\|_s^2 \cdot \|A' - A'_p\|_s}{1 - \|A' - A'_p\|_s \cdot \|B'\|_s} \\ &\leq \frac{\|B'\|_s^2}{1-\delta} \cdot \|A' - A'_p\|_s \end{aligned} \tag{8.20}$$

for all $p \geq p_2$. Since the first factor on the right-hand side of (8.20) does not depend on p and is finite, applying (8.19) to the second factor yields that there exists $C < \infty$ such that

$$\sum_{\underline{k} \in \Theta} (1 + |\underline{k}|_\infty)^s \, |b_{\underline{k}}(p) - b_{\underline{k}}| \leq C \cdot \sum_{\underline{k} \in \Theta \setminus \Theta(p)} (1 + |\underline{k}|_\infty)^{s+1} \, |a_{\underline{k}}| \qquad \forall \, p \geq p_2,$$

which completes the proof. $\square$

Proof of Lemma 3.5:
Due to Lemma 3.3 and Assumption 2, we can choose $\delta > 0$ and $n_0 \in \mathbb{N}$ large enough such that

$$\left|A_p(\underline{z})\right| \geq \delta, \quad \left|\widehat{A}_p(\underline{z})\right| \geq \delta \text{ in prob.} \tag{8.21}$$

for all $n \geq n_0$ and for all $\underline{z} \in S_p$. For those n, $B_p(\underline{z}) = 1 + \sum_{\underline{k} \in \Theta} b_{\underline{k}}(p) \, z_1^{k_1} z_2^{k_2}$ can be expanded as a power series in z_2 with coefficients depending on z_1. These coefficients can then be expanded as Laurent series in z_1, cf. the proof of Lemma 3.3 for a detailed explanation and and introduction of the notation which will also be used in this proof. To be precise, we have

$$B_p(z_1, z_2) = \sum_{k_2=0}^{\infty} \beta(z_1, k_2) \, z_2^{k_2}, \quad \text{where } \beta(z_1, k_2) = \sum_{k_1=-\infty}^{\infty} b_{(k_1,k_2)}(p) \, z_1^{k_1},$$

with $b_{(k_1,0)}(p) = 0$ for $k_1 < 0$ and $b_{(0,0)}(p) = 1$. Following exactly along the lines of the proof of Lemma 3.3, we get an expansion with the very same structure for $\widehat{B}_p(\underline{z}) = 1 + \sum_{\underline{k} \in \Theta} \widehat{b}_{\underline{k}}(p) \, z_1^{k_1} z_2^{k_2}$, in probability, as

$$\widehat{B}_p(z_1, z_2) = \sum_{k_2=0}^{\infty} \widehat{\beta}(z_1, k_2) \, z_2^{k_2}, \quad \text{where } \widehat{\beta}(z_1, k_2) = \sum_{k_1=-\infty}^{\infty} \widehat{b}_{(k_1,k_2)}(p) \, z_1^{k_1},$$

also with $\widehat{b}_{(k_1,0)}(p) = 0$ for $k_1 < 0$ and $\widehat{b}_{(0,0)}(p) = 1$. Then, for any $k_2 \geq 0$, we have the Laurent series expansion

$$\widehat{\beta}(z_1, k_2) - \beta(z_1, k_2) = \sum_{k_1=-\infty}^{\infty} \left(\widehat{b}_{(k_1,k_2)}(p) - b_{(k_1,k_2)}(p)\right) z_1^{k_1}$$

in probability, which converges absolutely on the ring $R_1 := p/(p+1) \leq z_1 \leq (p+1)/p$. Actually, following the same argument as for the function $\beta(z_1, k_2)$ in the proof of Lemma 3.3, the function $\widehat{\beta}(z_1, k_2) - \beta(z_1, k_2)$ is analytic (and the Laurent series expansion thus valid) on a slightly larger open set which contains the closed ring R_1 as a subset. Therefore, Cauchy's inequality for analytic functions yields the following bounds for the coefficients:

$$\begin{aligned} \left|\widehat{b}_{(k_1,k_2)}(p) - b_{(k_1,k_2)}(p)\right| &\le \left(\frac{p+1}{p}\right)^{-k_1} \sup_{|z_1|=(p+1)/p} \left|\widehat{\beta}(z_1,k_2) - \beta(z_1,k_2)\right| \quad \forall\, k_1 \ge 0, \\ \left|\widehat{b}_{(k_1,k_2)}(p) - b_{(k_1,k_2)}(p)\right| &\le \left(\frac{p}{p+1}\right)^{-k_1} \sup_{|z_1|=p/(p+1)} \left|\widehat{\beta}(z_1,k_2) - \beta(z_1,k_2)\right| \quad \forall\, k_1 < 0 \end{aligned}$$

in probability. These two bounds can be combined to obtain

$$\begin{aligned} &\left|\widehat{b}_{(k_1,k_2)}(p) - b_{(k_1,k_2)}(p)\right| \\ &\le \left(\frac{p+1}{p}\right)^{-|k_1|} \sup_{p/(p+1)\le|z_1|\le(p+1)/p} \left|\widehat{\beta}(z_1,k_2) - \beta(z_1,k_2)\right| \end{aligned} \tag{8.22}$$

in probability, for all $k_1 \in \mathbb{Z}$. Then, for any $z_1 \in R_1$, $\widehat{B}_p(z_1,z_2) - B_p(z_1,z_2)$, as a function in z_2, has the power series expansion

$$\widehat{B}_p(z_1,z_2) - B_p(z_1,z_2) = \sum_{k_2=0}^{\infty} \left(\widehat{\beta}(z_1,k_2) - \beta(z_1,k_2)\right) z_2^{k_2},$$

in probability, which converges absolutely on the closed disk $|z_2| \le (p+1)/p$. Hence, Cauchy's inequality yields the bound

$$\left|\widehat{\beta}(z_1,k_2) - \beta(z_1,k_2)\right| \le \left(\frac{p+1}{p}\right)^{-k_2} \sup_{|z_2|=(p+1)/p} \left|\widehat{B}_p(z_1,z_2) - B_p(z_1,z_2)\right| \quad \text{in prob.}$$

Inserting this bound into (8.22), and using (8.21), we get

$$\begin{aligned} &\left|\widehat{b}_{(k_1,k_2)}(p) - b_{(k_1,k_2)}(p)\right| \\ &\le \left(\frac{p+1}{p}\right)^{-|k_1|-k_2} \sup_{p/(p+1)\le|z_1|\le(p+1)/p,\ |z_2|=(p+1)/p} \left|\widehat{B}_p(z_1,z_2) - B_p(z_1,z_2)\right| \\ &\le \left(\frac{p+1}{p}\right)^{-|k_1|-k_2} \sup_{\underline{z}\in S_p} \left|\frac{\widehat{A}_p(\underline{z}) - A_p(\underline{z})}{\widehat{A}_p(\underline{z})\, A_p(\underline{z})}\right| \\ &\le \left(\frac{p+1}{p}\right)^{-|k_1|-k_2} \frac{1}{\delta^2} \sup_{\underline{z}\in S_p} \sum_{\underline{j}\in\Theta(p)} \left|\widehat{a}_{\underline{j}}(p) - a_{\underline{j}}(p)\right| |z_1|^{j_1}\, |z_2|^{j_2} \\ &\le \left(\frac{p+1}{p}\right)^{-|k_1|-k_2} \left(\frac{p+1}{p}\right)^{2p} \frac{1}{\delta^2} \sum_{\underline{j}\in\Theta(p)} \left|\widehat{a}_{\underline{j}}(p) - a_{\underline{j}}(p)\right| \\ &\le \left(\frac{p+1}{p}\right)^{-|k_1|-k_2} \frac{1}{p^4} \cdot C \end{aligned}$$

in probability, for some $C < \infty$, because of Assumption 2 and since $((p+1)/p)^{2p}$ is a sequence bounded by e^2. □

Proof of Lemma 7.1, assertion (7.2):
Throughout this proof, we consider only those n large enough such that $A_p(\underline{z})$ and $\widehat{A}_p(\underline{z})$ are bounded away from zero on S_p, the latter in probability, cf. Lemma 3.5. It holds

$$\begin{aligned}\sum_{\underline{k}\in\Theta}(1+|\underline{k}|_\infty)^2\left|\widehat{b}_{\underline{k}}(p)\right| \quad\leq\quad & \sum_{\underline{k}\in\Theta}(1+|\underline{k}|_\infty)^2\,|b_{\underline{k}}| + \sum_{\underline{k}\in\Theta}(1+|\underline{k}|_\infty)^2\,|b_{\underline{k}}(p)-b_{\underline{k}}| \\ & + \sum_{\underline{k}\in\Theta}(1+|\underline{k}|_\infty)^2\left|\widehat{b}_{\underline{k}}(p)-b_{\underline{k}}(p)\right|.\end{aligned}$$

The first summand on the right-hand side is finite due to (2.6) while the second summand can be bounded with Lemma 3.4 by

$$\sum_{\underline{k}\in\Theta}(1+|\underline{k}|_\infty)^2\,|b_{\underline{k}}(p)-b_{\underline{k}}| \leq C\cdot\sum_{\underline{k}\in\Theta}(1+|\underline{k}|_\infty)^3\,|a_{\underline{k}}|,$$

uniformly for all p (and thus all n). The right-hand side, again, is finite due to (2.6). Hence, the proof can be completed by showing

$$\sum_{\underline{k}\in\Theta}(1+|\underline{k}|_\infty)^2\left|\widehat{b}_{\underline{k}}(p)-b_{\underline{k}}(p)\right| = \mathcal{O}_P(1).$$

Due to

$$(1+|\underline{k}|_\infty)^2 \leq 1+3\,|\underline{k}|_\infty^2 \leq 3\,(1+|k_1|^2)+3\,(1+|k_2|^2)$$

it suffices to show

$$\sum_{\underline{k}\in\Theta}(1+|k_j|^2)\left|\widehat{b}_{\underline{k}}(p)-b_{\underline{k}}(p)\right| = \mathcal{O}_P(1), \quad j=1,2. \tag{8.23}$$

