STATISTICAL INFERENCE FOR SPATIAL AND SPATIO-TEMPORAL PROCESSES

A thesis submitted to
the London School of Economics and Political Science
and the
University of London
in the subject of
Statistics
for the degree of
Doctor of Philosophy

Chrysoula Dimitriou-Fakalou

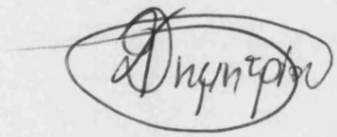
June 2006

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Signed declaration

I would like to declare that the work presented in this thesis is my own.

Chrysoula Dimitriou-Fakalou



Abstract

First, the time series analysis was widely introduced and used in the statistical world. Next, the analysis of spatio-temporal processes has followed, which is taking into account not only when, but also where the phenomenon under observation is taking place.

We mainly focus on stationary processes that are assumed to be taking place regularly over both time and space. We examine ways of estimating the parameters involved, without the risk of coming up with a very large bias for our estimators; the bias is the typical problem of estimation for the parameters of stationary processes on $\mathcal{Z}^d$, for any $d \geq 2$. We particularly study the cases of spatio-temporal ARMA processes and spatial auto-normal formulations on $\mathcal{Z}^d$. For both cases and any positive integer d, we propose estimators that are consistent, asymptotically unbiased and normal, if certain conditions are satisfied.

We do not only study the spatio-temporal processes that are observed regularly over space, but also those, for which we have recordings on a fixed number of locations anywhere. We might follow the route of a multivariate time series methodology then. Thus, the asymptotic behavior of the estimators proposed might be analyzed as the number of recordings over time only tends to infinity.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Spatio-temporal statistics is related to taking observations of a phenomenon at different times and different locations. It generalizes the notion of time series, by taking into account the space where the phenomenon takes place too. This implies that, in addition to the time axis, at least two more dimensions are added in the analysis, depending on whether the process takes place on the two or three-dimensional space. Thus, spatio-temporal processes are an application of the processes that take place on a d-dimensional space or processes with d-dimensional indices, where d is any positive integer. Nowadays, the statistical analysis of spatio-temporal processes has become very popular. It can be useful, for example, in geographical information systems, in meteorology, in seismology, in physics or for environmental applications over space and time.

It is very common for spatial statistics to record observations regularly over space, as it is for time series. As a result, it is meaningful to study in advance the theoretical spatial or spatio-temporal processes that take place on $\mathcal{Z}^d$, *i.e.* the regular d-dimensional lattice. Chapters 2, 3 and 4 altogether study at length some processes that take place on $\mathcal{Z}^d$. Section 2.2 generalizes the concept of unilateral ordering of any two locations on $\mathcal{Z}^d$, as this was introduced first by Whittle (1954) for $d = 2$ and continued by Guyon (1982) for $d \geq 2$. The Wold decomposition and the definitions of weak and strict stationarity in Section 2.3 are extended to include processes defined on more than one dimensions. In 2.3.2, a new definition of the so-called 'reverse strictly stationary process' is given; reverse strict stationarity allows for two $(q \times 1)$ random vectors from the process of interest to have the same distribution, if the relative placement of the q locations on $\mathcal{Z}^d$ within each vector is the same but the direction of the vectors is opposite.

Reverse strictly stationary processes are an example of a notion that has been introduced particularly for spatial statistics. Like on the time axis there is the 'past' and 'future', each dimension of space also occupies two different ends. Nevertheless, there can be no causal relationship to relate those two ends. For example, the 'past' and 'future' of the time axis are such that there is a natural order between the two, as anything that occurs in the 'past' could have an effect on the happenings of the 'future'. The one-dimensional spatial analogue of the time axis is the line transect, as this was described by Whittle (1954). For any two locations on the line transect, although those can be set in a unilateral order and they might be close enough to interact, there is usually no reason to assume that a causal relationship is taking place there.

In Chapters 3 and 4, we study the second-order properties of some (weakly) stationary processes that take place on the regular d-dimensional space. These processes might be spatial or spatio-temporal; this depends on whether all the dimensions involved are spatial, or whether the time axis is there as well. Studying their second-order properties is totally unconnected to the interpretation given to the d dimensions. Nevertheless, we will often refer to the inclusion or not of the time axis as a dimension, in order to study spatio-temporal and spatial processes on $\mathcal{Z}^d$ separately. For example, after we have defined the causal and invertible ARMA model on $\mathcal{Z}^d$ in Section 2.4, we have proposed different ways for the estimation of its parameters in Chapter 3. Whatever the number of dimensions d, an ARMA process on $\mathcal{Z}^d$ is a standard way to model data derived from a stationary process. Further as we are going to see in Section 3.6, causal and invertible ARMA models, compared to all other ARMA models, provide more simplicity for the methods used. The assumption of causality and invertibility might be directly related to the presence of the time axis. Thus, spatio-temporal ARMA models can often be better justified and understood than spatial ARMA models, since when a directional preference must be assumed, it can be attributed to the unidirectional flow of the time axis only.

On the other hand, Besag (1974) refused to compromise that an ARMA process is the best way to model observations from spatial and stationary processes on $\mathcal{Z}^2$. Especially under the assumption of causality or invertibility, such a model would risk to be unnatural and unable to provide an instinctive tool for prediction. His invention of the stationary auto-normal schemes, was made especially for the needs of spatial statistics. Following this example, in Chapter 4 we deal with stationary processes on $\mathcal{Z}^d$, which have a specific form of second-order dependence. The second-order properties of the processes of interest

can be discovered in their spectral density, which has a finite, symmetric and linear filter in the denominator. Under the assumption of normality, Besag (1974) expressed the second-order properties of the processes via a finite and linear conditional expectation of the value of the process on any location based on the values on all other locations of the lattice. When we manage to mask the second-order properties of the process into a conditional expectation without assuming that the process is Gaussian, then the process forms an auto-linear scheme. We will refer to such schemes as spatial auto-linear schemes, as including the time axis then would not be wise. This is because this formulation does not distinguish between the information from the 'past' and the information from the 'future', which should happen since, naturally, the information from the 'past' always comes first.

At this point, it should be made clear that one of the purposes of this thesis is not to highlight the gap between the different methods of estimation used for stationary spatial and spatio-temporal processes but to bridge that gap instead. When it comes to estimation, we try to establish in both Chapters 3 and 4, that any choice of parametrization for the second-order properties of the process on $\mathcal{Z}^d$ might be equally fruitful for the estimators of the parameters. In other words, the estimators we are proposing in the two chapters possess similar statistical properties. Thus, the reasons that make us consider Chapters 3 and 4 to be more related to spatio-temporal and spatial processes, respectively, is prediction and not estimation. For example, causal spatio-temporal auto-regressions, such as these analyzed in Section 3.7, could be very useful for prediction, since the assumed model is only using locations from past timings. On the other hand, models that use all the information around a location of interest, like the auto-linear schemes of Chapter 4, are more suitable for kriging (Cressie, 1993), which is the form of 'spatial prediction'. Of course, it can be that we have the time axis in our analysis, that we are missing an observation from the centre of our dataset and that we need to approximate its value. In this case known as smoothing, the parametrization adopted by an auto-linear formulation might be useful for a time series or a spatio-temporal process too.

Using a set of observations to estimate the parameters of a causal and invertible ARMA process on $\mathcal{Z}^d$, is not an easy task when $d \geq 2$; we explain why next. For one-dimensional ARMA models, the exact Gaussian likelihood estimators have all the desired statistical properties, as we can verify from Brockwell and Davis (1991). Unfortunately,

as the number of dimensions increases, we cannot conclude yet that maximizing the exact Gaussian likelihood of observations can produce both asymptotically unbiased and normal estimators. This problem, which is reflected in the order of the bias of the estimators, is known as the edge-effect and it has been very well described by Guyon (1982). The source of the edge-effect is the different setting of asymptotics that is taking place when $d \geq 2$. Indeed, although a set of observations on a finite set of $\mathcal{Z}^d$ is available, usually a hyper-rectangle or hyper-cube, we should allow that this set could grow towards all sides. All the second-order stationary processes studied in Chapters 3 and 4, use this setting to assess the quality of the estimators for the parameters of interest. Thus in both chapters, the estimators must be defined in such ways, which guarantee that their asymptotic normality can be established.

Defeating the edge-effect is one of the main challenges of this thesis. We have tried to tackle a very complex problem, for which the number of solutions proposed in the past has been limited. In Chapter 3, we have resorted to modifications of Gaussian likelihoods that may produce asymptotically unbiased and normal estimators of the parameters. This is the same tactic as the one followed by Guyon (1982) and Yao and Brockwell (2006), who referred to the estimation of the parameters of any stationary process on $\mathcal{Z}^d$ and the $(p+q)$ parameters of two-dimensional causal and invertible ARMA processes, respectively. We have studied the cases of auto-regressions, moving-averages and ARMA processes on $\mathcal{Z}^d$, separately. Section 3.3 deals with causal auto-regressions and proposes a conditional Gaussian likelihood for maximization. By contrast, Section 3.4 is dedicated to invertible moving-averages only. There are two new suggestions for estimation of the parameters and the second one is based on a modification of a Gaussian conditional likelihood. The way we have dealt with the moving-average there, is only a special case of the more general solution proposed next for the ARMA. Thus, Section 3.5 generalizes the results of 3.4.3 for the parameters of a causal and invertible ARMA(p,q) process. With a finite fourth moment of the error sequence of interest, the $(p+q)$ modified Gaussian likelihood estimators defined then are consistent, asymptotically unbiased and normal and they are efficient if the process under observation is Gaussian.

As a bilateral ARMA model might seem more meaningful than a causal and invertible ARMA model, when it refers to a spatial process, we would have liked to be able to extend our results for the case of any ARMA model. Furthermore, there have not been any real solutions for bilateral ARMA models so far. According to Section 3.6, the

modified Gaussian likelihood proposed for maximization in 3.5.3 is only a special case of the quantity that should have been maximized, in order to derive the estimators of the parameters of a bilateral ARMA(p, q) process. The path we have followed there is due to Whittle (1954), who, for two-dimensional processes, achieved a transition from the Gaussian likelihood of the observations from a finite bilateral auto-regression to the same likelihood expressed in terms of the parameters of the AR(∞) representation of the process. For bilateral ARMA models on $\mathcal{Z}^d$, we generalize his suggestion with a correction on the Gaussian likelihood, which affects its deterministic part only. This correction fixes the bias that the estimators of the auto-regressive and moving-average parameters would have, unless the process was causal and invertible, respectively.

Under no circumstances should that bias be considered to have any relation to the edge-effect. While the bilaterality of an ARMA process might add to the bias of the estimators even when $d = 1$, the edge-effect is very well disguised then, and makes its unpleasant appearance when $d \geq 2$, by causing the bias to move towards zero at equal $(d = 2)$ or slower $(d > 2)$ speed, compared to the speed of the standard error of the estimators. It might fairly be considered as the most difficult problem to tackle regarding the estimation of the parameters of a stationary process on $\mathcal{Z}^d$. This is the problem for which Guyon (1982) and Yao and Brockwell (2006) proposed solutions. Guyon used the form of Gaussian likelihood, which, according to Whittle (1954), involves the periodogram or sample auto-covariances in its random part. He corrected the edge-effect by using the unbiased estimators of theoretical auto-covariances there. On the other hand, Yao and Brockwell (2006) focused on two-dimensional ARMA models. Before modifying the genuine Gaussian likelihood, they used the innovations algorithm and a conventional unilateral ordering of locations in the sample; next they factorized the determinant involved into a product of prediction variances in the deterministic part, and they partitioned the random part into a sum of squares of prediction errors. Then, they put forward a selection of locations out of the ones available in the sample, and they used this information only in the product and sum of the deterministic and random part, respectively, of the proposed modified Gaussian likelihood.

In Section 3.5.3, we have suggested a new modification for a Gaussian likelihood, which is made especially for the ARMA on $\mathcal{Z}^d$. In other words, we have not restricted our number of dimensions d to be small, like Yao and Brockwell (2006). We have tried to justify that the quantity we have chosen to maximize is a modified Gaussian likelihood

using classical time domain arguments, rather than follow the route of Guyon (1982). The special characteristics of the ARMA have been highlighted and taken into account. Yao and Brockwell (2006) resorted to the AR(∞) representation of the ARMA process of interest; as a result, they introduced an infinite order to their problem and missed the opportunity to generalize their results to higher dimensionalities. Similarly, Guyon's (1982) suggestion would demand the computation of as many sample auto-covariances as possible, unless the ARMA was a finite auto-regression or a finite moving-average. We have tried to demonstrate that the ARMA deserves a solution, which takes into account its finite order. The finite order reflects both the finite auto-regressive and moving-average polynomials. Indeed, an Auto-Regressive Moving-Average can become a moving-average, if a finite linear transformation is applied to it. But what are these special advantages of these two characteristics, *i.e.* that the transformation used is finite and that the transformed process is a moving-average?

On the one hand, the finite transformation implies that, for any set of random variables from the ARMA of large enough cardinality, we may create a set of smaller cardinality of random variables from the moving-average and 'nothing is missing', *i.e.* information on more locations from the ARMA process could only contribute by offering more locations available from the moving-average, but not by augmenting the information on the sites already available, as we have everything we needed to know there. As the original set grows, so does its subset at equal speed. That is our first victory over the edge-effect, which clearly reflects the auto-regressive nature of the ARMA. Indeed, finite transformations work for the auto-regression as they might produce a sequence of uncorrelated random variables or they might produce a moving-average. Section 3.3 deals with problems of estimation for auto-regressions via transformations to white noise sequences, while estimating the parameters of an auto-regression using the moving-average path is a special case of Section 3.5. Special reference to the auto-regression transformed to a moving-average will also be made in Section 4.5.2.

On the other hand, as we are going to see in Section 3.4, the moving-average has another nice property. For a set of random variables from a moving-average of large enough cardinality, we may create a set of smaller cardinality, with random variables, which have in the larger set only and not any further, all their neighbours, in the sense that they share with them non-zero auto-covariances. Again, 'nothing is missing' in terms of information available, expressed now via the auto-covariance between two random

variables on two different sites, as more sites available cannot give any random variables that have non-zero auto-covariance with any member of the selected smaller set. Again, the cardinalities of the two sets move at the same speed and this signifies the second and final victory over the edge-effect, thanks to the moving-average nature of the ARMA.

To use correctly these two properties, we have proceeded with modifications on Gaussian likelihoods, rather than use them in their genuine form. As a result, the exponential functions of the modified likelihoods do not necessarily involve negative powers, and we cannot be sure that they can reach a minimum zero. This is a similar problem to the one that Guyon's (1982) proposed estimators had, as they were based on sample auto-covariances that did not necessarily have a positive-definite sample variance-covariance matrix or positive spectral estimates, as those last ones were to be computed for the likelihood version of Whittle (1954). Dahlhaus and Künsch (1987) dealt successfully with this problem by introducing 'data tapers', but paid the price of losing the efficiency of estimators for $d \geq 4$. Such corrections on our proposed estimators are beyond the interests of this thesis. It is remarkable that this problem does not concern the estimators of Yao and Brockwell (2006), as they make sure that a positive quantity is always to be minimized, involving a sum of squares of prediction errors.

Since most of our attempts to estimate the parameters of ARMA models are counted on Gaussian likelihoods and modifications made on them, we return to this subject again in Chapter 6 and examine it from a different scope. We focus there on two-dimensional ARMA processes only, although our results might be generalized when $d > 2$. First, for a special class of causal auto-regressions, which are linear-by-linear (Martin, 1979), we are able to write down explicitly the exact Gaussian likelihood of observations on a rectangle. In Chapter 3, we have only dealt with modifications on Gaussian likelihoods, but now the exact Gaussian likelihood version can be written down, if such an auto-regression provides a sensible representation of the second-order properties that are being studied. Then, for observations from an invertible moving-average, which uses two parameters only, since we cannot write the exact Gaussian likelihood then, we perform simulations to watch the performance of the exact Gaussian likelihood estimators and compare it to that of the modified estimators proposed by Yao and Brockwell (2006). We are trying to conclude if its worth to proceed with modifications when the dimensionality of the problem is still low.

Regarding the spatial auto-linear schemes of Chapter 4, we propose in 4.5.4 a new

method of estimation for the unknown coefficients involved. It is a method based on the moments of a new series, which may be produced from the original series, if a finite and linear filter is applied. This property, *i.e.* that with a finite transformation we may produce a series with an auto-covariance function which cuts off to zero outside a finite set of vector lags, sounds like the property of an auto-regression that can be transformed into a moving-average. Indeed, in Section 4.4, we show that, especially in terms of second-order properties rather than conditional expectations, it is always possible for an auto-regression to have an auto-linear representation. This same property of the auto-regression was used in Chapter 3 as a main tool against the edge-effect. Using that same tool, we have studied the spatial auto-linear schemes of any dimensionality d, as we can always produce the new series with a finite transformation.

The edge-effect has not been the real problem for the estimation of the parameters of a spatial auto-normal form, as this was introduced by Besag (1974) on $\mathcal{Z}^2$. On the contrary, the source of the confusion should be searched in the revolutionary work of Besag (1974) itself, which used a conditional moments representation of the process, rather than the auto-covariance function, in order to describe its characteristics, and that was difficult to interpret, though easy to comprehend. We have tried in Section 4.2 to demonstrate that, like the auto-covariance function can be found in the numerator of the spectral density, the denominator of the spectral density also generates the coefficients of the best linear predictors, which are the conditional expectations of the auto-normal formulations. In Section 4.5.2, we show that using a conditional likelihood, as this was proposed by Besag (1974), cannot guarantee any solution, unless we express our process as a unilateral or bilateral auto-regression first. In Section 4.5.3, we do prove the properties of the pseudo-likelihood estimators, also proposed by Besag (1974), which are consistent and asymptotically normal, if certain conditions are satisfied. For our method of moments estimators described in Section 4.5.4, we create the new series with a finite number of non-zero auto-correlations, for which each unknown coefficient of the auto-linear formulation of the original series, is equal to an auto-correlation between two locations. Since, we end up estimating our unknown coefficients as auto-correlations of a process, further to the consistency and asymptotic normality of the estimators, we also manage to come up with a variance matrix of the estimators that resurrects Bartlett's formula, as this was given by Brockwell and Davis (1991, p.221). In the past, we had not seen such a complete result for the estimation of the parameters of an auto-normal or

auto-linear scheme. A complete result would involve both defining new estimators and discovering their statistical properties.

It should be made clear now that both Chapters 3 and 4 try to model the second-order properties of (weakly) stationary processes on $\mathcal{Z}^d$; either this is for a unilateral spatio-temporal process or a spatial process on $\mathcal{Z}^d$, the same idea has been used repeatedly. In the end of Section 2.4.1, the subsection referring to the general Yule-Walker equations has given the answer to almost all our questions, regarding the estimation of parameters on $\mathcal{Z}^d$. The general Yule-Walker equations relate the second-order properties, *i.e.* the auto-covariance functions, of two processes. Moreover from a stationary process, it is always possible to apply a linear, 'time'-, or otherwise, invariant filter, in order to come up with a new stationary process, that is such that the two processes share together the general Yule-Walker representations. The filter one has to apply is none other than the one with coefficients equal to the auto-covariances of the second process that is about to be produced. As a result, if one of the two processes has the advantage of a finite number of non-zero auto-covariances, then all one has to do is apply a finite linear filter on the other process to use this advantage. Either we are dealing with an auto-regression or an ARMA or even a stationary process that has an auto-linear representation, a finite transformation will automatically make it a moving-average, or, in general, a process with similar second-order advantages. The question why these ideas were not that necessary and useful for processes that take place on $\mathcal{Z}$, can only lead us to one answer. It is the edge-effect that has made us look for finite filters to apply on data and finite auto-covariance functions to assume for the processes of interest. It is the edge-effect that has made us resort to the general Yule-Walker equations, instead of the standard techniques used for time series.

Finally, in Chapter 5 we have changed the general setting followed so far, for the analysis of stationary processes on $\mathcal{Z}^d$, and we have switched to spatio-temporal processes on $\mathcal{R}^d$ and $\mathcal{Z}$, respectively. It is a very common problem that the locations where the phenomenon is taking place might be anywhere. In those cases the inclusion of the time axis in the analysis might have a worthless contribution. More specifically, we follow the statistical analysis of observations recorded on any N locations of $\mathcal{R}^d$ and regularly over time. This is because, unless we record observations regularly over space, we cannot use any of theoretical background that has been studied in Chapters 3 and 4. We consider an unknown covariance structure between the variables on the N sites that does not change

over time. Next, using a multivariate time series setting and allowing for the number of regular recordings over time to tend to infinity, we fit multivariate auto-regressions and use a conditional Gaussian likelihood, in order to estimate the unknown spatial and time parameters and to assess the quality of our estimators.

In conclusion, either we study the spatial or spatio-temporal processes, either we have a set of regular or irregular recordings available, either we decide to approach the problem using causal formulations or not, in this thesis we have tried to obtain a profound understanding of the existing problems, and then we have tried to discover or even invent new ways for the statistical analysis of processes, which take place on d dimensions. All Chapters 3, 4 and 5 deal with estimation, hypotheses testing and, finally, with real data analysis. We hope that, altogether, this thesis could be regarded as a contribution for the statistical inference on spatial and spatio-temporal processes.

Chapter 2

Elementary results for processes on a d-dimensional lattice

2.1 Introduction

Before we move to the next two chapters that deal with some problems of statistical inference for processes on the regular d-dimensional lattice and before we propose various ways to solve them, we will need to summarize some basic definitions and results that have been given before, as well as to add some new results that will be extremely useful next. In Section 2.2, we recall the notion of unilateral ordering between any two locations $\mathbf{v}^\tau, \mathbf{v}^\tau + \mathbf{j}^\tau \in \mathcal{Z}^d$, which was given by Whittle (1954) when $d = 2$ and by Guyon (1982) when $d \geq 2$. Section 2.3 defines the weakly and strictly stationary processes and states the Wold decomposition, which provides a link between (weakly) stationary processes and linear processes. In that same section, we prove some properties of processes, which are linear functions of independent and identically distributed random variables. A new definition of the so-called reverse strict stationarity might also be found there, which is an attempt to extend the definition of strict stationarity in a way that does not favor any direction of each one of the d dimensions. Later in Proposition 2.5 and, consequently, in Chapter 3 and Sections 4.4 and 4.5, we have used conditions, which are satisfied if the process of interest is reverse strictly stationary. Thus, when we establish in the end of Section 2.3.2 that reverse strictly stationary processes exist, at the same time we allow for some of our conditions used in Chapters 3 and for 4 to be more realistic.

In Section 2.4, we define the causal and invertible ARMA processes on the

d-dimensional lattice and study their second-order properties. We focus on the special case of auto-regressions and moving-averages that not only share the same polynomial, but also they are generated by the same sequences of uncorrelated random variables. What we call the general Yule-Walker equations follow next, which provide a link between the auto-covariance functions of an auto-regression and a moving-average with the same polynomial. These equations will be further used in Chapter 3, which will only deal with ARMA processes, but they will also be used in Chapter 4. This is because they refer to the second-order properties of two processes, rather than any causal formulation considered to be taking place there. Not only will these equations be used as the theoretical base for a method of moments suggested in Sections 3.4.1 and 3.4.2, but also they are the key used, in order to find the forms of inverse conditional variance matrices for a set of random variables either from the auto-regression or the moving-average process of interest and mainly for Gaussian processes. Later in Chapter 3, this will allow us to use these matrices in Gaussian conditional likelihoods. Again, since the derivation of these matrices is based on the general Yule-Walker equations, we will also use these results to write conditional likelihoods in Chapter 4, even though the random variables there, might not have been generated from an auto-regression or a moving-average. We conclude the chapter with Section 2.5, in order to come up with a central limit theorem for processes on the regular d-dimensional lattice.

2.2 Unilateral orderings

We consider $\{X(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ to be a real valued process, where d is a positive integer and $\mathbf{v} = [v_1, \cdots, v_d]$ is a d-dimensional vector index. We denote with $>$ the lexicographic order on $\mathcal{Z}^d$; when $d = 1$ this is the same as the standard order on $\mathcal{Z}$. When $d = 2$, the notion of unilateral ordering was defined by Whittle (1954). For the general case of any positive integer d, we explain below the ordering due to Guyon (1982, p.96). We write

$$\mathbf{j} = [j_1, j_2, \cdots, j_d] > \mathbf{0} = [0, 0, \cdots, 0]$$

on $\mathcal{Z}^d$, if

$$j_1 > 0$$

or

$$j_1 = 0 \ \text{ and } \ [j_2, \cdots, j_d] > [0, \cdots, 0]$$

on $\mathbb{Z}^{d-1}$.

When $d > 1$, writing $\mathbf{j} > \mathbf{0}$ may have different meanings. For example, for two-dimensional processes

$$[j_1, j_2] > [0, 0]$$

if

$$j_1 > 0$$

or

$$j_1 = 0 \text{ and } j_2 > 0,$$

as described before. But we could also change the order of the indices and write

$$[j_2, j_1] > [0, 0]$$

if

$$j_2 > 0$$

or

$$j_2 = 0 \text{ and } j_1 > 0.$$

One interesting question would be how many such representations exist for general number of dimensions d. To answer that, we first consider the d distinct dimensions with two different ends. For the time axis, these would be the 'past' and the 'future' and would have a natural order. It could also be the 'west' and 'east' or the 'south' and 'north' for the dimensions of space. Next, we define an hierarchy between the dimensions indicated by the labels $k = 1, \cdots, d$. The most important dimension is labelled as 1 and the least important one as d. Dimension $k = 1, \cdots, d-1$, is considered more important than dimension $k^* = k+1, \cdots, d$, when moving its index towards any side has the same effect on the ordering of two locations, regardless of the way the other index has changed. For example, moving from time 1 and location labelled as 2 to either time 2 and location 3 or time 2 and location 1, is considered as moving to the future since time is going forwards in both cases. In general, there are $d!$ ways to label the different dimensions and the time axis is usually considered the most important of all and it is labelled as dimension 1.

Once the dimensions have been labelled, one has to do the same for the two ends of each dimension. As a result, there are 2 ways to decide about the direction of $j_k > 0$ for

each dimension $k = 1, \cdots, d$, and there are 2^d ways altogether. For example, for $d = 2$ we can define $2^2 = 4$ different orderings. Say there is the dimension 'west-east' first and the dimension 'south-north'; the 4 representations can be labelled as 'west-south' and 'east-north' or 'west-north' and 'east-south'. Of these, 2^{d-1} choices can be seen as the counterparts of the remaining 2^{d-1} representations. For example, 'east-north' is the counterpart of 'west-south', since it corresponds to the opposite quarter of $\mathcal{Z}^2$. Similarly, 'east-south' is the counterpart of 'west-north'.

2.3 Stationary processes

We extend the definitions of weak and strict stationarity for processes with d indices, where d is any positive integer.

Definition 2.1 (Weak stationarity). $\{X(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ is a (weakly) stationary process if $E\{X^2(\mathbf{v})\} < \infty$, and

1. $E\{X(\mathbf{v})\}$ is a constant independent of $\mathbf{v}$, and
2. $\mathrm{Cov}\{X(\mathbf{v}), X(\mathbf{v}+\mathbf{j})\}$ is independent of $\mathbf{v}$ for every $\mathbf{j}^\tau \in \mathcal{Z}^d$.

Without loss of generality, we will consider

$$E\{X(\mathbf{v})\} = 0 \tag{2.3.1}$$

unless stated otherwise. Then we will write the real-valued function

$$\gamma(\mathbf{j}) \equiv \mathrm{Cov}\{X(\mathbf{v}), X(\mathbf{v}+\mathbf{j})\} = E\{X(\mathbf{v})X(\mathbf{v}+\mathbf{j})\} \tag{2.3.2}$$

to be the auto-covariance function of the stationary process of interest defined for any lag $\mathbf{j}^\tau \in \mathcal{Z}^d$. This function is even in the sense that

$$\gamma(\mathbf{j}) = \gamma(-\mathbf{j}),\ \mathbf{j}^\tau \in \mathcal{Z}^d.$$

Under the condition that $\gamma(\cdot)$ is an absolutely summable function, we define the spectral density of $\gamma(\cdot)$ to be

$$g(\boldsymbol{\omega}) \equiv \frac{1}{(2\pi)^d} \sum_{\mathbf{j}^\tau \in \mathcal{Z}^d} e^{i\boldsymbol{\omega}\mathbf{j}^\tau} \gamma(\mathbf{j}),\ \boldsymbol{\omega}^\tau \in [-\pi, \pi]^d, \tag{2.3.3}$$

for $\boldsymbol{\omega} = [\omega_1, \cdots, \omega_d]$ and $i = \sqrt{-1}$.

Definition 2.2 (Strict stationarity). The process $\{X(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ is said to be strictly stationary if the joint distribution of $[X(\mathbf{v}_1), \cdots, X(\mathbf{v}_q)]^\tau$ and $[X(\mathbf{v}_1 + \mathbf{j}), \cdots, X(\mathbf{v}_q + \mathbf{j})]^\tau$ are the same for all positive integers q and for all $\mathbf{v}_1^\tau, \cdots, \mathbf{v}_q^\tau, \mathbf{j}^\tau \in \mathcal{Z}^d$.

2.3.1 Linear processes

We consider $\{u(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ to be a white noise sequence of random variables when they are generated on the points of $\mathcal{Z}^d$ and they are uncorrelated with each other. We may then state the Wold decomposition.

Theorem 2.1 (Wold decomposition). A zero-mean and (weakly) stationary process $\{X(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ with spectral density $g(\cdot)$, such that

$$\int_{[-\pi,\pi]^d} \log g(\boldsymbol{\omega}) d\boldsymbol{\omega} > -\infty, \tag{2.3.4}$$

can be expressed in the form

$$X(\mathbf{v}) = u(\mathbf{v}) + \sum_{\mathbf{j}>\mathbf{0}} \psi_{\mathbf{j}}\, u(\mathbf{v} - \mathbf{j}), \tag{2.3.5}$$

where

1. $\sum_{\mathbf{j}>\mathbf{0}} \psi_{\mathbf{j}}^2 < \infty$,
2. $\{u(\mathbf{v})\} \sim WN(0, \sigma^2)$.

Finally, $\sigma^2 = \exp\{(2\pi)^{-d} \int_{[-\pi,\pi]^d} \log f(\boldsymbol{\omega}) d\boldsymbol{\omega}\}$ is given by Kolmogorov's formula, where

$$f(\boldsymbol{\omega}) \equiv (2\pi)^d \cdot g(\boldsymbol{\omega}),\ \boldsymbol{\omega}^\tau \in [-\pi, \pi]^d.$$

For the proof of the theorem, see Rosanov (1967, p.64) for $d = 1$ and Helson and Lowdenslager (1958) for $d = 2$, the proof being similar for $d > 2$ (Guyon, 1982, p.96).

The Wold decomposition provides the link between (weak) stationarity of a process and linearity, *i.e.* that it may be represented as a linear combination of uncorrelated random variables. The second-order properties of a stationary sequence of random variables $\{X(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ may be fully described by the auto-covariance function $\gamma(\cdot)$ or the spectral density $g(\cdot)$. As the Wold decomposition allows for $X(\mathbf{v})$ to be written as a linear combination of $u(\mathbf{v} - \mathbf{j})$, $\mathbf{j} \geq \mathbf{0}$ only, a unilateral representation is allowed.

Now that the Wold decomposition has been established, it is useful to derive the asymptotic properties of linear processes for any $d \geq 1$ number of dimensions. Next, we prove two propositions that follow from Proposition 6.3.10 and Proposition 7.3.5 of Brockwell and Davis (1991).

Proposition 2.1 (Weak Law of Large Numbers for linear processes). Let $\{L(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ be the linear process defined by

$$L(\mathbf{v}) = \sum_{\mathbf{j} \geq \mathbf{0}} l_{\mathbf{j}}\, W(\mathbf{v}-\mathbf{j}),\ \sum_{\mathbf{j} \geq \mathbf{0}} |l_{\mathbf{j}}| < \infty,\ \{W(\mathbf{v})\} \sim IID(\mu, \sigma^2),$$

and $\mathcal{S} \subset \mathcal{Z}^d$ be a set of cardinality N. Then as $N \to \infty$, it holds that

$$L_N \equiv \frac{1}{N} \sum_{\mathbf{v}^\tau \in \mathcal{S}} L(\mathbf{v}) \xrightarrow{P} \left(\sum_{\mathbf{j} \geq \mathbf{0}} l_{\mathbf{j}} \right) \mu.$$

Proof. First note that

$$\begin{aligned} |E\{L(\mathbf{v})\}| &\leq E|L(\mathbf{v})| = E|\sum_{\mathbf{j} \geq \mathbf{0}} l_{\mathbf{j}}\, W(\mathbf{v}-\mathbf{j})| \leq E\{\sum_{\mathbf{j} \geq \mathbf{0}} |l_{\mathbf{j}}\, W(\mathbf{v}-\mathbf{j})|\} \\ &= \sum_{\mathbf{j} \geq \mathbf{0}} |l_{\mathbf{j}}| E|W(\mathbf{v}-\mathbf{j})| = E|W(\mathbf{v})| \sum_{\mathbf{j} \geq \mathbf{0}} |l_{\mathbf{j}}| < \infty, \end{aligned}$$

and the series is well-defined in the sense of convergence in probability. For positive integer K, we define the set

$$\begin{aligned} \mathcal{M}_K \equiv\ & \{[j_1, j_2, \cdots, j_d]^\tau : j_1 = 1, \cdots, K,\ j_k = 0, \pm 1, \cdots, \pm K,\ k = 2, \cdots, d\} \cup \\ \cup\ & \{[0, j_2, \cdots, j_d]^\tau : j_2 = 1, \cdots, K,\ j_k = 0, \pm 1, \cdots, \pm K,\ k = 3, \cdots, d\} \cup \cdots \cup \\ \cup\ & \{[0, 0, \cdots, j_d]^\tau : j_d = 1, \cdots, K\} \cup \{[0, \cdots, 0]^\tau\}. \end{aligned} \tag{2.3.6}$$

Then for any fixed K, as $N \to \infty$,

$$Y_{NK} \equiv \frac{1}{N} \sum_{\mathbf{v}^\tau \in \mathcal{S}} \sum_{\mathbf{j}^\tau \in \mathcal{M}_K} l_{\mathbf{j}}\, W(\mathbf{v}-\mathbf{j}) = \sum_{\mathbf{j}^\tau \in \mathcal{M}_K} l_{\mathbf{j}}\, \frac{1}{N} \sum_{\mathbf{v}^\tau \in \mathcal{S}} W(\mathbf{v}-\mathbf{j}) \xrightarrow{P} \mu \sum_{\mathbf{j}^\tau \in \mathcal{M}_K} l_{\mathbf{j}},$$

since for fixed $\mathbf{j}^\tau \in \mathcal{M}_K$, it holds that $\{W(\mathbf{v}-\mathbf{j}),\ \mathbf{v}^\tau \in \mathcal{S}\}$ are independent and identically distributed random variables. We also define the constants

$$\mu_L(K) \equiv \mu \sum_{\mathbf{j}^\tau \in \mathcal{M}_K} l_{\mathbf{j}}$$

and

$$\mu_L \equiv \mu \sum_{\mathbf{j} \geq \mathbf{0}} l_{\mathbf{j}}.$$

Then $Y_{NK} \xrightarrow{P} \mu_L(K)$ as $N \to \infty$ and $\mu_L(K) \to \mu_L$ as $K \to \infty$. We now only need to show that

$$\lim_{K\to\infty} \limsup_{N\to\infty} P\left(|L_N - Y_{NK}| > \epsilon\right) = 0, \text{ for any } \epsilon > 0.$$

Note that

$$\begin{aligned}
P(|L_N - Y_{NK}| > \epsilon) &= P(|\frac{1}{N}\sum_{\mathbf{v}^\tau\in\mathcal{S}} L(\mathbf{v}) - \frac{1}{N}\sum_{\mathbf{v}^\tau\in\mathcal{S}}\sum_{\mathbf{j}^\tau\in\mathcal{M}_K} l_{\mathbf{j}}\, W(\mathbf{v}-\mathbf{j})| > \epsilon) \\
&= P(|\frac{1}{N}\sum_{\mathbf{v}^\tau\in\mathcal{S}}\sum_{\mathbf{j}\geq\mathbf{0}} l_{\mathbf{j}}\, W(\mathbf{v}-\mathbf{j}) - \frac{1}{N}\sum_{\mathbf{v}^\tau\in\mathcal{S}}\sum_{\mathbf{j}^\tau\in\mathcal{M}_K} l_{\mathbf{j}}\, W(\mathbf{v}-\mathbf{j})| > \epsilon) \\
&= P(|\frac{1}{N}\sum_{\mathbf{v}^\tau\in\mathcal{S}}\sum_{\substack{\mathbf{j}^\tau\notin\mathcal{M}_K,\\ \mathbf{j}\geq\mathbf{0}}} l_{\mathbf{j}}\, W(\mathbf{v}-\mathbf{j})| > \epsilon) \\
&\leq \frac{1}{N\epsilon}\sum_{\mathbf{v}^\tau\in\mathcal{S}}\sum_{\substack{\mathbf{j}^\tau\notin\mathcal{M}_K,\\ \mathbf{j}\geq\mathbf{0}}} |l_{\mathbf{j}}| E|W(\mathbf{v}-\mathbf{j})| \\
&= \frac{1}{\epsilon}\sum_{\substack{\mathbf{j}^\tau\notin\mathcal{M}_K,\\ \mathbf{j}\geq\mathbf{0}}} |l_{\mathbf{j}}| E|W([1,\cdots,1]-\mathbf{j})|,
\end{aligned}$$

where the inequality is due to Chebychev. Then,

$$\sum_{\substack{\mathbf{j}^\tau\notin\mathcal{M}_K,\\ \mathbf{j}\geq\mathbf{0}}} |l_{\mathbf{j}}| = \sum_{\max_{k=2,\cdots,d}\{j_1,|j_k|\}>K} |l_{\mathbf{j}}| = \sum_{j_1+\sum_{k=2}^{d}|j_k|>K} |l_{\mathbf{j}}| \to 0,$$

as $K \to \infty$. ■

Proposition 2.2. If $\{C(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$, $\{D(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ are two linear processes such that

$$\begin{aligned}
C(\mathbf{v}) &= \sum_{\mathbf{i}\geq\mathbf{0}} c_{\mathbf{i}}\, W(\mathbf{v}-\mathbf{i}),\ \sum_{\mathbf{i}\geq\mathbf{0}} |c_{\mathbf{i}}| < \infty, \\
D(\mathbf{v}) &= \sum_{\mathbf{i}\geq\mathbf{0}} d_{\mathbf{i}}\, W(\mathbf{v}-\mathbf{i}),\ \sum_{\mathbf{i}\geq\mathbf{0}} |d_{\mathbf{i}}| < \infty,\ \{W(\mathbf{v})\} \sim IID(0,\sigma^2),
\end{aligned}$$

then for a set $\mathcal{S} \subset \mathcal{Z}^d$ of cardinality N and $\mathbf{j} \geq \mathbf{0}$, it holds that

$$\begin{aligned}
\frac{1}{N}\sum_{\mathbf{v}^\tau\in\mathcal{S}} C(\mathbf{v})C(\mathbf{v}+\mathbf{j}) &\xrightarrow{P} \left(\sum_{\mathbf{i}\geq\mathbf{0}} c_{\mathbf{i}}\, c_{\mathbf{i}+\mathbf{j}}\right)\sigma^2 = \mathrm{Cov}\{C(\mathbf{v}), C(\mathbf{v}+\mathbf{j})\}, \\
\frac{1}{N}\sum_{\mathbf{v}^\tau\in\mathcal{S}} C(\mathbf{v})D(\mathbf{v}+\mathbf{j}) &\xrightarrow{P} \left(\sum_{\mathbf{i}\geq\mathbf{0}} c_{\mathbf{i}}\, d_{\mathbf{i}+\mathbf{j}}\right)\sigma^2 = \mathrm{Cov}\{C(\mathbf{v}), D(\mathbf{v}+\mathbf{j})\}
\end{aligned}$$

as $N \to \infty$.