Let $j=1$. Lemma 3.5 yields

$$\begin{aligned}&\sum_{\underline{k}\in\Theta}(1+|k_1|^2)\left|\widehat{b}_{\underline{k}}(p)-b_{\underline{k}}(p)\right| \\ \leq\;& C\cdot\frac{1}{p^4}\sum_{\underline{k}\in\Theta}(1+|k_1|^2)\left(\frac{p+1}{p}\right)^{-|k_1|-k_2} \\ \leq\;& C\cdot\frac{1}{p^4}\sum_{k_1=-\infty}^{\infty}(1+|k_1|^2)\left(\frac{p}{p+1}\right)^{|k_1|}\cdot\sum_{k_2=0}^{\infty}\left(\frac{p}{p+1}\right)^{k_2} \quad \text{in prob.}\end{aligned} \tag{8.24}$$

For any $|x|<1$, differentiating the geometric series twice yields

$$\sum_{m=0}^{\infty}(m+2)\,(m+1)\,x^m = \frac{2}{(1-x)^3},$$

thus, the right-hand side of (8.24) can be bounded by

$$\begin{aligned}& C\cdot\frac{1}{p^4}\,2\sum_{k_1=0}^{\infty}(k_1+2)\,(k_1+1)\left(\frac{p}{p+1}\right)^{k_1}\cdot\sum_{k_2=0}^{\infty}\left(\frac{p}{p+1}\right)^{k_2} \\ =\;& C\cdot\frac{4}{p^4}\cdot\left(1-\frac{p}{p+1}\right)^{-3}\cdot\left(1-\frac{p}{p+1}\right)^{-1}\end{aligned}$$

$$= \; C \cdot \frac{4(p+1)^4}{p^4} = \mathcal{O}(1).$$

Since the inequality in (8.24) holds in probability, we have shown (8.23) for $j = 1$. The same calculation can be performed for $j = 2$, which completes the proof. $\square$

Proof of Lemma 7.1, assertion (7.3):
The random variables $\varepsilon^*_{\underline{t}}$ are, conditionally on the given data, uniformly distributed on $\{\widehat{\varepsilon}_{\underline{s}}(p) : \underline{s} \in \Pi(n,p)\}$. Hence, it holds

$$E^*\Big(|\varepsilon^*_{\underline{t}}|^{2w}\Big) = \frac{1}{|\Pi(n,p)|} \sum_{\underline{s}\in\Pi(n,p)} |\widehat{\varepsilon}_{\underline{s}}(p)|^{2w},$$

and the goal is to show that the right-hand side converges to $E\Big(|\varepsilon_{\underline{t}}|^{2w}\Big) = E\Big(|\varepsilon_{\underline{0}}|^{2w}\Big)$. Because of Assumption 4 it suffices to show

$$\frac{1}{|\Pi(n,p)|} \sum_{\underline{t}\in\Pi(n,p)} \Big(|\widehat{\varepsilon}_{\underline{t}}(p)|^{2w} - |\varepsilon_{\underline{t}}|^{2w}\Big) \xrightarrow{P} 0. \tag{8.25}$$

We have $\widehat{\varepsilon}_{\underline{t}}(p) = \varepsilon'_{\underline{t}}(p) - \overline{\varepsilon}$ with $\varepsilon'_{\underline{t}}(p) = X_{\underline{t}} - \sum_{\underline{k}\in\Theta(p)} \widehat{a}_{\underline{k}}(p)\, X_{\underline{t}-\underline{k}}$ and $\overline{\varepsilon} = (1/|\Pi(n,p)|)\sum_{\underline{t}\in\Pi(n,p)} \varepsilon'_{\underline{t}}(p)$. Thus, with representation (2.5), we have

$$\widehat{\varepsilon}_{\underline{t}}(p) = \varepsilon'_{\underline{t}}(p) - \overline{\varepsilon} = \varepsilon_{\underline{t}} + Q_{\underline{t}} + R_{\underline{t}} - \overline{\varepsilon},$$

where

$$\begin{aligned}
Q_{\underline{t}} \;&:=\; \sum_{\underline{k}\in\Theta(p)} \Big(a_{\underline{k}}(p) - \widehat{a}_{\underline{k}}(p)\Big)\, X_{\underline{t}-\underline{k}},\\
R_{\underline{t}} \;&:=\; \sum_{\underline{k}\in\Theta(p)} \Big(a_{\underline{k}} - a_{\underline{k}}(p)\Big)\, X_{\underline{t}-\underline{k}} + \sum_{\underline{k}\in\Theta\setminus\Theta(p)} a_{\underline{k}}\, X_{\underline{t}-\underline{k}}.
\end{aligned}$$

Decomposing $|\widehat{\varepsilon}_{\underline{t}}(p)|^{2w} = (\varepsilon_{\underline{t}} + Q_{\underline{t}} + R_{\underline{t}} - \overline{\varepsilon})^{2w}$ with a binomial expansion (with the notation $|\underline{d}| = d_1 + d_2 + d_3 + d_4$ for vectors $\underline{d} \in \mathbb{N}_0^4$), one can easily see that it holds for some $C < \infty$

$$\begin{aligned}
&\left|\frac{1}{|\Pi(n,p)|} \sum_{\underline{t}\in\Pi(n,p)} \Big(|\widehat{\varepsilon}_{\underline{t}}(p)|^{2w} - |\varepsilon_{\underline{t}}|^{2w}\Big)\right|\\
\le\;& C \cdot \sum_{|\underline{d}|=2w,\; d_1\neq 2w} \frac{1}{|\Pi(n,p)|} \sum_{\underline{t}\in\Pi(n,p)} |\varepsilon_{\underline{t}}|^{d_1}\, |Q_{\underline{t}}|^{d_2}\, |R_{\underline{t}}|^{d_3}\, |\overline{\varepsilon}|^{d_4}\\
\le\;& C \cdot \sum_{|\underline{d}|=2w,\; d_1\neq 2w} (I)^{d_1/2w}\, (II)^{d_2/2w}\, (III)^{d_3/2w}\, (IV)^{d_4/2w},
\end{aligned}$$

where Hölder's inequality was used in the final step and, moreover,

$$\begin{aligned}
(I) \;&=\; \frac{1}{|\Pi(n,p)|} \sum_{\underline{t}\in\Pi(n,p)} |\varepsilon_{\underline{t}}|^{2w},\; (II) = \frac{1}{|\Pi(n,p)|} \sum_{\underline{t}\in\Pi(n,p)} |Q_{\underline{t}}|^{2w},\\
(III) \;&=\; \frac{1}{|\Pi(n,p)|} \sum_{\underline{t}\in\Pi(n,p)} |R_{\underline{t}}|^{2w},\; (IV) = \frac{1}{|\Pi(n,p)|} \sum_{\underline{t}\in\Pi(n,p)} |\overline{\varepsilon}|^{2w} = |\overline{\varepsilon}|^{2w}.
\end{aligned}$$

Obviously, (8.25) holds true if we can show

$$(I) = \mathcal{O}_P(1),\ (II) = o_P(1),\ (III) = o_P(1),\ (IV) = o_P(1). \tag{8.26}$$

In the first step we show $(IV_n) = o_P(1)$. From the definition of $\overline{\varepsilon}$ we get

$$|\overline{\varepsilon}| \leq \left| \frac{1}{|\Pi(n,p)|} \sum_{\underline{t}\in\Pi(n,p)} \varepsilon_{\underline{t}} \right| + \frac{1}{|\Pi(n,p)|} \sum_{\underline{t}\in\Pi(n,p)} \left|Q_{\underline{t}}\right| + \frac{1}{|\Pi(n,p)|} \sum_{\underline{t}\in\Pi(n,p)} \left|R_{\underline{t}}\right|. \tag{8.27}$$

The first summand on the right-hand side obviously converges to zero in probability because of the WLLN (recall that the random variables $\varepsilon_{\underline{t}}$ are uncorrelated and have mean zero). Considering the second summand, we have from Assumption 1 that $|X_{\underline{t}}| = \mathcal{O}_P(1)$ uniformly for all $\underline{t} \in \mathbb{Z}^2$, and, since $|\Theta(p)| = 2p(p+1)$,

$$\sum_{\underline{k}\in\Theta(p)} \left|X_{\underline{t}-\underline{k}}\right| = 2p(p+1)\,\mathcal{O}_P(1) = \mathcal{O}_P(p^2). \tag{8.28}$$

It follows

$$\left|Q_{\underline{t}}\right| \leq \sum_{\underline{k}\in\Theta(p)} \left|a_{\underline{k}}(p) - \widehat{a}_{\underline{k}}(p)\right| \cdot \sum_{\underline{k}\in\Theta(p)} \left|X_{\underline{t}-\underline{k}}\right| \leq \frac{1}{p^4}\,\mathcal{O}_P(1)\cdot\mathcal{O}_P(p^2) = o_P(1), \tag{8.29}$$

where Assumption 2 was used. Since this bound does not depend on $\underline{t}$, we have

$$\frac{1}{|\Pi(n,p)|} \sum_{\underline{t}\in\Pi(n,p)} \left|Q_{\underline{t}}\right| = o_P(1). \tag{8.30}$$

Now consider the third summand on the right-hand side of (8.27). We will need the following preliminary results: From Theorem 3.2 and summability condition (2.6) (here, we assume $r = 4$) we get

$$\begin{aligned} & p^2 \cdot \sum_{\underline{k}\in\Theta(p)} |a_{\underline{k}}(p) - a_{\underline{k}}| \\ \leq\ & C \cdot \sum_{\underline{k}\in\Theta\setminus\Theta(p)} p^2\,(1+|\underline{k}|_\infty)\,|a_{\underline{k}}| \leq C \cdot \sum_{\underline{k}\in\Theta\setminus\Theta(p)} (1+|\underline{k}|_\infty)^3\,|a_{\underline{k}}| = o(1), \end{aligned}$$

because $p \leq |\underline{k}|_\infty$ for all $\underline{k} \in \Theta \setminus \Theta(p)$, and $\Theta(p) \to \Theta$, as $n \to \infty$. Hence we have $\sum_{\underline{k}\in\Theta(p)} |a_{\underline{k}}(p) - a_{\underline{k}}| = o(p^{-2})$. Moreover, since $|X_{\underline{t}}| = \mathcal{O}_P(1)$ uniformly for all $\underline{t} \in \mathbb{Z}^2$, we have

$$p \sum_{\underline{k}\in\Theta\setminus\Theta(p)} |a_{\underline{k}}|\,|X_{\underline{t}-\underline{k}}| \leq \mathcal{O}_P(1) \sum_{\underline{k}\in\Theta\setminus\Theta(p)} p\,|a_{\underline{k}}| \leq \mathcal{O}_P(1)\cdot\sum_{\underline{k}\in\Theta}(1+|\underline{k}|_\infty)\,|a_{\underline{k}}| = \mathcal{O}_P(1),$$

due to (2.6). This implies $\sum_{\underline{k}\in\Theta\setminus\Theta(p)} |a_{\underline{k}}|\,|X_{\underline{t}-\underline{k}}| = \mathcal{O}_P(p^{-1})$. Combining these results and (8.28) we get

$$\begin{aligned} \left|R_{\underline{t}}\right| &\leq \sum_{\underline{k}\in\Theta(p)} \left|a_{\underline{k}} - a_{\underline{k}}(p)\right| \cdot \sum_{\underline{k}\in\Theta(p)} \left|X_{\underline{t}-\underline{k}}\right| + \sum_{\underline{k}\in\Theta\setminus\Theta(p)} |a_{\underline{k}}|\,|X_{\underline{t}-\underline{k}}| \\ &\leq o(p^{-2})\cdot\mathcal{O}_P(p^2) + \mathcal{O}_P(p^{-1}) = o_P(1). \end{aligned} \tag{8.31}$$

Since this bound does not depend on $\underline{t}$, we have

$$\frac{1}{|\Pi(n,p)|}\sum_{\underline{t}\in\Pi(n,p)}\left|R_{\underline{t}}\right| = o_P(1). \tag{8.32}$$

Combining this with (8.27) and (8.28) gives $(IV) = o_P(1)$.
Since the bounds in (8.29) and (8.31) do not depend on $\underline{t}$, it follows immediately

$$\frac{1}{|\Pi(n,p)|}\sum_{\underline{t}\in\Pi(n,p)}|Q_{\underline{t}}|^{2w} = \mathcal{O}_P(p^{-4w}), \qquad \frac{1}{|\Pi(n,p)|}\sum_{\underline{t}\in\Pi(n,p)}|R_{\underline{t}}|^{2w} = o_P(1),$$

i.e. $(II) = o_P(1)$ and $(III) = o_P(1)$. Furthermore, Assumption 4 guarantees $(I) = \mathcal{O}_P(1)$, which delivers the final assertion of (8.26) and completes the proof of (7.3).
As a byproduct, we get a result about the empirical means of $(\widehat{\varepsilon}_{\underline{t}} - \varepsilon_{\underline{t}})^2$, which will be needed later on. Using the fact that $|\varepsilon_{\underline{t}}| = \mathcal{O}_P(1)$ uniformly for all $\underline{t}\in\mathbb{Z}^2$, we can derive

$$\begin{aligned}
&\left|\frac{1}{|\Pi(n,p)|}\sum_{\underline{t}\in\Pi(n,p)}\left(\widehat{\varepsilon}_{\underline{t}} - \varepsilon_{\underline{t}}\right)^2\right| \\
=\ &\left|\frac{1}{|\Pi(n,p)|}\sum_{\underline{t}\in\Pi(n,p)}\left(\widehat{\varepsilon}_{\underline{t}}^2 - \varepsilon_{\underline{t}}^2\right) - 2\cdot\frac{1}{|\Pi(n,p)|}\sum_{\underline{t}\in\Pi(n,p)}\varepsilon_{\underline{t}}\left(\widehat{\varepsilon}_{\underline{t}} - \varepsilon_{\underline{t}}\right)\right| \\
\leq\ &\left|\frac{1}{|\Pi(n,p)|}\sum_{\underline{t}\in\Pi(n,p)}\left(\widehat{\varepsilon}_{\underline{t}}^2 - \varepsilon_{\underline{t}}^2\right)\right| + \mathcal{O}_P(1)\cdot\frac{1}{|\Pi(n,p)|}\sum_{\underline{t}\in\Pi(n,p)}\left|Q_{\underline{t}} + R_{\underline{t}} - \overline{\varepsilon}\right| \\
=\ &o_P(1),
\end{aligned} \tag{8.33}$$

from (8.25) (with $w=1$), (8.30), (8.32) and $(IV) = o_P(1)$. □

Proof of Lemma 7.1, assertion (7.4):
This assertion can be obtained following exactly along the lines of the proof of Lemma 5.5 and Corollary 5.6 of Bühlmann (1997). The only difference to Bühlmann's proof is that we decompose

$$X_{\underline{t}}^* = X_{\underline{t},M}^* + U_{\underline{t}}^* + V_{\underline{t}}^*,$$

with $X_{\underline{t},M}^*$ as defined in (7.1) and

$$U_{\underline{t}}^* := \sum_{\underline{k}\in\Theta(M)}\left(\widehat{b}_{\underline{k}}(p) - b_{\underline{k}}\right)\varepsilon_{\underline{t}-\underline{k}}^*, \quad \text{and} \quad V_{\underline{t}}^* := \sum_{\underline{k}\in\Theta\setminus\Theta(M)}\widehat{b}_{\underline{k}}(p)\,\varepsilon_{\underline{t}-\underline{k}}^*,$$

analogously for $\widetilde{X}_{\underline{t}}$. The only assertions needed to adapt the proof of Bühlmann's Lemma 5.5 are given by (7.2) and (7.3), which correspond to Lemmas 5.1 and 5.3 in Bühlmann (1997), as well as

$$\sum_{\underline{k}\in\Theta}\left|\widehat{b}_{\underline{k}}(p) - b_{\underline{k}}\right| = o_P(1) \tag{8.34}$$

and

$$\varepsilon^*_{\underline{t}} \xrightarrow{d^*} \varepsilon_{\underline{t}} \quad \text{in prob.,} \tag{8.35}$$

which correspond to Lemmas 5.2 and 5.4 in Bühlmann (1997). In the following, we complete the proof by showing the latter two assertions.
It holds

$$\begin{aligned}
& \sum_{\underline{k}\in\Theta} \left|\widehat{b}_{\underline{k}}(p) - b_{\underline{k}}\right| \\
\leq\ & \sum_{\underline{k}\in\Theta} \left|\widehat{b}_{\underline{k}}(p) - b_{\underline{k}}(p)\right| + \sum_{\underline{k}\in\Theta} \left|b_{\underline{k}}(p) - b_{\underline{k}}\right| \\
\leq\ & \frac{1}{p^4}\,\mathcal{O}_P(1) \cdot \sum_{k_1=-\infty}^{\infty} \sum_{k_2=0}^{\infty} \left(\frac{p}{p+1}\right)^{|k_1|+k_2} + C \cdot \sum_{\underline{k}\in\Theta\setminus\Theta(p)} (1+|\underline{k}|_\infty)\,|a_{\underline{k}}| \\
\leq\ & \mathcal{O}_P(1) \cdot \frac{2(p+1)^2}{p^4} + C \cdot \sum_{\underline{k}\in\Theta\setminus\Theta(p)} (1+|\underline{k}|_\infty)\,|a_{\underline{k}}| \\
=\ & o_P(1),
\end{aligned}$$

due to Lemma 3.4, Lemma 3.5, (2.6) and Assumption 2. This yields (8.34).
As for (8.35), we adapt the proof of Lemma 5.4 of Bühlmann (1997). Let F with $F(x) := P\{\varepsilon_{\underline{t}} \leq x\} = P\{\widetilde{\varepsilon}_{\underline{t}} \leq x\}$ be the distribution function of $(\varepsilon_{\underline{t}})$ and let F_n be the empirical distribution function of $\{\varepsilon_{\underline{t}} : \underline{t} \in \Pi(n,p)\}$ as defined in Assumption 4. Furthermore, according to step 2 of the AR sieve bootstrap procedure, $(\varepsilon^*_{\underline{t}})$ is an i.i.d. sequence with marginal distribution function $\widehat{F}_n$, where

$$\widehat{F}_n(x) = \frac{1}{|\Pi(n,p)|} \sum_{\underline{t}\in\Pi(n,p)} \mathbb{1}\{\widehat{\varepsilon}_{\underline{t}}(p) \leq x\},$$

and $\widehat{\varepsilon}_{\underline{t}}(p) = \varepsilon'_{\underline{t}}(p) - \overline{\varepsilon}$ are the centered residuals of the autoregressive fit with $\varepsilon'_{\underline{t}}(p) = X_{\underline{t}} - \sum_{\underline{k}\in\Theta(p)} \widehat{a}_{\underline{k}}(p)\, X_{\underline{t}-\underline{k}}$ and $\overline{\varepsilon} = (1/|\Pi(n,p)|) \sum_{\underline{t}\in\Pi(n,p)} \varepsilon'_{\underline{t}}(p)$. We use the Mallows metric d_2, cf. Bickel and Freedman (1981), and derive

$$d_2(\widehat{F}_n, F) \leq d_2(\widehat{F}_n, F_n) + d_2(F_n, F).$$

From Assumption 4 we have convergence of second moments

$$\left|\int x^2\, dF_n(x) - \int x^2\, dF(x)\right| = \left|\frac{1}{|\Pi(n,p)|} \sum_{\underline{t}\in\Pi(n,p)} (\varepsilon_{\underline{t}})^2 - E(\varepsilon_{\underline{0}})^2\right| = o_P(1).$$