Proof. We only prove the first statement since the second one can be shown in a similar way. We write

$$\frac{1}{N}\sum_{\mathbf{v}^\tau\in\mathcal{S}} C(\mathbf{v})C(\mathbf{v}+\mathbf{j}) = \frac{1}{N}\sum_{\mathbf{v}^\tau\in\mathcal{S}}\sum_{\mathbf{i}\geq\mathbf{0}} c_{\mathbf{i}}\, W(\mathbf{v}-\mathbf{i})\sum_{\mathbf{i}^*\geq\mathbf{0}} c_{\mathbf{i}^*} W(\mathbf{v}+\mathbf{j}-\mathbf{i}^*)$$
$$= \frac{1}{N}\sum_{\mathbf{v}^\tau\in\mathcal{S}}\sum_{\mathbf{i}\geq\mathbf{0}} c_{\mathbf{i}}\, c_{\mathbf{i}+\mathbf{j}}\, W(\mathbf{v}-\mathbf{i})^2 + Y_{\mathbf{j}N},$$

where

$$Y_{\mathbf{j}N} \equiv \frac{1}{N}\sum_{\mathbf{v}^\tau\in\mathcal{S}}\sum_{\substack{\mathbf{i},\ \mathbf{i}^*\geq\mathbf{0},\\ \mathbf{i}^*\neq\mathbf{i}+\mathbf{j}}} c_{\mathbf{i}}\, c_{\mathbf{i}^*}\, W(\mathbf{v}-\mathbf{i})W(\mathbf{v}+\mathbf{j}-\mathbf{i}^*)$$
$$= \sum_{\substack{\mathbf{i},\ \mathbf{i}^*\geq\mathbf{0},\\ \mathbf{i}^*\neq\mathbf{i}+\mathbf{j}}} c_{\mathbf{i}}\, c_{\mathbf{i}^*}\left(N^{-1}\sum_{\mathbf{v}^\tau\in\mathcal{S}} W(\mathbf{v}-\mathbf{i})W(\mathbf{v}+\mathbf{j}-\mathbf{i}^*)\right)$$

For the first term $\sum_{\mathbf{v}^\tau\in\mathcal{S}}\sum_{\mathbf{i}\geq\mathbf{0}} c_{\mathbf{i}}\, c_{\mathbf{i}+\mathbf{j}}\, W(\mathbf{v}-\mathbf{i})^2/N$, it holds that $\{W(\mathbf{v})^2,\ \mathbf{v}^\tau\in\mathcal{Z}\}$ are independent and identically distributed with mean σ^2, and since $\sum_{\mathbf{i}\geq\mathbf{0}} |c_{\mathbf{i}}\, c_{\mathbf{i}+\mathbf{j}}| < \infty$, from the Weak Law of Large Numbers for the linear process $L_{\mathbf{j}}(\mathbf{v}) \equiv \sum_{\mathbf{i}\geq\mathbf{0}} c_{\mathbf{i}}\, c_{\mathbf{i}+\mathbf{j}}\, W(\mathbf{v}-\mathbf{i})^2$, it holds that

$$\frac{1}{N}\sum_{\mathbf{v}^\tau\in\mathcal{S}} L_{\mathbf{j}}(\mathbf{v}) \xrightarrow{P} \left(\sum_{\mathbf{i}\geq\mathbf{0}} c_{\mathbf{i}}\, c_{\mathbf{i}+\mathbf{j}}\right) E\{W(\mathbf{v})^2\} = \left(\sum_{\mathbf{i}\geq\mathbf{0}} c_{\mathbf{i}}\, c_{\mathbf{i}+\mathbf{j}}\right)\sigma^2,$$

as $N\to\infty$.

It suffices to show that $Y_{\mathbf{j}N} \xrightarrow{P} 0$. For $\mathbf{i}^*\neq\mathbf{i}+\mathbf{j}$, it holds that $\{W(\mathbf{v}-\mathbf{i})W(\mathbf{v}+\mathbf{j}-\mathbf{i}^*),\ \mathbf{v}^\tau\in\mathcal{Z}^d\} \sim WN(0,\sigma^4)$ and, hence,

$$\mathrm{Var}\left(N^{-1}\sum_{\mathbf{v}^\tau\in\mathcal{S}} W(\mathbf{v}-\mathbf{i})W(\mathbf{v}+\mathbf{j}-\mathbf{i}^*)\right) = N^{-1}\sigma^4 \to 0$$

as $N\to\infty$. For $\mathcal{M}_K$ as defined in (2.3.6), we may define for fixed K

$$Y_{\mathbf{j}NK} \equiv \sum_{\substack{\mathbf{i}^\tau,\ \mathbf{i}^{*\tau}\in\mathcal{M}_K,\\ \mathbf{i}^*\neq\mathbf{i}+\mathbf{j}}} c_{\mathbf{i}}\, c_{\mathbf{i}^*}\left(N^{-1}\sum_{\mathbf{v}^\tau\in\mathcal{S}} W(\mathbf{v}-\mathbf{i})W(\mathbf{v}+\mathbf{j}-\mathbf{i}^*)\right) \xrightarrow{P} 0,$$

as $N\to\infty$. So,

$$E|Y_{\mathbf{j}N} - Y_{\mathbf{j}NK}| \leq E|W[1,\cdots,1]W[1,\cdots,2]|\cdot$$
$$\left(\sum_{\substack{\mathbf{i}^\tau\in\mathcal{M}_K,\mathbf{i}^{*\tau}\notin\mathcal{M}_K,\\ \mathbf{i}^*\geq\mathbf{0},\ \mathbf{i}^*\neq\mathbf{i}+\mathbf{j}}} |c_{\mathbf{i}}|\,|c_{\mathbf{i}^*}| + \sum_{\substack{\mathbf{i}^\tau\notin\mathcal{M}_K,\mathbf{i}^{*\tau}\in\mathcal{M}_K,\\ \mathbf{i}\geq\mathbf{0},\ \mathbf{i}^*\neq\mathbf{i}+\mathbf{j}}} |c_{\mathbf{i}}|\,|c_{\mathbf{i}^*}| + \sum_{\substack{\mathbf{i}^\tau,\ \mathbf{i}^{*\tau}\notin\mathcal{M}_K,\\ \mathbf{i},\ \mathbf{i}^*\geq\mathbf{0},\ \mathbf{i}^*\neq\mathbf{i}+\mathbf{j}}} |c_{\mathbf{i}}|\,|c_{\mathbf{i}^*}|\right) \to 0,$$

as $K\to\infty$. ■

2.3.2 Reverse stationarity

As we have seen in Definition 2.1, weak stationarity relates any two random variables of the process of interest and it ensures that their auto-covariance is a function of the d-dimensional vector difference of the two locations. Further, if the auto-covariance function depends on this vector through its norm only, the process is called isotropic. Isotropic processes allow for more specific considerations and they are beyond the scope of this thesis.

For the definition of strict stationarity, we refer to any two random vectors and their distributions. The random vectors might be of any positive length, say q. In time series, strict stationarity means intuitively that the graphs over two time intervals of length q of a realization of the process should exhibit similar statistical characteristics. But it does not mean that those are the same characteristics as the ones exhibited within the same intervals if they are observed from future to past. For a process evolving on a line transect, as this was described by Whittle (1954, p.434), time is replaced by a dimension of space and this might not make sense. We may observe the process starting from any of the two ends towards the other end. Thus, we wish to define a form of stationarity that means intuitively that the two graphs over the intervals of length q that start from the same location, one from left to right and the other from right to left, exhibit similar statistical characteristics.

On the other hand, when we deal with conditional probabilities in a time series, the natural order of the indexes plays an important role. For example, for the two random variables $X(1)$ and $X(2)$, we will rarely introduce in our analysis the conditional probabilities of $X(1)$ given $X(2)$, unless we are asked to. In this last case, we usually convert to these probabilities from the conditional probabilities of $X(2)$ given $X(1)$, using the Bayesian formula. The answer might come much faster if we know that the distribution of the random vector $[X(1), X(2)]^\tau$ is the same as the distribution of the random vector $[X(2), X(1)]^\tau$, as this will imply both that the marginal distributions of $X(1)$ and $X(2)$ are the same and that

$$P(X(1) = u|X(2) = v) = P(X(2) = u|X(1) = v). \tag{2.3.7}$$

Equation (2.3.7) might make more sense when the indexes $1, 2$, refer to two locations on the line transect instead. The conditional probabilities of $X(1)$ given $X(2)$ might then be as meaningful as the conditional probabilities of $X(2)$ given $X(1)$.

According to Definition 2.2, for the general case of strictly stationary processes on the d-dimensional lattice, it holds for any $\mathbf{j}^\tau \in \mathcal{Z}^d$ that the two random vectors $[X(\mathbf{v}_1), \cdots, X(\mathbf{v}_q)]^\tau$ and $[X(\mathbf{v}_1+\mathbf{j}), \cdots, X(\mathbf{v}_q+\mathbf{j})]^\tau$ have the same distribution, because they refer to the same $\begin{pmatrix} q \\ 2 \end{pmatrix}$ differences, *i.e.* $\mathbf{v}_1 - \mathbf{v}_2, \cdots, \mathbf{v}_1 - \mathbf{v}_q, \cdots, \mathbf{v}_{q-1} - \mathbf{v}_q$. As a result, one can shift the random vector $[X(\mathbf{v}_1), \cdots, X(\mathbf{v}_q)]^\tau$ at any $\mathbf{j}^\tau \in \mathcal{Z}^d$ steps away, and the distributions of the new random vectors are still the same. But what if we are not interested in changing the location of the random vector but its direction? In other words, can we make sure that the random vector $[X(-\mathbf{v}_1), \cdots, X(-\mathbf{v}_q)]^\tau$ has the same distribution as the random vector $[X(\mathbf{v}_1), \cdots, X(\mathbf{v}_q)]^\tau$, although it refers to the differences of opposite sign, *i.e.* $-\mathbf{v}_1 + \mathbf{v}_2, \cdots, -\mathbf{v}_1 + \mathbf{v}_q, \cdots, -\mathbf{v}_{q-1} + \mathbf{v}_q$?

Definition 2.3 (Reverse strict stationarity). The process $\{X(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ is said to be reverse strictly stationary if the joint distribution of $[X(\mathbf{j} + \mathbf{v}_1), \cdots, X(\mathbf{j} + \mathbf{v}_q)]^\tau$ and $[X(\mathbf{j} - \mathbf{v}_1), \cdots, X(\mathbf{j} - \mathbf{v}_q)]^\tau$ are the same for all positive integers q and for all $\mathbf{j}^\tau,\ \mathbf{v}_1^\tau, \cdots,\ \mathbf{v}_q^\tau \in \mathcal{Z}^d$.

Now, it is clear that the differences between the locations of the second random vector are all of opposite sign from the ones of the first vector, but that has no effect on its distribution compared to the distribution of the first random vector. The two vectors this time, both have to originate on the same location $\mathbf{v}^\tau \in \mathcal{Z}^d$ without any shift. As a result, Definition 2.3 does not imply that for a reverse strictly stationary process $\{X(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$, the two random variables $X(\mathbf{v}_1)$ and $X(\mathbf{v}_2)$ have the same marginal distribution for any two locations $\mathbf{v}_1^\tau, \mathbf{v}_2^\tau \in \mathcal{Z}^d$. For example, if $d = 1$ the random variables $X(-1)$ and $X(1)$ have the same distribution since the two locations of interest are one step away from location 0. For the same reason, all the random variables $X(2j+1), j \in \mathcal{Z}$, share the same distribution and all the random variables $X(2j), j \in \mathcal{Z}$, also share the same distribution, but the two distributions do not have to be the same. This is established in the next proposition, which relates the two forms of stationarity. We denote with $\mathcal{O}$, $\mathcal{E}$ the odd and even integer number spaces, respectively. We also denote with $\mathcal{P}$ any of the 2^d orderings $\mathcal{O}^d$, $\mathcal{E} \times \mathcal{O}^{d-1}, \cdots, \mathcal{E}^d$.

Proposition 2.3. Let $\{X(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ be a reverse strictly stationary process and define

$$X_{\mathcal{P}}(\mathbf{v}) \equiv X(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{P}.$$

Then the process $\{X_{\mathcal{P}}(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{P}\}$ is strictly stationary.

Proof. Since the process is reverse strictly stationary, it holds for any positive integer q and any $\mathbf{v}_1^\tau, \cdots, \mathbf{v}_q^\tau \in \mathcal{Z}^d$ that the random vectors $[X(\mathbf{v}_1), \cdots, X(\mathbf{v}_q)]^\tau$ and $[X(-\mathbf{v}_1), \cdots, X(-\mathbf{v}_q)]^\tau$ have the same distribution. This stems from Definition 2.3 when we set $\mathbf{v} = \mathbf{0}$.

Similarly, it holds for any $\mathbf{v}^\tau \in \mathcal{Z}^d$ that the vectors $[X(\mathbf{j} - \mathbf{v}_1), \cdots, X(\mathbf{j} - \mathbf{v}_q)]^\tau$ and $[X(-\mathbf{j}+\mathbf{v}_1), \cdots, X(-\mathbf{j}+\mathbf{v}_q)]^\tau$ share the same distribution. On the other hand and again from Definition 2.3, since the process is reverse strictly stationary the first of previous random vectors has the same distribution as $[X(\mathbf{j} + \mathbf{v}_1), \cdots, X(\mathbf{j} + \mathbf{v}_q)]^\tau$. The two arguments combined together imply that the random vectors $[X(\mathbf{v}_1 - \mathbf{j}), \cdots, X(\mathbf{v}_q - \mathbf{j})]^\tau$ and $X(\mathbf{v}_1 + \mathbf{j}), \cdots, X(\mathbf{v}_q + \mathbf{j})]^\tau$, or $[X(\mathbf{v}_1), \cdots, X(\mathbf{v}_q)]^\tau$ and $[X(\mathbf{v}_1 + 2\mathbf{j}), \cdots, X(\mathbf{v}_q + 2\mathbf{j})]^\tau$ share the same distribution for any $\mathbf{j}^\tau, \mathbf{v}_1^\tau, \cdots, \mathbf{v}_q^\tau \in \mathcal{Z}^d$. We can see that for any $\mathbf{j}^{*\tau} \in \mathcal{E}^d$ there is a unique element $\mathbf{j}^\tau \in \mathcal{Z}^d$ such that $2\mathbf{j} = \mathbf{j}^*$ and vice versa. The proof is completed when we also see that for a specific $\mathbf{v}_1^\tau \in \mathcal{P}$, there is a unique element $\mathbf{v}_1^\tau + \mathbf{j}^{*\tau} \in \mathcal{P}$ for any $\mathbf{j}^{*\tau} \in \mathcal{E}^d$ and vice versa. ■

Proposition 2.3 shows that the way a reverse strictly stationary process has been defined does not allow us to necessarily conclude that it is strictly stationary as well. As a result, we first require that a process is strictly stationary and then look for its extra attributes. We may think of a simple example, where both properties exist and can be combined to derive useful results. We consider the case of a strictly and reverse strictly stationary process $\{X(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$. Then the distribution of $[X(\mathbf{v}_1), X(\mathbf{v}_2)]^\tau$ is the same as the distribution of $[X(-\mathbf{v}_1), X(-\mathbf{v}_2)]^\tau$, because of reverse strict stationarity, and this is the same as the distribution of $[X(-\mathbf{v}_1 + \mathbf{v}_1 + \mathbf{v}_2), X(-\mathbf{v}_2 + \mathbf{v}_1 + \mathbf{v}_2)]^\tau$ or $[X(\mathbf{v}_2), X(\mathbf{v}_1)]^\tau$, because of strict stationarity. In other words, we have reversed the order of the two locations $\mathbf{v}_1$ and $\mathbf{v}_2$. It is fairly easy to show, for example, that this property holds for any pair of identically distributed (but not necessarily independent) Bernoulli random variables.

But is it possible to start from a strictly stationary process and show that it is reverse strictly stationary for any positive integer q? The following proposition gives a sufficient condition for a strictly stationary process to be reverse strictly stationary as well.

Proposition 2.4. For a strictly stationary sequence of random variables $\{X(\mathbf{v}),\ \mathbf{v}^\tau \in$

$\mathcal{Z}^d\}$, if the joint probability function of $[X(\mathbf{v}_1), \cdots, X(\mathbf{v}_q)]^\tau$ is an even function of all the differences $\mathbf{v}_1 - \mathbf{v}_2, \cdots, \mathbf{v}_{q-1} - \mathbf{v}_q$, for any positive integer q and $\mathbf{v}_1^\tau, \cdots, \mathbf{v}_q^\tau \in \mathcal{Z}^d$, then it is a reverse strictly stationary process.

Proof. Since the sequence is strictly stationary, we know that the joint probability function depends on the possible differences of locations only. If we consider it an even function, in the sense that changing the sign of all differences results in the same distribution of the random vector on the new locations, then the sequence is reverse strictly stationary. ■

The joint probability function of q random variables, say $X_1, \cdots, X_q$, is such that its logarithm can be written as

$$\begin{aligned} \log\{P_{X_1,\cdots,X_q}(x_1, \cdots, x_q)\} &= K \cdot [\sum_i x_i \, g_i(x_i) + \sum_{i<j} x_i \cdot x_j \, g_{i,j}(x_i, x_j) + \cdots \\ &+ x_1 \cdots x_q \, g_{1,\cdots,q}(x_1, \cdots, x_q)]. \end{aligned} \tag{2.3.8}$$

Besag (1974, p.197) claimed the existence of such functions $g_i(\cdot), \cdots, g_{1,\cdots,q}(\cdot)$, under some very mild conditions. He focused on the special cases where

$$\log\{P_{X_1,\cdots,X_q}(x_1, \cdots, x_q)\} = K \cdot (\sum_i x_i \, g_i(x_i) + \sum_{i<j} x_i \cdot x_j \, \beta_{i,j}) \tag{2.3.9}$$

and called them auto-models. He then showed that $\beta_{i,j} = \beta_{j,i}$. Two examples are the auto-logistic and the auto-normal model. We will deal with the auto-models again in Chapter 4.

We let the strictly stationary sequence of random variables $\{X(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ and any positive integer q. For fixed, $\mathbf{v}_1^\tau, \cdots, \mathbf{v}_q^\tau \in \mathcal{Z}^d$, we write for convenience

$$X(\mathbf{v}_i) \equiv X_i,\ i = 1, \cdots, q,$$

and the joint probability function of the vector $[X(\mathbf{v}_1), \cdots, X(\mathbf{v}_q)]^\tau$ denoted as $P_{X_1,\cdots,X_q}(x_1, \cdots, x_q)$. Then it should hold in (2.3.8) that

$$K \equiv K(\mathbf{v}_1 - \mathbf{v}_2, \cdots, \mathbf{v}_1 - \mathbf{v}_q, \cdots, \mathbf{v}_{q-1} - \mathbf{v}_q),$$

and that

$$g_i(x) \equiv g^{(1)}(x),\ i = 1, \cdots, q,$$

and

$$g_{i,j}(x, y) \equiv g^{(2)}(\mathbf{v}_i - \mathbf{v}_j, x, y),\ i, j = 1, \cdots, q,\ i < j,$$

and

$$\vdots$$

and

$$g_{1,\cdots,q}(x_1,\cdots,x_q) \equiv g^{(q)}(\mathbf{v}_1-\mathbf{v}_2,\cdots,\mathbf{v}_1-\mathbf{v}_q,\cdots,\mathbf{v}_{q-1}-\mathbf{v}_q,x_1,\cdots,x_q),$$

for some functions $K(\cdot), g^{(1)}(\cdot),\cdots,g^{(q)}(\cdot)$. Reverse strict stationarity would require that these are even functions of the differences $\mathbf{v}_1-\mathbf{v}_2,\cdots,\mathbf{v}_1-\mathbf{v}_q,\cdots,\mathbf{v}_{q-1}-\mathbf{v}_q$, as it is described in Proposition 2.4.

Lemma 2.1. A strictly stationary sequence of Gaussian random variables is reverse strictly stationary.

Proof. The proof comes immediately from the fact that the joint density of any q identically distributed Gaussian random variables is a function of the $\begin{pmatrix} q \\ 2 \end{pmatrix}$ auto-covariances, *i.e.* an even function of all the possible differences. ■

Remark 2.1. Apart from the case of a Gaussian reverse strictly stationary process, as it was described in Lemma 2.1, are there any other reverse strictly stationary sequences? We explain here how reverse strictly stationary processes can be produced and reproduced. We know that when we have a strictly stationary process $\{X(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ and we apply for all $\mathbf{v}^\tau \in \mathcal{Z}^d$ the same linear filter, say

$$S(\mathbf{v}) \equiv \sum_{\mathbf{j}^\tau \in \mathcal{Z}^d} l_{\mathbf{j}}\, X(\mathbf{v}-\mathbf{j}),\ \sum_{\mathbf{j}^\tau \in \mathcal{Z}^d} |l_{\mathbf{j}}| < \infty,$$

then the new process $\{S(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ is also strictly stationary. Similarly, if $\{X(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ is reverse strictly stationary and we apply the symmetric linear filter

$$R(\mathbf{v}) \equiv l_0\, X(\mathbf{v}) + \sum_{\mathbf{j}>\mathbf{0}} l_{\mathbf{j}}\, [X(\mathbf{v}-\mathbf{j}) + X(\mathbf{v}+\mathbf{j})],\ \sum_{\mathbf{j}\geq\mathbf{0}} |l_{\mathbf{j}}| < \infty, \tag{2.3.10}$$

then $\{R(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ is reverse strictly stationary as well. For example, we may start from any sequence of independent and identically distributed random variables, since this is a reverse strictly stationary process. Then we may use (2.3.10) to produce another reverse strictly stationary process. The filters we apply on strictly or reverse strictly stationary processes do not necessarily have to be linear.

Consider the simple case of the reverse strictly stationary process $\{X(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ and the new process defined by the equation

$$R(\mathbf{v}) \equiv X(\mathbf{v}) + l\,[X(\mathbf{v}-\mathbf{j}) + X(\mathbf{v}+\mathbf{j})]$$

for some $\mathbf{j}^\tau \in \mathcal{Z}^d$ and $l \in \mathcal{R}$. Then for any $\mathbf{v}^\tau, \mathbf{j}_1^\tau \in \mathcal{Z}^d$, we may write the two-dimensional random vectors

$$\begin{bmatrix} R(\mathbf{v}) \\ R(\mathbf{v}-\mathbf{j}_1) \end{bmatrix} = \begin{bmatrix} 1 & 0 & l & 0 & 0 & l \\ 0 & l & 0 & 1 & l & 0 \end{bmatrix} \cdot \begin{bmatrix} X(\mathbf{v}) \\ X(\mathbf{v}-\mathbf{j}_1-\mathbf{j}) \\ X(\mathbf{v}-\mathbf{j}) \\ X(\mathbf{v}-\mathbf{j}_1) \\ X(\mathbf{v}-\mathbf{j}_1+\mathbf{j}) \\ X(\mathbf{v}+\mathbf{j}) \end{bmatrix}$$

and

$$\begin{bmatrix} R(\mathbf{v}) \\ R(\mathbf{v}+\mathbf{j}_1) \end{bmatrix} = \begin{bmatrix} 1 & 0 & l & 0 & 0 & l \\ 0 & l & 0 & 1 & l & 0 \end{bmatrix} \cdot \begin{bmatrix} X(\mathbf{v}) \\ X(\mathbf{v}+\mathbf{j}_1+\mathbf{j}) \\ X(\mathbf{v}+\mathbf{j}) \\ X(\mathbf{v}+\mathbf{j}_1) \\ X(\mathbf{v}+\mathbf{j}_1-\mathbf{j}) \\ X(\mathbf{v}-\mathbf{j}) \end{bmatrix}.$$

We can see immediately that the two vectors $[R(\mathbf{v}), R(\mathbf{v}-\mathbf{j}_1)]^\tau$ and $[R(\mathbf{v}), R(\mathbf{v}+\mathbf{j}_1)]^\tau$ have the same distribution, since the random vectors $[X(\mathbf{v}), X(\mathbf{v}-\mathbf{j}_1-\mathbf{j}), X(\mathbf{v}-\mathbf{j}), X(\mathbf{v}-\mathbf{j}_1), X(\mathbf{v}-\mathbf{j}_1+\mathbf{j}), X(\mathbf{v}+\mathbf{j})]^\tau$ and $[X(\mathbf{v}), X(\mathbf{v}+\mathbf{j}_1+\mathbf{j}), X(\mathbf{v}+\mathbf{j}), X(\mathbf{v}+\mathbf{j}_1), X(\mathbf{v}+\mathbf{j}_1-\mathbf{j}), X(\mathbf{v}-\mathbf{j})]^\tau$ have the same distribution too.

2.4 ARMA models

Definition 2.4. We define an ARMA process $\{Z(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ as

$$Z(\mathbf{v}) = \sum_{\mathbf{j}\in\mathcal{I}_p} b_\mathbf{j}\ Z(\mathbf{v}-\mathbf{j}) + \varepsilon(\mathbf{v}) + \sum_{\mathbf{j}\in\mathcal{J}_q} a_\mathbf{j}\ \varepsilon(\mathbf{v}-\mathbf{j}),\ \{\varepsilon(\mathbf{v})\} \sim WN(0,\sigma^2), \qquad (2.4.1)$$

where $\{b_\mathbf{j},\ \mathbf{j} \in \mathcal{I}_p\}$ and $\{a_\mathbf{j},\ \mathbf{j} \in \mathcal{J}_q\}$ are the auto-regressive and moving-average coefficients and both index sets $\mathcal{I}_p$ and $\mathcal{J}_q$ are contained in the set $\{\mathbf{j} > \mathbf{0}\}$. Both the sets $\mathcal{I}_p$ and $\mathcal{J}_q$ have finite cardinalities p and q, respectively.

For convenience, we introduce the vector back shift operator $\mathbf{B} \equiv [B_1, \cdots, B_d]$, such that

$$\mathbf{B}^{\mathbf{j}} Z(\mathbf{v}) \equiv Z(\mathbf{v} - \mathbf{j}),\ \mathbf{j}^\tau \in \mathcal{Z}^d.$$

This, of course, also implies that

$$\mathbf{B}^{-\mathbf{j}} Z(\mathbf{v}) = Z(\mathbf{v} + \mathbf{j}),\ \mathbf{j}^\tau \in \mathcal{Z}^d.$$

For $\mathbf{z} \equiv [z_1, \cdots, z_d]$, we write $\mathbf{z}^{\mathbf{j}} \equiv \prod_{k=1}^d z_k^{j_k}$. We can then define the polynomials

$$b(\mathbf{z}) \equiv 1 - \sum_{\mathbf{j} \in \mathcal{I}_p} b_{\mathbf{j}}\, \mathbf{z}^{\mathbf{j}} = 1 - \sum_{\mathbf{j} \in \mathcal{I}_p} b_{\mathbf{j}} \prod_{k=1}^d z_k^{j_k} \quad \text{and} \quad a(\mathbf{z}) \equiv 1 + \sum_{\mathbf{j} \in \mathcal{J}_q} a_{\mathbf{j}}\, \mathbf{z}^{\mathbf{j}} = 1 + \sum_{\mathbf{j} \in \mathcal{J}_q} a_{\mathbf{j}} \prod_{k=1}^d z_k^{j_k}. \tag{2.4.2}$$

Then model (2.4.1) can be written as

$$b(\mathbf{B}) Z(\mathbf{v}) = a(\mathbf{B}) \varepsilon(\mathbf{v}),\ \{\varepsilon(\mathbf{v})\} \sim WN(0, \sigma^2). \tag{2.4.3}$$

For this, we have assumed that $b(\mathbf{z})$ and $a(\mathbf{z})$ do not have common factors although they may still have common roots.

The process $\{Z(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ defined in (2.4.1) is *causal* if it admits a purely MA representation

$$Z(\mathbf{v}) = \varepsilon(\mathbf{v}) + \sum_{\mathbf{j} > \mathbf{0}} \psi_{\mathbf{j}}\, \varepsilon(\mathbf{v} - \mathbf{j}) \tag{2.4.4}$$

where $\sum_{\mathbf{j}>\mathbf{0}} |\psi_{\mathbf{j}}| < \infty$. A causal $\{Z(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}\}$ is always (weakly) stationary with mean 0 and the auto-covariance function

$$\gamma(\mathbf{j}) = E\{Z(\mathbf{v} + \mathbf{j}) Z(\mathbf{v})\} = \begin{cases} \sigma^2 \left\{\psi_{\mathbf{j}} + \sum_{\mathbf{i}>\mathbf{0}} \psi_{\mathbf{i}} \psi_{\mathbf{i}+\mathbf{j}}\right\}, & \mathbf{j} > \mathbf{0} \\ \sigma^2 \left\{1 + \sum_{\mathbf{i}>\mathbf{0}} \psi_{\mathbf{i}}^2\right\}, & \mathbf{j} = \mathbf{0} \\ \gamma(-\mathbf{j}), & \mathbf{j} < \mathbf{0} \end{cases} \tag{2.4.5}$$

The lemma below presents a sufficient condition for the causality.

Lemma 2.2. The process defined in (2.4.1) is causal if

$$b(0, \cdots, 0, 0, z_d) \neq 0 \text{ for all } |z_d| \leq 1 \text{ and } b(0, \cdots, 0, z_{d-1}, z_d) \neq 0 \text{ for all } |z_{d-1}| \leq 1$$
$$\text{and } |z_d| = 1 \text{ and } \cdots \text{ and } b(\mathbf{z}) \neq 0 \text{ for all } |z_1| \leq 1 \text{ and } |z_k| = 1,\ k = 2, \cdots, d, \tag{2.4.6}$$

where $z_1, \cdots, z_d \in \mathcal{C}$, *i.e.* the complex number space. Furthermore, condition (2.4.6) implies that the coefficients $\{\psi_{\mathbf{j}}\}$ defined in (2.4.4) decay at an exponential rate, and in particular

$$|\psi_{\mathbf{j}}| \leq C \alpha^{j_1 + \sum_{k=2}^d |j_k|}, \quad \text{for all } j_1 \geq 0, \tag{2.4.7}$$

where $\alpha \in (0,1)$, $C > 0$ are constants.

For the proof of the lemma, see Anderson and Jury (1974). Otherwise, we follow the same argument as Yao and Brockwell (2006) and the inequality (2.4.7) follows from the simple argument as follows. Let $\psi(\mathbf{z}) = 1 + \sum_{\mathbf{j}>\mathbf{0}} \psi_{\mathbf{j}}\ \mathbf{z}^{\mathbf{j}}$, where the coefficients $\psi_{\mathbf{j}}$ are given in (2.4.4). Then $\psi(\mathbf{z}) = a(\mathbf{z})/b(\mathbf{z})$. Due to the continuity of $b(\cdot)$, $b(\mathbf{z}) \neq 0$ for all $\mathbf{z} \in S_\delta \equiv \{[z_1, \cdots, z_d] : 1-\delta < |z_k| < 1+\delta,\ k = 1, \cdots, d\}$ under the causality condition, where $\delta > 0$ is a constant. Thus, $\psi(\cdot)$ is bounded on S_δ, *i.e.* $|\sum_{\mathbf{j}>\mathbf{0}} \psi_{\mathbf{j}}\ \mathbf{z}^{\mathbf{j}}| < \infty$ for any $\mathbf{z} \in S_\delta$ and, so $\psi_{\mathbf{j}}\ \alpha^{-j_1 - \sum_{k=2}^d |j_k|} \to 0$ as $\max_{k=2,\cdots,d}\{j_1, |j_k|\} \to \infty$, where $\alpha \in (0,1)$ is a constant.

Remark 2.2. (i) Inequality (2.4.7) also holds if we replace $\psi_{\mathbf{j}}$ by its derivative with respect to $b_{\mathbf{j}}$, $\mathbf{j} \in \mathcal{I}_p$, or $a_{\mathbf{j}}$, $\mathbf{j} \in \mathcal{J}_q$, under the condition (2.4.6). This can be justified via taking derivative on both sides of equation $\psi(\mathbf{z}) = a(\mathbf{z})/b(\mathbf{z})$, followed by the same argument as above.

(ii) The same condition guarantees that the auto-covariance function $\gamma(\cdot)$ decays at an exponential rate, *i.e.* $\gamma(\mathbf{j}) = O(\alpha^{j_1 + \sum_{k=2}^d |j_k|})$, where $\alpha \in (0,1)$ is a constant.

(iii) A partial derivative of $\gamma(\cdot)$ with respect to any of the $(p+q)$ parameters also decays at an exponential rate. This may be seen through combining (i) and the argument in (ii) together.

(iv) Condition (2.4.6) is not necessary for the causality when $d = 2, 3, \cdots$.

The process $\{Z(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}\}$ is *invertible* if it admits a purely AR representation

$$Z(\mathbf{v}) = \varepsilon(\mathbf{v}) + \sum_{\mathbf{j}>\mathbf{0}} \varphi_{\mathbf{j}}\ Z(\mathbf{v} - \mathbf{j}) \tag{2.4.8}$$

where $\sum_{\mathbf{j}>\mathbf{0}} |\varphi_{\mathbf{j}}| < \infty$. Like in Lemma 2.2, one can write down a sufficient condition for the invertibility of an ARMA process.

Lemma 2.3. The process defined in (2.4.1) is invertible if

$$\begin{aligned} &a(0, \cdots, 0, 0, z_d) \neq 0 \text{ for all } |z_d| \leq 1 \text{ and } a(0, \cdots, 0, z_{d-1}, z_d) \neq 0 \text{ for all } |z_{d-1}| \leq 1 \\ &\text{and } |z_d| = 1 \text{ and } \cdots \text{ and } a(\mathbf{z}) \neq 0 \text{ for all } |z_1| \leq 1 \text{ and } |z_k| = 1,\ k = 2, \cdots, d, \end{aligned} \tag{2.4.9}$$

where $z_1, \cdots, z_d \in \mathcal{C}$. Under this condition the coefficients $\{\varphi_{\mathbf{j}}\}$ and their partial derivatives (with respect to all parameters) decay at an exponential rate like before.

The spectral density of $\{Z(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ as defined in (2.3.3) is of the form

$$g(\omega) = \frac{\sigma^2}{(2\pi)^d} \left| \frac{a(e^{i\omega})}{b(e^{i\omega})} \right|^2,\ \omega^\tau \in [-\pi, \pi]^d. \tag{2.4.10}$$

If $\{Z(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ is causal and invertible, the spectral density $g(\cdot)$ is bounded away from ∞ and 0, respectively (Guyon, 1982). Note that the condition that $g(\cdot)$ is bounded away from ∞ and 0 is equivalent to the condition that $a(\mathbf{z})\ b(\mathbf{z}) \neq 0$ for all $z_1, \cdots, z_d \in \mathcal{C}$ with $|z_1| = |z_2| = \cdots = |z_d| = 1$. Under this condition $\{Z(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ is a (weakly) stationary process.

Remark 2.3. At this point, we may see how bounding the spectral density of any (weakly) stationary process away from ∞ and 0 may be very useful to make conclusions on the variance matrix of a set of random variables from this process. We consider any set $\mathcal{S} \subset \mathcal{Z}^d$ with N different elements, where N is a finite number, and the random variables $\{X(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{S}\}$ from a (weakly) stationary process with a bounded spectral density. Then for the $(N \times 1)$ random vector $\mathbf{X}$ with elements the random variables at any order, the variance matrix $\mathrm{Var}\{\mathbf{X}\}$ has all its eigenvalues also bounded away from ∞ and 0. For the case that $d = 1$ and a set of consecutive observations, the proof has been given by Proposition 4.5.3 of Brockwell and Davis (1991). For the case that $d = 2$ when the observations lie on a rectangle, we can find a similar proof in the paper by Yao and Brockwell (2006). For the general case of d number of dimensions and for any $\mathbf{v}^\tau, \mathbf{v}^{*\tau} \in \mathcal{Z}^d$ and $\mathbf{j} = \mathbf{v} - \mathbf{v}^* = [j_1, \cdots, j_d]$, it holds that $j_k \in \mathcal{Z}$ for all $k = 1, \cdots, d$. Then it is

$$\int_{[-\pi,\pi]^d} e^{i\sum_{k=1}^d \omega_k j_k} d\omega_1 \cdots d\omega_d = \prod_{k=1}^d \int_{-\pi}^{\pi} e^{i\omega_k j_k} d\omega_k = 0 \tag{2.4.11}$$

if and only if at least one $k = 1, \cdots, d$, is such that $j_k \neq 0$, or $\mathbf{v} \neq \mathbf{v}^*$. As a result, one can follow the same sequel as Proposition 4.5.3 of Brockwell and Davis and prove that all the eigenvalues are bounded, even though the set $\mathcal{S}$ may not have a specific structure on $\mathcal{Z}^d$.

2.4.1 Auto-Regressions and Moving-Averages

Two special cases of ARMA processes are studied here, the cases of the auto-regression and the moving-average. Later, in Chapter 3 we will be dealing again with ARMA processes. We will attempt then to describe the problems of estimation for the parameters of an ARMA model and to solve them. For this, we will need a profound understanding

of the much simpler world of the auto-regression and the moving-average, as they both have desirable properties that can be used to solve these problems. Thus, we will need to look for auto-regressions and moving-averages in the ARMA, rather than attack it directly.

Moreover in this section, we are interested in demonstrating various ways in which an auto-regression and a moving-average might be linked. Apart from the obvious connection that they may refer to the same polynomial, we are also interested in relating two such processes via the same white noise sequence of random variables. We then come up with results such as the general Yule-Walker equations, derived in the last part of this section, which involve the auto-covariance functions of an auto-regression and a moving-average with the same polynomial. Those equations have been used as the theoretical prototypes that should be imitated by data quantities to provide method of moments estimators in the next chapter.