This, together with Assumption 4, implies $d_2(F_n, F) = o_P(1)$, according to Lemma 8.3 of Bickel and Freedman (1981). Now let S be uniformly distributed on the finite set $\Pi(n,p)$. For any given realizations of $\{\varepsilon_{\underline{t}} : \underline{t} \in \Pi(n,p)\}$ and $\{\widehat{\varepsilon}_{\underline{t}}(p) : \underline{t} \in \Pi(n,p)\}$, $\widehat{F}_n$ and F_n are deterministic distribution functions, and it is easy to see that ε_S has distribution function F_n and $\widehat{\varepsilon}_S(p)$ has distribution function $\widehat{F}_n$. Hence, it holds

$$d_2(\widehat{F}_n, F_n) \leq E_S\Big(\widehat{\varepsilon}_S(p) - \varepsilon_S\Big)^2 \leq \frac{1}{|\Pi(n,p)|} \sum_{\underline{t}\in\Pi(n,p)} \Big(\widehat{\varepsilon}_{\underline{t}}(p) - \varepsilon_{\underline{t}}\Big)^2.$$

Therefore, we have for the random variable $d_2(\widehat{F}_n, F_n)$:

$$d_2(\widehat{F}_n, F_n) \leq \frac{1}{|\Pi(n,p)|} \sum_{\underline{t}\in\Pi(n,p)} \Big(\widehat{\varepsilon}_{\underline{t}}(p) - \varepsilon_{\underline{t}}\Big)^2 = o_P(1),$$

due to (8.33). This implies $d_2(\widehat{F}_n, F) = o_P(1)$, and, therefore, (8.35). $\square$

Proof of Lemma 7.1, assertion (7.5):

Let $\underline{c} \in \mathbb{R}^k$ be arbitrary. Using the notation $\|z\|_q = \Big(E(|z|^q)\Big)^{1/q}$, the goal is to prove $\|\underline{c}^T\, g(\widetilde{\mathbf{Y}}_{\underline{t},M})\|_{2+2/(h+1)} \leq C$ uniformly for all $\underline{t} \in \mathbb{Z}^2$. Due to

$$\Big\|\underline{c}^T\, g(\widetilde{\mathbf{Y}}_{\underline{t},M})\Big\|_{2+2/(h+1)} \leq \sum_{v=1}^{k} |c_v|\, \Big\|g_v(\widetilde{\mathbf{Y}}_{\underline{t},M})\Big\|_{2+2/(h+1)}$$

it suffices to show $\|g_v(\widetilde{\mathbf{Y}}_{\underline{t},M})\|_{2+2/(h+1)} \leq C_v$ for all $v = 1, \ldots, k$ (note that C may depend on $\underline{c}$). We will derive this assertion from a slight modification of Lemma 7.2. One can easily observe that the assertion of Lemma 7.2 remains true if one replaces $\widetilde{\mathbf{Y}}_{\underline{t}}$ with $\widetilde{\mathbf{Y}}_{\underline{t},M}$, i.e. it holds for each $W \subset \Theta \cup \{\underline{0}\}$

$$\Big\|g_v(\widetilde{\mathbf{Y}}_{\underline{t},M}) - g_v(\widetilde{\mathbf{Y}}^{(W)}_{\underline{t},M})\Big\|_2 \;\leq\; C_v \cdot \left(\sum_{\underline{k}\in\Theta\setminus W} |b_{\underline{k}}| + \mathbb{1}_{\{\underline{0}\notin W\}}\right).$$

Now we modify the proof of Lemma 7.2 by choosing $W = \emptyset$, which yields $\widetilde{\mathbf{Y}}^{(W)}_{\underline{t},M} = \underline{0}$, and by replacing the $\|\cdot\|_2$-norm with $\|\cdot\|_{2+2/(h+1)}$. Then, (8.42) reads:

$$\left\|\frac{D^{\boldsymbol{\alpha}} g_v(\underline{0})}{\boldsymbol{\alpha}!}\, (\widetilde{\mathbf{Y}}_{\underline{t},M})^{\boldsymbol{\alpha}}\right\|_{2+2/(h+1)} \leq C \cdot \left\|\frac{D^{\boldsymbol{\alpha}} g_v(\underline{0})}{\boldsymbol{\alpha}!}\right\|_{\left(2+\frac{2}{h+1}\right)\left(\frac{h+2}{h+2-|\boldsymbol{\alpha}|}\right)} \cdot \left(\sum_{\underline{k}\in\Theta} |b_{\underline{k}}| + 1\right)^{|\boldsymbol{\alpha}|}.$$

Note that the expression on the right-hand side does not depend on $\underline{t}$ and that

$$\left\|\frac{D^{\boldsymbol{\alpha}} g_v(\underline{0})}{\boldsymbol{\alpha}!}\right\|_{\left(2+\frac{2}{h+1}\right)\left(\frac{h+2}{h+2-|\boldsymbol{\alpha}|}\right)} = \left|\frac{D^{\boldsymbol{\alpha}} g_v(\underline{0})}{\boldsymbol{\alpha}!}\right| < \infty,$$

because the derivative of g_v at the origin is deterministic. Along the lines of the proof of Lemma 7.2, with the modifications mentioned above, one obtains

$$\Big\|g_v(\widetilde{\mathbf{Y}}_{\underline{t},M}) - g_v(\underline{0})\Big\|_{2+2/(h+1)} \;\leq\; C_v \cdot \left(\sum_{\underline{k}\in\Theta} |b_{\underline{k}}| + 1\right),$$

which completes the proof of the second assertion of (7.5) via

$$\begin{aligned}
\Big\|g_v(\widetilde{\mathbf{Y}}_{\underline{t},M})\Big\|_{2+2/(h+1)} &\leq \Big\|g_v(\underline{0})\Big\|_{2+2/(h+1)} + \Big\|g_v(\widetilde{\mathbf{Y}}_{\underline{t},M}) - g_v(\underline{0})\Big\|_{2+2/(h+1)} \\
&\leq \Big|g_v(\underline{0})\Big| + C_v \cdot \left(\sum_{\underline{k}\in\Theta} |b_{\underline{k}}| + 1\right).
\end{aligned}$$

An analogous modification for the bootstrap quantities in Lemma 7.2 yields

$$E^*\Big(\Big|\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M})\Big|^{2+2/(h+1)}\Big) = \Big(\Big\|\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M})\Big\|_{2+2/(h+1)}\Big)^{2+2/(h+1)} = \mathcal{O}_P(1),$$

with exactly the same arguments as for the non-bootstrap quantities. □

Proof of Lemma 7.1, assertion (7.6):
For arbitrary but fixed $\underline{c} \in \mathbb{R}^k$ we abbreviate $l(\underline{x}) := \underline{c}^T g(\underline{x})$. Let $0 < K < \infty$ be a constant that will be specified later on. We define the K-truncated version of function l by

$$\widetilde{l}(\underline{x}) := l(\underline{x}) \cdot \mathbb{1}\{|l(\underline{x})| \leq K\} + K \cdot \operatorname{sgn}(l(\underline{x})) \cdot \mathbb{1}\{|l(\underline{x})| > K\}.$$

For arbitrary $\varepsilon > 0$ we get from standard calculations

$$\begin{aligned} & P\Big\{\Big|\mathrm{Cov}^*\Big(l(\mathbf{Y}^*_{\underline{h},M}), l(\mathbf{Y}^*_{\underline{0},M})\Big) - \mathrm{Cov}\Big(l(\widetilde{\mathbf{Y}}_{\underline{h},M}), l(\widetilde{\mathbf{Y}}_{\underline{0},M})\Big)\Big| > \varepsilon\Big\} \\ \leq\ & P\{\,|I| > \varepsilon/3\,\} + P\{\,|II| > \varepsilon/3\,\} + P\{\,|III| > \varepsilon/3\,\}, \end{aligned} \tag{8.36}$$

where

$$\begin{aligned} I &:= \mathrm{Cov}^*\Big(\,l(\mathbf{Y}^*_{\underline{h},M}), l(\mathbf{Y}^*_{\underline{0},M})\Big) - \mathrm{Cov}^*\Big(\,\widetilde{l}(\mathbf{Y}^*_{\underline{h},M}), \widetilde{l}(\mathbf{Y}^*_{\underline{0},M})\Big), \\ II &:= \mathrm{Cov}^*\Big(\,\widetilde{l}(\mathbf{Y}^*_{\underline{h},M}), \widetilde{l}(\mathbf{Y}^*_{\underline{0},M})\Big) - \mathrm{Cov}\Big(\,\widetilde{l}(\widetilde{\mathbf{Y}}_{\underline{h},M}), \widetilde{l}(\widetilde{\mathbf{Y}}_{\underline{0},M})\Big), \\ III &:= \mathrm{Cov}\Big(\,\widetilde{l}(\widetilde{\mathbf{Y}}_{\underline{h},M}), \widetilde{l}(\widetilde{\mathbf{Y}}_{\underline{0},M})\Big) - \mathrm{Cov}\Big(\,l(\widetilde{\mathbf{Y}}_{\underline{h},M}), l(\widetilde{\mathbf{Y}}_{\underline{0},M})\Big). \end{aligned}$$

Hence, the desired assertion follows if we can, for each $\delta > 0$, specify $0 < K < \infty$ and $n_0 \in \mathbb{N}$ such that the right-hand side of (8.36) is smaller than δ for all $n \geq n_0$.

Per definition, $l(\underline{x})$ can be expanded as

$$l(\underline{x}) = \widetilde{l}(\underline{x}) + \Big[l(\underline{x}) - K \cdot \operatorname{sgn}(l(\underline{x}))\Big] \cdot \mathbb{1}\{|l(\underline{x})| > K\}.$$

Using this, we get

$$\begin{aligned} I \ =\ & \mathrm{Cov}^*\Big(\,\widetilde{l}(\mathbf{Y}^*_{\underline{h},M}), \Big[l(\mathbf{Y}^*_{\underline{0},M}) - K \cdot \operatorname{sgn}(l(\mathbf{Y}^*_{\underline{0},M}))\Big] \cdot \mathbb{1}\{|l(\mathbf{Y}^*_{\underline{0},M})| > K\}\Big) \\ & + \mathrm{Cov}^*\Big(\,\Big[l(\mathbf{Y}^*_{\underline{h},M}) - K \cdot \operatorname{sgn}(l(\mathbf{Y}^*_{\underline{h},M}))\Big] \cdot \mathbb{1}\{|l(\mathbf{Y}^*_{\underline{h},M})| > K\}, \widetilde{l}(\mathbf{Y}^*_{\underline{0},M})\Big) \\ & + \mathrm{Cov}^*\Big(\,\Big[l(\mathbf{Y}^*_{\underline{h},M}) - K \cdot \operatorname{sgn}(l(\mathbf{Y}^*_{\underline{h},M}))\Big] \cdot \mathbb{1}\{|l(\mathbf{Y}^*_{\underline{h},M})| > K\}, \\ & \qquad \Big[l(\mathbf{Y}^*_{\underline{0},M}) - K \cdot \operatorname{sgn}(l(\mathbf{Y}^*_{\underline{0},M}))\Big] \cdot \mathbb{1}\{|l(\mathbf{Y}^*_{\underline{0},M})| > K\}\Big). \end{aligned}$$

The first summand on the right-hand side can be bounded in absolute value with Hölder's and Markov's inequalities by

$$\Big(E^*\big(\widetilde{l}(\mathbf{Y}^*_{\underline{h},M})^2\big)\Big)^{1/2} \cdot \Big(E^*\Big(\Big|l(\mathbf{Y}^*_{\underline{0},M})\Big|^2 \cdot \mathbb{1}\{|l(\mathbf{Y}^*_{\underline{0},M})| > K\}\Big)\Big)^{1/2}$$

$$
\begin{aligned}
&\leq \ \mathcal{O}_P(1) \cdot \Big(E^*\Big(\Big|l(\mathbf{Y}^*_{\underline{0},M})\Big|^{2(h+2)/(h+1)}\Big)\Big)^{(h+1)/2(h+2)} \cdot \Big(P^*\Big\{|l(\mathbf{Y}^*_{\underline{0},M})| > K\Big\}\Big)^{1/2(h+2)} \\
&\leq \ \mathcal{O}_P(1) \cdot \Big(E^*\Big(\Big|l(\mathbf{Y}^*_{\underline{0},M})\Big|^{2(h+2)/(h+1)}\Big)\Big)^{1/2} \cdot \Big(\frac{1}{K^{2(h+2)/(h+1)}}\Big)^{1/2(h+2)} \\
&= \ K^{-1/(h+1)} \cdot \mathcal{O}_P(1),
\end{aligned}
$$

where the boundedness in probability of the moments is taken from (7.5), noting that $2(h+2)/(h+1) = 2 + 2/(h+1)$. The same calculations can be done for the second and third summand above which yields $I = K^{-1/(h+1)} \cdot \mathcal{O}_P(1)$. Hence, for the given $\delta > 0$, there exists $S(\delta) < \infty$ such that

$$
P\Big\{|I| > S(\delta)/K^{1/(h+1)}\Big\} \leq \delta/2 \quad \forall\, n \in \mathbb{N},
$$

and for each $K > (3\, S(\delta)/\varepsilon)^{h+1}$ we have

$$
P\Big\{|I| > \varepsilon/3\Big\} \leq P\Big\{|I| > S(\delta)/K^{1/(h+1)}\Big\} \leq \delta/2 \quad \forall\, n \in \mathbb{N}.
$$

By the very same calculations as for I, replacing E^* with E, one obtains

$$
|III| \leq \widetilde{C} \cdot K^{-1/(h+1)} \quad \forall\, n \in \mathbb{N}
$$

for some $\widetilde{C} < \infty$, using $E\,|l(\widetilde{\mathbf{Y}}_{\underline{t},M})|^{(2+2/(h+1))} \leq C$, cf. (7.5). Choosing $K > (3\,\widetilde{C}/\varepsilon)^{h+1}$ gives

$$
P\Big\{|III| > \varepsilon/3\Big\} \leq P\Big\{|III| > \widetilde{C}/K^{1/(h+1)}\Big\} = 0 \quad \forall\, n \in \mathbb{N},
$$

noting that III is deterministic. Combining the results for I and III, we get from choosing $K > (3\,(\widetilde{C} \vee S(\delta))/\varepsilon)^{h+1}$

$$
P\Big\{|I| > \varepsilon/3\Big\} + P\Big\{|III| > \varepsilon/3\Big\} \leq \delta/2 \quad \forall\, n \in \mathbb{N}. \tag{8.37}
$$

For this fixed $K < \infty$ we will now show $II = o_P(1)$. (7.4) implies

$$
\begin{pmatrix} \mathbf{Y}^*_{\underline{h},M} \\ \mathbf{Y}^*_{\underline{0},M} \end{pmatrix} \xrightarrow{d^*} \begin{pmatrix} \widetilde{\mathbf{Y}}_{\underline{h},M} \\ \widetilde{\mathbf{Y}}_{\underline{0},M} \end{pmatrix} \quad \text{in } P\text{-prob.}
$$

Hence, we have $E^* f(\mathbf{Y}^*_{\underline{h},M}, \mathbf{Y}^*_{\underline{0},M}) \to E\, f(\widetilde{\mathbf{Y}}_{\underline{h},M}, \widetilde{\mathbf{Y}}_{\underline{0},M})$ in P-probability for each continuous and bounded function f. It follows

$$
\begin{aligned}
II \ &= \ E^*\Big(\tilde{l}(\mathbf{Y}^*_{\underline{h},M})\, \tilde{l}(\mathbf{Y}^*_{\underline{0},M})\Big) - E\Big(\tilde{l}(\widetilde{\mathbf{Y}}_{\underline{h},M})\, \tilde{l}(\widetilde{\mathbf{Y}}_{\underline{0},M})\Big) \\
&\quad + E^*\Big(\tilde{l}(\mathbf{Y}^*_{\underline{h},M})\Big)\, E^*\Big(\tilde{l}(\mathbf{Y}^*_{\underline{0},M})\Big) - E\Big(\tilde{l}(\widetilde{\mathbf{Y}}_{\underline{h},M})\Big)\, E\Big(\tilde{l}(\widetilde{\mathbf{Y}}_{\underline{0},M})\Big) \\
&= \ o_P(1),
\end{aligned}
$$

since $\tilde{l}$ is continuous and bounded by K. We can therefore find $n_0 \in \mathbb{N}$ such that

$$
P\{\, |II| > \varepsilon/3\,\} \leq \delta/2 \quad \forall\, n \geq n_0,
$$

which together with (8.36) and (8.37) completes the proof. □

Proof of Lemma 7.1, assertion (7.7):
We prove that $\sum_{|h_1|,|h_2|\leq M}\left|\mathrm{Cov}\Big(g_u(\widetilde{\mathbf{Y}}_{\underline{h}}),g_v(\widetilde{\mathbf{Y}}_{\underline{0}})\Big)\right|$ converges to a finite limit as $M\to\infty$, by showing that the series tails

$$\sum_{h_1=M+1}^{\infty}\sum_{h_2=-M}^{M}\left|\mathrm{Cov}\Big(g_u(\widetilde{\mathbf{Y}}_{\underline{h}}),g_v(\widetilde{\mathbf{Y}}_{\underline{0}})\Big)\right|+\sum_{h_1=-\infty}^{-M-1}\sum_{h_2=-M}^{M}\left|\mathrm{Cov}\Big(g_u(\widetilde{\mathbf{Y}}_{\underline{h}}),g_v(\widetilde{\mathbf{Y}}_{\underline{0}})\Big)\right| \tag{8.38}$$

as well as the remaining tails

$$\sum_{h_1=-M}^{M}\sum_{|h_2|\geq M+1}\Big|\dots\Big|,\qquad \sum_{|h_1|,|h_2|=M+1}^{\infty}\Big|\dots\Big| \tag{8.39}$$

vanish for $M\to\infty$. We will only consider the first summand in (8.38) because all other expressions, including the ones in (8.39), can be treated with analogous arguments. In accordance with the definition of the vector $\widetilde{\mathbf{Y}}_{\underline{t}}$ we define for each $\underline{h}\in\mathbb{Z}^2$ the truncated versions $\widetilde{\mathbf{Y}}_{\underline{h}}^{(l)}$ (truncated at the left-hand side) and $\widetilde{\mathbf{Y}}_{\underline{0}}^{(r)}$ (truncated at the right-hand side) via

$$\widetilde{\mathbf{Y}}_{\underline{h}}^{(l)}:=(\widetilde{X}_{\underline{h}+\underline{s}(1)}^{(l)},\dots,\widetilde{X}_{\underline{h}+\underline{s}(m_1m_2)}^{(l)})^T,\quad \widetilde{\mathbf{Y}}_{\underline{0}}^{(r)}:=(\widetilde{X}_{\underline{0}+\underline{s}(1)}^{(r)},\dots,\widetilde{X}_{\underline{0}+\underline{s}(m_1m_2)}^{(r)})^T,$$

where

$$\begin{aligned}\widetilde{X}_{\underline{h}+\underline{s}(j)}^{(l)} &:= \sum_{\underline{k}\in\Theta}b_{\underline{k}}\cdot\mathbb{1}\{k_1\leq\lfloor h_1/2\rfloor-m_1\}\,\widetilde{\varepsilon}_{\underline{h}+\underline{s}(j)-\underline{k}}+\widetilde{\varepsilon}_{\underline{h}+\underline{s}(j)},\\ \widetilde{X}_{\underline{0}+\underline{s}(j)}^{(r)} &:= \sum_{\underline{k}\in\Theta}b_{\underline{k}}\cdot\mathbb{1}\{k_1\geq-\lfloor h_1/2\rfloor\}\,\widetilde{\varepsilon}_{\underline{0}+\underline{s}(j)-\underline{k}}+\widetilde{\varepsilon}_{\underline{0}+\underline{s}(j)}.\end{aligned}$$