For $\mathbf{z} = [z_1, \cdots, z_d]$ and $\mathbf{0} < \mathbf{j}_1 < \mathbf{j}_2 < \cdots < \mathbf{j}_q$, we define the polynomial

$$\theta(\mathbf{z}) \equiv 1 + \theta_{\mathbf{j}_1}\, \mathbf{z}^{\mathbf{j}_1} + \cdots + \theta_{\mathbf{j}_q}\, \mathbf{z}^{\mathbf{j}_q}, \tag{2.4.12}$$

where the coefficients $\theta_{\mathbf{j}_1}, \cdots, \theta_{\mathbf{j}_q}$, are such that we can write

$$\theta(\mathbf{z})^{-1} \equiv 1 + \sum_{\mathbf{j}>\mathbf{0}} \Theta_{\mathbf{j}}\, \mathbf{z}^{\mathbf{j}}, \ \sum_{\mathbf{j}>\mathbf{0}} |\Theta_{\mathbf{j}}| < \infty. \tag{2.4.13}$$

We first consider a white noise sequence $\{e(\mathbf{v})\} \sim WN(0,1)$ which generates the moving-average $\{Y(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$, such that

$$Y(\mathbf{v}) \equiv \theta(\mathbf{B})e(\mathbf{v}) \tag{2.4.14}$$

and the auto-regression $\{X(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$, such that

$$\theta(\mathbf{B}^{-1})X(\mathbf{v}) \equiv e(\mathbf{v}). \tag{2.4.15}$$

If we define the polynomial

$$\gamma(\mathbf{z}) \equiv \theta(\mathbf{z})\ \theta(\mathbf{z}^{-1}) \equiv \sum_{\mathbf{j}^\tau \in \mathcal{F}} \gamma(\mathbf{j})\ \mathbf{z}^{\mathbf{j}}, \tag{2.4.16}$$

then from (2.4.14),(2.4.15) and (2.4.16), we can see immediately that

$$Y(\mathbf{v}) = \gamma(\mathbf{B})X(\mathbf{v}). \tag{2.4.17}$$

The set $\mathcal{F} \subset \mathcal{Z}^d$ in (2.4.16) has finite cardinality and is such that $\mathbf{0}^\tau \in \mathcal{F}$ and that if $\mathbf{j}^\tau \in \mathcal{F}$ then $-\mathbf{j}^\tau \in \mathcal{F}$, as well. More specifically, we can write here

$$\mathcal{F} \equiv \{\mathbf{j}^\tau : \mathbf{j} = \mathbf{j}_n,\ -\mathbf{j}_n,\ \mathbf{j}_n - \mathbf{j}_m,\ n, m = 1, \cdots, q\}.$$

Later in Chapters 3 and 4, we will denote by $\mathcal{F}$ other finite sets of lags that include the zero lag and they are symmetric, but we will not know their exact elements then.

If we define the polynomial

$$c(\mathbf{z}) \equiv \gamma(\mathbf{z})^{-1} = \theta(\mathbf{z})^{-1}\ \theta(\mathbf{z}^{-1})^{-1} \equiv \sum_{\mathbf{j}^\tau \in \mathcal{Z}^d} c(\mathbf{j})\ \mathbf{z}^{\mathbf{j}},\ \sum_{\mathbf{j}^\tau \in \mathcal{Z}^d} |c(\mathbf{j})| < \infty, \tag{2.4.18}$$

we can re-write (2.4.17) as

$$X(\mathbf{v}) = c(\mathbf{B})Y(\mathbf{v}). \tag{2.4.19}$$

As a counterpart to the process defined by

$$e(\mathbf{v}) = X(\mathbf{v}) + \theta_{\mathbf{j}_1} X(\mathbf{v} + \mathbf{j}_1) + \cdots + \theta_{\mathbf{j}_q} X(\mathbf{v} + \mathbf{j}_q),$$

we define

$$u(\mathbf{v}) \equiv X(\mathbf{v}) + \theta_{\mathbf{j}_1} X(\mathbf{v} - \mathbf{j}_1) + \cdots + \theta_{\mathbf{j}_q} X(\mathbf{v} - \mathbf{j}_q) = \theta(\mathbf{B})X(\mathbf{v}). \tag{2.4.20}$$

Then

$$\{u(\mathbf{v})\} \sim WN(0, 1),$$

since it has exactly the same second-order properties as $\{e(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$. Indeed for any

$\mathbf{j}^\tau \in \mathcal{Z}^d$, it holds that

$$\begin{aligned}
E\{u(\mathbf{v})u(\mathbf{v}-\mathbf{j})\} &= E\{(X(\mathbf{v})+\sum_{n=1}^{q}\theta_{\mathbf{j}_n}X(\mathbf{v}-\mathbf{j}_n))(X(\mathbf{v}-\mathbf{j})+\sum_{m=1}^{q}\theta_{\mathbf{j}_m}X(\mathbf{v}-\mathbf{j}-\mathbf{j}_m))\} \\
&= E\{X(\mathbf{v})X(\mathbf{v}-\mathbf{j})\}+\sum_{n=1}^{q}\theta_{\mathbf{j}_n}E\{X(\mathbf{v}-\mathbf{j}_n)X(\mathbf{v}-\mathbf{j})\} \\
&+ \sum_{m=1}^{q}\theta_{\mathbf{j}_m}E\{X(\mathbf{v})X(\mathbf{v}-\mathbf{j}-\mathbf{j}_m))\} \\
&+ \sum_{n=1}^{q}\sum_{m=1}^{q}\theta_{\mathbf{j}_n}\theta_{\mathbf{j}_m}E\{X(\mathbf{v}-\mathbf{j}_n)X(\mathbf{v}-\mathbf{j}-\mathbf{j}_m)\} \\
&= E\{X(\mathbf{v})X(\mathbf{v}+\mathbf{j})\}+\sum_{n=1}^{q}\theta_{\mathbf{j}_n}E\{X(\mathbf{v}+\mathbf{j}_n)X(\mathbf{v}+\mathbf{j})\} \\
&+ \sum_{m=1}^{q}\theta_{\mathbf{j}_m}E\{X(\mathbf{v})X(\mathbf{v}+\mathbf{j}+\mathbf{j}_m)\} \\
&+ \sum_{n=1}^{q}\sum_{m=1}^{q}\theta_{\mathbf{j}_n}\theta_{\mathbf{j}_m}E\{X(\mathbf{v}+\mathbf{j}_n)X(\mathbf{v}+\mathbf{j}+\mathbf{j}_m)\} \\
&= E\{(X(\mathbf{v})+\sum_{n=1}^{q}\theta_{\mathbf{j}_n}X(\mathbf{v}+\mathbf{j}_n))(X(\mathbf{v}+\mathbf{j})+\sum_{m=1}^{q}\theta_{\mathbf{j}_m}X(\mathbf{v}+\mathbf{j}+\mathbf{j}_m))\} \\
&= E\{e(\mathbf{v})e(\mathbf{v}+\mathbf{j})\}.
\end{aligned}$$

From (2.4.16), (2.4.17) and (2.4.20), we can see that

$$Y(\mathbf{v}) = \theta(\mathbf{B}^{-1})u(\mathbf{v}). \tag{2.4.21}$$

The equations (2.4.14), (2.4.15), (2.4.20) and (2.4.21) show that

$$E\{X(\mathbf{v}+\mathbf{j})Y(\mathbf{v})\} = \begin{cases} 1, & \mathbf{j}=\mathbf{0} \\ 0, & \mathbf{j}\neq\mathbf{0} \end{cases} \tag{2.4.22}$$

and $Y(\mathbf{v})$ and $X(\mathbf{v}+\mathbf{j})$ are uncorrelated for any $\mathbf{j}\neq\mathbf{0}$.

On the other hand, if we write

$$X(\mathbf{v}) = u(\mathbf{v}) + \sum_{\mathbf{j}>\mathbf{0}}\Theta_{\mathbf{j}}\, u(\mathbf{v}-\mathbf{j}),$$

we can see immediately that

$$\Theta_{\mathbf{j}} = E\{X(\mathbf{v})u(\mathbf{v}-\mathbf{j})\},\ \mathbf{j}>\mathbf{0}. \tag{2.4.23}$$

If we multiply (2.4.20) by $u(\mathbf{v}-\mathbf{j}), \mathbf{j}>\mathbf{0}$, and then take expected values, using (2.4.23) we can derive

$$\Theta_{\mathbf{j}} + \theta_{\mathbf{j}_1}\Theta_{\mathbf{j}-\mathbf{j}_1} + \cdots + \theta_{\mathbf{j}_q}\Theta_{\mathbf{j}-\mathbf{j}_q} = 0, \tag{2.4.24}$$

where we consider $\Theta_0 = 1$ and $\Theta_{\mathbf{j}} = 0$, $\mathbf{j} < \mathbf{0}$, where necessary.

So far, we have seen how

$$\{e(\mathbf{v})\} \sim WN(0,1)$$

also implies that

$$\{u(\mathbf{v})\} \sim WN(0,1).$$

Moreover, if we can assume that the original sequence is Gaussian

$$\{e(\mathbf{v})\} \sim N(0,1),$$

then all the linear filters applied to it afterwards generate Gaussian sequences of random variables and

$$\{u(\mathbf{v})\} \sim N(0,1).$$

An interesting question would be what happens in the case that

$$\{e(\mathbf{v})\} \sim IID(0,1). \tag{2.4.25}$$

When (2.4.25) holds, all the sequences of random variables defined afterwards are strictly stationary. As a result, $\{u(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$, defined in (2.4.20), is a strictly stationary and white noise process.

Proposition 2.5. For $\{e(\mathbf{v})\} \sim WN(0,1)$, we consider the reverse strictly stationary process $\{X(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ defined by (2.4.15). We also define $\{u(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ by (2.4.20).

(i) Then for any $\mathbf{v}^\tau \in \mathcal{Z}^d$, the distribution of $u(\mathbf{v})$ is the same as the distribution of $e(\mathbf{v})$.

(ii) Further, if $\{e(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ are identically distributed, so are $\{u(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$.

(iii) For any $\mathbf{v}^\tau, \mathbf{j}^\tau \in \mathcal{Z}^d$ and $\mathbf{j} \neq \mathbf{0}$, if it holds that $e(\mathbf{v})$ and $e(\mathbf{v}+\mathbf{j})$ are two independent random variables, then $u(\mathbf{v})$ and $u(\mathbf{v}-\mathbf{j})$ are also two independent random variables.

(iv) If $\{e(\mathbf{v})\} \sim IID(0,1)$, then it holds that

$$\{u(\mathbf{v})\} \sim IID(0,1). \tag{2.4.26}$$

Proof. For (i), since $\{X(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ is reverse strictly stationary, the distribution of $[X(\mathbf{v}), X(\mathbf{v}+\mathbf{j}_1), \cdots, X(\mathbf{v}+\mathbf{j}_q)]^\tau$ is the same as the distribution of $[X(\mathbf{v}), X(\mathbf{v}-\mathbf{j}_1), \cdots, X(\mathbf{v}-\mathbf{j}_q)]^\tau$. We can conclude then that the distribution of $u(\mathbf{v})$ is the same as the distribution of $e(\mathbf{v})$. The proof for (ii) follows directly.

For (iii) and without loss of generality, we will consider continuous random variables and will denote by f a generic probability density function. It holds for any $\mathbf{j} \neq \mathbf{0}$ that $e(\mathbf{v})$ and $e(\mathbf{v}+\mathbf{j})$ are two independent random variables. As a result,

$$f_{e(\mathbf{v}),e(\mathbf{v}+\mathbf{j})}(w_1, w_2) = f_{e(\mathbf{v})}(w_1) \cdot f_{e(\mathbf{v}+\mathbf{j})}(w_2),$$

where $f_{e(\mathbf{v})}(w)$ and $f_{e(\mathbf{v}+\mathbf{j})}(w)$ might not be the same probability functions, since we have not assumed that $e(\mathbf{v})$ and $e(\mathbf{v}+\mathbf{j})$ are identically distributed here. We may re-write

$$\begin{aligned}
& \int_{R^{2q}} f_{X(\mathbf{v}),\cdots,X(\mathbf{v}+\mathbf{j}_q),X(\mathbf{v}+\mathbf{j}),\cdots,X(\mathbf{v}+\mathbf{j}+\mathbf{j}_q)}(x, x_1, \cdots, x_q, x^*, x_1^*, \cdots, x_q^*)\ dx_1 \cdots dx_q^* \\
= & \int_{R^q} f_{X(\mathbf{v}),\cdots,X(\mathbf{v}+\mathbf{j}_q)}(x, \cdots, x_q)\ dx_1 \cdots dx_q \cdot \\
& \int_{R^q} f_{X(\mathbf{v}+\mathbf{j}),\cdots,X(\mathbf{v}+\mathbf{j}+\mathbf{j}_q)}(x^*, \cdots, x_q^*)\ dx_1^* \cdots dx_q^*, \qquad (2.4.27)
\end{aligned}$$

where the integrations take place under the restrictions

$$x + \theta_1\ x_1 + \cdots + \theta_q\ x_q = w_1$$

and

$$x^* + \theta_1\ x_1^* + \cdots + \theta_q\ x_q^* = w_2,$$

according to (2.4.15). We may also write

$$\begin{aligned}
& f_{u(\mathbf{v}),u(\mathbf{v}-\mathbf{j})}(w_1, w_2) \\
= & \int_{R^{2q}} f_{X(\mathbf{v}),\cdots,X(\mathbf{v}-\mathbf{j}_q),X(\mathbf{v}-\mathbf{j}),\cdots,X(\mathbf{v}-\mathbf{j}-\mathbf{j}_q)}(x, x_1, \cdots, x_q, x^*, x_1^*, \cdots, x_q^*)\ dx_1 \cdots dx_q^* \\
= & \int_{R^{2q}} f_{X(\mathbf{v}),\cdots,X(\mathbf{v}+\mathbf{j}_q),X(\mathbf{v}+\mathbf{j}),\cdots,X(\mathbf{v}+\mathbf{j}+\mathbf{j}_q)}(x, x_1, \cdots, x_q, x^*, x_1^*, \cdots, x_q^*)\ dx_1 \cdots dx_q^* \\
= & \int_{R^q} f_{X(\mathbf{v}),\cdots,X(\mathbf{v}+\mathbf{j}_q)}(x, \cdots, x_q)\ dx_1 \cdots dx_q \cdot \\
& \int_{R^q} f_{X(\mathbf{v}+\mathbf{j}),\cdots,X(\mathbf{v}+\mathbf{j}+\mathbf{j}_q)}(x^*, \cdots, x_q^*)\ dx_1^* \cdots dx_q^* \\
= & \int_{R^q} f_{X(\mathbf{v}),\cdots,X(\mathbf{v}-\mathbf{j}_q)}(x, \cdots, x_q)\ dx_1 \cdots dx_q \cdot \\
& \int_{R^q} f_{X(\mathbf{v}-\mathbf{j}),\cdots,X(\mathbf{v}-\mathbf{j}-\mathbf{j}_q)}(x^*, \cdots, x_q^*)\ dx_1^* \cdots dx_q^* \\
= & f_{u(\mathbf{v})}(w_1) \cdot f_{u(\mathbf{v}-\mathbf{j})}(w_2), \qquad (2.4.28)
\end{aligned}$$

which implies that $u(\mathbf{v}), u(\mathbf{v}-\mathbf{j})$ are independent as well. Finally, (iv) follows directly from (ii) and (iii). ■

The following lemma generalizes (2.4.22) and turns the notion of uncorrelated random variables $Y(\mathbf{v})$ and $X(\mathbf{v}-\mathbf{j})$, $\mathbf{j} \neq \mathbf{0}$, to one of independent random variables.

Lemma 2.4. If $\{e(\mathbf{v})\} \sim IID(0,1)$ and $\{u(\mathbf{v})\} \sim IID(0,1)$, then it holds that $X(\mathbf{v}-\mathbf{j})$ and $Y(\mathbf{v})$ are two independent random variables for any $\mathbf{j} \neq \mathbf{0}$.

Proof. If it holds that $\{e(\mathbf{v})\} \sim IID(0,1)$, then the random variables $X(\mathbf{v}+\mathbf{j})$ and $Y(\mathbf{v})$ are independent for any $\mathbf{j} > \mathbf{0}$, since $X(\mathbf{v})$ is a linear function of $e(\mathbf{v}+\mathbf{j}), \mathbf{j} \geq \mathbf{0}$, and $Y(\mathbf{v})$ is a linear function of $e(\mathbf{v}-\mathbf{j}), \mathbf{j} \geq \mathbf{0}$.

On the other hand, if it holds that $\{u(\mathbf{v})\} \sim IID(0,1)$ too, then the random variables $X(\mathbf{v}-\mathbf{j})$ and $Y(\mathbf{v})$ are independent for any $\mathbf{j} > \mathbf{0}$. This comes from the fact that $X(\mathbf{v})$ is now a linear function of $u(\mathbf{v}-\mathbf{j}), \mathbf{j} \geq \mathbf{0}$, and $Y(\mathbf{v})$ is a linear function of $u(\mathbf{v}+\mathbf{j}), \mathbf{j} \geq \mathbf{0}$.

■

Proposition 2.6. We consider the zero-mean and strictly stationary process $\{Y(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ with absolutely summable auto-covariance function

$$\gamma^*(\mathbf{j}) \equiv E\{Y(\mathbf{v})Y(\mathbf{v}-\mathbf{j})\} \equiv \sigma^2\ \gamma(\mathbf{j}),\ \mathbf{j}^\tau \in \mathcal{Z}^d, \tag{2.4.29}$$

where σ^2 is given from Kolmogorov's formula. Then we define the polynomials

$$\gamma(\mathbf{z}) \equiv \sum_{\mathbf{j}^\tau \in \mathcal{Z}^d} \gamma(\mathbf{j})\ \mathbf{z}^{\mathbf{j}}, \quad \sum_{\mathbf{j}^\tau \in \mathcal{Z}^d} |\gamma(\mathbf{j})| < \infty, \tag{2.4.30}$$

$$\gamma^*(\mathbf{z}) \equiv \sum_{\mathbf{j}^\tau \in \mathcal{Z}^d} \gamma^*(\mathbf{j})\ \mathbf{z}^{\mathbf{j}}, \quad \sum_{\mathbf{j}^\tau \in \mathcal{Z}^d} |\gamma^*(\mathbf{j})| < \infty, \tag{2.4.31}$$

and consider

$$\gamma(\mathbf{z}) \equiv \theta(\mathbf{z})\ \theta(\mathbf{z}^{-1}), \tag{2.4.32}$$

such that

$$\theta(\mathbf{z}) \equiv 1 + \sum_{\mathbf{j}>\mathbf{0}} \theta_{\mathbf{j}}\ \mathbf{z}^{\mathbf{j}},\ \sum_{\mathbf{j}>\mathbf{0}} |\theta_{\mathbf{j}}| < \infty, \tag{2.4.33}$$

and

$$\theta(\mathbf{z})^{-1} \equiv 1 + \sum_{\mathbf{j}>\mathbf{0}} \Theta_{\mathbf{j}}\ \mathbf{z}^{\mathbf{j}},\ \sum_{\mathbf{j}>\mathbf{0}} |\Theta_{\mathbf{j}}| < \infty. \tag{2.4.34}$$

For the polynomial

$$c^*(\mathbf{z}) \equiv \gamma^*(\mathbf{z})^{-1}, \tag{2.4.35}$$

we define the new process

$$X^*(\mathbf{v}) \equiv c^*(\mathbf{z})Y(\mathbf{v}). \tag{2.4.36}$$

If $|E\{Y(\mathbf{v})^3\}| < \infty$, $1 + \sum_{\mathbf{j}>\mathbf{0}} \theta_{\mathbf{j}}^3 \neq 0$, and $X^*(\mathbf{v}-\mathbf{j})$ and $Y(\mathbf{v})$ are two independent random variables for any $\mathbf{j} \neq \mathbf{0}$, then for any $\mathbf{v}^\tau \in \mathcal{Z}^d$, any $\mathbf{i} \geq \mathbf{0}$ and any two locations

$\mathbf{v}_1 \leq \mathbf{v}_2$, it holds that

$$\mathrm{Cov}\{Y(\mathbf{v})Y(\mathbf{v}+\mathbf{i}), X^*(\mathbf{v}_1)X^*(\mathbf{v}_2)\} = \begin{cases} 0, & \mathbf{i} > \mathbf{0},\ \mathbf{v}_1 \neq \mathbf{v} \text{ or } \mathbf{v}_2 \neq \mathbf{v}+\mathbf{i} \\ 1, & \mathbf{i} > \mathbf{0},\ \mathbf{v}_1 = \mathbf{v},\ \mathbf{v}_2 = \mathbf{v}+\mathbf{i} \\ 0, & \mathbf{i} = \mathbf{0},\ \mathbf{v}_1 \text{ or } \mathbf{v}_2 \neq \mathbf{v} \end{cases} . \quad (2.4.37)$$

Proof. We first define the polynomial

$$c(\mathbf{z}) \equiv \gamma(\mathbf{z})^{-1} = \frac{1}{\theta(\mathbf{z})\ \theta(\mathbf{z}^{-1})} = \sigma^2 c^*(\mathbf{z}) \quad (2.4.38)$$

and then the process

$$X(\mathbf{v}) \equiv c(\mathbf{B})Y(\mathbf{v}) = \sigma^2\ X^*(\mathbf{v}). \quad (2.4.39)$$

We also define the strictly stationary processes

$$\varepsilon(\mathbf{v}) \equiv \theta(\mathbf{B})^{-1}Y(\mathbf{v}) = \theta(\mathbf{B}^{-1})X(\mathbf{v}) \quad (2.4.40)$$

and

$$u(\mathbf{v}) \equiv \theta(\mathbf{B}^{-1})^{-1}Y(\mathbf{v}) = \theta(\mathbf{B})X(\mathbf{v}). \quad (2.4.41)$$

It holds that $\varepsilon(\mathbf{v})$ is independent of $\varepsilon(\mathbf{v}+\mathbf{j})$ for any $\mathbf{j} > \mathbf{0}$, since the former is a linear combination of $Y(\mathbf{v}-\mathbf{j}^*), \mathbf{j}^* \geq \mathbf{0}$, and the latter a linear combination of $X(\mathbf{v}+\mathbf{j}+\mathbf{j}^*), \mathbf{j}^* \geq \mathbf{0}$. A similar argument can be used for the two random variables $u(\mathbf{v})$ and $u(\mathbf{v}-\mathbf{j})$ for any $\mathbf{j} > \mathbf{0}$. Of course, the two processes $\{\varepsilon(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ and $\{u(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ are sequences of independent and identically distributed random variables, since we originated the proposition with a strictly stationary sequence $\{Y(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$. It also holds that

$$E\{\varepsilon(\mathbf{v})^2\} = E\{u(\mathbf{v})^2\} = \sigma^2,\ \mathbf{v}^\tau \in \mathcal{Z}^d. \quad (2.4.42)$$

We can also write that

$$\begin{aligned} E\{Y(\mathbf{v})^3\} &= [1 + \sum_{\mathbf{j}>\mathbf{0}} \theta_{\mathbf{j}}^3]\ E\{\varepsilon(\mathbf{v})^3\} + 3 \cdot E\{\varepsilon(\mathbf{v})^2\}E\{\varepsilon(\mathbf{v})\} \sum_{\mathbf{j}>\mathbf{0}} \theta_{\mathbf{j}} \\ &+ 3 \cdot E\{\varepsilon(\mathbf{v})\}E\{\varepsilon(\mathbf{v})^2\} \sum_{\mathbf{j}>\mathbf{0}} \theta_{\mathbf{j}}^2 \\ &+ 3 \cdot E\{\varepsilon(\mathbf{v})\}^3 \sum_{\substack{\mathbf{j}_1,\mathbf{j}_2>\mathbf{0}, \\ \mathbf{j}_1\neq\mathbf{j}_2}} \theta_{\mathbf{j}_1}\theta_{\mathbf{j}_2} + 3 \cdot E\{\varepsilon(\mathbf{v})^2\}E\{\varepsilon(\mathbf{v})\} \sum_{\substack{\mathbf{j}_1,\mathbf{j}_2>\mathbf{0}, \\ \mathbf{j}_1\neq\mathbf{j}_2}} \theta_{\mathbf{j}_1}^2\theta_{\mathbf{j}_2} \\ &+ 3 \cdot E\{\varepsilon(\mathbf{v})\}^3 \sum_{\substack{\mathbf{j}_1,\mathbf{j}_2,\mathbf{j}_3>\mathbf{0}, \\ \mathbf{j}_1\neq\mathbf{j}_2,\ \mathbf{j}_1\neq\mathbf{j}_3,\ \mathbf{j}_2\neq\mathbf{j}_3}} \theta_{\mathbf{j}_1}\theta_{\mathbf{j}_2}\theta_{\mathbf{j}_3} \\ &= [1 + \sum_{\mathbf{j}>\mathbf{0}} \theta_{\mathbf{j}}^3]\ E\{\varepsilon(\mathbf{v})^3\}. \end{aligned}$$

As a result, if $|E\{Y(\mathbf{v})^3\}| < \infty$ and $1 + \sum_{\mathbf{j}>\mathbf{0}} \theta_{\mathbf{j}}^3 \neq 0$, then it holds that

$$|E\{\varepsilon(\mathbf{v})^3\}| < \infty$$

and, similarly, it holds that

$$|E\{u(\mathbf{v})^3\}| < \infty.$$

The finite third moments of the noise sequences will be useful next. To prove (2.4.37), we will start with the cases when $\mathbf{i} > \mathbf{0}$. We separate into the following cases:

(i) If $\mathbf{v}_1 \leq \mathbf{v}_2 < \mathbf{v} < \mathbf{v}+\mathbf{i}$, then it holds that the random variables $X(\mathbf{v}_1)X(\mathbf{v}_2)$ and $Y(\mathbf{v})Y(\mathbf{v}+\mathbf{i})$ are independent.

(ii) If $\mathbf{v}_1 < \mathbf{v} = \mathbf{v}_2 < \mathbf{v}+\mathbf{i}$, then we may write

$$\begin{aligned}
& E\{X(\mathbf{v}_1)X(\mathbf{v})Y(\mathbf{v})Y(\mathbf{v}+\mathbf{i})\} = E\{E\{X(\mathbf{v}_1)X(\mathbf{v})Y(\mathbf{v})Y(\mathbf{v}+\mathbf{i})|u(\mathbf{v}_1-\mathbf{j}),\mathbf{j}\geq\mathbf{0}\}\} \\
= \; & E\{\{X(\mathbf{v}_1)[u(\mathbf{v}) + \sum_{\mathbf{0}<\mathbf{j}<\mathbf{v}-\mathbf{v}_1} \Theta_{\mathbf{j}}\, u(\mathbf{v}-\mathbf{j}) \\
+ \; & \sum_{\mathbf{j}\geq\mathbf{v}-\mathbf{v}_1} \Theta_{\mathbf{j}}\, u(\mathbf{v}-\mathbf{j})]Y(\mathbf{v})Y(\mathbf{v}+\mathbf{i})|u(\mathbf{v}_1-\mathbf{j}),\mathbf{j}\geq\mathbf{0}\}\} \\
= \; & E\{X(\mathbf{v}_1)[\sum_{\mathbf{j}\geq\mathbf{v}-\mathbf{v}_1} \Theta_{\mathbf{j}}\, u(\mathbf{v}-\mathbf{j})]\}E\{Y(\mathbf{v})Y(\mathbf{v}+\mathbf{i})\} \\
+ \; & E\{X(\mathbf{v}_1)\}\; E\{[u(\mathbf{v}) + \sum_{\mathbf{0}<\mathbf{j}<\mathbf{v}-\mathbf{v}_1} \Theta_{\mathbf{j}}\, u(\mathbf{v}-\mathbf{j})]Y(\mathbf{v})Y(\mathbf{v}+\mathbf{i})\} \\
= \; & E\{X(\mathbf{v}_1)X(\mathbf{v})\}\; E\{Y(\mathbf{v})Y(\mathbf{v}+\mathbf{i})\},
\end{aligned}$$

and we have used the fact that $|E\{u(\mathbf{v})^3\}| < \infty$ for the second term.

(iii) If $\mathbf{v}_1 < \mathbf{v} < \mathbf{v}_2 < \mathbf{v}+\mathbf{i}$, then we may use the same argument as in (i).

(iv) If $\mathbf{v}_1 < \mathbf{v} < \mathbf{v}_2 = \mathbf{v}+\mathbf{i}$, then we may write

$$\begin{aligned}
& E\{X(\mathbf{v}_1)X(\mathbf{v}+\mathbf{i})Y(\mathbf{v})Y(\mathbf{v}+\mathbf{i})\} \\
= \; & E\{E\{X(\mathbf{v}_1)X(\mathbf{v}+\mathbf{i})Y(\mathbf{v})Y(\mathbf{v}+\mathbf{i})|u(\mathbf{v}+\mathbf{i}+\mathbf{j}),\mathbf{j}\geq\mathbf{0}\}\} \\
= \; & E\{X(\mathbf{v}_1)[\sum_{\mathbf{j}>\mathbf{0}} \Theta_{\mathbf{j}}\, u(\mathbf{v}+\mathbf{i}-\mathbf{j})]\}\; E\{[\sum_{\mathbf{j}\geq\mathbf{i}} \theta_{\mathbf{j}}\, u(\mathbf{v}+\mathbf{j})]Y(\mathbf{v}+\mathbf{i})\} \\
+ \; & E\{X(\mathbf{v}_1)[\sum_{\mathbf{j}>\mathbf{0}} \Theta_{\mathbf{j}}\, u(\mathbf{v}+\mathbf{i}-\mathbf{j})][u(\mathbf{v}) + \sum_{\mathbf{0}<\mathbf{j}<\mathbf{i}} \theta_{\mathbf{j}}\, u(\mathbf{v}+\mathbf{j})]\}E\{Y(\mathbf{v}+\mathbf{i})\} \\
+ \; & E\{X(\mathbf{v}_1)\}\; E\{u(\mathbf{v}+\mathbf{i})[\sum_{\mathbf{j}\geq\mathbf{i}} \theta_{\mathbf{j}}\, u(\mathbf{v}+\mathbf{j})]Y(\mathbf{v}+\mathbf{i})\} \\
+ \; & E\{X(\mathbf{v}_1)[u(\mathbf{v}) + \sum_{\mathbf{0}<\mathbf{j}<\mathbf{i}} \theta_{\mathbf{j}}\, u(\mathbf{v}+\mathbf{j})]\}\; E\{u(\mathbf{v}+\mathbf{i})Y(\mathbf{v}+\mathbf{i})\} \\
= \; & E\{X(\mathbf{v}_1)X(\mathbf{v}+\mathbf{i})\}\; E\{Y(\mathbf{v})Y(\mathbf{v}+\mathbf{i})\},
\end{aligned}$$

and the finite third moment has been used for the two middle terms.

(v) If $\mathbf{v}_1 < \mathbf{v} < \mathbf{v}+\mathbf{i} < \mathbf{v}_2$, then we may use the same argument as in (i).

(vi) If $\mathbf{v}_1 = \mathbf{v}_2 = \mathbf{v} < \mathbf{v}+\mathbf{i}$, then we may write

$$\begin{aligned}
& E\{X(\mathbf{v})^2 Y(\mathbf{v}) Y(\mathbf{v}+\mathbf{i})\} = E\{E\{X(\mathbf{v})^2 Y(\mathbf{v}) Y(\mathbf{v}+\mathbf{i}) | u(\mathbf{v}+\mathbf{i}+\mathbf{j}), \mathbf{j} \geq \mathbf{0}\}\} \\
= \; & E\{X(\mathbf{v})^2 [u(\mathbf{v}) + \sum_{\mathbf{0}<\mathbf{j}<\mathbf{i}} \theta_{\mathbf{j}}\, u(\mathbf{v}+\mathbf{j})]\}\; E\{Y(\mathbf{v}+\mathbf{i})\} \\
+ \; & E\{X(\mathbf{v})^2\}\; E\{[\sum_{\mathbf{j}\geq\mathbf{i}} \theta_{\mathbf{j}}\, u(\mathbf{v}+\mathbf{j})]\; Y(\mathbf{v}+\mathbf{i})\} \\
= \; & E\{X(\mathbf{v})^2\}\; E\{Y(\mathbf{v}) Y(\mathbf{v}+\mathbf{i})\},
\end{aligned}$$

and the finite third moment has been used for the first term.

(vii) If $\mathbf{v}_1 = \mathbf{v} < \mathbf{v}_2 < \mathbf{v}+\mathbf{i}$, then we may write

$$\begin{aligned}
& E\{X(\mathbf{v}) X(\mathbf{v}_2) Y(\mathbf{v}) Y(\mathbf{v}+\mathbf{i})\} = E\{E\{X(\mathbf{v}) X(\mathbf{v}_2) Y(\mathbf{v}) Y(\mathbf{v}+\mathbf{i}) | u(\mathbf{v}-\mathbf{j}), \mathbf{j} \geq \mathbf{0}\}\} \\
= \; & E\{X(\mathbf{v})[\sum_{\mathbf{j}\geq\mathbf{v}_2-\mathbf{v}} \Theta_{\mathbf{j}}\, u(\mathbf{v}_2-\mathbf{j})]\}\; E\{[\sum_{\mathbf{j}>\mathbf{0}} \theta_{\mathbf{j}}\, u(\mathbf{v}+\mathbf{j})] Y(\mathbf{v}+\mathbf{i})\} \\
+ \; & E\{X(\mathbf{v})[\sum_{\mathbf{j}\geq\mathbf{v}_2-\mathbf{v}} \Theta_{\mathbf{j}}\, u(\mathbf{v}_2-\mathbf{j})] u(\mathbf{v})\}\; E\{Y(\mathbf{v}+\mathbf{i})\} \\
+ \; & E\{X(\mathbf{v})\}\; E\{[u(\mathbf{v}_2) + \sum_{\mathbf{0}<\mathbf{j}<\mathbf{v}_2-\mathbf{v}} \Theta_{\mathbf{j}}\, u(\mathbf{v}_2-\mathbf{j})][\sum_{\mathbf{j}>\mathbf{0}} \theta_{\mathbf{j}}\, u(\mathbf{v}+\mathbf{j})] Y(\mathbf{v}+\mathbf{i})\} \\
+ \; & E\{X(\mathbf{v}) u(\mathbf{v})\}\; E\{[u(\mathbf{v}_2) + \sum_{\mathbf{0}<\mathbf{j}<\mathbf{v}_2-\mathbf{v}} \Theta_{\mathbf{j}}\, u(\mathbf{v}_2-\mathbf{j})] Y(\mathbf{v}+\mathbf{i})\} \\
= \; & E\{X(\mathbf{v}) X(\mathbf{v}_2)\}\; E\{Y(\mathbf{v}) Y(\mathbf{v}+\mathbf{i})\},
\end{aligned}$$

and the finite third moment has been used for the two middle terms.

(viii) If $\mathbf{v}_1 = \mathbf{v} < \mathbf{v}_2 = \mathbf{v}+\mathbf{i}$, then we may write

$$\begin{aligned}
& E\{X(\mathbf{v}) X(\mathbf{v}+\mathbf{i}) Y(\mathbf{v}) Y(\mathbf{v}+\mathbf{i})\} \\
= \; & E\{\{X(\mathbf{v}) X(\mathbf{v}+\mathbf{i}) Y(\mathbf{v}) Y(\mathbf{v}+\mathbf{i}) | u(\mathbf{v}-\mathbf{j}), \mathbf{j} \geq \mathbf{0}\}\} \cdot \\
= \; & E\{X(\mathbf{v})[\sum_{\mathbf{j}\geq\mathbf{i}} \Theta_{\mathbf{j}}\, u(\mathbf{v}+\mathbf{i}-\mathbf{j})]\}\; E\{[\sum_{\mathbf{j}>\mathbf{0}} \theta_{\mathbf{j}}\, u(\mathbf{v}+\mathbf{j})] Y(\mathbf{v}+\mathbf{i})\} \\
+ \; & E\{X(\mathbf{v})\}\; E\{[u(\mathbf{v}+\mathbf{i}) + \sum_{\mathbf{0}<\mathbf{j}<\mathbf{i}} \Theta_{\mathbf{j}}\, u(\mathbf{v}+\mathbf{i}-\mathbf{j})][\sum_{\mathbf{j}>\mathbf{0}} \theta_{\mathbf{j}}\, u(\mathbf{v}+\mathbf{j})] Y(\mathbf{v}+\mathbf{i})\} \\
+ \; & E\{X(\mathbf{v})[\sum_{\mathbf{j}\geq\mathbf{i}} \Theta_{\mathbf{j}}\, u(\mathbf{v}+\mathbf{i}-\mathbf{j})] u(\mathbf{v})\}\; E\{Y(\mathbf{v}+\mathbf{i})\} \\
+ \; & E\{X(\mathbf{v}) u(\mathbf{v})\}\; E\{[u(\mathbf{v}+\mathbf{i}) + \sum_{\mathbf{0}<\mathbf{j}<\mathbf{i}} \Theta_{\mathbf{j}}\, u(\mathbf{v}+\mathbf{i}-\mathbf{j})] Y(\mathbf{v}+\mathbf{i})\} \\
= \; & E\{X(\mathbf{v}) X(\mathbf{v}+\mathbf{i})\}\; E\{Y(\mathbf{v}) Y(\mathbf{v}+\mathbf{i})\} + E\{u(\mathbf{v})^2\}\; E\{u(\mathbf{v}+\mathbf{i})^2\} \\
= \; & E\{X(\mathbf{v}) X(\mathbf{v}+\mathbf{i})\}\; E\{Y(\mathbf{v}) Y(\mathbf{v}+\mathbf{i})\} + \sigma^4,
\end{aligned}$$

and the finite third moment has been used for the two middle terms. Similarly, if we use (2.4.39) we may write

$$E\{X^*(\mathbf{v})X^*(\mathbf{v}+\mathbf{i})Y(\mathbf{v})Y(\mathbf{v}+\mathbf{i})\} = 1 + E\{X^*(\mathbf{v})X^*(\mathbf{v}+\mathbf{i})\}\ E\{Y(\mathbf{v})Y(\mathbf{v}+\mathbf{i})\}.$$

With exactly the same arguments and the change of roles of the two sequences $\{X(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ and $\{Y(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ and the two sequences $\{u(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ and $\{\varepsilon(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$, we may proceed as follows:

(ix) If $\mathbf{v}_1 = \mathbf{v} < \mathbf{v}+\mathbf{i} < \mathbf{v}_2$, then we may use the same argument as in (vii).

(x) If $\mathbf{v} < \mathbf{v}_1 \leq \mathbf{v}_2 < \mathbf{v}+\mathbf{i}$, then we may use the same argument as in (i).

(xi) If $\mathbf{v} < \mathbf{v}_1 < \mathbf{v}_2 = \mathbf{v}+\mathbf{i}$, then we may use the same argument as in (iv).

(xii) If $\mathbf{v} < \mathbf{v}_1 < \mathbf{v}+\mathbf{i} < \mathbf{v}_2$, then we may use the same argument as in (i).

(xiii) If $\mathbf{v} < \mathbf{v}_1 = \mathbf{v}_2 = \mathbf{v}+\mathbf{i}$, then we may use the same argument as in (vi).

(xiv) If $\mathbf{v} < \mathbf{v}_1 = \mathbf{v}+\mathbf{i} < \mathbf{v}_2$, then we may use the same argument as in (ii).

(xv) If $\mathbf{v} < \mathbf{v}+\mathbf{i} < \mathbf{v}_1 \leq \mathbf{v}_2$, then we may use the same argument as in (i).

Finally, for the case of $\mathbf{i} = \mathbf{0}$, we want to verify that

$$E\{X(\mathbf{v}_1)X(\mathbf{v}_2)Y(\mathbf{v})^2\} = E\{X(\mathbf{v}_1)X(\mathbf{v}_2)\}\ E\{Y(\mathbf{v})^2\},$$

when either $\mathbf{v}_1 \neq \mathbf{v}$ or $\mathbf{v}_2 \neq \mathbf{v}$. If $\mathbf{v}_1,\ \mathbf{v}_2 \neq \mathbf{v}$, then we may use the same argument as in (i). If either $\mathbf{v}_1 = \mathbf{v}$ or $\mathbf{v}_2 = \mathbf{v}$ (but not both), then a similar argument like in (vi) might be used. ■

General Yule-Walker equations

For a given polynomial

$$\theta(\mathbf{z}) = 1 + \theta_{\mathbf{j}_1}\mathbf{z}^{\mathbf{j}_1} + \cdots + \theta_{\mathbf{j}_q}\mathbf{z}^{\mathbf{j}_q},$$

we derive in this section the general Yule-Walker equations. These equations refer to the coefficients $\theta_{\mathbf{j}_1}, \cdots, \theta_{\mathbf{j}_q}$, of the polynomial $\theta(\mathbf{z})$ rather than the auto-regression with polynomial $\theta(\mathbf{z})$, like the original Yule-Walker equations. For a more detailed description of the original Yule-Walker equations, one may refer to Section 3.3.1. Here, the auto-covariance functions of both the two processes, the auto-regression and the moving-average that use $\theta(\mathbf{z})$, are involved, in contrast to the original Yule-Walker equations that use the auto-covariance function of the auto-regression only.

If we define the spectral density of the auto-regression $\{X(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ defined in (2.4.15) as

$$g_X(\boldsymbol{\omega}) \equiv \frac{1}{(2\pi)^d} \cdot \frac{1}{\theta(e^{i\boldsymbol{\omega}})\theta(e^{-i\boldsymbol{\omega}})} = \frac{1}{(2\pi)^d} \cdot \frac{1}{\gamma(e^{i\boldsymbol{\omega}})},\ \boldsymbol{\omega}^\tau \in [-\pi,\pi]^d \tag{2.4.43}$$

then it is clear that the polynomial $c(\mathbf{z})$ defined in (2.4.18) generates the auto-covariances of the auto-regression

$$c(\mathbf{j}) = \mathrm{Cov}\{X(\mathbf{v}), X(\mathbf{v}+\mathbf{j})\} = E\{X(\mathbf{v})X(\mathbf{v}+\mathbf{j})\},\ \mathbf{j}^\tau \in \mathcal{Z}^d. \tag{2.4.44}$$

Similarly, if the spectral density of the moving-average $\{Y(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ defined in (2.4.14) is

$$g_Y(\boldsymbol{\omega}) = \frac{1}{(2\pi)^d}\ \theta(e^{i\boldsymbol{\omega}})\ \theta(e^{-i\boldsymbol{\omega}}) = \frac{1}{(2\pi)^d}\ \gamma(e^{i\boldsymbol{\omega}}),\ \boldsymbol{\omega}^\tau \in [-\pi,\pi]^d, \tag{2.4.45}$$

then combining this with (2.3.3) implies that the polynomial

$$\gamma(\mathbf{z}) = \sum_{\mathbf{j}^\tau \in \mathcal{F}} \gamma(\mathbf{j})\ \mathbf{z}^{\mathbf{j}} \tag{2.4.46}$$

generates the auto-covariances

$$\gamma(\mathbf{j}) = \mathrm{Cov}\{Y(\mathbf{v}), Y(\mathbf{v}+\mathbf{j})\} = E\{Y(\mathbf{v})Y(\mathbf{v}+\mathbf{j})\},\ \mathbf{j}^\tau \in \mathcal{Z}^d. \tag{2.4.47}$$

From (2.4.17) and (2.4.22), we can derive

$$\sum_{\mathbf{i}^\tau \in \mathcal{F}} \gamma(\mathbf{i})\ c(\mathbf{i}-\mathbf{j}) = \begin{cases} 1, & \mathbf{j} = \mathbf{0} \\ 0, & \mathbf{j} \neq \mathbf{0} \end{cases}. \tag{2.4.48}$$

We will refer to equations (2.4.48) as the general Yule-Walker equations.

As we are going to see later in Chapter 4, the invention of the general Yule-Walker equations does not only concern the ARMA processes. Looking at (2.4.48), one might see that the auto-covariance functions of two processes are involved and there are two fundamental elements needed, in order to build these equations. The first is that the set of lags $\mathcal{F}$, where the auto-covariance function of one of the processes is non-zero, is of finite cardinality. The second is that the two processes must be linked via a specific transformation to achieve such a relation between their auto-covariance functions. In the specific case of the auto-regression $\{X(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ and the moving-average $\{Y(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$, this transformation was expressed in

$$Y(\mathbf{v}) = \gamma(\mathbf{B})X(\mathbf{v})$$

or

$$X(\mathbf{v}) = c(\mathbf{B})Y(\mathbf{v}).$$

Thus, the two processes are connected via transformations that use their auto-covariance functions. If $\mathcal{F}$ is not of finite cardinality, then the equations still hold but the summations extend over infinity; this is of little use when we observe processes on many dimensions $d \geq 2$, due to a problem called the edge-effect, which will be further analyzed in Chapters 3 and 6. In other words, processes that have a non-zero auto-covariance function over a finite set of lags may have a worthless contribution in the solution of the problems of estimation for the parameters of processes on the d-dimensional regular lattice. Further, for a given polynomial $\gamma(\mathbf{z})$, a finite set $\mathcal{F}$ allows us to create a set of random variables $\{Y(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{S}^*\}$ from another set of random variables $\{X(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{S}\}$, where $\mathcal{S}^* \subseteq \mathcal{S}$, if we apply the finite transformation involved.

We do not worry about observing any process yet, this will be important in the next chapters of estimation, so we refer to sets of random variables rather than sets of observations. The general Yule-Walker equations will next be used to demonstrate that both the inverse theoretical variance matrix of a vector of random variables from an auto-regression and the inverse theoretical variance matrix of a vector of random variables from a moving-average are conditional variance matrices. Gaussian likelihoods, which will be used later for estimation, involve the inverse variance matrices and their determinants and now we know their form. Moreover, since the inverse conditional variance matrices involve the auto-covariance functions of the auto-regression and the moving-average, it might be useful to refer to conditional Gaussian likelihoods. Indeed, the inverse variance matrix of a set of random variables from the auto-regression is a conditional variance matrix referring to the same set of random variables from the moving-average and vice versa. The property we are showing is dual but we are only using later one of its two sides. In Chapters 3 and 4 we will refer to conditional likelihoods and we will come up with inverse conditional variance matrices, *i.e.* variance matrices with elements the auto-covariance functions of simple processes such as an auto-regression or a moving-average.

Conditional variance matrix for the auto-regression

For any set $\mathcal{S} \subset \mathcal{Z}^d$ of finite cardinality N, and the set $\mathcal{F} \subset \mathcal{Z}^d$ as defined in Section 2.4.1, we consider $\mathcal{S}^*$ to be the maximal set such that for every $\mathbf{v}^\tau \in \mathcal{S}^*$, it holds that $\mathbf{v}^\tau - \mathbf{j}^\tau \in \mathcal{S}$ for all $\mathbf{j}^\tau \in \mathcal{F}$. We write then N^* for the cardinality of $\mathcal{S}^*$. Of course, we may also

write that $\mathcal{S}^* \subset \mathcal{S}$ and $N^* < N$. We let $\mathbf{X}^*, \mathbf{Y}^*$ to be the $(N^* \times 1)$ random vectors from processes (2.4.15) and (2.4.14), respectively, of the elements of $\mathcal{S}^*$ in any order. Similarly, we consider the random vector $\mathbf{X}_0$ with members from the process (2.4.15) on the locations of the set $\mathcal{S} - \mathcal{S}^*$. The polynomial $\theta(\mathbf{z})$ we are using is such that the two processes are causal and invertible, respectively. It is then clear from (2.4.17) that we can write

$$\mathbf{Y}^* = \mathbf{\Gamma} \cdot \mathbf{X}^* + \mathbf{\Gamma}_0 \cdot \mathbf{X}_0, \tag{2.4.49}$$

for an $(N^* \times N^*)$ matrix $\mathbf{\Gamma}$ and an $(N^* \times (N - N^*))$ matrix $\mathbf{\Gamma}_0$. It holds according to (2.4.22) that

$$\text{Cov}\{\mathbf{X}^*, \mathbf{Y}^*\} = \mathbf{I}_{N^*}, \tag{2.4.50}$$

i.e. it is the identity matrix. Again from (2.4.17), it holds that

$$\mathbf{\Gamma} \equiv \text{Var}\{\mathbf{Y}^*\}, \tag{2.4.51}$$

since the vectors $\mathbf{X}^*, \mathbf{Y}^*$ refer to all the locations $\mathbf{v}^\tau \in \mathcal{S}^*$ set in the same order.

Thanks to Remark 2.3, the eigenvalues of $\text{Var}\{\mathbf{Y}^*\}$ are positive and its inverse exists. We can now write

$$\mathbf{X}^* = \text{Var}\{\mathbf{Y}^*\}^{-1}\mathbf{Y}^* - \text{Var}\{\mathbf{Y}^*\}^{-1}\mathbf{\Gamma}_0 \, \mathbf{X}_0. \tag{2.4.52}$$

Equation (2.4.52) reveals that $\mathbf{X}^*$ can be written as a linear combination of the two random vectors $\mathbf{Y}^*$ and $\mathbf{X}_0$, which are uncorrelated using (2.4.22). As a result and using (2.4.50), it should hold that

$$\mathbf{X}^* = \text{Var}\{\mathbf{Y}^*\}^{-1}\mathbf{Y}^* + \text{Cov}\{\mathbf{X}^*, \mathbf{X}_0\} \, \text{Var}\{\mathbf{X}_0\}^{-1}\mathbf{X}_0, \tag{2.4.53}$$

from which we can write

$$\text{Var}\{\mathbf{Y}^*\}^{-1} = \text{Var}\{\mathbf{X}^*\} - \text{Cov}\{\mathbf{X}^*, \mathbf{X}_0\} \, \text{Var}\{\mathbf{X}_0\}^{-1}\text{Cov}\{\mathbf{X}_0, \mathbf{X}^*\}. \tag{2.4.54}$$

Moreover, if we can assume that $\{e(\mathbf{v})\} \sim IID(0,1)$ and that $\{u(\mathbf{v})\} \sim IID(0,1)$ too, then the random vectors $\mathbf{Y}^*$ and $\mathbf{X}_0$ are independent, according to Lemma 2.4. It holds from (2.4.53) that

$$\text{Var}\{\mathbf{X}^*|\mathbf{X}_0\} = \text{Var}\{\mathbf{Y}^*\}^{-1} \tag{2.4.55}$$

and also

$$E\{\mathbf{X}^*|\mathbf{X}_0 = \mathbf{x}_0\} = \text{Cov}\{\mathbf{X}^*, \mathbf{X}_0\} \, \text{Var}\{\mathbf{X}_0\}^{-1}\mathbf{x}_0. \tag{2.4.56}$$

Conditional variance matrix for the moving average

For the same sets $\mathcal{S}$, $\mathcal{S}^*$, and cardinalities N, N^* as defined before, we now consider the $(N \times 1)$ random vectors $\mathbf{X}$ and $\mathbf{Y}$ from processes (2.4.15) and (2.4.14), respectively, of the elements of $\mathcal{S}$ in any order. We write

$$\mathbf{X} = \mathbf{C} \cdot \mathbf{Y} + \mathbf{Y}_0, \tag{2.4.57}$$

where

$$\mathbf{C} \equiv \text{Var}\{\mathbf{X}\} \tag{2.4.58}$$

and $\mathbf{Y}_0$ is the $(N \times 1)$ random vector with element corresponding to the specific location $\mathbf{v}^\tau \in \mathcal{S}$ equal to

$$\sum_{\mathbf{v}^\tau - \mathbf{j}^\tau \notin \mathcal{S}} c(\mathbf{j})\, Y(\mathbf{v} - \mathbf{j}). \tag{2.4.59}$$

Thanks to Remark 2.3, the inverse $\text{Var}\{\mathbf{X}\}^{-1}$ exists. We can write

$$\mathbf{Y} = \text{Var}\{\mathbf{X}\}^{-1}\mathbf{X} - \text{Var}\{\mathbf{X}\}^{-1}\mathbf{Y}_0. \tag{2.4.60}$$

Since $\mathbf{X}$ and $\mathbf{Y}_0$ are two uncorrelated random vectors, we can write

$$\text{Var}\{\mathbf{X}\}^{-1} = \text{Var}\{\mathbf{Y}\} - \text{Cov}\{\mathbf{Y}, \mathbf{Y}_0\}\text{Var}\{\mathbf{Y}_0\}^{-1}\text{Cov}\{\mathbf{Y}_0, \mathbf{Y}\}, \tag{2.4.61}$$

which, if both $\{e(\mathbf{v})\} \sim IID(0,1)$ and $\{u(\mathbf{v})\} \sim IID(0,1)$, can be expressed in terms of a conditional variance that is later given in (2.4.65). But that would imply that we know the value $\mathbf{y}_0$ of the random vector $\mathbf{Y}_0$ which depends on all values $y(\mathbf{v})$, $\mathbf{v}^\tau \notin \mathcal{S}$.

We define

$$\mathbf{w} \equiv -[\mathbf{Y} - \text{Var}\{\mathbf{X}\}^{-1}\mathbf{X}] = \text{Var}\{\mathbf{X}\}^{-1}[\mathbf{X} - \text{Var}\{\mathbf{X}\}\ \mathbf{Y}] = \text{Var}\{\mathbf{X}\}^{-1}\mathbf{Y}_0. \tag{2.4.62}$$

and re-express (2.4.60) as

$$\mathbf{Y} = \text{Var}\{\mathbf{X}\}^{-1}\mathbf{X} - \mathbf{w}. \tag{2.4.63}$$

The general Yule-Walker equations guarantee that the elements of the symmetric matrix $\text{Var}\{\mathbf{X}\}^{-1} = [\gamma^*_{k,l}]_{k,l=1}^N$ are

$$\gamma^*_{k,l} = \gamma(\mathbf{v}_k - \mathbf{v}_l) \tag{2.4.64}$$

if at least one $\mathbf{v}_k \in \mathcal{S}^*$ or $\mathbf{v}_l \in \mathcal{S}^*$ for all the elements of $\mathcal{S}$ labeled as $k, l = 1, \cdots, N$. As a result, the random vector

$$\text{Var}\{\mathbf{X}\}^{-1}\mathbf{X}$$

gives the random variables $Y(\mathbf{v}), \mathbf{v}^\tau \in \mathcal{S}^*$. Similarly, the random vector $\mathbf{w}$ has zero elements for $\mathbf{v}^\tau \in \mathcal{S}^*$. The remaining elements of the vector $\mathbf{w}$ on locations $\mathbf{v}^\tau \in \mathcal{S} - \mathcal{S}^*$ can be seen as functions of the values $Y(\mathbf{v})$, $\mathbf{v}^\tau \notin \mathcal{S}$, or functions of the errors $e(\mathbf{v}) + \sum_{n=1}^{q} \theta_{\mathbf{j}_n}\ e(\mathbf{v} - \mathbf{j}_n)$, or $u(\mathbf{v}) + \sum_{n=1}^{q} \theta_{\mathbf{j}_n}\ u(\mathbf{v} + \mathbf{j}_n)$, $n = 1, \cdots, q$, for all $\mathbf{v}^\tau \notin \mathcal{S}$. Thus, a reasonable assumption is that $\mathbf{w} = \mathbf{0}$, which when $\mathbf{w}$ is independent of $\mathbf{X}$ can result in

$$\mathrm{Var}\{\mathbf{Y}|\mathbf{w} = \mathbf{0}\} = \mathrm{Var}\{\mathbf{Y}|\mathbf{w}\} = \mathrm{Var}\{\mathbf{X}\}^{-1} \tag{2.4.65}$$

and

$$E\{\mathbf{Y}|\mathbf{w} = \mathbf{0}\} = \mathbf{0}. \tag{2.4.66}$$

It should be made clear now that, although the variance matrices and their inverses only involve the second-order properties of the processes of interest and they are totally unconnected to any causal formulations assumed to be taking place there, before that we have made a fundamental assumption for the polynomial $\theta(\mathbf{z})$, without which we cannot proceed in that way. We have assumed that

$$\theta(\mathbf{z})^{-1} \equiv 1 + \sum_{\mathbf{j}>\mathbf{0}} \Theta_{\mathbf{j}}\ \mathbf{z}^{\mathbf{j}},\ \sum_{\mathbf{j}>\mathbf{0}} |\Theta_{\mathbf{j}}| < \infty.$$

It is only then that we may move smoothly between an auto-regression $\{X(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ and a moving-average $\{Y(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$, such that

$$\mathrm{Cov}\{X(\mathbf{v} + \mathbf{j}), Y(\mathbf{v})\} = 0,\ \mathbf{j} \neq \mathbf{0}.$$

As it will be made clear in Section 3.6, this is extremely important in estimation if we are interested in estimating the parameters of a bilateral rather than a unilateral ARMA process. More details about how we define unilateral and bilateral ARMA models will be provided in Section 3.2.1.

2.5 K-dependent processes

First we write the following condition.

(C1) (i) For a set $\mathcal{S} \equiv \mathcal{S}_N \subset \mathcal{Z}^d$ of cardinality N, we write $N \to \infty$ if the length M of the minimal hypercube including $\mathcal{S}$, say $\mathcal{S} \subseteq \mathcal{C}_M$, and the length m of the maximal hypercube included in $\mathcal{S}$, say $\mathcal{C}_m \subseteq \mathcal{S}$, are such that $M, m \to \infty$.

(ii) As $M, m \to \infty$, it holds that M/m is bounded away from ∞.

A typical example for (C1)(i) is when our observations lie on a hyper-rectangle $\mathcal{S} = \{1, \cdots, N_1\} \times \cdots \times \{1, \cdots, N_d\}$ and $\min_{k=1}^{d}\{N_k\} \to \infty$. We will often refer to this example for simplicity, although the results hold for all increasing sets satisfying (C1). For (C1)(ii), the minimum and maximum number of recordings in $\mathcal{S}$ per dimension $k = 1, \cdots, d$, increase at the same speed. In the case of the rectangle, we may write

$$0 < \alpha \leq \frac{N_k}{N_j} \leq \beta < \infty, \ k, j = 1, \cdots, d. \tag{2.5.1}$$

Condition (C1) was used by Guyon (1982, p.95). Part (ii) also implies that we can write

$$N = O(M^d) = O(m^d). \tag{2.5.2}$$

Next we give the following definition.

Definition 2.5. For the minimum non-negative integer K, a strictly stationary sequence of random variables $\{U(\mathbf{v}), \ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ is said to be K-dependent if for every $v_k, v_k^* \in \mathcal{Z}$, $k = 1, \cdots, d$, the two sets of random variables $\{U(v_1, \cdots, v_{k-1}, j_k, v_{k+1}, \cdots, v_d), \ j_k \leq v_k\}$ and
$\{U(v_1^*, \cdots, v_{k-1}^*, j_k, v_{k+1}^*, \cdots, v_d^*), \ j_k \geq v_k + K + 1\}$ are independent.

For any positive integer K, we define the set

$$\mathcal{B}_K \equiv \mathcal{M}_K - \{[0, \cdots, 0]^\tau\}, \tag{2.5.3}$$

where $\mathcal{M}_K$ was defined in (2.3.6). We also define the set

$$\mathcal{F}_K \equiv \{\mathbf{j}^\tau : \ \mathbf{j}^\tau \in \mathcal{M}_K, \ -\mathbf{j}^\tau \in \mathcal{M}_K\}. \tag{2.5.4}$$

Theorem 2.2 (Central Limit Theorem for strictly stationary K-dependent sequences). Let $\{U(\mathbf{v}), \ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ be a strictly stationary K-dependent sequence of random variables with zero mean and auto-covariance function $\gamma(\cdot)$, and let $\mathcal{S} \subset \mathcal{Z}^d$ be a set of cardinality N. Write

$$\nu_K \equiv \sum_{\mathbf{j}^\tau \in \mathcal{F}_K} \gamma(\mathbf{j}), \tag{2.5.5}$$

and

$$U_N \equiv \sum_{\mathbf{v}^\tau \in \mathcal{S}} U(\mathbf{v})/N. \tag{2.5.6}$$

Then under condition (C1), it holds that

(i) $N \ \mathrm{Var}\{U_N\} \to \nu_K$ and

(ii) $N^{1/2}U_N \xrightarrow{D} N(0, \nu_K)$

as $N \to \infty$.

Proof. For simplicity and without loss of generality, we will consider the case of the hyper-rectangle.

(i) We may write

$$N \text{ Var}\{U_N\} = \frac{1}{N} \sum_{\mathbf{v}^\tau, \mathbf{v}^{*\tau} \in \mathcal{S}} \gamma(\mathbf{v} - \mathbf{v}^*) = \sum_{\mathbf{j}^\tau \in \mathcal{F}_K} \frac{f_\mathbf{j}}{N} \gamma(\mathbf{j}) \tag{2.5.7}$$

where $f_\mathbf{j}$ is the number of times we may find the lag $\mathbf{j}^\tau \in \mathcal{Z}^d$ from the difference of locations $\mathbf{v} - \mathbf{v}^*$ when both $\mathbf{v}^\tau, \mathbf{v}^{*\tau} \in \mathcal{S}$. For example, it always holds that $f_\mathbf{0} = N$. For the case of the rectangle, Proposition 3.1 that appears in the next chapter shows that

$$f_\mathbf{j} = \prod_{k=1}^{d} (N_k - |j_k|) = N \cdot \prod_{k=1}^{d} \left(1 - \frac{|j_k|}{N_k}\right) \to N, \tag{2.5.8}$$

as $\min_{k=1}^{d}\{N_k\} \to \infty$, which proves the first part of the theorem.

(ii) For each fixed integer $m > 2K$, we define the sets $\mathcal{S}^{(m)}_{p_1,\cdots,p_d} \subset \mathcal{S}$, such that

$$\mathcal{S}^{(m)}_{p_1,\cdots,p_d} \equiv \{\mathbf{v}^\tau = [v_1, \cdots, v_d] : v_k = (p_k - 1)m + 1, \cdots, p_k \cdot m - K,\ k = 1, \cdots, d\} \tag{2.5.9}$$

for $p_k = 1, \cdots, r$, for all $k = 1, \cdots, d$, and $r = \left[\min_{k=1}^{d}\{N_k\}/m\right]$. We also define the set

$$\mathcal{R}^{(m)} \equiv \cup_{p_1=1}^{r} \cdots \cup_{p_d=1}^{r} \mathcal{S}^{(m)}_{p_1,\cdots,p_d} \tag{2.5.10}$$

and the random variable

$$V_{Nm} \equiv N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{R}^{(m)}} U(\mathbf{v}). \tag{2.5.11}$$

Then the random variable $N^{1/2}V_{Nm}$ is a sum of r^d independent and identically distributed random variables each with zero mean and variance equal to

$$\begin{aligned} R_{m-K} &\equiv \text{Var}\{U(1, \cdots, 1) + \cdots + U(m-K, \cdots, m-K)\} \\ &= \sum_{\mathbf{j}^\tau \in \mathcal{F}_{m-K}} \prod_{k=1}^{d} (m - K - |j_k|)\ \gamma(\mathbf{j}) \\ &= (m-K)^d \sum_{\mathbf{j}^\tau \in \mathcal{F}_{m-K}} \prod_{k=1}^{d} \left(1 - \frac{|j_k|}{m-K}\right) \gamma(\mathbf{j}). \end{aligned} \tag{2.5.12}$$

In (2.5.12), we are using (2.5.8) for the case of the hyper-cube. For fixed m, as $\min_{k=1}^{d}\{N_k\} \to \infty$, it holds that $r \to \infty$ and we can apply the central limit theorem and

write that

$$[r^d]^{-1/2}\left(\sum_{\mathbf{v}^\tau\in\mathcal{R}^{(m)}}U(\mathbf{v})\right)=[r^d]^{-1/2}\left(N^{1/2}V_{Nm}\right)\xrightarrow{D}N(0,R_{m-K}). \tag{2.5.13}$$

Under (C1)(ii), it holds that

$$\frac{N}{r^d}\sim\left(\frac{\min_{k=1}^d\{N_k\}}{r}\right)^d\sim m^d.$$

As a result, we may re-write (2.5.13) as

$$(m^d)^{1/2}V_{Nm}\xrightarrow{D}N(0,R_{m-K})$$

or

$$V_{Nm}\xrightarrow{D}V_m\sim N\left(0,\frac{R_{m-K}}{m^d}\right), \tag{2.5.14}$$

where it holds that

$$\frac{R_{m-K}}{m^d}=\left(\frac{m-K}{m}\right)^d\sum_{\mathbf{j}^\tau\in\mathcal{F}_{m-K}}\prod_{k=1}^d\left(1-\frac{|j_k|}{m-K}\right)\gamma(\mathbf{j})\to\sum_{\mathbf{j}^\tau\in\mathcal{F}_K}\gamma(\mathbf{j})=\nu_K, \tag{2.5.15}$$

as $m\to\infty$. Finally, it holds that

$$N^{1/2}U_N=N^{-1/2}\sum_{\mathbf{v}^\tau\in\mathcal{S}}U(\mathbf{v})$$

and we may look at the variance of the random variable

$$N^{-1/2}\sum_{\mathbf{v}^\tau\in\mathcal{S}}U(\mathbf{v})-N^{-1/2}\sum_{\mathbf{v}^\tau\in\mathcal{R}^{(m)}}U(\mathbf{v}).$$

In a similar way like Theorem 6.4.2 of Brockwell and Davis (1991), we may find that

$$\limsup_{\min\{N_1,\cdots,N_d\}\to\infty}N^{-1}\mathrm{Var}\{\sum_{\substack{\mathbf{v}^\tau\in\mathcal{S},\\ \mathbf{v}^\tau\notin\mathcal{R}^{(m)}}}U(\mathbf{v})\}=\frac{R_K}{m^d} \tag{2.5.16}$$

which tends to 0 as $m\to\infty$. ■

Chapter 3

Estimation for ARMA models on a d-dimensional lattice

3.1 Introduction

In this chapter, we are concerned with the estimation of the parameters of ARMA models on the d-dimensional lattice. We take one step at a time, as we first consider the case of observations from a causal auto-regression and then the cases of the moving-average and ARMA processes. Due to a problem known as the edge-effect, which is described in Section 3.2.2, the standard methods used for time series cannot be applied when $d \geq 2$; the exact Gaussian likelihood estimators computed from N observations on the d-dimensional lattice have an absolute bias of order $N^{-1/d}$, as Guyon (1982, p.95) has explained. Thus, we have looked for modified versions of Gaussian likelihoods that can generate consistent and asymptotically unbiased and normal estimators. A similar methodology was used by Yao and Brockwell (2006) for the case when $d = 2$. Before that, Guyon (1982) had proposed a modification on the quantity, for which Whittle (1954, p.440) proved that it is a modified version of Gaussian likelihood when $d = 2$.

We start in Section 3.3 with the case of observations from a causal auto-regression, and we consider the Yule-Walker estimators of the parameters. These estimators are conditional likelihood estimators, as we explain later in Section 3.3.3. Moreover, they are the least squares estimators that Guyon (1982, p.103) was suggesting for the case of a unilateral auto-regression. We prove that they have the properties that we mentioned before and also that they are efficient for Gaussian random variables.

In Section 3.4 we study the case of a moving-average. Though invertible moving-averages are considered there, our results can be generalized, as we explain later in Section 3.6. We propose a Yule-Walker method of estimation that imitates the general Yule-Walker equations from the previous chapter for the moving-average. The estimators are consistent, asymptotically unbiased and normal, though we have managed to write their variance matrix under a condition, which is definitely satisfied for Gaussian random variables. Even if the original process is Gaussian, the estimators are still not efficient. Thus following the Yule-Walker method, we maximize a modification of a conditional Gaussian likelihood that improves our results and gives efficient estimators under exactly the same conditions.

The good results achieved for the moving-average are still not surprising. As Guyon (1982, p.100) noted, when the range of summation in the numerator or the denominator of the spectral density of the process of interest is finite, we should expect to find ways to deal with the edge-effect. An ARMA process though, does not have an auto-covariance function that cuts off to zero outside a finite set of lags like a moving-average, nor is it possible to apply a finite filter on the ARMA process and come up with a sequence of uncorrelated random variables, like it is for an auto-regression. Nevertheless, a finite transformation applied on the ARMA process does produce a moving-average and these two features, *i.e.* the finite transformation applied on the ARMA and the finite auto-covariance function of the moving-average that is produced then, are the special features of an ARMA process that we will take advantage of. As we will show later, three different finite filters can produce three moving-averages in this way, but only one of them is appropriate for use and capable of defeating the edge-effect. The case of the causal and invertible ARMA is studied in Section 3.5 and generalizes the results of Section 3.4, rather than 3.3. Thus, we propose a modified version of Gaussian likelihood estimators that are consistent, asymptotically unbiased and normal. The variance matrix of the estimators is known under an extra assumption, and in the Gaussian case the estimators are guaranteed to be efficient. Even if the ARMA process is not causal and invertible, we may define estimators with same properties, according to Section 3.6, after we correct the deterministic part of the modified likelihood properly.

This also applies to the conditional likelihood proposed in Section 3.3, which is appropriate for causal auto-regressions only. It is an obvious generalization of Whittle's (1954) contribution for the estimation of parameters of bilateral auto-regressions on $\mathbb{Z}^2$,

as Whittle originally approached the random part of the Gaussian likelihood as a sum of squares of uncorrelated random variables. Since the likelihood proposed in 3.3 is not clearly a special case of the likelihood proposed in Section 3.5, we might find a link later in Section 4.5.2, when it will be possible to estimate the parameters of a unilateral or bilateral auto-regression using a modification on a Gaussian conditional likelihood, *i.e.* the special case of the likelihood of 3.5. The results of Section 3.3 demand that the auto-regression of interest is causal but they also provide the form of the variance matrix of the estimators, even though the random variables might not be Gaussian. In Section 3.5, apart from the assumption of causality which is also necessary, we have shown that the form of variance matrix is known under an extra condition, which is automatically satisfied for Gaussian processes. Thus, in Section 3.7, we establish the usefulness of causal auto-regressions when one of the dimensions is the time axis and we use further the results of Section 3.3, without making any specific assumption on the distribution of the process of interest. Tests for the unknown coefficients might be performed and all the results are applied on a spatio-temporal dataset.

3.2 The problems of the ARMA

The multi-dimensional ARMA process was defined in the previous chapter in a similar way to the one-dimensional ARMA used for time series. In this section, we present two main problems that arise when we deal with the ARMA defined on the regular d-dimensional lattice.

3.2.1 Unilaterality, causality and invertibility

For any location $\mathbf{v}^\tau \in \mathcal{Z}^d$ on which the process takes place, any other location belongs either to its 'past' $\{\mathbf{v}-\mathbf{i}, \mathbf{i}>\mathbf{0}\}$, or 'future' $\{\mathbf{v}+\mathbf{i}, \mathbf{i}>\mathbf{0}\}$. We call 'past' and 'future' for convenience, when we refer to the two opposite orderings of interest, as described in Section 2.2. After we have decided which orderings represent the 'past' and 'future', we may define a causal and invertible ARMA process as in Section 2.4. In this section, we will attempt to explain the idea of unilaterality for the case of ARMA processes. For the simple case of an auto-regression first; a causal auto-regression is always unilateral, in the sense that the process on any point $\mathbf{v}^\tau \in \mathcal{Z}^d$ can be expressed as a linear combination

of $\varepsilon(\mathbf{v}-\mathbf{i}), \mathbf{i} \geq \mathbf{0}$, with absolutely summable coefficients, and

$$\{\varepsilon(\mathbf{v})\} \sim WN(0, \sigma^2)$$

is the related sequence of random errors. Then for the simple case of a moving-average; an invertible moving-average is always unilateral, in the sense that for any $\mathbf{v}^\tau \in \mathcal{Z}^d$ we may express the related random error as a linear combination of the 'present' and 'past' values of the process itself, with absolutely summable coefficients.

For the case of an ARMA process, the relation between causality, invertibility and unilaterality may be extended in an obvious way. We give the following definition.

Definition 3.1. A finite ARMA process $\{Z(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$, which has been expressed in terms of the random sequence

$$\{\varepsilon(\mathbf{v})\} \sim WN(0, \sigma^2),$$

is unilateral if

1. $Z(\mathbf{v})$ is a linear function of $\varepsilon(\mathbf{v}-\mathbf{i}),\ \mathbf{i} \geq \mathbf{0}$, and $\varepsilon(\mathbf{v})$ is a linear function of $Z(\mathbf{v}-\mathbf{i}),\ \mathbf{i} \geq \mathbf{0}$, both with absolutely summable coefficients, or

2. $Z(\mathbf{v})$ is a linear function of $\varepsilon(\mathbf{v}+\mathbf{i}),\ \mathbf{i} \geq \mathbf{0}$, and $\varepsilon(\mathbf{v})$ is a linear function of $Z(\mathbf{v}+\mathbf{i}),\ \mathbf{i} \geq \mathbf{0}$, both with absolutely summable coefficients.

According to the definition, a causal and invertible ARMA process is always unilateral. Moreover, if we have an ARMA process that is not causal and invertible but the value of the process on $\mathbf{v}^\tau \in \mathcal{Z}^d$ is a function of 'present' and 'future' random errors, from the related error sequence, and the value of the random error on $\mathbf{v}^\tau \in \mathcal{Z}^d$ is a function of 'present' and 'future' values of the process, both with absolutely summable coefficients, then this is a unilateral ARMA process too. An interesting question is what happens when the two polynomials are oriented in an opposite way but do generate absolutely summable coefficients. For the sake of example, we write the one-dimensional ARMA(1,1) process $\{Z_t,\ t \in \mathcal{Z}\}$ defined by the equation

$$Z_t - 0.5\ Z_{t-1} \equiv u_t + 0.2\ u_{t+1},\ \{u_t\} \sim WN(0, \sigma^2).$$

According to Definition 3.1, we will not consider that $\{Z_t,\ t \in \mathcal{Z}\}$ form a unilateral ARMA process, although in terms of parametrization nothing would really change when

writing it down as a causal and invertible ARMA process. In other words, there is a sequence of uncorrelated random variables $\{\varepsilon_t,\ t \in \mathcal{Z}\}$, which we can define by the equation

$$Z_t - 0.5\ Z_{t-1} \equiv \varepsilon_t + 0.2\ \varepsilon_{t-1},\ \{\varepsilon_t\} \sim WN(0, \sigma^2)$$

and which now allows to our ARMA process to be considered causal and invertible. The problem of bilaterality of an ARMA process is far more complex than this. In order to describe it better, in order to understand Definition 3.1 and to follow the example of Whittle (1954), we will focus next on the cases of causal or unilateral auto-regressions and we will try to demonstrate how these two forms are connected.

Although a causal auto-regression is unilateral, a unilateral auto-regression is not necessarily causal. Similarly, an auto-regression that is not causal does not have to be bilateral. There are causal auto-regressions, there are unilateral and not causal auto-regressions and there are bilateral auto-regressions. We may give a simple example of a two-dimensional bilateral auto-regression, say $\{X^{(1)}(u,v),\ u,v \in \mathcal{Z}\}$, defined by the equation

$$\begin{aligned} & X^{(1)}(u,v) \equiv 5.1\ X^{(1)}(u-1,v) - 0.5\ X^{(1)}(u-2,v) + 0.3\ X^{(1)}(u,v-1) \\ - \ & 1.53\ X^{(1)}(u-1,v-1) + 0.15\ X^{(1)}(u-2,v-1) + w^{(1)}(u,v), \end{aligned} \tag{3.2.1}$$

where

$$\{w^{(1)}(u,v)\} \sim WN(0, \sigma_w^2).$$

We may re-write

$$(1 - 0.1\ B_1) \cdot (1 - 5\ B_1) \cdot (1 - 0.3\ B_2)\ X^{(1)}(u,v) = w^{(1)}(u,v) \tag{3.2.2}$$

and it holds that

$$\begin{aligned} & (1 - 0.1\ z_1) \cdot (1 - 5\ z_1) \cdot (1 - 0.3\ z_2) = (1 - 5.1\ z_1 + 0.5\ z_1^2) \cdot (1 - 0.3\ z_2) \\ = \ & 1 - 5.1\ z_1 + 0.5\ z_1^2 - 0.3\ z_2 + 1.53\ z_1 z_2 - 0.15\ z_1^2 z_2. \end{aligned} \tag{3.2.3}$$

Moreover, since

$$(1 - 5\ z_1) \cdot (1 - 5\ z_1^{-1}) = 25\ (1 - 0.2\ z_1^{-1})(1 - 0.2\ z_1), \tag{3.2.4}$$

the auto-regression defined in (3.2.1) shares exactly the same second-order properties as the causal auto-regression $\{X^{(2)}(u,v),\ u,v \in \mathcal{Z}\}$ defined by the equation

$$(1-0.1\ B_1)\cdot(1-0.2\ B_1)\cdot(1-0.3\ B_2)\ X^{(2)}(u,v) \equiv w^{(2)}(u,v),\ \{w^{(2)}(u,v)\} \sim WN(0, \sigma_w^2/25). \tag{3.2.5}$$

Both the auto-regressions defined by (3.2.1) and (3.2.5) are linear-by-linear, in the sense that the auto-regressive polynomials have been factorized into two parts, one referring to the dimension $u \in \mathcal{Z}$ only and the other to the other dimension $v \in \mathcal{Z}$, while this is not always possible. Linear-by-linear processes will be defined properly in Chapter 6. Here, they have been used as examples for mathematical convenience only.