Note that the dependence of $\widetilde{\mathbf{Y}}_{\underline{0}}^{(r)}$ on $\underline{h}$ is suppressed in the notation. One can easily check that $\widetilde{\mathbf{Y}}_{\underline{h}}^{(l)}$ and $\widetilde{\mathbf{Y}}_{\underline{0}}^{(r)}$ are independent random variables. Hence, the first summand in (8.38) can be bounded by

$$\begin{aligned}&\sum_{h_1=M+1}^{\infty}\sum_{h_2=-M}^{M}\left|\mathrm{Cov}\Big(g_u(\widetilde{\mathbf{Y}}_{\underline{h}})-g_u(\widetilde{\mathbf{Y}}_{\underline{h}}^{(l)}),g_v(\widetilde{\mathbf{Y}}_{\underline{0}})\Big)\right|\\ +\;&\sum_{h_1=M+1}^{\infty}\sum_{h_2=-M}^{M}\left|\mathrm{Cov}\Big(g_u(\widetilde{\mathbf{Y}}_{\underline{h}}^{(l)}),g_v(\widetilde{\mathbf{Y}}_{\underline{0}})-g_v(\widetilde{\mathbf{Y}}_{\underline{0}}^{(r)})\Big)\right|.\end{aligned}$$

Both of these expressions can be treated in the same way. Therefore, we will only consider the latter summand which can be bounded by

$$C\cdot\sum_{h_1=M+1}^{\infty}\sum_{h_2=-M}^{M}\left\|g_v(\widetilde{\mathbf{Y}}_{\underline{0}})-g_v(\widetilde{\mathbf{Y}}_{\underline{0}}^{(r)})\right\|_2,$$

because $\|g_u(\widetilde{\mathbf{Y}}_{\underline{h}}^{(l)})\|_2\leq C$ follows as in (7.5) (here, $\|\cdot\|_w$ denotes the usual L^w-norm). For the remainder of this proof, C will denote a generic constant that may change from line to line. We get from Lemma 7.2

$$\left\|g_v(\widetilde{\mathbf{Y}}_{\underline{0}})-g_v(\widetilde{\mathbf{Y}}_{\underline{0}}^{(r)})\right\|_2\leq C\cdot\sum_{k_1=-\infty}^{-\lfloor h_1/2\rfloor-1}\sum_{k_2=1}^{\infty}|b_{\underline{k}}|. \tag{8.40}$$

Hence, the previously derived expression is bounded by

$$\begin{aligned}
& C \cdot \sum_{h_1=M+1}^{\infty} \sum_{h_2=-M}^{M} \sum_{k_1=-\infty}^{-\lfloor h_1/2 \rfloor -1} \sum_{k_2=1}^{\infty} |b_{\underline{k}}| \\
\leq \;& C \cdot (2M+1) \sum_{k_1=-\infty}^{-\lfloor (M+1)/2 \rfloor -1} 2\Big(- \lfloor (M+1)/2 \rfloor - k_1 \Big) \sum_{k_2=1}^{\infty} |b_{\underline{k}}| \\
\leq \;& C \cdot \sum_{k_1=-\infty}^{-\lfloor (M+1)/2 \rfloor -1} \sum_{k_2=1}^{\infty} M\, |k_1|\, |b_{\underline{k}}| \leq C \cdot \sum_{k_1=-\infty}^{-\lfloor (M+1)/2 \rfloor -1} \sum_{k_2=1}^{\infty} (1+|\underline{k}|_\infty)^2\, |b_{\underline{k}}|,
\end{aligned}$$

since it holds $M \leq 2\,|k_1| \leq 2\,|\underline{k}|_\infty$ for all $k_1 \leq -\lfloor (M+1)/2 \rfloor - 1$. Note that the right-hand side converges to zero as $M \to \infty$, because Lemma 2.1 ensures $\sum_{\underline{k}\in\Theta} (1+|\underline{k}|_\infty)^{r-1}\, |b_{\underline{k}}| < \infty$, and we assume $r = 4$ in Lemma 7.1. The remaining expressions in (8.38) and (8.39) can be treated analogously, using the summability conditions $\sum_{\underline{k}\in\Theta} |k_2|^2\, |b_{\underline{k}}| < \infty$ and $\sum_{\underline{k}\in\Theta} |k_1\, k_2|\, |b_{\underline{k}}| < \infty$ which are fulfilled since $|k_2|^2 \leq (1+|\underline{k}|_\infty)^2$ and $|k_1\, k_2| \leq (1+|\underline{k}|_\infty)^2$. This completes the proof. □

Proof of Lemma 7.2:
We will perform a Taylor expansion of order h of $g_v(\widetilde{\mathbf{Y}}_{\underline{t}})$ around $\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)}$. Let $m := m_1 m_2$. We use the common multi-index notation $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m) \in \mathbb{N}_0^m$ with $|\boldsymbol{\alpha}| = \sum_{i=1}^m \alpha_i$ and $\boldsymbol{\alpha}! = \alpha_1! \cdot \ldots \cdot \alpha_m!$. Furthermore, we abbreviate

$$D^{\boldsymbol{\alpha}} g_v(\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)}) := \frac{\partial^{|\boldsymbol{\alpha}|} g_v(\underline{x})}{\partial {x_1}^{\alpha_1} \ldots \partial {x_m}^{\alpha_m}} \Big|_{\underline{x} = \widetilde{\mathbf{Y}}_{\underline{t}}^{(W)}},$$

and get

$$\begin{aligned}
& \Big\| g_v(\widetilde{\mathbf{Y}}_{\underline{t}}) - g_v(\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)}) \Big\|_2 \\
\leq & \sum_{1 \leq |\boldsymbol{\alpha}| < h} \left\| \frac{D^{\boldsymbol{\alpha}} g_v(\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)})}{\boldsymbol{\alpha}!} (\widetilde{\mathbf{Y}}_{\underline{t}} - \widetilde{\mathbf{Y}}_{\underline{t}}^{(W)})^{\boldsymbol{\alpha}} \right\|_2 + \sum_{|\boldsymbol{\alpha}| = h} \left\| \frac{D^{\boldsymbol{\alpha}} g_v(\boldsymbol{\xi}_{\underline{t}})}{\boldsymbol{\alpha}!} (\widetilde{\mathbf{Y}}_{\underline{t}} - \widetilde{\mathbf{Y}}_{\underline{t}}^{(W)})^{\boldsymbol{\alpha}} \right\|_2, \quad (8.41)
\end{aligned}$$

where $\boldsymbol{\xi}_{\underline{t}}$ is between $\widetilde{\mathbf{Y}}_{\underline{t}}$ and $\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)}$. Note that for each $\boldsymbol{\alpha}$ we find suitable integers $1 \leq j(1), j(2), \ldots, j(|\boldsymbol{\alpha}|) \leq m$ such that

$$(\widetilde{\mathbf{Y}}_{\underline{t}} - \widetilde{\mathbf{Y}}_{\underline{t}}^{(W)})^{\boldsymbol{\alpha}} = (\widetilde{X}_{\underline{t}+\underline{s}(j(1))} - \widetilde{X}^{(W)}_{\underline{t}+\underline{s}(j(1))}) \cdot \ldots \cdot (\widetilde{X}_{\underline{t}+\underline{s}(j(|\boldsymbol{\alpha}|))} - \widetilde{X}^{(W)}_{\underline{t}+\underline{s}(j(|\boldsymbol{\alpha}|))})$$

and, thus, Hölder's inequality yields

$$\begin{aligned}
& \left\| \frac{D^{\boldsymbol{\alpha}} g_v(\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)})}{\boldsymbol{\alpha}!} (\widetilde{\mathbf{Y}}_{\underline{t}} - \widetilde{\mathbf{Y}}_{\underline{t}}^{(W)})^{\boldsymbol{\alpha}} \right\|_2 \\
\leq \;& \left\| \frac{D^{\boldsymbol{\alpha}} g_v(\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)})}{\boldsymbol{\alpha}!} \right\|_{2(h+2)/(h+2-|\boldsymbol{\alpha}|)} \cdot \prod_{k=1}^{|\boldsymbol{\alpha}|} \Big\| \widetilde{X}_{\underline{t}+\underline{s}(j(k))} - \widetilde{X}^{(W)}_{\underline{t}+\underline{s}(j(k))} \Big\|_{2(h+2)}
\end{aligned}$$

$$\leq \; C \cdot \left\| \frac{D^{\boldsymbol{\alpha}} g_v(\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)})}{\boldsymbol{\alpha}!} \right\|_{2(h+2)/(h+2-|\boldsymbol{\alpha}|)} \cdot \left(\sum_{\underline{k} \in \Theta \setminus W} |b_{\underline{k}}| + \mathbb{1}_{\{\underline{0} \notin W\}} \right)^{|\boldsymbol{\alpha}|}. \tag{8.42}$$

Here, we have used that it follows, per definition, for any index $\underline{u}$

$$\begin{aligned} \left\| \widetilde{X}_{\underline{u}} - \widetilde{X}_{\underline{u}}^{(W)} \right\|_{2(h+2)} &= \left\| \sum_{\underline{k} \in \Theta \setminus W} b_{\underline{k}} \, \widetilde{\varepsilon}_{\underline{u}-\underline{k}} + \widetilde{\varepsilon}_{\underline{u}} \, \mathbb{1}_{\{\underline{0} \notin W\}} \right\|_{2(h+2)} \\ &\leq \|\widetilde{\varepsilon}_{\underline{0}}\|_{2(h+2)} \left(\sum_{\underline{k} \in \Theta \setminus W} |b_{\underline{k}}| + \mathbb{1}_{\{\underline{0} \notin W\}} \right) \end{aligned} \tag{8.43}$$

from strict stationarity of $(\widetilde{\varepsilon}_{\underline{t}})$. Note that $\|\widetilde{\varepsilon}_{\underline{0}}\|_{2(h+2)} < \infty$ follows from Assumption 4. On the other hand, abbreviating $q := 2(h+2)/(h+2-|\boldsymbol{\alpha}|)$, the first factor in (8.42) can be bounded via another Taylor expansion of order $h - |\boldsymbol{\alpha}|$ around the zero vector as