But what if we are interested in an auto-regression that is not causal but it is unilateral? A simple example of a linear-by-linear unilateral auto-regression, which is not causal, is the auto-regression $\{A^{(1)}(u,v),\ u,v \in \mathcal{Z}\}$ defined by the equation

$$(1 - 0.2\ B_1^{-1}) \cdot (1 - 0.3\ B_2)\ A^{(1)}(u,v) \equiv e^{(1)}(u,v),\ \{e^{(1)}(u,v)\} \sim WN(0,\sigma_e^2). \quad (3.2.6)$$

Now, (3.2.6) shares exactly the same second-order properties as the causal auto-regression $\{A^{(2)}(u,v),\ u,v \in \mathcal{Z}\}$ defined by

$$(1 - 0.2\ B_1) \cdot (1 - 0.3\ B_2)\ A^{(2)}(u,v) \equiv e^{(2)}(u,v),\ \{e^{(2)}(u,v)\} \sim WN(0,\sigma_e^2). \quad (3.2.7)$$

The auto-regression defined in (3.2.6) is unilateral and a simple justification for this is that it would be causal if a different selection of unilateral orderings had been made to represent the 'past' and 'future'. Thus in our first example, what actually makes (3.2.2) a bilateral auto-regression is not that the root 0.2 is less than 1 in $(1 - 5\ z_1) = 0$, as this is only the reason why the auto-regression is not causal. The process defined in (3.2.2) is bilateral because, for the first dimension $u \in \mathcal{Z}$, there is a root 10 and a root 0.2, outside and inside the unit circle, respectively, in $(1 - 0.1\ z_1) \cdot (1 - 5\ z_1) = 0$. This means that the process runs over both sides of the axis, and since this is the primary axis, we may rest immediately that the auto-regression is bilateral. Following the same argument as Whittle (1954, p.436) for time series, we may safely say that there are $2^2 = 4$ unilateral processes that share exactly the same second-order properties as (3.2.2), as they occupy the four different quarters. One of them is the causal auto-regression defined in (3.2.5). Let $\{X^{(3)}(u,v), X^{(4)}(u,v), X^{(5)}(u,v),\ u,v \in \mathcal{Z}\}$ be the other three defined by

$$(1 - 0.1\ B_1) \cdot (1 - 0.2\ B_1) \cdot (1 - 0.3\ B_2^{-1})\ X^{(3)}(u,v) \equiv w^{(3)}(u,v), \quad (3.2.8)$$

$$(1 - 0.1\ B_1^{-1}) \cdot (1 - 0.2\ B_1^{-1}) \cdot (1 - 0.3\ B_2)\ X^{(4)}(u,v) \equiv w^{(4)}(u,v), \quad (3.2.9)$$

$$(1 - 0.1\ B_1^{-1}) \cdot (1 - 0.2\ B_2^{-1}) \cdot (1 - 0.3\ B_2^{-1})\ X^{(5)}(u,v) \equiv w^{(5)}(u,v), \quad (3.2.10)$$

where

$$\{w^{(3)}(u,v)\}, \{w^{(4)}(u,v)\}, \{w^{(5)}(u,v)\} \sim WN(0,\sigma_w^2/25).$$

While we can see immediately that the processes defined in (3.2.9) and (3.2.10) are not causal, since they both run over the 'future' side of the primary axis, it is harder to figure out whether the process defined in (3.2.8) is causal or not, as this depends on the secondary rather than primary axis. For any $u, v \in \mathcal{Z}$, it holds that $X^{(3)}(u,v)$ is a linear function of $w^{(3)}(u-k, v+l), k, l \geq 0$. While $w^{(3)}(u-k, v+l), k > 0, l \in \mathcal{Z}$, are allowed in a causal representation, the same cannot be said for $w^{(3)}(u, v+l), l > 0$, and the auto-regression is not causal. This verifies that $\{X^{(2)}(u,v),\ u, v \in \mathcal{Z}\}$ is the unique causal auto-regression corresponding to the spectral density of interest.

The two examples that we have used to understand the notion of causal versus unilateral auto-regressions suggest two different ways of finding out whether the process of interest has any of these properties. The first example writes the auto-regressive polynomial as

$$1 - \sum_{\mathbf{j} \in \mathcal{I}_p} \theta_{\mathbf{j}}\ \mathbf{z}^{\mathbf{j}}, \tag{3.2.11}$$

with the restriction that $\mathcal{I}_p \subseteq \{\mathbf{j} > \mathbf{0}\}$ exactly like in Section 2.4. Then it checks whether the process is causal according to Lemma 2.2. The second example allows the auto-regressive polynomial to be written as

$$\theta_p(\mathbf{z})\ \theta_f(\mathbf{z}^{-1}), \tag{3.2.12}$$

where

$$\theta_p(\mathbf{z}) \equiv 1 - \sum_{\mathbf{j} \in \mathcal{U}_p} \theta_{\mathbf{j}}^{(p)}\ \mathbf{z}^{\mathbf{j}}, \tag{3.2.13}$$

$$\theta_f(\mathbf{z}) \equiv 1 - \sum_{\mathbf{j} \in \mathcal{U}_f} \theta_{\mathbf{j}}^{(f)}\ \mathbf{z}^{\mathbf{j}}, \tag{3.2.14}$$

but puts the restrictions that $\mathcal{U}_p, \mathcal{U}_f \subseteq \{\mathbf{j} > \mathbf{0}\}$ and that the polynomials $\theta_p(\mathbf{z}), \theta_f(\mathbf{z})$ could be used to define causal auto-regressions. The process is causal then if and only if $\theta_f(\mathbf{z}) \equiv 1$. Of course, when is it that the auto-regressive polynomial can be factorized in the form (3.2.12) remains a question of interest.

If it is so, it would be interesting to find a way to re-express an auto-regression with polynomial of the form (3.2.12) as an auto-regression with a polynomial of the form (3.2.11). For example, we have defined the auto-regression $\{A^{(1)}(u,v),\ u, v \in \mathcal{Z}\}$ from (3.2.6). We may re-write it as

$$-0.2\ A^{(1)}(u+1, v) + 0.06\ A^{(1)}(u+1, v-1) + A^{(1)}(u,v) - 0.3\ A^{(1)}(u, v-1) = e^{(1)}(u,v) \tag{3.2.15}$$

or

$$A^{(1)}(u+1,v) - 0.3\ A^{(1)}(u+1,v-1) - 5\ A^{(1)}(u,v) + 1.5\ A^{(1)}(u,v-1) = -5\ e^{(1)}(u,v), \tag{3.2.16}$$

which makes sure that since the location $[u+1,v]$ is in the future of all other locations $[u+1,v-1]$, $[u,v]$ and $[u,v-1]$, then the coefficient multiplied by $A^{(1)}(u+1,v)$ is unity. Next, we redefine the e sequence by translation, according to Whittle (1954, p.436). Thus,

$$e^{(1)^*}(u+1,v) \equiv -5\ e^{(1)}(u,v),\ u,v \in \mathcal{Z}, \tag{3.2.17}$$

which implies that

$$\{e^{(1)^*}(u,v)\} \sim WN(0, 25\ \sigma_e^2) \tag{3.2.18}$$

and that

$$A^{(1)}(u,v) - 0.3\ A^{(1)}(u,v-1) - 5\ A^{(1)}(u-1,v) + 1.5\ A^{(1)}(u-1,v-1) = e^{(1)^*}(u,v). \tag{3.2.19}$$

As a result, the polynomial

$$1 - 0.3\ z_2 - 5\ z_1 + 1.5\ z_1 z_2 = (1 - 5\ z_1) \cdot (1 - 0.3\ z_2) \tag{3.2.20}$$

has the roots $z_1 = 0.2$ and $z_2 = 10/3$ and the first one is inside the unit circle, which implies that the auto-regression is not causal. Still, it is a unilateral auto-regression as for every dimension there is only one root available. This is the same argument that could be used to justify that any AR(1) in the standard time series is unilateral. We need at least two roots to decide whether they force the process to run over just one or both sides of the time axis.

The difficulty that appears when we generalize the concept of unilateral and bilateral processes, from the case that $d = 1$ to the cases $d \geq 2$, has its source on the fact that every dimension introduces two ends. For example, when there are two dimensions, there exist four quarters rather than two sides. Instead of defining multilateral processes, Whittle (1954, p.438) preferred to simplify this problem and, for any $[u,v]^\tau \in \mathcal{Z}^2$, to separate all the other points of the regular lattice $\mathcal{Z}^2$ into two equal parts. This has allowed the definition of unilateral and bilateral processes. Moreover, with a distinction made between these two parts, it is possible to define causal processes.

It can be easily understood that the transition from the concept of a unilateral auto-regression to that of a causal auto-regression is closely related to the interpretation

given to the dimensions. If those are the dimensions of space, we would not have any particular reason to expect that the dependence structure follows a direction. Indeed, it is not the high dimensionality that might be a problem when defining a causal auto-regression but the natural meaning that the dimensions have. Later in Section 3.7, we will see how meaningful causal auto-regressions can be defined when one dimension only is the time axis, though there might be spatial dimensions also involved. On the other hand, comparing a time series versus a process taking place on the line transect, makes apparent that defining a causal auto-regression might be problematic even if there is only one dimension. A fundamental assumption used to prove the results of this chapter is that all ARMA models of interest must be causal and invertible. Thus, this method has been severely criticized and has made many analysts of spatial statistics to resort to other methods, such as those described in the next chapter. Nevertheless, as we are going to see in Section 3.6, we may consider that the results hold even if the ARMA of interest is bilateral and we may find the elements that link Chapters 3 and 4.

3.2.2 Two sides of the edge-effect

We collect N observations from a causal and invertible ARMA process taking place on the d-dimensional lattice and we wish to estimate the unknown auto-regressive and moving-average parameters. In order to assess the asymptotic behavior of our estimators, we need to imagine ways that could have generated more observations for the statistical analysis. The most reasonable assumption is that all dimensions can give more and more locations and we usually assume that we obtain an increasing sequence of sampling sets $\mathcal{S}_N$, which satisfy (C1). A good asymptotic behavior would be achieved by the estimators if the square of their bias reduces to zero at a faster speed than their variance, as the number of observations increases.

When we deal with $d \geq 2$ indices, this relation between the order of the bias and variance does not seem to take place for the classical maximum Gaussian likelihood estimators of the parameters. In general, bias of order $N^{-1/d}$ occurs, unimportant if $d = 1$, but of the same order as the standard error if $d = 2$, and of greater order if $d \geq 3$ (Guyon, 1982, p.95). This problem that becomes existent for $d \geq 2$, is known as the edge-effect. The edge-effect is an obstacle for good estimation, either the Gaussian likelihood has been expressed in terms of the time domain or the spectral domain quantities. The transition of the Gaussian likelihood from one form to the other was first achieved by Whittle

(1954). A paper by Yao and Brockwell (2006) for the case $d = 2$, has chosen the original form of the likelihood and has used the innovations algorithm to unfold its quadratic part and compute the likelihood. The innovations algorithm imitates the AR(∞) representation of the process of interest for each observation available in the sample, based on all observations in the sample from its 'past'. In the paper, a modification of the likelihood is proposed, which follows an adequate selection of observations, confines the edge-effect and ensures that the absolute bias of the estimators is of smaller order than $N^{-1/2}$.

On the other hand and long before that, Guyon (1982) revealed the presence of the edge-effect for the general case $d \geq 2$, using the periodogram to access the random part of the likelihood. We consider the Gaussian likelihood to be a product of the random part, *i.e.* a function of the parameters and the data, and the deterministic part, which does not depend on the data. Guyon (1982, pp.96-7) also referred to the quadratic and deterministic parts of an approximation of the likelihood. Since the sample auto-covariances are the only random variables involved in the periodogram, plugging-in their unbiased estimators has cancelled the edge-effect there. Next, we examine closer how the edge-effect dominates against asymptotic unbiasedness of the estimators, either we approximate the likelihood in terms of the innovations algorithm and follow the time domain methodology or in terms of the sample auto-covariances, the periodogram and the spectral domain representations.

First, suppose that we are using the innovations algorithm. For each observation in the sample, the algorithm creates the best linear predictor based on all the observations from its 'past' in the sample and the prediction variance. This strongly resembles the AR(∞) representation of the process, as the information from the 'past' becomes more and more. In time series, every observation in the sample used for the estimation of the parameters plays two roles; it increases the sample size to reduce the variance and it also serves as past of other observations generated after it. For example, if T observations have been collected, the observation labelled as $t = 1, \cdots, T$, has $(t-1)$ observations available giving information about its recent past. As a result, every new observation generated has more observations available from its past than the previous one. Therefore, the absolute bias of the estimators would be reduced quite fast, as the number of observations increases.

Unfortunately, the same cannot be said for processes with more than one indices. A

new observation may lie close to an edge of the hyper-rectangle. It always serves its first role with success and contributes to reducing the variances of the estimators, but fails to succeed in the second task, that of reducing the absolute bias fast enough. The 'past' now has a different meaning and involves more than one dimensions. It cannot be made sure that increasing the number of observations decreases the absolute value of the bias with any order smaller than $N^{-1/d}$; a new observation does not always have more 'past' information available.

Even if we do not follow a causal formulation that mimics the AR(∞) representation of the process, but we approximate the likelihood in terms of the sample auto-covariances only, the problems remain. For observations $\{X_t,\ t = 1, \cdots, T\}$ from a (weakly) stationary time series, the theoretical variance matrix of the observations consists of exactly $(T-1)$ different elements apart from the element of the main diagonal, $\mathrm{Var}\{X_t\}$. In other words, any T consecutive observations introduce $(T-1)$ lags and every new observation introduces one new lag and one new element in the variance matrix only. Moreover, by the time a new lag, say $i > 0$, is introduced, it holds that there are $(T-i)$ pairs of observations in the sample that can give information about it. On the other hand and as $T \to \infty$, for a given lag $i > 0$, there is always a constant number of i observations $X_t, t = 1, \cdots, i$ that cannot be paired with X_{t-i}, since the latter does not exist in the sample.

Let us now consider the simplest case of a two-dimensional process and observations on a square lattice $\{X(u,v),\ u,v = 1, \cdots, n\}$. If we consider the fixed lag $[i,0]$ for an $i > 0$ that has been introduced by the observations in the sample and large enough n, then there are $n(n-i)$ pairs in the sample to give information about it, but there are also $n \cdot i$ observations that cannot be paired, *i.e.* $X(u,v), u = 1, \cdots, i, v = 1, \cdots, n$. These observations lie on one edge of the square and they become infinitely many as the edge of the square goes to infinity, which was not the case in time series. Moreover, after putting the observations in a unilateral order, the first n observations $X(1,1), \cdots, X(1,n)$, move on a straight line and they have introduced $(n-1)$ different auto-covariances, apart from the variance of the process. By the time the next observation $X(2,1)$ is set in order, there are n new lags introduced with all the previous observations. As a result, if we increase the surface that the square occupies on the lattice, the number of new lags introduced by some observations also increases. The number of sample auto-covariances involved in the periodogram this time is significantly different from $n^2 - 1$, *i.e.* the number of

observations minus one element corresponding to the variance; for the answer to that see Remark 3.1(ii) when $d = 2$. Before that, the following proposition indicates how this number of elements changes as the dimensionality d increases.

Proposition 3.1 (Inclusion-Exclusion formula). For $\mathcal{S} = \{1, \cdots, N_1\} \times \cdots \times \{1, \cdots, N_d\}$, let observations $\{X(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{S}\}$ from a (weakly) stationary process on the regular d-dimensional lattice and let $\mathbf{X}$ be a column vector consisting of the observations in any order. Then, the variance matrix $Var(\mathbf{X})$ consists of exactly

$$Q^{(d)}_{N_1 \cdots N_d} = \begin{cases} 2^{d-1}S^{(d)} - 2^{d-2}S^{(d-1)} + \cdots + (-1)^{d-1}2^0 S^{(1)} - 1, & \text{if } d \text{ is odd} \\ 2^{d-1}S^{(d)} - 2^{d-2}S^{(d-1)} + \cdots + (-1)^{d-1}2^0 S^{(1)}, & \text{if } d \text{ is even} \end{cases} \tag{3.2.21}$$

different auto-covariances apart from the element of the main diagonal, where $S^{(d)} = \prod_{k=1}^{d} N_k$ and $S^{(d-k)} = \sum_{1 \le p_1 < p_2 < \cdots < p_k \le d} P^{(d-k)}_{(p_1, p_2, \cdots, p_k)}$, with $P^{(d-k)}_{(p_1, p_2, \cdots, p_k)} = \frac{N}{N_{p_1} N_{p_2} \cdots N_{p_k}}$ for $k = 1, 2, \cdots, d-1$, and the positive integers $p_1, p_2, \cdots, p_k$.

Proof. We want to find the number of different lags generated by the observations of a stationary process $\{X(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{S}\}$, without forgetting the property $\gamma(\mathbf{i}) = \gamma(-\mathbf{i})$. We construct Table 3.1.

Table 3.1: Form and number of different elements in $Var(\mathbf{X})$.

Form of lag $\mathbf{i}$	Number of different lags	Number of appearances
$[0, \cdots, 0, 0]$	1	$N_1 \cdots N_{d-1} N_d$
$[0, \cdots, 0, i_d]$, $i_d > 0$	$Q^{(1)}_{N_d} = N_d - 1$	$N_1 \cdots N_{d-1}(N_d - \lvert i_d \rvert)$
$[0, \cdots, 0, i_d]$, $i_d < 0$	repeated lags	$N_1 \cdots N_{d-1}(N_d - \lvert i_d \rvert)$
$[0, \cdots, i_{d-1}, i_d]$, $i_{d-1} > 0$	$Q^{(2)}_{N_{d-1}N_d} - Q^{(1)}_{N_d} =$ $(N_{d-1} - 1)(2N_d - 1)$	$N_1 \cdots N_{d-2} \prod_{k=d-1}^{d}(N_k - \lvert i_k \rvert)$
$[0, \cdots, i_{d-1}, i_d]$, $i_{d-1} < 0$	repeated lags	$N_1 \cdots N_{d-2} \prod_{k=d-1}^{d}(N_k - \lvert i_k \rvert)$
$\vdots$	$\vdots$	$\vdots$
$[i_1, \cdots, i_{d-1}, i_d]$, $i_1 > 0$	$Q^{(d)}_{N_1 \cdots N_d} - Q^{(d-1)}_{N_2 \cdots N_d} =$ $(N_1 - 1)\prod_{k=2}^{d}(2N_k - 1)$	$\prod_{k=1}^{d}(N_k - \lvert i_k \rvert)$
$[i_1, \cdots, i_{d-1}, i_d]$, $i_1 < 0$	repeated lags	$\prod_{k=1}^{d}(N_k - \lvert i_k \rvert)$

The first column demonstrates all the different types of lags in the covariance matrix. For each one of these types, we may find in the second column the number of different auto-covariances. Thus, the sum of the elements of the second column is exactly what we are looking for. For example, in the variance matrix we can find the auto-covariance

at lag $[0,\cdots,0]$. According to the third column, the variance of the process appears in the main diagonal of the matrix $(N_1\cdots N_{d-1}N_d)$ times. Now, if we want to find how many lags of the form $[0,\cdots,0,i_d]$, $i_d>0$ can be generated by the data, that means we can only select (N_d-1) lags, *i.e.* $i_d=1,\cdots,N_d-1$. On the other hand, when the lag i_d is unrestricted, like in the next case where its previous lag i_{d-1} is restricted, we may have $i_d=0,\pm1,\cdots,\pm(N_d-1)$ and, so, $2(N_d-1)+1=2N_d-1$ different lags for each fixed i_{d-1}. We may write

$$Q^{(d)}_{N_1\cdots N_d}=(N_d-1)+(N_{d-1}-1)(2N_d-1)+\cdots+(N_1-1)\prod_{k=2}^{d}(2N_k-1). \quad (3.2.22)$$

Although finding $Q^{(d)}$ seems complicated, all someone needs to know is that the formula treats all N_k, $k=1,2,\cdots,d$, equally. As a result, we may look at the last term that is the only one involving N_1 and derive (3.2.21). ■

Remark 3.1. (i) We may see what happens when at least one of the N_k, $k=1,2,\cdots,d$, is equal to one. Without loss of generality, we may consider $N_1=1$. Then the last term of (3.2.22) is equal to zero and the rest of the formula remains unaffected and equal to $Q^{(d-1)}_{N_2\cdots N_d}$.

(ii) In the special case where the observations lie on a hyper-cube and $N_1=\cdots=N_d=n$ the formula simplifies to

$$Q^{(d)}_n=\frac{\left[(2n-1)^d-1\right]}{2}, \quad (3.2.23)$$

either if d is odd or even. Indeed, $(2n-1)$ is an odd number and so is $(2n-1)^2=4n^2-4n+1$, and, by induction, the numbers $(2n-1)^d$, $d=1,2,\cdots$, are odd in general.

3.3 Estimation for AR processes

3.3.1 Original Yule-Walker equations for the auto-regression

For $\mathbf{0}<\mathbf{i}_1<\cdots<\mathbf{i}_p$, we consider the causal auto-regression $\{X(\mathbf{v}),\ \mathbf{v}^\tau\in\mathcal{Z}^d\}$, such that it satisfies

$$X(\mathbf{v})\equiv\sum_{n=1}^{p}\varphi_{\mathbf{i}_n}X(\mathbf{v}-\mathbf{i}_n)+\varepsilon(\mathbf{v}),\ \{\varepsilon(\mathbf{v})\}\sim WN(0,\sigma^2). \quad (3.3.1)$$

If we multiply by $X(\mathbf{v}-\mathbf{i}_m)$, $m=1,\cdots,p$, and find the expected values, then due to the assumption of causality, it holds for $\gamma(\mathbf{i})=E\{X(\mathbf{v})X(\mathbf{v}+\mathbf{i})\}$, that

$$\gamma(\mathbf{i}_m)=\sum_{n=1}^{p}\varphi_{\mathbf{i}_n}\gamma(\mathbf{i}_m-\mathbf{i}_n),\ m=1,\cdots,p, \tag{3.3.2}$$

or, similarly,

$$\boldsymbol{\gamma}_p=\boldsymbol{\Gamma}_p\cdot\boldsymbol{\varphi}, \tag{3.3.3}$$

where

$$\boldsymbol{\gamma}_p\equiv[\gamma(\mathbf{i}_1),\cdots,\gamma(\mathbf{i}_p)]^\tau \tag{3.3.4}$$

and

$$\boldsymbol{\Gamma}_p\equiv\begin{bmatrix}\gamma(\mathbf{0}) & \gamma(\mathbf{i}_1-\mathbf{i}_2) & \cdots & \gamma(\mathbf{i}_1-\mathbf{i}_p)\\ \gamma(\mathbf{i}_2-\mathbf{i}_1) & \gamma(\mathbf{0}) & \cdots & \gamma(\mathbf{i}_2-\mathbf{i}_p)\\ \vdots & & & \vdots\\ \gamma(\mathbf{i}_p-\mathbf{i}_1) & \gamma(\mathbf{i}_p-\mathbf{i}_2) & \cdots & \gamma(\mathbf{0})\end{bmatrix} \tag{3.3.5}$$

and, finally,

$$\boldsymbol{\varphi}=[\varphi_{\mathbf{i}_1},\cdots,\varphi_{\mathbf{i}_p}]^\tau. \tag{3.3.6}$$

Finally, if we multiply (3.3.1) by $X(\mathbf{v})$ and find the expected value, we write the equation

$$\gamma(\mathbf{0})=\boldsymbol{\gamma}_p^\tau\cdot\boldsymbol{\varphi}+\sigma^2. \tag{3.3.7}$$

The equations (3.3.2) are the theoretical Yule-Walker equations for a causal autoregression on the d-dimensional lattice. In the next section and for a given set of observations from (3.3.1), we use these equations to estimate the parameters $\varphi_{\mathbf{i}_n}, n=1,\cdots,p$. Since the Yule-Walker equations involve the moments $\gamma(\mathbf{i}_n-\mathbf{i}_m), n,m=1,\cdots,p$, the Yule-Walker estimators are method of moments estimators.

3.3.2 Method of moments estimators

We observe $\{X(\mathbf{v}),\ \mathbf{v}^\tau\in\mathcal{S}\}$, where $\mathcal{S}\subset\mathcal{Z}^d$ is a set of finite cardinality. We wish to estimate the unknown parameters $\boldsymbol{\varphi}_0=[\varphi_{\mathbf{i}_1,0},\cdots,\varphi_{\mathbf{i}_p,0}]^\tau$. We consider the maximal set $\mathcal{S}^*$, such that $\mathbf{v}^\tau\in\mathcal{S}^*$ if $\mathbf{v}^\tau\in\mathcal{S}$ and $\mathbf{v}^\tau-\mathbf{i}_n^\tau\in\mathcal{S}$ for all $n=1,\cdots,p$. We assume that $\mathcal{S}$ is large enough, so that $\mathcal{S}^*$ is not the empty set. Also, we consider N and N^* the cardinalities of the sets $\mathcal{S}$ and $\mathcal{S}^*$, respectively. We assume that $\Theta_1\subset\mathcal{R}^p$ is the parameter space and that the following condition holds.

(C2) The parameter space Θ_1 is a compact set containing the true value φ_0 as an inner point. Further, for any $\varphi \in \Theta_1$, the auto-regression (3.3.1) is causal.

For the elements of $\mathcal{S}^*$ in the ascending order, say $\mathbf{v}_1^\tau < \cdots < \mathbf{v}_{N^*}^\tau$, we define the Yule-Walker estimators $\varphi^* = [\varphi_{\mathbf{i}_1}^*, \cdots, \varphi_{\mathbf{i}_p}^*]^\tau$ to be such that

$$\varphi^* \equiv (\mathbf{X}^{*\tau}\mathbf{X}^*)^{-1}(\mathbf{X}^{*\tau}\mathbf{Y}^*), \tag{3.3.8}$$

where

$$\mathbf{X}^* \equiv \begin{bmatrix} X(\mathbf{v}_1 - \mathbf{i}_1) & X(\mathbf{v}_1 - \mathbf{i}_2) & \cdots & X(\mathbf{v}_1 - \mathbf{i}_p) \\ X(\mathbf{v}_2 - \mathbf{i}_1) & X(\mathbf{v}_2 - \mathbf{i}_2) & \cdots & X(\mathbf{v}_2 - \mathbf{i}_p) \\ \vdots & & & \vdots \\ X(\mathbf{v}_{N^*} - \mathbf{i}_1) & X(\mathbf{v}_{N^*} - \mathbf{i}_2) & \cdots & X(\mathbf{v}_{N^*} - \mathbf{i}_p) \end{bmatrix} \tag{3.3.9}$$

and

$$\mathbf{Y}^* \equiv [X(\mathbf{v}_1), \cdots, X(\mathbf{v}_{N^*})]^\tau. \tag{3.3.10}$$

We can see immediately that (3.3.8) are least squares estimators as well. This shows when we consider the linear equations

$$X(\mathbf{v}) = \sum_{n=1}^{p} \varphi_{\mathbf{i}_n} X(\mathbf{v} - \mathbf{i}_n) + \varepsilon(\mathbf{v}), \ \mathbf{v}^\tau \in \mathcal{S}^*. \tag{3.3.11}$$

Remark 3.2 (Consistency). We can re-write (3.3.8) as

$$\varphi^* = \left(\frac{\mathbf{X}^{*\tau}\mathbf{X}^*}{N}\right)^{-1} \left(\frac{\mathbf{X}^{*\tau}\mathbf{Y}^*}{N}\right). \tag{3.3.12}$$

This representation justifies why we call (3.3.8) the Yule-Walker estimators, as

$$\frac{\mathbf{X}^{*\tau}\mathbf{X}^*}{N^*}$$

imitates the theoretical matrix $\mathbf{\Gamma}_p$, and so does the vector

$$\frac{\mathbf{X}^{*\tau}\mathbf{Y}^*}{N^*}$$

for the vector $\gamma_{p,0}$. We use the zero sub-index for the quantities corresponding to the true parameter vector φ_0, except for the case of the variance matrix $\mathbf{\Gamma}_p$, which will also correspond to the true parameter vector.

If $\{\varepsilon(\mathbf{v})\} \sim IID(0, \sigma^2)$, then as we increase the number of observations $N \to \infty$, it holds that $N^*/N \to 1$, which implies that

$$\frac{1}{N} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} X(\mathbf{v} - \mathbf{i}_n) X(\mathbf{v}) \xrightarrow{P} \gamma_0(\mathbf{i}_n),$$

according to Proposition 2.2, and so

$$\widehat{\gamma}_p \equiv \frac{\mathbf{X}^{*\tau}\mathbf{Y}^*}{N} \xrightarrow{P} \gamma_{p,0}. \tag{3.3.13}$$

For the same reasons,

$$\widehat{\mathbf{\Gamma}}_p \equiv \frac{\mathbf{X}^{*\tau}\mathbf{X}^*}{N} \xrightarrow{P} \mathbf{\Gamma}_p, \tag{3.3.14}$$

so that

$$\boldsymbol{\varphi}^* = \widehat{\mathbf{\Gamma}}_p^{-1}\widehat{\gamma}_p \xrightarrow{P} \mathbf{\Gamma}_p^{-1}\,\gamma_{p,0} = \boldsymbol{\varphi}_0. \tag{3.3.15}$$

Finally, we define

$$\sigma^{2^*} \equiv \widehat{\gamma}(\mathbf{0}) - \widehat{\gamma}_p^{\tau}\,\boldsymbol{\varphi}^* \xrightarrow{P} \gamma_0(\mathbf{0}) - \gamma_{p,0}^{\tau}\,\boldsymbol{\varphi}_0 = \sigma^2, \tag{3.3.16}$$

where $\widehat{\gamma}(\mathbf{0}) \equiv \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} X(\mathbf{v})^2/N \xrightarrow{P} \gamma_0(\mathbf{0})$.

Asymptotic normality

Theorem 3.1. Let the variance matrix

$$\mathbf{W}_p \equiv \frac{1}{\sigma^2}\,\mathbf{\Gamma}_p.$$

If $\{\varepsilon(\mathbf{v})\} \sim IID(0, \sigma^2)$, then under conditions (C1) and (C2), it holds that

$$N^{1/2}[\boldsymbol{\varphi}^* - \boldsymbol{\varphi}_0] \xrightarrow{D} N(\mathbf{0}, \mathbf{W}_p^{-1})$$

as $N \to \infty$.

Proof. First we define

$$\boldsymbol{\varepsilon}^* = [\varepsilon(\mathbf{v}_1), \cdots, \varepsilon(\mathbf{v}_{N^*})]^\tau, \tag{3.3.17}$$

so that we can write the linear model

$$\mathbf{Y}^* = \mathbf{X}^*\boldsymbol{\varphi}_0 + \boldsymbol{\varepsilon}^*. \tag{3.3.18}$$

It holds that

$$\begin{aligned} N^{1/2}[\boldsymbol{\varphi}^* - \boldsymbol{\varphi}_0] &= N^{1/2}\{(\mathbf{X}^{*\tau}\mathbf{X}^*)^{-1}\mathbf{X}^{*\tau}\mathbf{Y}^* - \boldsymbol{\varphi}_0\} \\ &= N^{1/2}\{(\mathbf{X}^{*\tau}\mathbf{X}^*)^{-1}\mathbf{X}^{*\tau}(\mathbf{X}^*\boldsymbol{\varphi}_0 + \boldsymbol{\varepsilon}^*) - \boldsymbol{\varphi}_0\} \\ &= N^{1/2}\{(\mathbf{X}^{*\tau}\mathbf{X}^*)^{-1}\mathbf{X}^{*\tau}\boldsymbol{\varepsilon}^*\} = \{N(\mathbf{X}^{*\tau}\mathbf{X}^*)^{-1}\}\{N^{-1/2}\mathbf{X}^{*\tau}\boldsymbol{\varepsilon}^*\}. \end{aligned} \tag{3.3.19}$$

For the first part, we know from Remark 3.2 that

$$N(\mathbf{X}^{*\tau}\mathbf{X}^{*})^{-1} \xrightarrow{P} \mathbf{\Gamma}_p^{-1} \tag{3.3.20}$$

as $N \to \infty$. For the second part, we will show that for any fixed vector $\boldsymbol{\lambda} \in \mathcal{R}^p$ as $N \to \infty$ and (C1) holds, then

$$N^{-1/2}\boldsymbol{\lambda}^{\tau}\mathbf{X}^{*\tau}\boldsymbol{\varepsilon}^{*} \xrightarrow{D} N(0, \sigma^2\boldsymbol{\lambda}^{\tau}\mathbf{\Gamma}_p\boldsymbol{\lambda}), \tag{3.3.21}$$

which will imply the truth of the theorem from the Cramer-Wold device.

We write

$$N^{-1/2}\mathbf{X}^{*\tau}\boldsymbol{\varepsilon}^{*} \equiv N^{1/2} \cdot \frac{1}{N} \sum_{\mathbf{v}^{\tau} \in \mathcal{S}^{*}} \mathbf{U}(\mathbf{v}), \tag{3.3.22}$$

where

$$\mathbf{U}(\mathbf{v}) \equiv \begin{bmatrix} X(\mathbf{v}-\mathbf{i}_1) \\ X(\mathbf{v}-\mathbf{i}_2) \\ \vdots \\ X(\mathbf{v}-\mathbf{i}_p) \end{bmatrix} \varepsilon(\mathbf{v}), \ \mathbf{v}^{\tau} \in \mathcal{Z}^d. \tag{3.3.23}$$

We can see immediately that $E\{\mathbf{U}(\mathbf{v})\} = \mathbf{0}$ and that

$$\mathrm{Cov}\{\mathbf{U}(\mathbf{v}), \mathbf{U}(\mathbf{v}+\mathbf{j})\} = E\{\mathbf{U}(\mathbf{v})\mathbf{U}^{\tau}(\mathbf{v}+\mathbf{j})\} = \begin{cases} \sigma^2 \cdot \mathbf{\Gamma}_p, & \mathbf{j} = \mathbf{0} \\ \mathbf{O}_{p\times p}, & \text{otherwise} \end{cases}.$$

We may then write $\{\boldsymbol{\lambda}^{\tau}\mathbf{U}(\mathbf{v})\} \sim WN(0, \sigma^2\boldsymbol{\lambda}^{\tau}\mathbf{\Gamma}_p\ \boldsymbol{\lambda})$. We recall the MA($\infty$) representation of the auto-regression,

$$X(\mathbf{v}) = \varepsilon(\mathbf{v}) + \sum_{\mathbf{j}>\mathbf{0}} \Phi_{\mathbf{j},0}\ \varepsilon(\mathbf{v}-\mathbf{j})$$

and, for fixed positive integer K, we define a new process $\{X^{(K)}(\mathbf{v}),\ \mathbf{v}^{\tau} \in \mathcal{Z}^d\}$ by the equation

$$X^{(K)}(\mathbf{v}) \equiv \varepsilon(\mathbf{v}) + \sum_{\mathbf{j}^{\tau} \in \mathcal{B}_K} \Phi_{\mathbf{j},0}\ \varepsilon(\mathbf{v}-\mathbf{j}),$$

where the set $\mathcal{B}_K$ was defined back in (2.5.3). We also define

$$\mathbf{U}^{(K)}(\mathbf{v}) \equiv \begin{bmatrix} X^{(K)}(\mathbf{v}-\mathbf{i}_1) \\ X^{(K)}(\mathbf{v}-\mathbf{i}_2) \\ \vdots \\ X^{(K)}(\mathbf{v}-\mathbf{i}_p) \end{bmatrix} \varepsilon(\mathbf{v}), \ \mathbf{v}^{\tau} \in \mathcal{Z}^d, \tag{3.3.24}$$

where we can write again $\{\boldsymbol{\lambda}^\tau \mathbf{U}^{(K)}(\mathbf{v})\} \sim WN(0, \sigma^2 \boldsymbol{\lambda}^\tau \boldsymbol{\Gamma}_p^{(K)} \boldsymbol{\lambda})$, with the variance matrix

$$\boldsymbol{\Gamma}_p^{(K)} \equiv \frac{1}{\sigma^2} \mathrm{Var}\{\mathbf{U}^{(K)}(\mathbf{v})\}.$$

Moreover $\{\boldsymbol{\lambda}^\tau \mathbf{U}^{(K)}(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ is a strictly stationary K^*-dependent process for some finite and positive integer K^*. Then as $N \to \infty$ and (C1) holds, according to Theorem 2.2, we can write

$$N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \boldsymbol{\lambda}^\tau \mathbf{U}^{(K)}(\mathbf{v}) \xrightarrow{D} \boldsymbol{\lambda}^\tau \mathbf{V}^{(K)} \sim N(0, \sigma^2 \boldsymbol{\lambda}^\tau \boldsymbol{\Gamma}_p^{(K)} \boldsymbol{\lambda}). \tag{3.3.25}$$

Also as $K \to \infty$

$$\boldsymbol{\lambda}^\tau \boldsymbol{\Gamma}_p^{(K)} \boldsymbol{\lambda} \to \boldsymbol{\lambda}^\tau \boldsymbol{\Gamma}_p\ \boldsymbol{\lambda},$$

which implies that

$$\boldsymbol{\lambda}^\tau \mathbf{V}^{(K)} \xrightarrow{D} \boldsymbol{\lambda}^\tau \mathbf{V} \sim N(0, \sigma^2 \boldsymbol{\lambda}^\tau \boldsymbol{\Gamma}_p\ \boldsymbol{\lambda}). \tag{3.3.26}$$

We may conclude using Chebychev's inequality, which guarantees that

$$\begin{aligned} & P\left(|N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \boldsymbol{\lambda}^\tau \mathbf{U}(\mathbf{v}) - N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \boldsymbol{\lambda}^\tau \mathbf{U}^{(K)}(\mathbf{v})| > \epsilon \right) \\ \leq\ & (1/\epsilon^2) \cdot (N^*/N)\ \boldsymbol{\lambda}^\tau \mathrm{Var}\{\mathbf{U}(\mathbf{v}) - \mathbf{U}^{(K)}(\mathbf{v})\} \boldsymbol{\lambda} \to 0, \end{aligned}$$

as $K \to \infty$, since the (n, m)-th element of the matrix

$$E\{(\mathbf{U}(\mathbf{v}) - \mathbf{U}^{(K)}(\mathbf{v}))(\mathbf{U}(\mathbf{v}) - \mathbf{U}^{(K)}(\mathbf{v}))^\tau\}$$

is such that

$$\sigma^4 \sum_{\substack{\mathbf{j}, \mathbf{j}^* > 0,\ \mathbf{j}^\tau \mathbf{j}^{*\tau} \notin \mathcal{B}_K, \\ \mathbf{j} + \mathbf{j}_n = \mathbf{j}^* + \mathbf{j}_m}} \Phi_{\mathbf{j},0}\ \Phi_{\mathbf{j}^*,0} \to 0,$$

as $K \to \infty$. ■

3.3.3 Conditional likelihood estimation

For observations $\{X(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{S}\}$ from the causal auto-regression defined in (3.3.1), we may write the conditional Gaussian likelihood

$$L^*(\boldsymbol{\varphi}, \sigma^2) \propto \frac{1}{(\sigma^2)^{N^*/2}} \exp\{-\frac{1}{2\sigma^2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} [X(\mathbf{v}) - \sum_{n=1}^{p} \varphi_{\mathbf{j}_n}\ X(\mathbf{v} - \mathbf{j}_n)]^2\},\ \boldsymbol{\varphi} \in \Theta_1,\ \sigma^2 > 0, \tag{3.3.27}$$

where the notation was introduced back in Sections 3.3.1 and 3.3.2. If we consider the maximum likelihood estimators $\widehat{\varphi}$ and $\widehat{\sigma^2}$ that maximize it, we can see immediately that

$$\widehat{\varphi} = \varphi^*, \qquad \widehat{\sigma^2} = \sigma^{2*}$$

and these are exactly the same as the Yule-Walker estimators defined by (3.3.8) and (3.3.16). Thus, we know all about their asymptotic behavior from the previous section.