$$\begin{aligned} &\left\| \frac{D^{\boldsymbol{\alpha}} g_v(\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)})}{\boldsymbol{\alpha}!} \right\|_q \\ \leq \; &\sum_{0 \leq |\boldsymbol{\beta}| < h-|\boldsymbol{\alpha}|} \left\| \frac{D^{\boldsymbol{\alpha}+\boldsymbol{\beta}} g_v(\underline{0})}{\boldsymbol{\alpha}!\,\boldsymbol{\beta}!} (\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)})^{\boldsymbol{\beta}} \right\|_q + \sum_{|\boldsymbol{\beta}| = h-|\boldsymbol{\alpha}|} \left\| \frac{D^{\boldsymbol{\alpha}+\boldsymbol{\beta}} g_v(\boldsymbol{\tau}_{\underline{0}})}{\boldsymbol{\alpha}!\,\boldsymbol{\beta}!} (\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)})^{\boldsymbol{\beta}} \right\|_q , \end{aligned}$$

where $\boldsymbol{\tau}_{\underline{0}}$ is between $\underline{0}$ and $\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)}$. The first summand on the right-hand side, analogous to (8.42) and (8.43), is bounded by

$$C \cdot \sum_{|\boldsymbol{\beta}| < h-|\boldsymbol{\alpha}|} \left(1 + \sum_{\underline{k} \in \Theta} |b_{\underline{k}}| \right)^{|\boldsymbol{\beta}|} < \infty,$$

since the derivative at zero is constant and $q|\boldsymbol{\beta}| \leq 2(h+2)$. Using the Lipschitz property of the h-th derivatives of g_v, the second summand is bounded by

$$\begin{aligned} &\sum_{|\boldsymbol{\beta}| = h-|\boldsymbol{\alpha}|} \left(\left\| \frac{D^{\boldsymbol{\alpha}+\boldsymbol{\beta}} g_v(\underline{0})}{\boldsymbol{\alpha}!\,\boldsymbol{\beta}!} (\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)})^{\boldsymbol{\beta}} \right\|_q + \left\| \left| \frac{D^{\boldsymbol{\alpha}+\boldsymbol{\beta}} g_v(\boldsymbol{\tau}_{\underline{0}})}{\boldsymbol{\alpha}!\,\boldsymbol{\beta}!} - \frac{D^{\boldsymbol{\alpha}+\boldsymbol{\beta}} g_v(\underline{0})}{\boldsymbol{\alpha}!\,\boldsymbol{\beta}!} \right| (\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)})^{\boldsymbol{\beta}} \right\|_q \right) \\ \leq \; &\sum_{|\boldsymbol{\beta}| = h-|\boldsymbol{\alpha}|} \left(\left\| \frac{D^{\boldsymbol{\alpha}+\boldsymbol{\beta}} g_v(\underline{0})}{\boldsymbol{\alpha}!\,\boldsymbol{\beta}!} (\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)})^{\boldsymbol{\beta}} \right\|_q + C \cdot \left\| \left(\sum_{j=1}^{m} \left| \widetilde{X}_{\underline{t}+\underline{s}(j)}^{(W)} \right| \right) (\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)})^{\boldsymbol{\beta}} \right\|_q \right), \end{aligned}$$

which is finite due to similar arguments as for the first summand. With the same calculation we can also treat the second sum in (8.41) analogous to the first sum. Together with (8.42) and (8.43), we finally get

$$\begin{aligned} \left\| g_v(\widetilde{\mathbf{Y}}_{\underline{t}}) - g_v(\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)}) \right\|_2 &\leq C \cdot \sum_{1 \leq |\boldsymbol{\alpha}| \leq h} \left(\sum_{\underline{k} \in \Theta \setminus W} |b_{\underline{k}}| + \mathbb{1}_{\{\underline{0} \notin W\}} \right)^{|\boldsymbol{\alpha}|} \\ &\leq C \cdot \left(\sum_{\underline{k} \in \Theta \setminus W} |b_{\underline{k}}| + \mathbb{1}_{\{\underline{0} \notin W\}} \right), \end{aligned} \tag{8.44}$$

with a generic constant $C < \infty$, which depends merely on $\|\widetilde{\varepsilon}_{\underline{0}}\|_{2(h+2)}$ and on $\sum_{\underline{k}\in\Theta} |b_{\underline{k}}|$. Therefore, one can follow along these lines for the second assertion in Lemma 7.2 concerning the bootstrap versions $\mathbf{Y}^*_{\underline{t}}$ and $\mathbf{Y}^{*(W)}_{\underline{t}}$. Since (7.3), Assumption 4 and (7.2) ensure $\|\varepsilon^*_{\underline{t}}\|_{*2(h+2)} = \mathcal{O}_P(1)$ and $\sum_{\underline{k}\in\Theta} \left|\widehat{b}_{\underline{k}}(p)\right| = \mathcal{O}_P(1)$, it follows with the same calculation as for (8.44)

$$\left\|g_v(\mathbf{Y}^*_{\underline{t}}) - g_v(\mathbf{Y}^{*(W)}_{\underline{t}})\right\|_{*2} \;\leq\; \mathcal{O}_P(1)\cdot\left(\sum_{\underline{k}\in\Theta\setminus W}\left|\widehat{b}_{\underline{k}}(p)\right| + \mathbb{1}_{\{\underline{0}\notin W\}}\right),$$

which completes the proof. □

Proof of Lemma 5.2:
Suppose $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ is a linear spatial process as given by (5.3) with some suitable absolutely summable coefficients $(\alpha_{\underline{\nu}})_{\underline{\nu}\in\mathbb{Z}^2}$ and an i.i.d. white noise process $(u_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ with $E(u^2_{\underline{t}}) = \sigma^2 \in (0,\infty)$ and $E(u^4_{\underline{t}}) = \eta\sigma^4 \in (0,\infty)$. For the comparative quantity $\check{\gamma}(\underline{h})$ defined in Lemma 5.2 it holds

$$\check{\gamma}(\underline{h}) := \frac{1}{n^2}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n} X_{\underline{t}+\underline{h}}\, X_{\underline{t}},$$

which is asymptotically equivalent to $\widehat{\gamma}(\underline{h})$. Then, standard calculations as in the time series case yield for all $\underline{h},\underline{k}\in\mathbb{Z}^2$

$$\begin{aligned}
& n^2\,\mathrm{Cov}(\check{\gamma}(\underline{h}),\check{\gamma}(\underline{k})) \\
= \;& \sum_{\underline{r}\in\mathbb{Z}^2:|r_1|<n,|r_2|<n}\frac{(n-|r_1|)(n-|r_2|)}{n^2}\Big(\big(\gamma(\underline{r}-\underline{k}+\underline{h})\gamma(\underline{r})+\gamma(\underline{r}+\underline{h})\gamma(\underline{r}-\underline{k})\big) \\
& \qquad + (\eta-3)\sum_{\underline{\nu}\in\mathbb{Z}^2}\alpha_{\underline{\nu}}\alpha_{\underline{\nu}-\underline{h}}\alpha_{\underline{\nu}-\underline{r}+\underline{k}-\underline{h}}\alpha_{\underline{\nu}-\underline{r}-\underline{h}}\,\sigma^4\Big) \\
= \;& \sum_{\underline{r}\in\mathbb{Z}^2}\big(\gamma(\underline{r}-\underline{k}+\underline{h})\gamma(\underline{r})+\gamma(\underline{r}+\underline{h})\gamma(\underline{r}-\underline{k})\big) + (\eta-3)\,\gamma(\underline{h})\gamma(\underline{k}) + o(1) \\
=: \;& V(\underline{h},\underline{k}) + o(1),
\end{aligned}$$

which leads for the sample autocovariances to

$$n^2\mathrm{Var}\left(\begin{pmatrix}\check{\gamma}(\underline{0})\\ \check{\gamma}(\underline{h})\\ \check{\gamma}(\underline{k})\end{pmatrix}\right) = \begin{pmatrix} V(\underline{0},\underline{0}) & V(\underline{0},\underline{h}) & V(\underline{0},\underline{k})\\ V(\underline{h},\underline{0}) & V(\underline{h},\underline{h}) & V(\underline{h},\underline{k})\\ V(\underline{k},\underline{0}) & V(\underline{k},\underline{h}) & V(\underline{k},\underline{k})\end{pmatrix} + o(1) =: V + o(1).$$

For the quantities $\check{\rho}(\underline{h}) = \check{\gamma}(\underline{h})/\check{\gamma}(\underline{0})$, we get

$$\begin{pmatrix}\check{\rho}(\underline{h})\\ \check{\rho}(\underline{k})\end{pmatrix} = \begin{pmatrix} f_1(\check{\gamma}(\underline{0}),\check{\gamma}(\underline{h}),\check{\gamma}(\underline{k}))\\ f_2(\check{\gamma}(\underline{0}),\check{\gamma}(\underline{h}),\check{\gamma}(\underline{k}))\end{pmatrix},$$

where $f(x_1,x_2,x_3) = (f_1(x_1,x_2,x_3), f_2(x_1,x_2,x_3))^T = (x_2/x_1, x_3/x_1)^T$. An application of the Δ-method leads to

$$n^2 \mathrm{Var}\left(\begin{pmatrix} \check{\rho}(\underline{h}) \\ \check{\rho}(\underline{k}) \end{pmatrix}\right) = J_f V J_f^T + o(1),$$

where

$$J_f = \begin{pmatrix} -\gamma(\underline{h})\gamma(\underline{0})^{-2} & \gamma(\underline{0})^{-1} & 0 \\ -\gamma(\underline{k})\gamma(\underline{0})^{-2} & 0 & \gamma(\underline{0})^{-1} \end{pmatrix},$$