3.4 Estimation for MA processes

3.4.1 General Yule-Walker equations for the moving-average

For $\mathbf{0} < \mathbf{j}_1 < \cdots < \mathbf{j}_q$, we consider the invertible moving-average $\{Y(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$, such that it satisfies

$$Y(\mathbf{v}) \equiv \varepsilon(\mathbf{v}) + \sum_{m=1}^{q} \theta_{\mathbf{j}_m} \varepsilon(\mathbf{v} - \mathbf{j}_m),\ \{\varepsilon(\mathbf{v})\} \sim WN(0, \sigma^2). \tag{3.4.1}$$

For $\gamma(\mathbf{j}) \equiv E\{Y(\mathbf{v})Y(\mathbf{v}+\mathbf{j})\}$, the general Yule-Walker equations are given in (2.4.48). We can re-write them as

$$\sum_{\mathbf{j}^\tau \in \mathcal{F}} \gamma(\mathbf{j})\ c(\mathbf{j} - \mathbf{j}_m) = 0,\ m = 1, \cdots, q, \tag{3.4.2}$$

and

$$\sum_{\mathbf{j}^\tau \in \mathcal{F}} \gamma(\mathbf{j})\ c(\mathbf{j}) = \sigma^2. \tag{3.4.3}$$

The notation for the set $\mathcal{F} \subset \mathcal{Z}^d$ and the polynomial $c(\mathbf{z})$ was introduced back in Section 2.4.1. We have considered $\mathcal{F}$ the set of all vector lags, for which the auto-covariance function of $\{Y(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ is non-zero. Also, from the polynomial

$$\theta(\mathbf{z}) \equiv 1 + \sum_{m=1}^{q} \theta_{\mathbf{j}_m}\ \mathbf{z}^{\mathbf{j}_m}, \tag{3.4.4}$$

we have defined

$$c(\mathbf{z}) \equiv \theta(\mathbf{z})^{-1}\theta(\mathbf{z}^{-1})^{-1}. \tag{3.4.5}$$

The equations (3.4.2) are the theoretical Yule-Walker equations, which are used here for an invertible moving-average on the d-dimensional lattice. In the next section and for a given set of observations from (3.4.1), we use these equations to estimate the parameters $\theta_{\mathbf{j}_m}, m = 1, \cdots, q$. Since the Yule-Walker equations involve the moments $\gamma(\mathbf{j}), \mathbf{j}^\tau \in \mathcal{F}$, the Yule-Walker estimators are method of moments estimators.

3.4.2 Method of moments estimators

We observe $\{Y(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{S}\}$ and wish to estimate the unknown parameters $\boldsymbol{\theta}_0 = [\theta_{\mathbf{j}_1,0}, \cdots, \theta_{\mathbf{j}_q,0}]^\tau$. We consider the maximal set $\mathcal{S}^*$, such that for every $\mathbf{v}^\tau \in \mathcal{S}^*$, it holds that $\mathbf{v}^\tau - \mathbf{j}^\tau \in \mathcal{S}$ for all $\mathbf{j}^\tau \in \mathcal{F}$. We assume that $\mathcal{S}$ is large enough, so that $\mathcal{S}^*$ is not the empty set. Also, we consider N and N^* the cardinalities of the sets $\mathcal{S}$ and $\mathcal{S}^*$, respectively. For any $\mathbf{v}^\tau \in \mathcal{Z}^d$, we define the set $\mathcal{F}_\mathbf{v} \subset \mathcal{Z}$ to be such that $\mathbf{j}^\tau \in \mathcal{F}_\mathbf{v}$ if $\mathbf{v}^\tau - \mathbf{j}^\tau \in \mathcal{S}$. We assume that $\Theta_2 \subset \mathcal{R}^q$ is the parameter space and that the following condition holds.

(C3) The parameter space Θ_2 is a compact set containing the true value $\boldsymbol{\theta}_0$ as an inner point. Further, for any $\boldsymbol{\theta} \in \Theta_2$, the moving-average (3.4.1) is invertible.

We define the estimators $\boldsymbol{\theta}^* = [\theta^*_{\mathbf{j}_1}, \cdots, \theta^*_{\mathbf{j}_q}]^\tau$ to be the solutions of the equations

$$\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \{ \sum_{\mathbf{j}^\tau + \mathbf{j}_m^\tau \in \mathcal{F}_\mathbf{v}} c^*(\mathbf{j})\ Y(\mathbf{v} - \mathbf{j}_m - \mathbf{j})\} Y(\mathbf{v}) \equiv 0,\ m = 1, \cdots, q, \tag{3.4.6}$$

where we consider

$$c^*(\mathbf{z}) = \theta^*(\mathbf{z})^{-1}\theta^*(\mathbf{z}^{-1})^{-1}$$

and

$$\theta^*(\mathbf{z}) \equiv 1 + \sum_{m=1}^{q} \theta^*_{\mathbf{j}_m} \mathbf{z}^{\mathbf{j}_m}.$$

Also, we set the estimator of the variance

$$\sigma^{2^*} \equiv \sum_{\mathbf{j}^\tau \in \mathcal{F}} c^*(\mathbf{j}) \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y(\mathbf{v})Y(\mathbf{v} - \mathbf{j})/N. \tag{3.4.7}$$

Remark 3.3 (Consistency). In general, we will denote with zero sub-index the quantities corresponding to the true parameter vector $\boldsymbol{\theta}_0$. We can re-write (3.4.6) as

$$\begin{aligned} &\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \sum_{\mathbf{j}^\tau + \mathbf{j}_m^\tau \in \mathcal{F}_\mathbf{v}} c^*(\mathbf{j})\ Y(\mathbf{v} - \mathbf{j}_m - \mathbf{j})Y(\mathbf{v})/N = \\ &\sum_{\mathbf{j}^\tau + \mathbf{j}_m^\tau \in \mathcal{Z}^d} c^*(\mathbf{j}) \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y(\mathbf{v} - \mathbf{j}_m - \mathbf{j})Y(\mathbf{v})/N \\ -\ &\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \sum_{\mathbf{j}^\tau + \mathbf{j}_m^\tau \notin \mathcal{F}_\mathbf{v}} c^*(\mathbf{j})\ Y(\mathbf{v} - \mathbf{j}_m - \mathbf{j})Y(\mathbf{v})/N = 0. \end{aligned} \tag{3.4.8}$$

As we increase the number of observations $N \to \infty$, then it holds that $N^*/N \to 1$, which combined with the fact that $\{\varepsilon(\mathbf{v})\} \sim IID(0, \sigma^2)$ implies that

$$\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y(\mathbf{v} - \mathbf{j}_m - \mathbf{j})Y(\mathbf{v})/N \xrightarrow{P} \gamma_0(\mathbf{j} + \mathbf{j}_m), \tag{3.4.9}$$

according to Proposition 2.2. Then the first term of the left hand-side of (3.4.8) converges to

$$\sum_{\mathbf{j}^\tau+\mathbf{j}_m^\tau\in\mathcal{Z}^d} \gamma_0(\mathbf{j}+\mathbf{j}_m)\ c^*(\mathbf{j}) = \sum_{\mathbf{j}^\tau+\mathbf{j}_m^\tau\in\mathcal{F}} \gamma_0(\mathbf{j}+\mathbf{j}_m)\ c^*(\mathbf{j}).$$

For the second term of the left hand-side in (3.4.8), it holds that

$$\begin{aligned}
& E|\sum_{\mathbf{v}^\tau\in\mathcal{S}^*}\sum_{\mathbf{j}^\tau+\mathbf{j}_m^\tau\notin\mathcal{F}_\mathbf{v}} c^*(\mathbf{j})\ Y(\mathbf{v}-\mathbf{j}_m-\mathbf{j})Y(\mathbf{v})/N| \\
\leq\ & \frac{1}{N}\sum_{\mathbf{v}^\tau\in\mathcal{S}^*}\sum_{\mathbf{j}^\tau+\mathbf{j}_m^\tau\notin\mathcal{F}_\mathbf{v}} E|c^*(\mathbf{j})\ Y(\mathbf{v}-\mathbf{j}_m-\mathbf{j})Y(\mathbf{v})| \\
\leq\ & \frac{1}{N}\sum_{\mathbf{v}^\tau\in\mathcal{S}^*}\sum_{\mathbf{j}^\tau+\mathbf{j}_m^\tau\notin\mathcal{F}_\mathbf{v}} E\{c^*(\mathbf{j})^2\}^{1/2}E\{Y(\mathbf{v}-\mathbf{j}_m-\mathbf{j})^2Y(\mathbf{v})^2\}^{1/2} \\
=\ & E\{Y(\mathbf{v})^2\}\ \frac{1}{N}\sum_{\mathbf{v}^\tau\in\mathcal{S}^*}\sum_{\mathbf{j}^\tau+\mathbf{j}_m^\tau\notin\mathcal{F}_\mathbf{v}} E\{c^*(\mathbf{j})^2\}^{1/2},
\end{aligned}$$

due to the Cauchy-Schwartz inequality and the independence of $Y(\mathbf{v})$, $Y(\mathbf{v}-\mathbf{j})$, $\mathbf{j}^\tau\notin\mathcal{F}$. Now for any vector $\boldsymbol{\theta}\in\Theta_2$, it holds that $c(\cdot)$ is the auto-covariance function of a causal auto-regression. According to Remark 2.2(ii), we can always find constants $C(\boldsymbol{\theta})>0$ and $\alpha(\boldsymbol{\theta})\in(0,1)$, such that

$$c(\mathbf{j})^2 \leq C(\boldsymbol{\theta})\ \alpha(\boldsymbol{\theta})^{\sum_{k=1}^d|j_k|}. \tag{3.4.10}$$

Similarly, for the estimator $\boldsymbol{\theta}^*\in\Theta_2$, we can write

$$c^*(\mathbf{j})^2 \leq C(\boldsymbol{\theta}^*)\alpha(\boldsymbol{\theta}^*)^{\sum_{k=1}^d|j_k|} \leq \sup_{\boldsymbol{\theta}\in\Theta_2} C(\boldsymbol{\theta})\alpha(\boldsymbol{\theta})^{\sum_{k=1}^d|j_k|} \leq \sup_{\boldsymbol{\theta}\in\Theta_2} C(\boldsymbol{\theta})\{\sup_{\boldsymbol{\theta}\in\Theta_2}\alpha(\boldsymbol{\theta})\}^{\sum_{k=1}^d|j_k|}, \tag{3.4.11}$$

with probability 1 and

$$E\{c^*(\mathbf{j})^2\} \leq \sup_{\boldsymbol{\theta}\in\Theta_2} C(\boldsymbol{\theta})\{\sup_{\boldsymbol{\theta}\in\Theta_2}\alpha(\boldsymbol{\theta})\}^{\sum_{k=1}^d|j_k|}. \tag{3.4.12}$$

For the case of observations on a hyper-rectangle when (C1)(ii) holds, one can easily verify that

$$\sum_{\mathbf{v}^\tau\in\mathcal{S}^*}\sum_{\mathbf{j}^\tau+\mathbf{j}_m^\tau\notin\mathcal{F}_\mathbf{v}} E\{c^*(\mathbf{j})^2\} = O(N^{(d-1)/d}). \tag{3.4.13}$$

For example, we can see the arguments of Yao and Brockwell (2006) for the case $d=2$. In general, we can write that

$$\frac{1}{N}\sum_{\mathbf{v}^\tau\in\mathcal{S}^*}\sum_{\mathbf{j}^\tau+\mathbf{j}_m^\tau\notin\mathcal{F}_\mathbf{v}} E\{c^*(\mathbf{j})^2\}^{1/2} \to 0 \tag{3.4.14}$$

and that

$$\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \sum_{\mathbf{j}^\tau + \mathbf{j}_m^\tau \notin \mathcal{F}_\mathbf{v}} c^*(\mathbf{j})\ Y(\mathbf{v} - \mathbf{j}_m - \mathbf{j}) Y(\mathbf{v})/N \xrightarrow{P} 0, \tag{3.4.15}$$

as (C1)(i) holds.

After combining the two results for the terms of (3.4.8), we may write

$$0 = \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \sum_{\mathbf{j}^\tau + \mathbf{j}_m^\tau \in \mathcal{F}_\mathbf{v}} c^*(\mathbf{j})\ Y(\mathbf{v} - \mathbf{j}_m - \mathbf{j}) Y(\mathbf{v})/N \xrightarrow{P} \sum_{\mathbf{j}^\tau + \mathbf{j}_m^\tau \in \mathcal{F}} \gamma_0(\mathbf{j} + \mathbf{j}_m)\ c^*(\mathbf{j}) \tag{3.4.16}$$

exactly like the theoretical equivalent (3.4.2) dictates. Thus

$$\boldsymbol{\theta}^* \xrightarrow{P} \boldsymbol{\theta}_0, \tag{3.4.17}$$

as $N \to \infty$ and (C1)(i) holds. Finally, from (3.4.7) we can see immediately that

$$\sigma^{2*} \xrightarrow{P} \sum_{\mathbf{j}^\tau \in \mathcal{F}} c_0(\mathbf{j})\ \gamma_0(\mathbf{j}) = \sigma^2, \tag{3.4.18}$$

since

$$\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y(\mathbf{v}) Y(\mathbf{v} - \mathbf{j})/N \xrightarrow{P} \gamma_0(\mathbf{j}).$$

Asymptotic normality

For mathematical convenience, we define the new variable

$$H_Y(\mathbf{v}) \equiv \begin{cases} Y(\mathbf{v}), & \mathbf{v}^\tau \in \mathcal{S} \\ 0, & \text{otherwise} \end{cases} \tag{3.4.19}$$

and we re-write (3.4.6) as

$$\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \{ \sum_{\mathbf{j}^\tau \in \mathcal{Z}^d} c^*(\mathbf{j})\ H_Y(\mathbf{v} - \mathbf{j}_m - \mathbf{j}) \} Y(\mathbf{v}) \equiv 0,\ m = 1, \cdots, q, \tag{3.4.20}$$

or

$$\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \{ \sum_{\mathbf{j}^\tau \in \mathcal{Z}^d} c_0(\mathbf{j})\ H_Y(\mathbf{v} - \mathbf{j}_m - \mathbf{j}) \} Y(\mathbf{v}) - \mathbf{J}_m[\boldsymbol{\theta}^* - \boldsymbol{\theta}_0] = 0,\ m = 1, \cdots, q, \tag{3.4.21}$$

where we define

$$\mathbf{J}_m \equiv [J_{m,1}, \cdots, J_{m,q}],\ m = 1, \cdots, q, \tag{3.4.22}$$

with elements $m, n = 1, \cdots, q$, equal to

$$\begin{aligned} J_{m,n} \equiv & \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \{ c_0(\mathbf{B}) [\theta_0(\mathbf{B})^{-1} H_Y(\mathbf{v} - \mathbf{j}_m - \mathbf{j}_n) + \theta_0(\mathbf{B}^{-1})^{-1} H_Y(\mathbf{v} - \mathbf{j}_m + \mathbf{j}_n)] \} Y(\mathbf{v}) \\ + & O_P(N ||\boldsymbol{\theta}^* - \boldsymbol{\theta}_0||). \end{aligned} \tag{3.4.23}$$

Equations (3.4.21) may also be re-written as

$$\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \{c_0(\mathbf{B}) H_Y(\mathbf{v} - \mathbf{j}_m)\} Y(\mathbf{v}) - \mathbf{J}_m[\boldsymbol{\theta}^* - \boldsymbol{\theta}_0] = 0, \ m = 1, \cdots, q. \tag{3.4.24}$$

If we stack all the q equations together, we can write

$$N^{1/2}[\boldsymbol{\theta}^* - \boldsymbol{\theta}_0] = \{\mathbf{J}/N\}^{-1}\{N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \mathbf{H}_Y(\mathbf{v})\}, \tag{3.4.25}$$

where

$$\mathbf{J}^\tau \equiv [\mathbf{J}_1^\tau, \cdots, \mathbf{J}_q^\tau] \tag{3.4.26}$$

and also for any $\mathbf{v}^\tau \in \mathcal{Z}^d$,

$$\mathbf{H}_Y(\mathbf{v}) \equiv \begin{bmatrix} c_0(\mathbf{B}) H_Y(\mathbf{v} - \mathbf{j}_1) \\ c_0(\mathbf{B}) H_Y(\mathbf{v} - \mathbf{j}_2) \\ \vdots \\ c_0(\mathbf{B}) H_Y(\mathbf{v} - \mathbf{j}_q) \end{bmatrix} Y(\mathbf{v}). \tag{3.4.27}$$

Proposition 3.2. Let the polynomial

$$\theta_0(\mathbf{z})^{-1} \equiv 1 + \sum_{\mathbf{j}>\mathbf{0}} \Theta_{\mathbf{j},0} \, \mathbf{z}^{\mathbf{j}}.$$

If $\{\varepsilon(\mathbf{v})\} \sim IID(0, \sigma^2)$, then under conditions (C1)(i) and (C3), it holds that

$$\mathbf{J}/N \xrightarrow{P} \sigma^2 \cdot \boldsymbol{\Theta}_0 \equiv \sigma^2 \cdot \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \Theta_{\mathbf{j}_2 - \mathbf{j}_1, 0} & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \\ \Theta_{\mathbf{j}_q - \mathbf{j}_1, 0} & \Theta_{\mathbf{j}_q - \mathbf{j}_2, 0} & \Theta_{\mathbf{j}_q - \mathbf{j}_3, 0} & & 1 \end{bmatrix} \tag{3.4.28}$$

as $N \to \infty$.

Proof. Looking back at the (m,n)-th element of $\mathbf{J}/N$, $m, n = 1, \cdots, q$, from (3.4.23) and due to the consistency of the estimators, it suffices to look at

$$\begin{aligned} & 1/N \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \{c_0(\mathbf{B})[\theta_0(\mathbf{B})^{-1} H_Y(\mathbf{v} - \mathbf{j}_m - \mathbf{j}_n) + \theta_0(\mathbf{B}^{-1})^{-1} H_Y(\mathbf{v} - \mathbf{j}_m + \mathbf{j}_n)]\} Y(\mathbf{v}) \\ = \ & 1/N \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \{\theta_0(\mathbf{B})^{-1} c_0(\mathbf{B}) Y(\mathbf{v} - \mathbf{j}_m - \mathbf{j}_n) + \theta_0(\mathbf{B}^{-1})^{-1} c_0(\mathbf{B}) Y(\mathbf{v} - \mathbf{j}_m + \mathbf{j}_n)\} Y(\mathbf{v}) \\ + \ & o_P(1) \end{aligned} \tag{3.4.29}$$

If we consider the polynomial

$$d_0(\mathbf{z}) \equiv \theta_0(\mathbf{z})^{-1} c_0(\mathbf{z}) \equiv \sum_{\mathbf{i}^\tau \in \mathcal{Z}^d} d_{\mathbf{i},0} \, \mathbf{z}^{\mathbf{i}},$$

then the last part of (3.4.29) follows from the fact that

$$\begin{aligned} & E|1/N \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \{ \sum_{\mathbf{i}^\tau \notin \mathcal{F}_\mathbf{v}} d_{\mathbf{i},0}\, Y(\mathbf{v}-\mathbf{i})\} Y(\mathbf{v})| \\ \leq\ & 1/N \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \sum_{\mathbf{i}^\tau \notin \mathcal{F}_\mathbf{v}} |d_{\mathbf{i},0}|\, E|Y(\mathbf{v}-\mathbf{i})Y(\mathbf{v})| \\ =\ & \{E|Y(\mathbf{v})|\}^2\, 1/N \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \sum_{\mathbf{i}^\tau \notin \mathcal{F}_\mathbf{v}} |d_{\mathbf{i},0}| \to 0 \end{aligned}$$

as $N \to \infty$ and (C1)(i) holds. The last limit comes from the same argument as before. For example, if (C1)(ii) is true, we can write $\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \sum_{\mathbf{i}^\tau \notin \mathcal{F}_\mathbf{v}} |d_{\mathbf{i},0}| = O(N^{(d-1)/d})$, since for any $\mathbf{i}^\tau \in \mathcal{Z}^d$ it holds that $|d_{\mathbf{i},0}| \leq C\alpha^{\sum_{k=1}^d |i_k|}$ for constants $C > 0$ and $\alpha \in (0,1)$ as well. We may take similar action for the polynomial $\theta_0(\mathbf{z}^{-1})^{-1}c_0(\mathbf{z})$.

Next, we proceed by defining some new processes, like in Section 2.4.1. From $\{\varepsilon(\mathbf{v})\} \sim IID(0,\sigma^2)$, we have generated the moving-average process

$$Y(\mathbf{v}) = \theta_0(\mathbf{B})\varepsilon(\mathbf{v}), \tag{3.4.30}$$

but also the auto-regressive process

$$\theta_0(\mathbf{B}^{-1})X(\mathbf{v}) \equiv \varepsilon(\mathbf{v}) \tag{3.4.31}$$

and it holds that

$$X(\mathbf{v}) = \theta_0(\mathbf{B})^{-1}\theta_0(\mathbf{B}^{-1})^{-1}Y(\mathbf{v}) = c_0(\mathbf{B})Y(\mathbf{v}), \tag{3.4.32}$$

and $X(\mathbf{v}+\mathbf{j})$ is uncorrelated with $Y(\mathbf{v})$ for any $\mathbf{j} \neq \mathbf{0}$ according to (2.4.22). Therefore, we can re-write the (m,n)-th element referred in (3.4.29) as

$$1/N \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \{\theta_0(\mathbf{B})^{-1}X(\mathbf{v}-\mathbf{j}_m-\mathbf{j}_n) + \theta_0(\mathbf{B}^{-1})^{-1}X(\mathbf{v}-\mathbf{j}_m+\mathbf{j}_n)\}Y(\mathbf{v}) + o_P(1). \tag{3.4.33}$$

According to (3.4.30) and (3.4.31), the processes $\{Y(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ and $\{X(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ can be written as linear combinations of independent and identically distributed random variables, and it holds, according to Proposition 2.2, that

$$\frac{1}{N} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} X(\mathbf{v}-\mathbf{j})\, Y(\mathbf{v}) \xrightarrow{P} E\{X(\mathbf{v}-\mathbf{j})\, Y(\mathbf{v})\} = \begin{cases} \sigma^2, & \mathbf{j} = \mathbf{0} \\ 0, & \mathbf{j} \neq \mathbf{0} \end{cases} \tag{3.4.34}$$

as $N \to \infty$. As a result,

$$E\{\theta_0(\mathbf{B})^{-1}X(\mathbf{v}-\mathbf{j}_m-\mathbf{j}_n)\, Y(\mathbf{v})\} = 0 \tag{3.4.35}$$

and

$$E\{\theta_0(\mathbf{B}^{-1})^{-1}X(\mathbf{v}-\mathbf{j}_m+\mathbf{j}_n)\ Y(\mathbf{v})\} = \sigma^2 \cdot \Theta_{\mathbf{j}_m-\mathbf{j}_n,0},\ \mathbf{j}_m \geq \mathbf{j}_n. \tag{3.4.36}$$

Since we consider $\mathbf{j}_1 < \mathbf{j}_2 < \cdots < \mathbf{j}_q$, the proposition has been proven. ■

Before we move on to the next theorem, we define the process

$$u(\mathbf{v}) \equiv \theta_0(\mathbf{B})X(\mathbf{v}), \tag{3.4.37}$$

for which it holds that

$$\{u(\mathbf{v})\} \sim WN(0, \sigma^2).$$

This was justified properly in Section 2.4.1.

Theorem 3.2. Let the processes $\{W(\mathbf{v})\} \sim IID(0,1)$ and

$$\theta_0(\mathbf{B})\eta(\mathbf{v}) \equiv W(\mathbf{v}).$$

Also let the vector $\boldsymbol{\xi} \equiv [\eta(-\mathbf{j}_1), \cdots, \eta(-\mathbf{j}_q)]^\tau$ and the variance matrix

$$\mathbf{W}_q^* \equiv \mathrm{Var}\{\boldsymbol{\xi} \mid W(-\mathbf{j}_1 - \mathbf{j}),\ \mathbf{j} > \mathbf{0},\ \mathbf{j} \neq \mathbf{j}_2 - \mathbf{j}_1, \cdots, \mathbf{j}_q - \mathbf{j}_1\}.$$

If $\{\varepsilon(\mathbf{v})\} \sim IID(0, \sigma^2)$ and $E\{\varepsilon(\mathbf{v})^4\} < \infty$, then under conditions (C1) and (C3), it holds that

$$N^{1/2}[\boldsymbol{\theta}^* - \boldsymbol{\theta}_0] \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Delta}) \tag{3.4.38}$$

as $N \to \infty$. Otherwise, if $\{\varepsilon(\mathbf{v})\}, \{u(\mathbf{v})\} \sim IID(0, \sigma^2)$ and $|E\{\varepsilon(\mathbf{v})^3\}| < \infty$, then under conditions (C1) and (C3), it holds that

$$N^{1/2}[\boldsymbol{\theta}^* - \boldsymbol{\theta}_0] \xrightarrow{D} N(\mathbf{0}, \mathbf{W}_q^{*-1}) \tag{3.4.39}$$

as $N \to \infty$.

Proof. First, we write for $m = 1, \cdots, q$,

$$\begin{aligned} & N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \sum_{\mathbf{j}^\tau \in \mathbb{Z}^d} c_0(\mathbf{j}) H_Y(\mathbf{v} - \mathbf{j}_m - \mathbf{j}) Y(\mathbf{v}) \\ = \ & N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \sum_{\mathbf{j}^\tau \in \mathbb{Z}^d} c_0(\mathbf{j}) Y(\mathbf{v} - \mathbf{j}_m - \mathbf{j}) Y(\mathbf{v}) + o_P(1), \end{aligned} \tag{3.4.40}$$

which might be justified first by the simple argument that $E\{Y(\mathbf{v})Y(\mathbf{v} - \mathbf{j}_m - \mathbf{j})\} = 0$ for any $\mathbf{v}^\tau \in \mathcal{S}^*$ and $\mathbf{j}_m^\tau + \mathbf{j}^\tau \notin \mathcal{F}_\mathbf{v}$. Then, we may look at the variance

$$\mathrm{Var}\{N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \sum_{\mathbf{j}^\tau + \mathbf{j}_m^\tau \notin \mathcal{F}_\mathbf{v}} c_0(\mathbf{j}) Y(\mathbf{v} - \mathbf{j}_m - \mathbf{j}) Y(\mathbf{v})\} = \frac{1}{N} \mathrm{Var}\{ \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \widetilde{u}_m(\mathbf{v})\}, \tag{3.4.41}$$

where we define for $\mathbf{v}^\tau \in \mathcal{S}^*$ the random variables

$$\widetilde{u}_m(\mathbf{v}) \equiv \sum_{\mathbf{j}^\tau + \mathbf{j}_m^\tau \notin \mathcal{F}_\mathbf{v}} c_0(\mathbf{j}) Y(\mathbf{v} - \mathbf{j}_m - \mathbf{j}) Y(\mathbf{v}). \tag{3.4.42}$$

Since (C1)(ii) holds, we can write that

$$\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \mathrm{Var}\{\widetilde{u}_m(\mathbf{v})\} = \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} E\{\widetilde{u}_m(\mathbf{v})^2\} = O(N^{(d-1)/d}) \tag{3.4.43}$$

and a similar argument can be written for the cross-terms due to the Cauchy-Schwartz inequality. For the case $d = 2$ and observations on a rectangle, we may find a justification for that in the paper of Yao and Brockwell (2006). We can then write

$$\mathrm{Var}\{N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \sum_{\mathbf{j}^\tau + \mathbf{j}_m^\tau \notin \mathcal{F}_\mathbf{v}} c_0(\mathbf{j}) Y(\mathbf{v} - \mathbf{j}_m - \mathbf{j}) Y(\mathbf{v})\} \to 0,$$

as $N \to \infty$ and (C1) holds, which results in the convergence in probability to 0.

Since $X(\mathbf{v}) = c_0(\mathbf{B}) Y(\mathbf{v})$, we may re-write (3.4.40) as

$$\begin{aligned} & N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \sum_{\mathbf{j}^\tau \in \mathcal{Z}^d} c_0(\mathbf{j}) H_Y(\mathbf{v} - \mathbf{j}_m - \mathbf{j}) Y(\mathbf{v}) \\ = \; & N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} X(\mathbf{v} - \mathbf{j}_m) Y(\mathbf{v}) + o_p(1). \end{aligned} \tag{3.4.44}$$

As a result, equation (3.4.25) can also be re-expressed as

$$N^{1/2}[\boldsymbol{\theta}^* - \boldsymbol{\theta}_0] = \{\mathbf{J}/N\}^{-1} \{N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \mathbf{U}(\mathbf{v}) + o_p(1)\}, \tag{3.4.45}$$

where

$$\mathbf{U}(\mathbf{v}) \equiv \begin{bmatrix} X(\mathbf{v} - \mathbf{j}_1) \\ X(\mathbf{v} - \mathbf{j}_2) \\ \vdots \\ X(\mathbf{v} - \mathbf{j}_q) \end{bmatrix} Y(\mathbf{v}). \tag{3.4.46}$$

For any positive integer K, we defined back in (2.5.4) the set $\mathcal{F}_K$. We also define the vector

$$\mathbf{U}^{(K)}(\mathbf{v}) \equiv \begin{bmatrix} \sum_{\mathbf{j}^\tau + \mathbf{j}_1^\tau \in \mathcal{F}_K} c_0(\mathbf{j}) Y(\mathbf{v} - \mathbf{j}_1 - \mathbf{j}) \\ \sum_{\mathbf{j}^\tau + \mathbf{j}_2^\tau \in \mathcal{F}_K} c_0(\mathbf{j}) Y(\mathbf{v} - \mathbf{j}_2 - \mathbf{j}) \\ \vdots \\ \sum_{\mathbf{j}^\tau + \mathbf{j}_q^\tau \in \mathcal{F}_K} c_0(\mathbf{j}) Y(\mathbf{v} - \mathbf{j}_q - \mathbf{j}) \end{bmatrix} Y(\mathbf{v}). \tag{3.4.47}$$

For any $n, m = 1, \cdots, q$, the (n, m)-th element of $E\{\mathbf{U}^{(K)}(\mathbf{v})\mathbf{U}^{(K)^T}(\mathbf{v}-\mathbf{j})\}$ is equal to

$$
\begin{aligned}
& \sum_{\mathbf{i}^T+\mathbf{j}_n^T \in \mathcal{F}_K} c_0(\mathbf{i}) \sum_{\mathbf{i}^{*T}+\mathbf{j}_m^T \in \mathcal{F}_K} c_0(\mathbf{i}^*)\, E\{Y(\mathbf{v}-\mathbf{j}_n-\mathbf{i})Y(\mathbf{v})Y(\mathbf{v}-\mathbf{j}-\mathbf{j}_m-\mathbf{i}^*)Y(\mathbf{v}-\mathbf{j})\} \\
= \; & E\{X(\mathbf{v}-\mathbf{j}_n)Y(\mathbf{v})X(\mathbf{v}-\mathbf{j}-\mathbf{j}_m)Y(\mathbf{v}-\mathbf{j})\} \\
- \; & \sum_{\mathbf{i}^T+\mathbf{j}_n^T \notin \mathcal{F}_K} c_0(\mathbf{i})\, E\{Y(\mathbf{v}-\mathbf{j}_n-\mathbf{i})Y(\mathbf{v})X(\mathbf{v}-\mathbf{j}-\mathbf{j}_m)Y(\mathbf{v}-\mathbf{j})\} \\
- \; & \sum_{\mathbf{i}^{*T}+\mathbf{j}_m^T \notin \mathcal{F}_K} c_0(\mathbf{i}^*)\, E\{X(\mathbf{v}-\mathbf{j}_n)Y(\mathbf{v})Y(\mathbf{v}-\mathbf{j}-\mathbf{j}_m-\mathbf{i}^*)Y(\mathbf{v}-\mathbf{j})\} \\
+ \; & \sum_{\mathbf{i}^T+\mathbf{j}_n^T \notin \mathcal{F}_K} c_0(\mathbf{i}) \sum_{\mathbf{i}^{*T}+\mathbf{j}_m^T \notin \mathcal{F}_K} c_0(\mathbf{i}^*)\, E\{Y(\mathbf{v}-\mathbf{j}_n-\mathbf{i})Y(\mathbf{v})Y(\mathbf{v}-\mathbf{j}-\mathbf{j}_m-\mathbf{i}^*)Y(\mathbf{v}-\mathbf{j})\} \\
\equiv \; & E\{X(\mathbf{v}-\mathbf{j}_n)Y(\mathbf{v})X(\mathbf{v}-\mathbf{j}-\mathbf{j}_m)Y(\mathbf{v}-\mathbf{j})\} \\
- \; & E\{r_n^*(\mathbf{v})X(\mathbf{v}-\mathbf{j}-\mathbf{j}_m)Y(\mathbf{v}-\mathbf{j})\} - E\{r_m^*(\mathbf{v}-\mathbf{j})X(\mathbf{v}-\mathbf{j}_n)Y(\mathbf{v})\} \\
+ \; & E\{r_n^*(\mathbf{v})r_m^*(\mathbf{v}-\mathbf{j})\},
\end{aligned}
\tag{3.4.48}
$$

where we have defined

$$
r_n^*(\mathbf{v}) \equiv \sum_{\mathbf{i}^T+\mathbf{j}_n^T \notin \mathcal{F}_K} c_0(\mathbf{i})\; Y(\mathbf{v}-\mathbf{j}_n-\mathbf{i})Y(\mathbf{v}),\; n = 1, \cdots, q.
$$

We consider that K is a positive integer. Then, for $\mathcal{F} \subseteq \mathcal{F}_K$ it holds that $Y(\mathbf{v})$ is independent of $Y(\mathbf{v}-\mathbf{j}), \mathbf{j}^T \notin \mathcal{F}_K$, and

$$
\begin{aligned}
E\{r_n^*(\mathbf{v})^2\} &= \sum_{\mathbf{i}^T+\mathbf{j}_n^T \notin \mathcal{F}_K} c_0(\mathbf{i}) \sum_{\mathbf{i}^{*T}+\mathbf{j}_n^T \notin \mathcal{F}_K} c_0(\mathbf{i}^*)\, E\{Y(\mathbf{v}-\mathbf{j}_n-\mathbf{i})Y(\mathbf{v}-\mathbf{j}_n-\mathbf{i}^*)Y^2(\mathbf{v})\} \\
&= E\{Y^2(\mathbf{v})\} \sum_{\mathbf{i}^T+\mathbf{j}_n^T \notin \mathcal{F}_K} c_0(\mathbf{i}) \sum_{\mathbf{i}^{*T}+\mathbf{j}_n^T \notin \mathcal{F}_K} c_0(\mathbf{i}^*)\, E\{Y(\mathbf{v}-\mathbf{j}_n-\mathbf{i})Y(\mathbf{v}-\mathbf{j}_n-\mathbf{i}^*)\} \\
&\leq C_1 \cdot \alpha_1^K,
\end{aligned}
\tag{3.4.49}
$$

for constants $C_1 > 0$ and $\alpha_1 \in (0, 1)$.

The first of four terms in (3.4.48) is the (n, m)-th element of $E\{\mathbf{U}(\mathbf{v})\mathbf{U}^T(\mathbf{v}-\mathbf{j})\}$. For

the other three terms, it holds, due to the Cauchy-Schwartz inequality and (3.4.49), that

$$\begin{aligned} |E\{r_n^*(\mathbf{v})X(\mathbf{v}-\mathbf{j}-\mathbf{j}_m)Y(\mathbf{v}-\mathbf{j})\}| &\leq [E\{r_n^*(\mathbf{v})^2\}]^{1/2}[E\{X(\mathbf{v}-\mathbf{j}-\mathbf{j}_m)^2Y(\mathbf{v}-\mathbf{j})^2\}]^{1/2} \\ &\leq C_2 \cdot \alpha_2^K, \qquad (3.4.50) \\ |E\{r_m^*(\mathbf{v}-\mathbf{j})X(\mathbf{v}-\mathbf{j}_n)Y(\mathbf{v})\}| &\leq [E\{r_m^*(\mathbf{v}-\mathbf{j})^2\}]^{1/2}[E\{X(\mathbf{v}-\mathbf{j}_n)^2Y(\mathbf{v})^2\}]^{1/2} \\ &\leq C_3 \cdot \alpha_3^K, \qquad (3.4.51) \\ |E\{r_n^*(\mathbf{v})r_m^*(\mathbf{v}-\mathbf{j})\}| &\leq [E\{r_n^*(\mathbf{v})^2\}]^{1/2}[E\{r_m^*(\mathbf{v}-\mathbf{j})^2\}]^{1/2} \\ &\leq C_4 \cdot \alpha_4^K, \qquad (3.4.52) \end{aligned}$$

for some constants $C_2, C_3, C_4 > 0$ and $\alpha_2, \alpha_3, \alpha_4 \in (0,1)$. For (3.4.50) and (3.4.51), we have assumed that either $E\{\varepsilon(\mathbf{v})^4\} < \infty$ or that $X(\mathbf{v}-\mathbf{j})$ and $Y(\mathbf{v})$ are independent for any $\mathbf{j} \neq \mathbf{0}$.

For any $\boldsymbol{\lambda} \in \mathcal{R}^q$, it holds in general that

$$E\{\boldsymbol{\lambda}^\tau \mathbf{U}^{(K)}(\mathbf{v})\} = 0 \qquad (3.4.53)$$

and that $\{\boldsymbol{\lambda}^\tau \mathbf{U}^{(K)}(\mathbf{v}), \mathbf{v}^\tau \in \mathcal{Z}^d\}$ is a strictly stationary and K^*-dependent process for some fixed finite integer number K^*. This implies that as $N \to \infty$ and (C1) holds,

$$N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \boldsymbol{\lambda}^\tau \mathbf{U}^{(K)}(\mathbf{v}) \xrightarrow{D} \boldsymbol{\lambda}^\tau \mathbf{V}_K \sim N(0, \boldsymbol{\lambda}^\tau \mathbf{M}_K \boldsymbol{\lambda}), \qquad (3.4.54)$$

where

$$\mathbf{M}_K \equiv \sum_{\mathbf{j}^\tau \in \mathcal{Z}^d} \boldsymbol{\Gamma}_K(\mathbf{j}), \qquad (3.4.55)$$

with

$$\boldsymbol{\Gamma}_K(\mathbf{j}) \equiv E\{\mathbf{U}^{(K)}(\mathbf{v})\mathbf{U}^{(K)^\tau}(\mathbf{v}-\mathbf{j})\}. \qquad (3.4.56)$$

Similarly, if we define

$$\boldsymbol{\Gamma}(\mathbf{j}) \equiv E\{\mathbf{U}(\mathbf{v})\mathbf{U}^\tau(\mathbf{v}-\mathbf{j})\} \qquad (3.4.57)$$

and

$$\mathbf{M} \equiv \sum_{\mathbf{j}^\tau \in \mathcal{Z}^d} \boldsymbol{\Gamma}(\mathbf{j}), \qquad (3.4.58)$$

it holds that

$$\boldsymbol{\lambda}^\tau \boldsymbol{\Gamma}_K(\mathbf{j})\boldsymbol{\lambda} \to \boldsymbol{\lambda}^\tau \boldsymbol{\Gamma}(\mathbf{j})\boldsymbol{\lambda}, \ \mathbf{j}^\tau \in \mathcal{Z}^d, \qquad (3.4.59)$$

as $K \to \infty$. This is thanks to (3.4.48), (3.4.50), (3.4.51) and (3.4.52). Then we may write

$$\boldsymbol{\lambda}^\tau \mathbf{V}_K \xrightarrow{D} \boldsymbol{\lambda}^\tau \mathbf{V} \sim N(0, \boldsymbol{\lambda}^\tau \mathbf{M}\, \boldsymbol{\lambda}) \qquad (3.4.60)$$

as $K \to \infty$. We may conclude using the same argument as in Theorem 3.1, where thanks to the Chebychev inequality, all we need to show is that

$$\boldsymbol{\lambda}^\tau \text{Var}\{\mathbf{U}(\mathbf{v}) - \mathbf{U}^{(K)}(\mathbf{v})\}\boldsymbol{\lambda} \leq C \cdot \alpha^K \to 0, \tag{3.4.61}$$

as $K \to \infty$ for some $C > 0$ and $\alpha \in (0,1)$. From the Cramer-Wold device we then write

$$N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \mathbf{U}(\mathbf{v}) \xrightarrow{D} N(\mathbf{0}, \mathbf{M}), \tag{3.4.62}$$

which combined with (3.4.45) and Proposition 3.2 gives

$$N^{1/2}[\boldsymbol{\theta}^* - \boldsymbol{\theta}_0] \xrightarrow{D} N(\mathbf{0}, \frac{1}{\sigma^4}\boldsymbol{\Theta}_0^{-1}\mathbf{M}\boldsymbol{\Theta}_0^{\tau-1}). \tag{3.4.63}$$

As a result, the first part of the theorem has been proven and

$$\boldsymbol{\Delta} = \frac{1}{\sigma^4}\boldsymbol{\Theta}_0^{-1}\mathbf{M}\boldsymbol{\Theta}_0^{\tau-1}.$$

We now let the vector $\mathbf{W} = [W(-\mathbf{j}_1), \cdots, W(-\mathbf{j}_q)]^\tau$ and then write

$$\begin{aligned} \boldsymbol{\xi} &= \begin{bmatrix} \eta(-\mathbf{j}_1) \\ \eta(-\mathbf{j}_2) \\ \vdots \\ \eta(-\mathbf{j}_q) \end{bmatrix} = \begin{bmatrix} W(-\mathbf{j}_1) + \Theta_{\mathbf{j}_2-\mathbf{j}_1,0}W(-\mathbf{j}_2) + \cdots + \Theta_{\mathbf{j}_q-\mathbf{j}_1,0}W(-\mathbf{j}_q) \\ W(-\mathbf{j}_2) + \cdots + \Theta_{\mathbf{j}_q-\mathbf{j}_2,0}W(-\mathbf{j}_q) \\ \vdots \\ W(-\mathbf{j}_q) \end{bmatrix} + \mathbf{R} \\ &= \boldsymbol{\Theta}_0^\tau \mathbf{W} + \mathbf{R}, \end{aligned} \tag{3.4.64}$$

where $\mathbf{R}$ is a $(q \times 1)$ random vector that is independent of $\mathbf{W}$ since it is a linear function of $W(-\mathbf{j}_1 - \mathbf{j})$, $\mathbf{j} > \mathbf{0}$, $\mathbf{j} \neq \mathbf{j}_2 - \mathbf{j}_1, \cdots, \mathbf{j}_q - \mathbf{j}_1$. As a result,

$$\begin{aligned} \mathbf{W}_q^* &= \text{Var}\{\boldsymbol{\xi} \mid W(-\mathbf{j}_1 - \mathbf{j}),\ \mathbf{j} > \mathbf{0},\ \mathbf{j} \neq \mathbf{j}_2 - \mathbf{j}_1, \cdots, \mathbf{j}_q - \mathbf{j}_1\} \\ &= \boldsymbol{\Theta}_0^\tau\ \text{Var}\{\mathbf{W}\}\ \boldsymbol{\Theta}_0 = \boldsymbol{\Theta}_0^\tau\ \mathbf{I}_q\ \boldsymbol{\Theta}_0 = \boldsymbol{\Theta}_0^\tau \boldsymbol{\Theta}_0. \end{aligned} \tag{3.4.65}$$

Finally, for any $\mathbf{j} \geq \mathbf{0}$, we can write the (n,m)-th element of $\boldsymbol{\Gamma}(\mathbf{j}) = E\{\mathbf{U}(\mathbf{v})\mathbf{U}^\tau(\mathbf{v}-\mathbf{j})\}$ to be equal to

$$\begin{aligned} & E\{X(\mathbf{v} - \mathbf{j}_n)X(\mathbf{v} - \mathbf{j}_m - \mathbf{j})Y(\mathbf{v})Y(\mathbf{v} - \mathbf{j})\} \\ &= E\{X(\mathbf{v} - \mathbf{j}_n)X(\mathbf{v} - \mathbf{j}_m - \mathbf{j})\}\ E\{Y(\mathbf{v})Y(\mathbf{v} - \mathbf{j})\} \\ &= \sigma^2\ c_0(\mathbf{j}_n - \mathbf{j}_m - \mathbf{j})\gamma_0(\mathbf{j}), \end{aligned} \tag{3.4.66}$$

for any $n, m = 1, \cdots, q$, where the first equality in (3.4.66) is due to Proposition 2.6, since $\{\varepsilon(\mathbf{v})\}, \{u(\mathbf{v})\} \sim IID(0, \sigma^2)$ and $|E\{\varepsilon(\mathbf{v})^3\}| < \infty$. Then, from (3.4.58), we may

re-write the variance matrix $\mathbf{M}$ as

$$\mathbf{M} = \sum_{\mathbf{j}^\tau \in \mathcal{F}} \mathbf{\Gamma}(\mathbf{j}) = \mathbf{\Gamma}(\mathbf{0}) + \sum_{\substack{\mathbf{j}^\tau \in \mathcal{F}, \\ \mathbf{j} > \mathbf{0}}} [\mathbf{\Gamma}(\mathbf{j}) + \mathbf{\Gamma}^\tau(\mathbf{j})], \tag{3.4.67}$$

and find its (n, m)-th element for any $n, m = 1, \cdots, q$, to be equal to

$$\begin{aligned}
& \sigma^2[c_0(\mathbf{j}_n - \mathbf{j}_m)\gamma_0(\mathbf{0}) + \sum_{\substack{\mathbf{j}^\tau \in \mathcal{F}, \\ \mathbf{j} > \mathbf{0}}} (c_0(\mathbf{j}_n - \mathbf{j}_m - \mathbf{j})\gamma_0(\mathbf{j}) + c_0(\mathbf{j}_m - \mathbf{j}_n - \mathbf{j})\gamma_0(\mathbf{j}))] \\
= \; & \sigma^2[c_0(\mathbf{j}_n - \mathbf{j}_m)\gamma_0(\mathbf{0}) + \sum_{\substack{\mathbf{j}^\tau \in \mathcal{F}, \\ \mathbf{j} > \mathbf{0}}} c_0(\mathbf{j}_n - \mathbf{j}_m - \mathbf{j})\gamma_0(\mathbf{j}) + \sum_{\substack{\mathbf{j}^\tau \in \mathcal{F}, \\ \mathbf{j} < \mathbf{0}}} c_0(\mathbf{j}_m - \mathbf{j}_n + \mathbf{j})\gamma_0(-\mathbf{j})] \\
= \; & \sigma^2[c_0(\mathbf{j}_n - \mathbf{j}_m)\gamma_0(\mathbf{0}) + \sum_{\substack{\mathbf{j}^\tau \in \mathcal{F}, \\ \mathbf{j} > \mathbf{0}}} c_0(\mathbf{j}_n - \mathbf{j}_m - \mathbf{j})\gamma_0(\mathbf{j}) + \sum_{\substack{\mathbf{j}^\tau \in \mathcal{F}, \\ \mathbf{j} < \mathbf{0}}} c_0(\mathbf{j}_n - \mathbf{j}_m - \mathbf{j})\gamma_0(\mathbf{j})] \\
= \; & \sigma^2 \sum_{\mathbf{j}^\tau \in \mathcal{F}} c_0(\mathbf{j}_n - \mathbf{j}_m - \mathbf{j})\gamma_0(\mathbf{j}) = \begin{cases} \sigma^4, & n = m \\ 0, & n \neq m \end{cases},
\end{aligned}$$

thanks to (3.4.66) and the general Yule-Walker equations. Thus

$$\mathbf{M} = \sigma^4 \cdot \mathbf{I}_q. \tag{3.4.68}$$

After combining (3.4.63), (3.4.65) and (3.4.68), we conclude that

$$N^{1/2}[\boldsymbol{\theta}^* - \boldsymbol{\theta}_0] \xrightarrow{D} N(\mathbf{0}, \mathbf{W}_q^{*-1}). \tag{3.4.69}$$

■

3.4.3 Modified likelihood estimation

The edge-effect is the source of the order of the bias in the exact likelihood estimators of the parameters of an ARMA process and it is the reason why modified versions of the Gaussian likelihood have been used for estimation before. We particularly refer to the papers by Guyon (1982) and Yao and Brockwell (2006), in which the modifications proposed have produced asymptotically unbiased and normal estimators. Moreover, if the original process is Gaussian, the estimators are efficient. This is another nice property that we did not manage to achieve with the Yule-Walker estimators of the parameters of a moving-average process in the previous section. Thus, in this section we resort to a new modification of the Gaussian likelihood that produces asymptotically unbiased, normal and efficient estimators for the parameters of a moving-average process. This

will equalize the results of the Yule-Walker or conditional likelihood estimators of the parameters of an auto-regression, that we managed to achieve before.

The paper by Guyon (1982) suggests a modification on that version of the likelihood, which uses the periodogram or the sample auto-covariances and which was introduced by Whittle (1954). The edge-effect disappears when the periodogram is computed based on the unbiased estimators of the theoretical auto-covariances at all possible lags. Since the paper refers to the estimation of the parameters of almost all (weakly) stationary processes, the case of a causal and invertible ARMA process is also included. According to Remark 2.2(ii), we know then that the auto-covariance function dies out at an exponential rate. That means that the number of sample auto-covariances to be computed for the likelihood increases, as more observations are obtained. For example, if the observations lie on the hyper-rectangle, this number comes from Proposition 3.1. For the special case of a moving-average process, the auto-covariances are not zero for a finite set of vector lags only. Thus, increasing the number of observations in the sample only increases the amount of information on the auto-covariances for this fixed set of lags. We may conclude that the representation of the Gaussian likelihood in terms of the sample auto-covariances, which was achieved by Whittle (1954), clearly favors a moving-average process, since the computation is simple and fast then. Nevertheless, when the representation is used for the estimation of the parameters of an ARMA process on the regular d-lattice, and after the modification proposed by Guyon (1982) has taken place, the estimators are not deprived of any of the desired properties.

On the other hand, the paper by Yao and Brockwell (2006) uses the innovations algorithm to factorize the random part of the Gaussian likelihood. As the algorithm imitates the AR(∞) representation of the process of interest, it uses a classical time domain methodology to prove the properties of the Gaussian likelihood estimators. The modification of the likelihood is effective for the special case when $d = 2$ only. Otherwise, the edge-effect has not been confined and the estimators have a bias, which tends to zero more slowly than their standard error; it seems like a hopeless case then.

The innovations algorithm is based on a conventional ordering between the observations in the sample; each observation offers information on every observation that is coming 'next'. The algorithm computes a triangular matrix of the coefficients of best linear predictors of each observation based on all the observations generated 'before' it in the sample. Then it composes this matrix and its transpose to create the inverse

theoretical variance matrix that is needed in the quadratic form of the likelihood. One has to combine the matrices of coefficients and the observations, in order to write the random part of the likelihood, explicitly. In this case, it is not as straightforward as it was before, that the sample auto-covariances were involved and using the unbiased estimators of their theoretical equivalents would defeat the edge-effect, but the decomposition used by the innovations algorithm is easy to interpret and gives speedy results. Moreover, if the process of interest is an auto-regression, applying the innovations algorithm is preferred, as it takes a smaller number of steps to complete. The number of steps is not fixed for $d \geq 2$, but it is asymptotically negligible compared to the number of observations, *i.e.* the number of steps that the algorithm would take if the process was not an auto-regression of finite order.

In this section, we try to conclude on a modification of the Gaussian likelihood that will produce asymptotically unbiased, normal and efficient estimators for the parameters of an invertible moving-average, which takes place on any positive integer d number of dimensions, by combining the two concepts in the two different papers. On the one hand, our suggestion for the likelihood is not related to Whittle's (1954) suggestion involving the periodogram and, consequently, it is not directly related to the sample auto-covariances. It is a suggestion that is only justified when the inverse of a conditional variance matrix of the observations generated by the moving-average process takes a convenient form. As we very well know, discovering the form of an inverse variance matrix can be the key to computing a Gaussian likelihood, as its elements can be found in both the quadratic form and the determinant involved in the likelihood, *i.e.* the two sources of computational struggle. Thus, we suggest a conditional likelihood to be maximized, based on classical time domain arguments, similarly to the paper of Yao and Brockwell (2006). On the other hand, we do not forget that we are dealing with a moving-average and most advantageously use the fact that the auto-covariances are not equal to zero for a finite set of lags only. That strongly resembles the picture presented by Whittle's (1954) likelihood, which immediately involves the sample auto-covariances. In the conditional likelihood suggested, we will draw the picture of how each observation in the sample is paired with every other observation in the sample and how some of these observations and their pairs will be rejected and left out of the quantity maximized. The rejection will take place only to ensure that these are the observations which lie on the edges, *i.e.* the ones that not only miss the pairs that are 'far away', but also the pairs that are

close to them and which correspond to the lags of non-zero auto-covariances, and might slow down the bias and harm the results.

For observations $\{Y(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{S}\}$ from the invertible moving-average defined in (3.4.1), we write the modified version of the conditional Gaussian likelihood

$$L^*(\boldsymbol{\theta},\sigma^2) \propto \frac{1}{(\sigma^2)^{N^*/2}} \exp\{-\frac{1}{2\sigma^2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y(\mathbf{v}) \sum_{\mathbf{j}^\tau \in \mathcal{F}_\mathbf{v}} c(\mathbf{j})\ Y(\mathbf{v}-\mathbf{j})\},\ \boldsymbol{\theta} \in \Theta_2,\ \sigma^2 > 0, \tag{3.4.70}$$

where the notation was introduced back in Sections 3.4.1 and 3.4.2. Again, we have considered

$$c(\mathbf{z}) = \theta(\mathbf{z})^{-1}\theta(\mathbf{z}^{-1})^{-1} \tag{3.4.71}$$

with

$$\theta(\mathbf{z}) = 1 + \sum_{m=1}^{q} \theta_{\mathbf{j}_m}\ \mathbf{z}^{\mathbf{j}_m}. \tag{3.4.72}$$

A justification why (3.4.70) is a modified Gaussian likelihood comes from (2.4.65) and (2.4.66). According to them, a true conditional likelihood is

$$L(\boldsymbol{\theta},\sigma^2) \propto \frac{|\mathrm{Var}\{\mathbf{X}\}|^{1/2}}{(\sigma^4)^{N/2}} \exp\{-\frac{1}{2\sigma^2} \sum_{\mathbf{v}^\tau \in \mathcal{S}} Y(\mathbf{v}) \sum_{\mathbf{j}^\tau \in \mathcal{F}_\mathbf{v}} c(\mathbf{j})\ Y(\mathbf{v}-\mathbf{j})\},\ \boldsymbol{\theta} \in \Theta_2,\ \sigma^2 > 0, \tag{3.4.73}$$

where $\mathbf{X}$ is the vector of random variables $\{X(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{S}\}$ from the causal auto-regression

$$X(\mathbf{v}) + \sum_{m=1}^{q} \theta_{\mathbf{j}_m} X(\mathbf{v}+\mathbf{j}_m) = \varepsilon(\mathbf{v}),\ \{\varepsilon(\mathbf{v})\} \sim WN(0,\sigma^2). \tag{3.4.74}$$

Now, we explain why (3.4.73) is correct. First, we need to see that we have involved $\mathbf{X}$ in the likelihood only via its variance matrix $\mathrm{Var}\{\mathbf{X}\}$. Indeed, the likelihood must be a function of the data $\mathbf{Y}$, which are the observations from the moving-average of interest, and a function of the parameters $\theta_{\mathbf{j}_m}, m = 1,\cdots,q$. The determinant $|\mathrm{Var}\{\mathbf{X}\}|$ clearly belongs to the second category and it is a piece of the deterministic, not the random part of the likelihood. Why have we chosen to write this determinant using the variance matrix of random variables that we do not observe at all? Simply because the sampling set $\mathcal{S}$, on which we have observed $\{Y(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{S}\}$, also determines which exactly is this random vector $\mathbf{X}$ that will contribute with its variance matrix. This is not very surprising, considering that the variance matrix itself depends on the locations of observations and so does its inverse.

Secondly, one might notice that $(\sigma^4)^{N/2}$ is in the denominator of the likelihood $L(\boldsymbol{\theta},\sigma^2)$, while $(\sigma^2)^{N^*/2}$ is in the denominator of the modification $L^*(\boldsymbol{\theta},\sigma^2)$. This is

because $|\text{Var}\{\mathbf{X}\}|^{1/2}$ is also in the numerator of $L(\boldsymbol{\theta}, \sigma^2)$ and we can re-write

$$|\text{Var}\{\mathbf{X}\}|^{1/2} = (\sigma^2)^{N/2}|\text{Var}\{\mathbf{X}/\sigma\}|^{1/2}$$

and

$$(X(\mathbf{v})/\sigma) + \sum_{m=1}^{q} \theta_{\mathbf{j}_m}(X(\mathbf{v}+\mathbf{j}_m)/\sigma) = (\varepsilon(\mathbf{v})/\sigma),\ \{\varepsilon(\mathbf{v})/\sigma\} \sim WN(0,1).$$

Now that we have achieved a unit variance of the error sequence, we can use the results proven in Chapter 2.

Finally, we should recall from the same chapter which conditional likelihood is (3.4.73). Thanks to an argument mentioned in the end of Section 2.4.1, (3.4.73) is the conditional Gaussian likelihood of $\mathbf{Y}$ given that $\mathbf{w} = \mathbf{0}$, where $\mathbf{w}$ is a vector with fixed zero elements when $\mathbf{v}^\tau \in \mathcal{S}^*$, while when $\mathbf{v}^\tau \in \mathcal{S} - \mathcal{S}^*$, the elements of the vector are random variables that are linear combinations of the unobserved values of the process of interest $Y(\mathbf{v})$, $\mathbf{v}^\tau \notin \mathcal{S}$. Those values might also be written as $\varepsilon(\mathbf{v}) + \sum_{m=1}^{q} \theta_{\mathbf{j}_m}\ \varepsilon(\mathbf{v}-\mathbf{j}_m)$ for $\mathbf{v}^\tau \notin \mathcal{S}$. Thus, $L(\boldsymbol{\theta}, \sigma^2)$ is a conditional likelihood of $\mathbf{Y}$ given that these values are equal to their mean value zero.

Nevertheless, we have proceeded with a selection of locations $\mathbf{v}^\tau \in \mathcal{S}^*$. It is true that if we attempt to factorize the determinant $|\text{Var}\{\mathbf{X}\}|$ into the prediction variances produced from the innovations algorithm, say $r(\mathbf{v}, \boldsymbol{\theta}), \mathbf{v}^\tau \in \mathcal{S}$, then for the standard ordering of locations, we would have come up with

$$r(\mathbf{v}, \boldsymbol{\theta}) = \sigma^2,\ \mathbf{v}^\tau \in \mathcal{S}^*. \tag{3.4.75}$$

This is because we are using the vector $\mathbf{X}$ and the nice properties of a causal auto-regression. If it was necessary to perform the innovations algorithm to compute $|\text{Var}\{\mathbf{Y}\}|$, we would not see that property there, and the prediction variances would keep changing, since a moving-average possesses an AR(∞) representation. We would need to observe the full 'past' of each observation to derive a prediction variance equal to σ^2. As a result for the finite auto-regression, replacing the original set $\mathcal{S}$ by a subset $\mathcal{S}^*$, results in replacing the determinant

$$|\text{Var}\{\mathbf{X}\}| = \prod_{\mathbf{v}^\tau \in \mathcal{S}} r(\mathbf{v}, \boldsymbol{\theta})$$

by

$$\prod_{\mathbf{v}^\tau \in \mathcal{S}^*} r(\mathbf{v}, \boldsymbol{\theta}) = (\sigma^2)^{N^*}.$$

For a more detailed description of the innovations algorithm for the d-dimensional ARMA, one should see the paper by Yao and Brockwell (2006).

But the determinant involved in the likelihood is not the main reason why we have excluded some observations. Like for the Yule-Walker estimators, for any $\mathbf{v}^\tau \in \mathcal{S}^*$, it holds that $\mathcal{F} \subseteq \mathcal{F}_{\mathbf{v}}$ and so,

$$E\{\sum_{\mathbf{i}^\tau \in \mathcal{Z}^d} d_{\mathbf{i}}\, Y(\mathbf{v})Y(\mathbf{v}-\mathbf{i})\} = E\{\sum_{\mathbf{i}^\tau \in \mathcal{F}} d_{\mathbf{i}}\, Y(\mathbf{v})Y(\mathbf{v}-\mathbf{i})\} \tag{3.4.76}$$

for any polynomial

$$d(\mathbf{z}) = \sum_{\mathbf{i}^\tau \in \mathcal{Z}^d} d_{\mathbf{i}}\, \mathbf{z}^{\mathbf{i}}.$$

Thus, lacking the infinite information from both the 'past' and 'future' of any observation in the sample, does not have any effect on the estimation regarding, the expected values of the random variables involved. In other words, the selection of observations $\mathbf{v}^\tau \in \mathcal{S}^*$ guarantees that both the Yule-Walker estimators defined in 3.4.2 and the modified likelihood estimators will have an unimportant bias, like we have seen in time series when $d = 1$.

As for the variance of the estimators, this is now related to the coefficients of the polynomial $c(\mathbf{z})$ that we have used in the quadratic form of the likelihood (3.4.70), rather than anything else. This is the main difference from the Yule-Walker estimators, which according to Theorem 3.2, did not achieve the inverse of the variance matrix of the random vector $\boldsymbol{\xi}$, but the inverse of a conditional variance matrix of the same vector, *i.e.* something 'bigger' in terms of quadratic forms. Why we are using $\{c(\mathbf{j}),\ \mathbf{j}^\tau \in \mathcal{Z}^d\}$ is justified by (2.4.65) and (2.4.66), as mentioned before. For more complicated models, though, such as the ARMA, we do not, in general, have results available for the variance matrices of the observations in the sample. In such cases, we may resort to the theoretical properties reflected in the spectral density and look at its denominator. For example, for the moving-average of interest that would be

$$g_Y(\boldsymbol{\omega}) \equiv \sigma^2\, \theta(e^{i\boldsymbol{\omega}})\theta(e^{-i\boldsymbol{\omega}}) = \frac{\sigma^2}{c(e^{i\boldsymbol{\omega}})},\ \boldsymbol{\omega}^\tau \in [-\pi,\pi]^d. \tag{3.4.77}$$

This is part of a general argument, which will be shown later in Chapters 4 and 5. The argument in Section 5.2.1 says that in the inverse of a variance matrix, one can find the coefficients of best linear predictors of each observation based on all other observations used from the sample. On the other hand, according to Proposition 4.1, the denominator of a spectral density of a (weakly) stationary process taking place on the regular

lattice, generates the coefficients of best linear predictors of any observation based on all other information on the lattice. We expect, as we increase the number of observations available, that these two coefficients come closer. Of course, the auto-regression defined in (3.4.74) enjoys the privilege of a finite representation in the denominator

$$g_X(\omega) \equiv \frac{\sigma^2}{\theta(e^{i\omega})\theta(e^{-i\omega})}, \ \omega^\tau \in [-\pi, \pi]^d. \tag{3.4.78}$$

Properties of estimators

We define the quantity

$$Q^*(\boldsymbol{\theta}) \equiv \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y(\mathbf{v}) \sum_{\mathbf{j}^\tau \in \mathcal{F}_\mathbf{v}} c(\mathbf{j})\, Y(\mathbf{v}-\mathbf{j}) = \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y(\mathbf{v})[\theta(\mathbf{B})^{-1}\theta(\mathbf{B}^{-1})^{-1}H_Y(\mathbf{v})], \ \boldsymbol{\theta} \in \Theta_2, \tag{3.4.79}$$

where $H_Y(\mathbf{v}), \mathbf{v}^\tau \in \mathcal{Z}^d$, was defined back in (3.4.19). We consider $\widehat{\boldsymbol{\theta}}$ and $\widehat{\sigma^2}$ to be the maximum likelihood estimators, such that

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta} \in \Theta_2} Q^*(\boldsymbol{\theta}), \tag{3.4.80}$$

and

$$\widehat{\sigma^2} = Q^*(\widehat{\boldsymbol{\theta}})/N^*. \tag{3.4.81}$$

In general, we will denote with 'hats' the functions of the estimators and with zero sub-indexes the quantities corresponding to the true parameter vector $\boldsymbol{\theta}_0 \in \Theta_2$. We consider the processes $\{X(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ and $\{u(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ as defined back in (3.4.31) and (3.4.37), respectively.

Theorem 3.3 (Consistency). If $\{\varepsilon(\mathbf{v})\} \sim IID(0, \sigma^2)$, then under conditions (C1)(i) and (C3), it holds that

$$\widehat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}_0$$

and

$$\widehat{\sigma^2} \xrightarrow{P} \sigma^2$$

as $N \to \infty$.

Proof. From (3.4.80), we can write immediately that

$$\limsup_{N\to\infty} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y(\mathbf{v}) \sum_{\mathbf{j}^\tau \in \mathcal{F}_\mathbf{v}} \widehat{c}(\mathbf{j})\, Y(\mathbf{v}-\mathbf{j})/N \leq \lim_{N\to\infty} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y(\mathbf{v}) \sum_{\mathbf{j}^\tau \in \mathcal{F}_\mathbf{v}} c_0(\mathbf{j})\, Y(\mathbf{v}-\mathbf{j})/N = \sigma^2, \tag{3.4.82}$$

where the last equality holds from the fact that

$$\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y(\mathbf{v}) \sum_{\mathbf{j}^\tau \in \mathcal{F}_\mathbf{v}} c_0(\mathbf{j})\, Y(\mathbf{v}-\mathbf{j})/N = \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y(\mathbf{v}) \sum_{\mathbf{j}^\tau \in \mathcal{Z}^d} c_0(\mathbf{j})\, Y(\mathbf{v}-\mathbf{j})/N + o_P(1),$$

thanks to the same argument as the one used in (3.4.29) when (C1)(i) holds and the fact that

$$\frac{1}{N} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y(\mathbf{v}) Y(\mathbf{v}-\mathbf{j}) \xrightarrow{P} E\{Y(\mathbf{v})Y(\mathbf{v}-\mathbf{j})\} = \gamma_0(\mathbf{j}), \tag{3.4.83}$$

since $\{\varepsilon(\mathbf{v})\} \sim IID(0,\sigma^2)$. On the other hand, we can consider that for any $\boldsymbol{\theta} \in \Theta_2$, it holds that

$$c(\mathbf{j})^2 \leq C(\boldsymbol{\theta})\ \alpha(\boldsymbol{\theta})^{\sum_{k=1}^d |j_k|},$$

for $C(\boldsymbol{\theta}) > 0$ and $a(\boldsymbol{\theta}) \in (0,1)$, and we can write that

$$\frac{1}{N} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y(\mathbf{v}) \sum_{\mathbf{j}^\tau \in \mathcal{F}_\mathbf{v}} c(\mathbf{j})\, Y(\mathbf{v}-\mathbf{j}) = \frac{1}{N} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y(\mathbf{v}) \sum_{\mathbf{j}^\tau \in \mathcal{Z}^d} c(\mathbf{j})\, Y(\mathbf{v}-\mathbf{j}) + o_P(1). \tag{3.4.84}$$

Then for any $\boldsymbol{\theta} \in \Theta_2$,

$$\begin{aligned}
\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y(\mathbf{v}) \sum_{\mathbf{j}^\tau \in \mathcal{Z}^d} c(\mathbf{j})\, Y(\mathbf{v}-\mathbf{j})/N &= \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y(\mathbf{v})[\theta(\mathbf{B})^{-1}\theta(\mathbf{B}^{-1})^{-1}Y(\mathbf{v})]/N \\
&\xrightarrow{P} E\{Y(\mathbf{v})[\theta(\mathbf{B})^{-1}\theta(\mathbf{B}^{-1})^{-1}Y(\mathbf{v})]\} \\
&= E\{Y(\mathbf{v})[\theta(\mathbf{B})^{-1}\theta(\mathbf{B}^{-1})^{-1}\theta_0(\mathbf{B})\theta_0(\mathbf{B}^{-1})X(\mathbf{v})]\} \\
&\geq E\{Y(\mathbf{v})X(\mathbf{v})\} = \sigma^2,
\end{aligned} \tag{3.4.85}$$

since $Y(\mathbf{v})$ and $X(\mathbf{v}-\mathbf{j})$ are uncorrelated for any $\mathbf{j} \neq \mathbf{0}$. The equality in (3.4.85) holds if and only if $\boldsymbol{\theta} \equiv \boldsymbol{\theta}_0$. Finally, if we combine (3.4.82), (3.4.84) and (3.4.85), we can see immediately that

$$\widehat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}_0. \tag{3.4.86}$$

Straight from (3.4.81)

$$\begin{aligned}
\widehat{\sigma^2} \xrightarrow{P} Q^*(\boldsymbol{\theta}_0)/N^* &= \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y(\mathbf{v}) \sum_{\mathbf{j}^\tau \in \mathcal{F}_\mathbf{v}} c_0(\mathbf{j})\, Y(\mathbf{v}-\mathbf{j})/N^* \\
&\xrightarrow{P} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y(\mathbf{v}) \sum_{\mathbf{j}^\tau \in \mathcal{Z}^d} c_0(\mathbf{j})\, Y(\mathbf{v}-\mathbf{j})/N^* \\
&= \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y(\mathbf{v})X(\mathbf{v})/N^* \xrightarrow{P} E\{Y(\mathbf{v})X(\mathbf{v})\} = \sigma^2,
\end{aligned} \tag{3.4.87}$$

for the same reasons as before. ■

To find the minimum of $Q^*(\boldsymbol{\theta})$, we need to write the derivatives

$$-\partial Q^*(\boldsymbol{\theta})/\partial\theta_{\mathbf{j}_m} = \sum_{\mathbf{v}^\tau\in\mathcal{S}^*} Y(\mathbf{v})[\theta(\mathbf{B})^{-2}\theta(\mathbf{B}^{-1})H_Y(\mathbf{v}-\mathbf{j}_m)+\theta(\mathbf{B})^{-1}\theta(\mathbf{B}^{-1})^{-2}H_Y(\mathbf{v}+\mathbf{j}_m)], \tag{3.4.88}$$

for all $m = 1, \cdots, q$. We may then set equal to zero

$$\begin{aligned}&\sum_{\mathbf{v}^\tau\in\mathcal{S}^*} Y(\mathbf{v})[\theta_0(\mathbf{B})^{-2}\theta_0(\mathbf{B}^{-1})^{-1}H_Y(\mathbf{v}-\mathbf{j}_m)+\theta_0(\mathbf{B})^{-1}\theta_0(\mathbf{B}^{-1})^{-2}H_Y(\mathbf{v}+\mathbf{j}_m)]\\ -\ &\mathbf{J}_m[\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0] = 0,\end{aligned} \tag{3.4.89}$$

where we define

$$\mathbf{J}_m = [J_{m,1}, \cdots, J_{m,q}],\ m = 1, \cdots, q, \tag{3.4.90}$$

and elements

$$\begin{aligned}J_{m,n} \equiv &\sum_{\mathbf{v}^\tau\in\mathcal{S}^*} Y(\mathbf{v})[2\ \theta_0(\mathbf{B})^{-3}\theta_0(\mathbf{B}^{-1})H_Y(\mathbf{v}-\mathbf{j}_m-\mathbf{j}_n)\\ &+\ \theta_0(\mathbf{B})^{-2}\theta_0(\mathbf{B}^{-1})^{-2}H_Y(\mathbf{v}-\mathbf{j}_m+\mathbf{j}_n)\\ &+\ \theta_0(\mathbf{B})^{-2}\theta_0(\mathbf{B}^{-1})^{-2}H_Y(\mathbf{v}+\mathbf{j}_m-\mathbf{j}_n)\\ &+\ 2\ \theta_0(\mathbf{B}^{-1})^{-3}\theta_0(\mathbf{B})^{-1}H_Y(\mathbf{v}+\mathbf{j}_m+\mathbf{j}_n)] + O_P(N||\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0||),\end{aligned} \tag{3.4.91}$$

for all $n, m = 1, \cdots, q$. If we stack all the q equations together, we can write

$$N^{1/2}[\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0] = \{\mathbf{J}/N\}^{-1}\{N^{-1/2}\sum_{\mathbf{v}^\tau\in\mathcal{S}^*}\mathbf{H}_U(\mathbf{v})\}, \tag{3.4.92}$$

where

$$\mathbf{J}^\tau \equiv [\mathbf{J}_1^\tau, \cdots, \mathbf{J}_q^\tau] \tag{3.4.93}$$

and also for any $\mathbf{v}^\tau \in \mathcal{Z}^d$,

$$\mathbf{H}_U(\mathbf{v}) \equiv \begin{bmatrix}\theta_0(\mathbf{B})^{-2}\theta_0(\mathbf{B}^{-1})^{-1}H_Y(\mathbf{v}-\mathbf{j}_1)+\theta_0(\mathbf{B})^{-1}\theta_0(\mathbf{B}^{-1})^{-2}H_Y(\mathbf{v}+\mathbf{j}_1)\\ \vdots\\ \theta_0(\mathbf{B})^{-2}\theta_0(\mathbf{B}^{-1})^{-1}H_Y(\mathbf{v}-\mathbf{j}_q)+\theta_0(\mathbf{B})^{-1}\theta_0(\mathbf{B}^{-1})^{-2}H_Y(\mathbf{v}+\mathbf{j}_q)\end{bmatrix} Y(\mathbf{v}). \tag{3.4.94}$$

Before we go on to prove the asymptotic normality of the estimators, similarly to Theorem 3.2, for a process $\{W(\mathbf{v})\} \sim WN(0,1)$, we define the auto-regression $\{\eta(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ by the equation

$$\theta_0(\mathbf{B})\eta(\mathbf{v}) \equiv W(\mathbf{v}), \tag{3.4.95}$$

and we let the vector $\boldsymbol{\xi} \equiv [\eta(-\mathbf{j}_1), \cdots, \eta(-\mathbf{j}_q)]^\tau$ and the variance matrix

$$\mathbf{W}_q \equiv \mathrm{Var}\{\boldsymbol{\xi}\}. \tag{3.4.96}$$

Proposition 3.3. If $\{\varepsilon(\mathbf{v})\} \sim IID(0, \sigma^2)$, then under conditions (C1)(i) and (C3), it holds that

$$\mathbf{J}/N \xrightarrow{P} 2 \cdot \sigma^2 \mathbf{W}_q \tag{3.4.97}$$

as $N \to \infty$.