such that

$$\begin{aligned}
& n^2 \operatorname{Cov}(\check{\rho}(\underline{h}), \check{\rho}(\underline{k})) \\
= \; & [J_f V J_f^T]_{2,1} + o(1) \\
= \; & [J_f]_{2,\bullet} V [J_f]_{\bullet,1}^T + o(1) \\
= \; & \left(-\gamma(\underline{k})\gamma(\underline{0})^{-2} V(\underline{0},\underline{0}) + \gamma(\underline{0})^{-1} V(\underline{k},\underline{0})\right)\left(-\gamma(\underline{h})\gamma(\underline{0})^{-2}\right) \\
& + \left(-\gamma(\underline{k})\gamma(\underline{0})^{-2} V(\underline{0},\underline{h}) + \gamma(\underline{0})^{-1} V(\underline{k},\underline{h})\right)\gamma(\underline{0})^{-1} + o(1) \\
= \; & \rho(\underline{k})\rho(\underline{h}) \sum_{\underline{r}\in\mathbb{Z}^2} (\rho(\underline{r})\rho(\underline{r}) + \rho(\underline{r})\rho(\underline{r})) - \rho(\underline{h}) \sum_{\underline{r}\in\mathbb{Z}^2} (\rho(\underline{r}+\underline{k})\rho(\underline{r}) + \rho(\underline{r}+\underline{k})\rho(\underline{r})) \\
& -\rho(\underline{k}) \sum_{\underline{r}\in\mathbb{Z}^2} (\rho(\underline{r}-\underline{h})\rho(\underline{r}) + \rho(\underline{r}-\underline{h})\rho(\underline{r})) \\
& + \sum_{\underline{r}\in\mathbb{Z}^2} (\rho(\underline{r}-\underline{h}+\underline{k})\rho(\underline{r}) + \rho(\underline{r}+\underline{k})\rho(\underline{r}-\underline{h})) + o(1) \\
= \; & \sum_{\underline{r}\in\mathbb{Z}^2} \Big\{ 2\rho(\underline{r})^2\rho(\underline{k})\rho(\underline{h}) - 2\rho(\underline{r}+\underline{k})\rho(\underline{r})\rho(\underline{h}) - 2\rho(\underline{r}-\underline{h})\rho(\underline{r})\rho(\underline{k}) \\
& \qquad + \rho(\underline{r}-\underline{h}+\underline{k})\rho(\underline{r}) + \rho(\underline{r}+\underline{k})\rho(\underline{r}-\underline{h}) \Big\} + o(1).
\end{aligned}$$

In particular, the latter quantity depends exclusively on the second order structure of the linear process $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$. □

Acknowledgments

We gratefully acknowledge financial support by the German Research Foundation (DFG). The research of Marco Meyer was partly supported by the Research Training Group RTG 1953 and the position of Carsten Jentsch was financed by the Research Center 884 "Political Economy of Reform" (Project B6).

References

Alonso, A. M., Pena, D., and Romo, J. (2002): Forecasting time series with sieve bootstrap. *Journal of Statistical Planning and Inference* **100**, 1–11.

Basu, S. and Reinsel, G. C. (1992): A note on properties of spatial Yule-Walker -estimators. *J. Statist. Comput. Simul.* **41**, 243–255.

Basu, S. and Reinsel, G. C. (1993): Properties of the Spatial Unilateral First-Order ARMA Model. *Advances in Applied Probability* **25**, 631–648.

Baxter, G. (1962): An Asymptotic Result for the Finite Predictor *Math. Scand.* **10**, 137–144.

Baxter, G. (1963): A Norm Inequality for a Finite-Section Wiener-Hopf Equation. *Illinois J. Math.* **7**, 97–103.

Besag, J. E. (1972): On the correlation structure of some two dimensional stationary processes. *Biometrika* **59**, 43–48.

Bhatt, S.J. and Dedania, H.V. (2003): Beurling algebra analogues of the classical theorems of Wiener and Lévy on absolutely convergent Fourier series. *Proc. Indian Acad. Sci. (Math. Sci.)* **113**, 179–182.

Bickel, P.J. and Freedman, D.A. (1981): Some Asymptotic Theory for the Bootstrap. *The Annals of Statistics* **9**, 1196–1217.

Brockwell, P.J. and Davis, R.A. (1991): *Time Series: Theory and Methods (2nd Ed.)*. Springer, New York.

Bühlmann, P. (1997): Sieve Bootstrap for Time Series. *Bernoulli* **3**, 123–148.

Chang, Y. and Park, J.Y. (2003): A sieve bootstrap for the test of a unit root. *Journal of Time Series Analysis*, **24**, 379–400.

Cheng, R. and Pourahmadi, M. (1993): Baxter's inequality and convergence of finite predictors of multivariate stochastic processes. *Probability Theory and Related Fields* **95**, 115–124.

Choi, B.S. and Politis, D.N. (2007): Modeling 2-D AR Processes with Various Regions of Support. *IEEE Trans. Signal Proc.* **55**, no. 5, pp. 1696–1707.

Cressie, N. (1993): *Statistics for Spatial Data.* Wiley, New York.

Freedman, D. (1984): On bootstrapping two-stage least squares estimates in stationary linear models. . *The Annals of Statistics* **12**, 827–842.

Gröchenig, K. (2007): Weight functions in time-frequency analysis. In: *Pseudodifferential Operators: Partial Differential Equations and Time-Frequency Analysis.* Rodino, L. et al. (Eds.). *Fields Institute Comm.* **52**, 343–366.

Guyon, X. (1982): Parameter estimation for a stationary process on a d-dimensional lattice. *Biometrika* **69**, 95–105.

Guyon, X. (1995): *Random Fields on a Network – Modelling, Statistics and Applications.* Springer, New York.

Ha, E. and Newton, H. J. (1993): The bias of estimators of causal spatial autoregressive processes. *Biometrika* **80**, 242–245.

Hall, P. (1985): Resampling a coverage pattern. *Stoch Process Appl* **20**, 231–246.

Hannan, E.J. and Deistler, M. (1988): *The Statistical Theory of Linear Systems.* Wiley Series in Prob. and Math. Stat.

Helson, H. and Lowdenslager, D. (1958): Prediction Theory and Fourier Series in Several Variables. *Acta Math.* **99**, 165–202.

Kreiss, J.-P. (1988): Asymptotical Inference for a Class of Stochastic Processes. Habilitationsschrift, Universität Hamburg.

Kreiss, J.-P. (1992): Bootstrap Procedures for AR(∞) Processes. In: Bootstrapping and Related Techniques. Jöckel, K.H., Rothe, G. and Sender, W. (Eds.). Lecture Notes in Economics and Mathematical Systems 376 Springer, Heidelberg, 107–113.

Kreiss, J. P. (1997): Asymptotic properties of residual bootstrap for autoregressions. Mimeo, Institute for Mathematical Stochastics, Technical University of Braunschweig, Germany.

Kreiss, J.-P., Paparoditis, E. and Politis, D. (2011): On the Range of Validity of the Autoregressive Sieve Bootstrap. *The Annals of Statistics* **39**, 2103–2130.

Künsch, H.R. (1989): The Jackknife and the Bootstrap for General Stationary Observations. *The Annals of Statistics* **17**, 1217–1241.

Lütkepohl, H. (1996): *Handbook of Matrices*. Wiley, Chichester.

Meyer, M. and Kreiss, J.-P. (2014+): On the vector Autoregressive Sieve Bootstrap. *Journal of Time Series Analysis*, to appear.

Nordman, D.J. and Lahiri, S.N. (2004): On optimal spatial subsample size for variance estimation. *Ann. Statist.*, **32**, 1981–2027.

Nordman, D. J. and Lahiri, S. N. (2007): Optimal Block Size for Variance Estimation by a Spatial Block Bootstrap Method. *Sankhya : The Indian Journal of Statistics Special Issue on Statistics in Biology and Health Sciences* **69**, 468–493.

Paparoditis, E. (1996): Bootstrapping Autoregressive and Moving Average Parameter Estimates of Infinite Order Vector Autoregressive Processes. *Journal of Multivariate Analysis* **57**, 277–296.

Paparoditis, E. and Politis, D. N. (2005): Bootstrapping Unit Root Tests for Autoregressive Time Series. *J. Amer. Statist. Assoc.* **100**, 545–553.

Paparoditis, E. and Streitberg, B. (1991): Order Identification Statistics in Stationary Autoregressive Moving Average Models: Vector Autocorrelations and the Bootstrap. *Journal of Time Series Analysis* **13**, 415–435.

Politis, D. N. and Romano, J. P. (1993): Nonparametric resampling for homogeneous strong mixing random fields. *J. Multivariate Anal.* **47**, 301–328.

Pourahmadi, M. (2001): *Foundations of Time Series Analysis and Prediction Theory*. Wiley, 2001.

Smeekes, S. and J.-P. Urbain (2013): On the applicability of the sieve bootstrap in time series panels. *Oxford Bulletin of Economics and Statistics* **76** no. (1), pp. 139–151.

Tjøstheim, D. (1978): Statistical Spatial Series Modelling. *Advances in Applied Probability* **10**, 130–154.

Tjøstheim, D. (1981): Autoregressive Modeling and Spectral Analysis of Array Data in the Plane. *IEEE Transactions on Geoscience and Remote Sensing* **19**, 15–24.

Tjøstheim, D. (1983): Statistical Spatial Series Modelling II: Some fur- ther Results in Unilateral Processes. *Advances in Applied Probability* **15**, 562–684.

Whittle, P. (1954): On Stationary Processes in the Plane. *Biometrika* **41**, 434–449.

Yokoyama, R. (1980): Moment Bounds for Stationary Mixing Sequences. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **52**, 45–57.

Zhu, J. and Lahiri, S.N. (2007): Bootstrapping the Empirical Distribution Function of a Spatial Process. *Statistical Inference for Stochastic Processes*, **10**, 107–145.

Zygmund, A. (2002): *Trigonometric Series – Third Edition*. Cambridge University Press, Cambridge.

Marco Meyer, Technische Universität Braunschweig, Inst. f. Math. Stochastik, Pockelsstrasse 14, D–38106 Braunschweig, Germany. Email: marco.meyer@tu-bs.de

Carsten Jentsch, University of Mannheim, Department of Economics, L7, 3–5, D–68131 Mannheim, Germany. Email: cjentsch@mail.uni-mannheim.de

Jens-Peter Kreiss, Technische Universität Braunschweig, Inst. f. Math. Stochastik, Pockelsstrasse 14, D–38106 Braunschweig, Germany. Email: j.kreiss@tu-bs.de