Proof. For any $n, m = 1, \cdots, q$, we can write

$$\begin{aligned} J_{m,n}/N &= \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y(\mathbf{v})[2\ \theta_0(\mathbf{B})^{-3}\theta_0(\mathbf{B}^{-1})^{-1}Y(\mathbf{v}-\mathbf{j}_m-\mathbf{j}_n) \\ &+ \theta_0(\mathbf{B})^{-2}\theta_0(\mathbf{B}^{-1})^{-2}Y(\mathbf{v}-\mathbf{j}_m+\mathbf{j}_n) \\ &+ \theta_0(\mathbf{B})^{-2}\theta_0(\mathbf{B}^{-1})^{-2}Y(\mathbf{v}+\mathbf{j}_m-\mathbf{j}_n) \\ &+ 2\ \theta_0(\mathbf{B}^{-1})^{-3}\theta_0(\mathbf{B})^{-1}Y(\mathbf{v}+\mathbf{j}_m+\mathbf{j}_n)]/N + o_P(1), \end{aligned} \tag{3.4.98}$$

using the same argument again. Equation (3.4.98) may be re-written as

$$\begin{aligned} J_{m,n}/N &= \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y(\mathbf{v})[2\ \theta_0(\mathbf{B})^{-2}X(\mathbf{v}-\mathbf{j}_m-\mathbf{j}_n) \\ &+ \theta_0(\mathbf{B})^{-1}\theta_0(\mathbf{B}^{-1})^{-1}X(\mathbf{v}-\mathbf{j}_m+\mathbf{j}_n) + \theta_0(\mathbf{B})^{-1}\theta_0(\mathbf{B}^{-1})^{-1}X(\mathbf{v}+\mathbf{j}_m-\mathbf{j}_n) \\ &+ 2\ \theta_0(\mathbf{B}^{-1})^{-2}X(\mathbf{v}+\mathbf{j}_m+\mathbf{j}_n)]/N + o_P(1). \end{aligned} \tag{3.4.99}$$

It holds that

$$\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y(\mathbf{v})[2\ \theta_0(\mathbf{B})^{-2}X(\mathbf{v}-\mathbf{j}_m-\mathbf{j}_n)]/N \xrightarrow{P} E\{Y(\mathbf{v})[2\ \theta_0(\mathbf{B})^{-2}X(\mathbf{v}-\mathbf{j}_m-\mathbf{j}_n)]\} = 0 \tag{3.4.100}$$

and, similarly,

$$\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y(\mathbf{v})[2\ \theta_0(\mathbf{B}^{-1})^{-2}X(\mathbf{v}+\mathbf{j}_m+\mathbf{j}_n)]/N \xrightarrow{P} E\{Y(\mathbf{v})[2\ \theta_0(\mathbf{B}^{-1})^{-2}X(\mathbf{v}+\mathbf{j}_m+\mathbf{j}_n)]\} = 0, \tag{3.4.101}$$

since $Y(\mathbf{v})$ and $X(\mathbf{v}-\mathbf{j})$ are uncorrelated for any $\mathbf{j} \neq \mathbf{0}$. We now look at the polynomial

$$c_0(\mathbf{z}) = (1 + \sum_{\mathbf{j}>\mathbf{0}} \Theta_{\mathbf{j},0}\ \mathbf{z}^{\mathbf{j}}) \cdot (1 + \sum_{\mathbf{j}>\mathbf{0}} \Theta_{\mathbf{j},0}\ \mathbf{z}^{-\mathbf{j}}), \tag{3.4.102}$$

which generates

$$\Theta_{\mathbf{j}_m-\mathbf{j}_n,0} + \sum_{\mathbf{j}>\mathbf{0}} \Theta_{\mathbf{j},0}\Theta_{\mathbf{j}_m-\mathbf{j}_n+\mathbf{j},0}$$

to be the coefficient of $\mathbf{z}^{\mathbf{j}_m-\mathbf{j}_n}$ and $\mathbf{z}^{\mathbf{j}_n-\mathbf{j}_m}$ when $\mathbf{j}_m \geq \mathbf{j}_n$. Looking back at (3.4.99), we can conclude that

$$J_{m,n}/N \xrightarrow{P} 2 \cdot \sigma^2(\Theta_{\mathbf{j}_m-\mathbf{j}_n,0} + \sum_{\mathbf{j}>\mathbf{0}} \Theta_{\mathbf{j},0}\Theta_{\mathbf{j}_m-\mathbf{j}_n+\mathbf{j},0}), \ m \geq n \tag{3.4.103}$$

and the same holds for $J_{n,m}/N$, which proves the proposition. ■

Before we state and prove the next theorem, we should note that like we did for the Yule-Walker estimators of the previous section, we are going to use again the sequence of uncorrelated random variables $\{u(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$, for which it holds

$$Y(\mathbf{v}) = u(\mathbf{v}) + \sum_{m=1}^{q} \theta_{\mathbf{j}_m,0}\ u(\mathbf{v}+\mathbf{j}_m),\ \{u(\mathbf{v})\} \sim WN(0,\sigma^2),$$

and

$$X(\mathbf{v}) + \sum_{m=1}^{q} \theta_{\mathbf{j}_m,0}\ X(\mathbf{v}-\mathbf{j}_m) = u(\mathbf{v}),\ \{u(\mathbf{v})\} \sim WN(0,\sigma^2).$$

Theorem 3.4. If $\{\varepsilon(\mathbf{v})\} \sim IID(0,\sigma^2)$ and $E\{\varepsilon(\mathbf{v})^4\} < \infty$, then under conditions (C1) and (C3), it holds that

$$N^{1/2}[\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0] \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Delta}) \tag{3.4.104}$$

as $N \to \infty$. Otherwise, if $\{\varepsilon(\mathbf{v})\}, \{u(\mathbf{v})\} \sim IID(0,\sigma^2)$ and $|E\{\varepsilon(\mathbf{v})^3\}| < \infty$, then under conditions (C1) and (C3), it holds that

$$N^{1/2}[\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0] \xrightarrow{D} N(\mathbf{0}, \mathbf{W}_q^{-1}) \tag{3.4.105}$$

as $N \to \infty$.

Proof. First, we define

$$\mathbf{U}(\mathbf{v}) \equiv \begin{bmatrix} \theta_0(\mathbf{B})^{-2}\theta_0(\mathbf{B}^{-1})^{-1}Y(\mathbf{v}-\mathbf{j}_1) + \theta_0(\mathbf{B})^{-1}\theta_0(\mathbf{B}^{-1})^{-2}Y(\mathbf{v}+\mathbf{j}_1) \\ \vdots \\ \theta_0(\mathbf{B})^{-2}\theta_0(\mathbf{B}^{-1})^{-1}Y(\mathbf{v}-\mathbf{j}_q) + \theta_0(\mathbf{B})^{-1}\theta_0(\mathbf{B}^{-1})^{-2}Y(\mathbf{v}+\mathbf{j}_q) \end{bmatrix} Y(\mathbf{v}) \tag{3.4.106}$$

or

$$\mathbf{U}(\mathbf{v}) = \begin{bmatrix} \theta_0(\mathbf{B})^{-1}X(\mathbf{v}-\mathbf{j}_1) + \theta_0(\mathbf{B}^{-1})^{-1}X(\mathbf{v}+\mathbf{j}_1) \\ \vdots \\ \theta_0(\mathbf{B})^{-1}X(\mathbf{v}-\mathbf{j}_q) + \theta_0(\mathbf{B}^{-1})^{-1}X(\mathbf{v}+\mathbf{j}_q) \end{bmatrix} Y(\mathbf{v}). \tag{3.4.107}$$

Using the same argument as in (3.4.40), we may write for any $m = 1, \cdots, q$,

$$\begin{aligned} N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} & [\theta_0(\mathbf{B})^{-2}\theta_0(\mathbf{B}^{-1})^{-1} H_Y(\mathbf{v} - \mathbf{j}_m) \\ + & \ \theta_0(\mathbf{B})^{-1}\theta_0(\mathbf{B}^{-1})^{-2} H_Y(\mathbf{v} + \mathbf{j}_m)] Y(\mathbf{v}) \\ = & \ N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} [\theta_0(\mathbf{B})^{-2}\theta_0(\mathbf{B}^{-1})^{-1} Y(\mathbf{v} - \mathbf{j}_m) \\ + & \ \theta_0(\mathbf{B})^{-1}\theta_0(\mathbf{B}^{-1})^{-2} Y(\mathbf{v} + \mathbf{j}_m)] Y(\mathbf{v}) + o_P(1) \\ = & \ N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} [\theta_0(\mathbf{B})^{-1} X(\mathbf{v} - \mathbf{j}_m) + \theta_0(\mathbf{B}^{-1})^{-1} X(\mathbf{v} + \mathbf{j}_m)] Y(\mathbf{v}) + o_P(1). \end{aligned} \tag{3.4.108}$$

Then for any $\boldsymbol{\lambda} \in \mathcal{R}^q$, we can write

$$N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \boldsymbol{\lambda}^\tau \mathbf{H}_U(\mathbf{v}) = N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \boldsymbol{\lambda}^\tau \mathbf{U}(\mathbf{v}) + o_P(1). \tag{3.4.109}$$

For any positive integer K, we defined back in (2.5.4) the set $\mathcal{F}_K$. For convenience, we define the polynomial

$$h_0(\mathbf{z}) \equiv \theta_0(\mathbf{z})\{\theta_0(\mathbf{z})\theta_0(\mathbf{z}^{-1})\}^{-1} \equiv \sum_{\mathbf{j}^\tau \in \mathcal{Z}^d} h_{\mathbf{j},0}\, \mathbf{z}^{\mathbf{j}}. \tag{3.4.110}$$

We also define the vector

$$\mathbf{U}^{(K)}(\mathbf{v}) \equiv \begin{bmatrix} \sum_{\mathbf{j}^\tau \in \mathcal{F}_K} h_{\mathbf{j},0}\, Y(\mathbf{v} - \mathbf{j}_1 - \mathbf{j}) + \sum_{\mathbf{j}^\tau \in \mathcal{F}_K} h_{\mathbf{j},0}\, Y(\mathbf{v} + \mathbf{j}_1 + \mathbf{j}) \\ \vdots \\ \sum_{\mathbf{j}^\tau \in \mathcal{F}_K} h_{\mathbf{j},0}\, Y(\mathbf{v} - \mathbf{j}_q - \mathbf{j}) + \sum_{\mathbf{j}^\tau \in \mathcal{F}_K} h_{\mathbf{j},0}\, Y(\mathbf{v} + \mathbf{j}_q + \mathbf{j}) \end{bmatrix} Y(\mathbf{v}). \tag{3.4.111}$$

We define the random variables

$$r_n(\mathbf{v}) \equiv \sum_{\mathbf{j}^\tau \notin \mathcal{F}_K} h_{\mathbf{j},0}\, Y(\mathbf{v} - \mathbf{j}_n - \mathbf{j}) Y(\mathbf{v}) + \sum_{\mathbf{j}^\tau \notin \mathcal{F}_K} h_{\mathbf{j},0}\, Y(\mathbf{v} + \mathbf{j}_n + \mathbf{j}) Y(\mathbf{v}),\ n = 1, \cdots, q, \tag{3.4.112}$$

such that

$$\begin{aligned} & [\theta_0(\mathbf{B})^{-1} X(\mathbf{v} - \mathbf{j}_n)] Y(\mathbf{v}) + [\theta_0(\mathbf{B}^{-1})^{-1} X(\mathbf{v} + \mathbf{j}_n)] Y(\mathbf{v}) \\ = & \sum_{\mathbf{j}^\tau \in \mathcal{F}_K} h_{\mathbf{j},0}\, Y(\mathbf{v} - \mathbf{j}_n - \mathbf{j}) Y(\mathbf{v}) + \sum_{\mathbf{j}^\tau \in \mathcal{F}_K} h_{\mathbf{j},0}\, Y(\mathbf{v} + \mathbf{j}_n + \mathbf{j}) Y(\mathbf{v}) + r_n(\mathbf{v}), \end{aligned}$$

for all $n = 1, \cdots, q$. Then the (n,m)-th element of $E\{\mathbf{U}^{(K)}(\mathbf{v})\mathbf{U}^{(K)^\tau}(\mathbf{v}-\mathbf{i})\}$ is equal to

$$\begin{aligned}
& E\{[[\theta_0(\mathbf{B})^{-1}X(\mathbf{v}-\mathbf{j}_n)]Y(\mathbf{v}) + [\theta_0(\mathbf{B}^{-1})^{-1}X(\mathbf{v}+\mathbf{j}_n)]Y(\mathbf{v})] \\
& [[\theta_0(\mathbf{B})^{-1}X(\mathbf{v}-\mathbf{j}_m-\mathbf{i})]Y(\mathbf{v}-\mathbf{i}) + [\theta_0(\mathbf{B}^{-1})^{-1}X(\mathbf{v}+\mathbf{j}_m-\mathbf{i})]Y(\mathbf{v}-\mathbf{i})]\} \\
- \; & E\{r_n(\mathbf{v})[\theta_0(\mathbf{B})^{-1}X(\mathbf{v}-\mathbf{j}_m-\mathbf{i})Y(\mathbf{v}-\mathbf{i}) + \theta_0(\mathbf{B}^{-1})^{-1}X(\mathbf{v}+\mathbf{j}_m-\mathbf{i})Y(\mathbf{v}-\mathbf{i})]\} \\
- \; & E\{r_m(\mathbf{v}-\mathbf{i})[\theta_0(\mathbf{B})^{-1}X(\mathbf{v}-\mathbf{j}_n)Y(\mathbf{v}) + \theta_0(\mathbf{B}^{-1})^{-1}X(\mathbf{v}+\mathbf{j}_n)Y(\mathbf{v})]\} \\
+ \; & E\{r_n(\mathbf{v})r_m(\mathbf{v}-\mathbf{i})\}, \qquad (3.4.113)
\end{aligned}$$

for any $n, m = 1, \cdots, q$. It holds that

$$\begin{aligned}
E\{r_n(\mathbf{v})^2\} &= \sum_{\mathbf{j}^\tau \mathbf{j}^{*\tau} \notin \mathcal{F}_K} h_{\mathbf{j},0}h_{\mathbf{j}^*,0}\, E\{Y(\mathbf{v}-\mathbf{j}_n-\mathbf{j})Y(\mathbf{v}-\mathbf{j}_n-\mathbf{j}^*)Y(\mathbf{v})^2\} \\
&+ \sum_{\mathbf{j}^\tau \mathbf{j}^{*\tau} \notin \mathcal{F}_K} h_{\mathbf{j},0}h_{\mathbf{j}^*,0}\, E\{Y(\mathbf{v}+\mathbf{j}_n+\mathbf{j})Y(\mathbf{v}+\mathbf{j}_n+\mathbf{j}^*)Y(\mathbf{v})^2\} \\
&+ \sum_{\mathbf{j}^\tau \mathbf{j}^{*\tau} \notin \mathcal{F}_K} h_{\mathbf{j},0}h_{\mathbf{j}^*,0}\, E\{Y(\mathbf{v}-\mathbf{j}_n-\mathbf{j})Y(\mathbf{v}+\mathbf{j}_n+\mathbf{j}^*)Y(\mathbf{v})^2\} \cdot 2 \\
&\leq C \cdot \alpha^K, \qquad (3.4.114)
\end{aligned}$$

for constants $C > 0$ and $\alpha \in (0,1)$. Indeed, that comes immediately from the argument that for large enough and fixed, positive integer K, due to the independence of $Y(\mathbf{v})$ with $Y(\mathbf{v}-\mathbf{j}), \mathbf{j}^\tau \notin \mathcal{F}$, we may write for any $\mathbf{j}^\tau, \mathbf{j}^{*\tau} \notin \mathcal{F}_K$

$$E\{Y(\mathbf{v}-\mathbf{j}_n-\mathbf{j})Y(\mathbf{v}-\mathbf{j}_n-\mathbf{j}^*)Y(\mathbf{v})^2\} = E\{Y(\mathbf{v})^2\}\; E\{Y(\mathbf{v}-\mathbf{j}_n-\mathbf{j})Y(\mathbf{v}-\mathbf{j}_n-\mathbf{j}^*)\},$$

and similarly for all the expected values involved in (3.4.114).

Back to (3.4.113) now, the first term there, is the (n,m)-th element of $E\{\mathbf{U}(\mathbf{v})\mathbf{U}^\tau(\mathbf{v}-\mathbf{i})\}$. The absolute values of the other three terms tend to 0, as $K \to \infty$, thanks to the Cauchy-Schwartz inequality and (3.4.114). Of course, we are also using the fact that $E\{\varepsilon(\mathbf{v})^4\} < \infty$ or that $Y(\mathbf{v})$ and $X(\mathbf{v}-\mathbf{j})$ are independent for any $\mathbf{j} \neq \mathbf{0}$.

It holds, in general, that $\{\boldsymbol{\lambda}^\tau \mathbf{U}^{(K)}(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ is a zero-mean and strictly stationary K^*-dependent process, for some fixed positive and finite integer K^*. This implies that as $N \to \infty$ and (C1) holds,

$$N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \boldsymbol{\lambda}^\tau \mathbf{U}^{(K)}(\mathbf{v}) \xrightarrow{D} \boldsymbol{\lambda}^\tau \mathbf{V}_K \sim N(0, \boldsymbol{\lambda}^\tau \mathbf{M}_K \boldsymbol{\lambda}), \qquad (3.4.115)$$

where

$$\mathbf{M}_K \equiv \sum_{\mathbf{i}^\tau \in \mathcal{Z}^d} \boldsymbol{\Gamma}_K(\mathbf{i}), \qquad (3.4.116)$$

with

$$\mathbf{\Gamma}_K(\mathbf{i}) = E\{\mathbf{U}^{(K)}(\mathbf{v})\mathbf{U}^{(K)^\tau}(\mathbf{v}-\mathbf{i})\}. \tag{3.4.117}$$

Similarly, we may define

$$\mathbf{\Gamma}(\mathbf{i}) \equiv E\{\mathbf{U}(\mathbf{v})\mathbf{U}^\tau(\mathbf{v}-\mathbf{i})\} \tag{3.4.118}$$

and

$$\mathbf{M} \equiv \sum_{\mathbf{i}^\tau \in \mathcal{Z}^d} \mathbf{\Gamma}(\mathbf{i}). \tag{3.4.119}$$

Thanks to (3.4.113) and (3.4.114), it holds that

$$\boldsymbol{\lambda}^\tau \mathbf{\Gamma}_K(\mathbf{i})\boldsymbol{\lambda} \to \boldsymbol{\lambda}^\tau \mathbf{\Gamma}(\mathbf{i})\boldsymbol{\lambda} \tag{3.4.120}$$

and

$$\boldsymbol{\lambda}^\tau \mathbf{M}_K \boldsymbol{\lambda} \to \boldsymbol{\lambda}^\tau \mathbf{M}\boldsymbol{\lambda}, \tag{3.4.121}$$

as $K \to \infty$. We can write that

$$\boldsymbol{\lambda}^\tau \mathbf{V} \xrightarrow{D} \boldsymbol{\lambda}^\tau \mathbf{V} \sim N(0, \boldsymbol{\lambda}^\tau \mathbf{M}\boldsymbol{\lambda}) \tag{3.4.122}$$

as $K \to \infty$. Finally, after checking that

$$\boldsymbol{\lambda}^\tau \mathrm{Var}\{\mathbf{U}(\mathbf{v}) - \mathbf{U}^{(K)}(\mathbf{v})\}\boldsymbol{\lambda} \to 0 \tag{3.4.123}$$

as $K \to \infty$, we may combine (3.4.115), (3.4.122), (3.4.123) to write that

$$N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \boldsymbol{\lambda}^\tau \mathbf{U}(\mathbf{v}) \xrightarrow{D} N(0, \boldsymbol{\lambda}^\tau \mathbf{M}\boldsymbol{\lambda}) \tag{3.4.124}$$

as $N \to \infty$. Using (3.4.109) and the Cramer-Wold device we can convert (3.4.124) to

$$N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \mathbf{H}_U(\mathbf{v}) \xrightarrow{D} N(\mathbf{0}, \mathbf{M}). \tag{3.4.125}$$

Looking back at (3.4.92) and Proposition 3.3, we can see that the first part of the theorem has been proven and it holds that

$$\mathbf{\Delta} = \frac{1}{4 \cdot \sigma^4} \mathbf{W}_q^{-1} \mathbf{M} \mathbf{W}_q^{-1}. \tag{3.4.126}$$

Moreover, when

$$\{u(\mathbf{v})\} \sim IID(0, \sigma^2)$$

and $|E\{\varepsilon(\mathbf{v})^3\}| < \infty$, we need to show that

$$\mathbf{M} = 4 \cdot \sigma^4 \mathbf{W}_q. \tag{3.4.127}$$

Indeed, it holds then that $X(\mathbf{v}-\mathbf{j})$ and $Y(\mathbf{v})$ are two independent random variables for any $\mathbf{j} \neq \mathbf{0}$ and we may use Proposition 2.6.

We may write the (n,m)-th element of $\mathbf{\Gamma(i)}$ as

$$\begin{aligned}
& E\{[[\theta_0(\mathbf{B})^{-1}X(\mathbf{v}-\mathbf{j}_n)]Y(\mathbf{v}) + [\theta_0(\mathbf{B}^{-1})^{-1}X(\mathbf{v}+\mathbf{j}_n)]Y(\mathbf{v})] \\
& [[\theta_0(\mathbf{B})^{-1}X(\mathbf{v}-\mathbf{j}_m-\mathbf{i})]Y(\mathbf{v}-\mathbf{i}) + [\theta_0(\mathbf{B}^{-1})^{-1}X(\mathbf{v}+\mathbf{j}_m-\mathbf{i})]Y(\mathbf{v}-\mathbf{i})]\} \\
= \; & E\{[\theta_0(\mathbf{B})^{-1}X(\mathbf{v}-\mathbf{j}_n)][\theta_0(\mathbf{B})^{-1}X(\mathbf{v}-\mathbf{j}_m-\mathbf{i})]Y(\mathbf{v})Y(\mathbf{v}-\mathbf{i})\} \qquad (3.4.128) \\
+ \; & E\{[\theta_0(\mathbf{B})^{-1}X(\mathbf{v}-\mathbf{j}_n)][\theta_0(\mathbf{B}^{-1})^{-1}X(\mathbf{v}+\mathbf{j}_m-\mathbf{i})]Y(\mathbf{v})Y(\mathbf{v}-\mathbf{i})\} \qquad (3.4.129) \\
+ \; & E\{[\theta_0(\mathbf{B}^{-1})^{-1}X(\mathbf{v}+\mathbf{j}_n)][\theta_0(\mathbf{B})^{-1}X(\mathbf{v}-\mathbf{j}_m-\mathbf{i})]Y(\mathbf{v})Y(\mathbf{v}-\mathbf{i})\} \qquad (3.4.130) \\
+ \; & E\{[\theta_0(\mathbf{B}^{-1})^{-1}X(\mathbf{v}+\mathbf{j}_n)][\theta_0(\mathbf{B}^{-1})^{-1}X(\mathbf{v}+\mathbf{j}_m-\mathbf{i})]Y(\mathbf{v})Y(\mathbf{v}-\mathbf{i})\}, \qquad (3.4.131)
\end{aligned}$$

for any $n,m = 1,\cdots,q$. Without loss of generality, we will consider that $m \geq n$. We may recall the form of the variance matrix $\mathbf{M}$ from (3.4.119) or re-write it as

$$\mathbf{M} = \mathbf{\Gamma(0)} + \sum_{\mathbf{i}>\mathbf{0}}\{\mathbf{\Gamma(i)} + \mathbf{\Gamma}^{\tau}\mathbf{(i)}\} = \mathbf{\Gamma(0)} + \sum_{\mathbf{i}<\mathbf{0}}\{\mathbf{\Gamma(i)} + \mathbf{\Gamma}^{\tau}\mathbf{(i)}\}. \qquad (3.4.132)$$

For the first term (3.4.128), we write for $\mathbf{i} \geq \mathbf{0}$

$$\sum_{\mathbf{j},\mathbf{j}^*\geq\mathbf{0}} \Theta_{\mathbf{j},0}\Theta_{\mathbf{j}^*,0}\; E\{X(\mathbf{v}-\mathbf{j}_n-\mathbf{j})X(\mathbf{v}-\mathbf{j}_m-\mathbf{i}-\mathbf{j}^*)Y(\mathbf{v})Y(\mathbf{v}-\mathbf{i})\}.$$

It holds then that $\mathbf{v}-\mathbf{j}_n-\mathbf{j} < \mathbf{v}$ and $\mathbf{v}-\mathbf{j}_m-\mathbf{i}-\mathbf{j}^* < \mathbf{v}$, so we may re-write it as

$$\begin{aligned}
& \sum_{\mathbf{j},\mathbf{j}^*\geq\mathbf{0}} \Theta_{\mathbf{j},0}\Theta_{\mathbf{j}^*,0}\; E\{X(\mathbf{v}-\mathbf{j}_n-\mathbf{j})X(\mathbf{v}-\mathbf{j}_m-\mathbf{i}-\mathbf{j}^*)Y(\mathbf{v})Y(\mathbf{v}-\mathbf{i})\} \\
= \; & \sum_{\mathbf{j},\mathbf{j}^*\geq\mathbf{0}} \Theta_{\mathbf{j},0}\Theta_{\mathbf{j}^*,0}\; E\{X(\mathbf{v}-\mathbf{j}_n-\mathbf{j})X(\mathbf{v}-\mathbf{j}_m-\mathbf{i}-\mathbf{j}^*)\}\; E\{Y(\mathbf{v})Y(\mathbf{v}-\mathbf{i})\} \\
= \; & \sigma^2\cdot\gamma_0(\mathbf{i}) \sum_{\mathbf{j},\mathbf{j}^*\geq\mathbf{0}} \Theta_{\mathbf{j},0}\Theta_{\mathbf{j}^*,0}\; c_0(\mathbf{j}_m-\mathbf{j}_n+\mathbf{i}+\mathbf{j}^*-\mathbf{j}). \qquad (3.4.133)
\end{aligned}$$

We should remember here that for the coefficients of the polynomials $\gamma_0(\mathbf{z})$, $c_0(\mathbf{z})$, it holds that

$$\begin{aligned}
E\{Y(\mathbf{v})Y(\mathbf{v}-\mathbf{j})\} &= \gamma_0(\mathbf{j}), \\
E\{X(\mathbf{v})X(\mathbf{v}-\mathbf{j})\} &= \sigma^2\cdot c_0(\mathbf{j}).
\end{aligned}$$

On the other hand, in the transpose $\mathbf{\Gamma}^{\tau}\mathbf{(i)}$ and still for $\mathbf{i} > \mathbf{0}$, we will have

$$\begin{aligned}
& \sigma^2\cdot\gamma_0(\mathbf{i}) \sum_{\mathbf{j},\mathbf{j}^*\geq\mathbf{0}} \Theta_{\mathbf{j},0}\Theta_{\mathbf{j}^*,0}\; c_0(\mathbf{j}_n-\mathbf{j}_m+\mathbf{i}+\mathbf{j}^*-\mathbf{j}) \\
= \; & \sigma^2\cdot\gamma_0(\mathbf{i}) \sum_{\mathbf{j},\mathbf{j}^*\geq\mathbf{0}} \Theta_{\mathbf{j},0}\Theta_{\mathbf{j}^*,0}\; c_0(\mathbf{j}_m-\mathbf{j}_n-\mathbf{i}-\mathbf{j}^*+\mathbf{j}),
\end{aligned}$$

which, if we re-write it for $\mathbf{i} < \mathbf{0}$, would be equal to

$$\begin{aligned} & \sigma^2 \cdot \gamma_0(\mathbf{i}) \sum_{\mathbf{j},\mathbf{j}^* \geq \mathbf{0}} \Theta_{\mathbf{j},0}\Theta_{\mathbf{j}^*,0}\, c_0(\mathbf{j}_m - \mathbf{j}_n + \mathbf{i} - \mathbf{j}^* + \mathbf{j}) \\ = \; & \sigma^2 \cdot \gamma_0(\mathbf{i}) \sum_{\mathbf{j},\mathbf{j}^* \geq \mathbf{0}} \Theta_{\mathbf{j},0}\Theta_{\mathbf{j}^*,0}\, c_0(\mathbf{j}_m - \mathbf{j}_n + \mathbf{i} + \mathbf{j}^* - \mathbf{j}). \end{aligned} \tag{3.4.134}$$

Taking the sum over all $\mathbf{i}^\tau \in \mathcal{Z}^d$ according to (3.4.132), would produce according to (3.4.133) and (3.4.134)

$$\begin{aligned} & \sum_{\mathbf{i}^\tau \in \mathcal{Z}^d} \sigma^2 \cdot \gamma_0(\mathbf{i}) \sum_{\mathbf{j},\mathbf{j}^* \geq \mathbf{0}} \Theta_{\mathbf{j},0}\Theta_{\mathbf{j}^*,0}\, c_0(\mathbf{j}_m - \mathbf{j}_n + \mathbf{i} + \mathbf{j}^* - \mathbf{j}) \\ = \; & \sigma^2 \sum_{\mathbf{j},\mathbf{j}^* \geq \mathbf{0}} \Theta_{\mathbf{j},0}\Theta_{\mathbf{j}^*,0} \sum_{\mathbf{i}^\tau \in \mathcal{Z}^d} \gamma_0(\mathbf{i})\, c_0(\mathbf{j}_m - \mathbf{j}_n + \mathbf{i} + \mathbf{j}^* - \mathbf{j}) \\ = \; & \sigma^2 \sum_{\mathbf{j} \geq \mathbf{0}} \Theta_{\mathbf{j},0}\Theta_{\mathbf{j}+\mathbf{j}_m-\mathbf{j}_n,0} \sum_{\mathbf{i}^\tau \in \mathcal{Z}^d} \gamma_0(\mathbf{i})\, c_0(\mathbf{i}) \\ + \; & \sigma^2 \sum_{\substack{\mathbf{j},\mathbf{j}^* \geq \mathbf{0}, \\ \mathbf{j}-\mathbf{j}^* \neq \mathbf{j}_m-\mathbf{j}_n}} \Theta_{\mathbf{j},0}\Theta_{\mathbf{j}^*,0} \sum_{\mathbf{i}^\tau \in \mathcal{Z}^d} \gamma_0(\mathbf{i})\, c_0(\mathbf{j}_m - \mathbf{j}_n + \mathbf{i} + \mathbf{j}^* - \mathbf{j}) \\ = \; & \sigma^4[\Theta_{\mathbf{j}_m-\mathbf{j}_n,0} + \sum_{\mathbf{j} > \mathbf{0}} \Theta_{\mathbf{j}_m-\mathbf{j}_n+\mathbf{j},0}\Theta_{\mathbf{j},0}], \end{aligned} \tag{3.4.135}$$

thanks to the general Yule-Walker equations. The last term (3.4.131) can be shown in exactly the same way to produce the same result.

For the second term (3.4.129), we write for $\mathbf{i} \leq \mathbf{0}$

$$\sum_{\mathbf{j},\mathbf{j}^* \geq \mathbf{0}} \Theta_{\mathbf{j},0}\Theta_{\mathbf{j}^*,0}\, E\{X(\mathbf{v} - \mathbf{j}_n - \mathbf{j})X(\mathbf{v} + \mathbf{j}_m - \mathbf{i} + \mathbf{j}^*)Y(\mathbf{v})Y(\mathbf{v} - \mathbf{i})\}.$$

It holds that $\mathbf{v} - \mathbf{j}_n - \mathbf{j} < \mathbf{v}$ and $\mathbf{v} - \mathbf{j}_n - \mathbf{j} < \mathbf{v} - \mathbf{i}$, and so we can write

$$\begin{aligned} & \sum_{\mathbf{j},\mathbf{j}^* \geq \mathbf{0}} \Theta_{\mathbf{j},0}\Theta_{\mathbf{j}^*,0}\, E\{X(\mathbf{v} - \mathbf{j}_n - \mathbf{j})X(\mathbf{v} + \mathbf{j}_m - \mathbf{i} + \mathbf{j}^*)Y(\mathbf{v})Y(\mathbf{v} - \mathbf{i})\} \\ = \; & \sum_{\mathbf{j},\mathbf{j}^* \geq \mathbf{0}} \Theta_{\mathbf{j},0}\Theta_{\mathbf{j}^*,0}\, E\{X(\mathbf{v} - \mathbf{j}_n - \mathbf{j})X(\mathbf{v} + \mathbf{j}_m - \mathbf{i} + \mathbf{j}^*)\}E\{Y(\mathbf{v})Y(\mathbf{v} - \mathbf{i})\} \\ = \; & \sigma^2 \cdot \gamma_0(\mathbf{i}) \sum_{\mathbf{j},\mathbf{j}^* \geq \mathbf{0}} \Theta_{\mathbf{j},0}\Theta_{\mathbf{j}^*,0}\, c_0(\mathbf{j}_m + \mathbf{j}_n + \mathbf{j} + \mathbf{j}^* - \mathbf{i}). \end{aligned} \tag{3.4.136}$$

On the other hand, from the same term when $\mathbf{i} > \mathbf{0}$, we write

$$\sum_{\mathbf{j},\mathbf{j}^* \geq \mathbf{0}} \Theta_{\mathbf{j},0}\Theta_{\mathbf{j}^*,0}\, E\{X(\mathbf{v} - \mathbf{j}_n - \mathbf{j})X(\mathbf{v} + \mathbf{j}_m - \mathbf{i} + \mathbf{j}^*)Y(\mathbf{v})Y(\mathbf{v} - \mathbf{i})\}.$$

Now, it can be that $\mathbf{v} - \mathbf{i} + \mathbf{j}_m + \mathbf{j}^* = \mathbf{v}$ and $\mathbf{v} - \mathbf{j}_n - \mathbf{j} = \mathbf{v} - \mathbf{i}$, or equally that $\mathbf{j}^* = \mathbf{i} - \mathbf{j}_m$

and $\mathbf{j} = \mathbf{i} - \mathbf{j}_n$. As a result, we can write for $\mathbf{i} \geq \mathbf{j}_m$, that

$$\begin{aligned}
&\sum_{\mathbf{j},\mathbf{j}^* \geq \mathbf{0}} \Theta_{\mathbf{j},0}\Theta_{\mathbf{j}^*,0}\, E\{X(\mathbf{v}-\mathbf{j}_n-\mathbf{j})X(\mathbf{v}+\mathbf{j}_m-\mathbf{i}+\mathbf{j}^*)Y(\mathbf{v})Y(\mathbf{v}-\mathbf{i})\} \\
&= \sigma^4 \cdot \Theta_{\mathbf{i}-\mathbf{j}_n,0}\Theta_{\mathbf{i}-\mathbf{j}_m,0} \\
&+ \sum_{\mathbf{j},\mathbf{j}^* \geq \mathbf{0}} \Theta_{\mathbf{j},0}\Theta_{\mathbf{j}^*,0}\, E\{X(\mathbf{v}-\mathbf{j}_n-\mathbf{j})X(\mathbf{v}+\mathbf{j}_m-\mathbf{i}+\mathbf{j}^*)\}E\{Y(\mathbf{v})Y(\mathbf{v}-\mathbf{i})\} \\
&= \sigma^4 \cdot \Theta_{\mathbf{i}-\mathbf{j}_n,0}\Theta_{\mathbf{i}-\mathbf{j}_m,0} + \sigma^2 \cdot \gamma_0(\mathbf{i}) \sum_{\mathbf{j},\mathbf{j}^* \geq \mathbf{0}} \Theta_{\mathbf{j},0}\Theta_{\mathbf{j}^*,0}\, c_0(\mathbf{j}_m+\mathbf{j}_n+\mathbf{j}+\mathbf{j}^*-\mathbf{i}),
\end{aligned} \tag{3.4.137}$$

while when $\mathbf{0} < \mathbf{i} < \mathbf{j}_m$, it is just

$$\begin{aligned}
&\sum_{\mathbf{j},\mathbf{j}^* \geq \mathbf{0}} \Theta_{\mathbf{j},0}\Theta_{\mathbf{j}^*,0}\, E\{X(\mathbf{v}-\mathbf{j}_n-\mathbf{j})X(\mathbf{v}+\mathbf{j}_m-\mathbf{i}+\mathbf{j}^*)Y(\mathbf{v})Y(\mathbf{v}-\mathbf{i})\} \\
&= \sigma^2 \cdot \gamma_0(\mathbf{i}) \sum_{\mathbf{j},\mathbf{j}^* \geq \mathbf{0}} \Theta_{\mathbf{j},0}\Theta_{\mathbf{j}^*,0}\, c_0(\mathbf{j}_m+\mathbf{j}_n+\mathbf{j}+\mathbf{j}^*-\mathbf{i}).
\end{aligned} \tag{3.4.138}$$

Putting (3.4.136), (3.4.137) and (3.4.138) together for all $\mathbf{i}^\tau \in \mathcal{Z}^d$, we may write

$$\begin{aligned}
&\sum_{\mathbf{i} \geq \mathbf{j}_m} \sigma^4 \cdot \Theta_{\mathbf{i}-\mathbf{j}_n,0}\Theta_{\mathbf{i}-\mathbf{j}_m,0} \\
&+ \sum_{\mathbf{i}^\tau \in \mathcal{Z}^d} \sigma^2 \cdot \gamma_0(\mathbf{i}) \sum_{\mathbf{j},\mathbf{j}^* \geq \mathbf{0}} \Theta_{\mathbf{j},0}\Theta_{\mathbf{j}^*,0}\, c_0(\mathbf{j}_m+\mathbf{j}_n+\mathbf{j}+\mathbf{j}^*-\mathbf{i}) \\
&= \sigma^4[\Theta_{\mathbf{j}_m-\mathbf{j}_n,0} + \sum_{\mathbf{j}>\mathbf{0}} \Theta_{\mathbf{j}_m-\mathbf{j}_n+\mathbf{j},0}\Theta_{\mathbf{j},0}] \\
&+ \sigma^2 \sum_{\mathbf{j},\mathbf{j}^* \geq \mathbf{0}} \Theta_{\mathbf{j},0}\Theta_{\mathbf{j}^*,0} \sum_{\mathbf{i}^\tau \in \mathcal{Z}^d} \gamma_0(\mathbf{i})\, c_0(\mathbf{j}_m+\mathbf{j}_n+\mathbf{j}+\mathbf{j}^*-\mathbf{i}) \\
&= \sigma^4[\Theta_{\mathbf{j}_m-\mathbf{j}_n,0} + \sum_{\mathbf{j}>\mathbf{0}} \Theta_{\mathbf{j}_m-\mathbf{j}_n+\mathbf{j},0}\Theta_{\mathbf{j},0}],
\end{aligned} \tag{3.4.139}$$

since it cannot be that $\mathbf{j}_m + \mathbf{j}_n + \mathbf{j} + \mathbf{j}^* = \mathbf{0}$. The third term (3.4.130) can be shown in a similar way.

As a result, for the (n,m)-th element of $\mathbf{M}$ is according to (3.4.135) and (3.4.139) equal to

$$4 \cdot \sigma^4[\Theta_{\mathbf{j}_m-\mathbf{j}_n,0} + \sum_{\mathbf{j}>\mathbf{0}} \Theta_{\mathbf{j}_m-\mathbf{j}_n+\mathbf{j},0}\Theta_{\mathbf{j},0}],$$

when $m \geq n$. Then

$$\mathbf{M} = 4 \cdot \sigma^4 \mathbf{W}_q, \tag{3.4.140}$$

which combined with (3.4.125) and Proposition 3.3 proves the theorem.

■

3.5 Modified Gaussian likelihood estimation for ARMA processes

3.5.1 Introduction

The edge-effect does not allow us to maximize the exact Gaussian likelihood with the hope to produce both asymptotically unbiased and normal estimators. Even when the observations have been derived from an auto-regression or a moving-average, we need to maximize a modified version of the Gaussian likelihood rather than the exact likelihood itself, to rest that the estimators have both these properties. When it comes to the ARMA though, it seems that no modified version of the likelihood, which uses the special characteristics of the ARMA, has been proposed. Guyon's (1982) modified likelihood might be maximized, but it is based on a quantity that uses all the possible auto-covariances in the sample and keeps growing as the number of observations increases, since the ARMA has non-zero auto-covariances at all possible lags. This is the quantity used in the case of any process with an auto-covariance function that does not necessarily cut off after a finite number of lags and, thus, it does not make any use of the special structure of the ARMA, as it does for a moving-average. Similarly, the modified likelihood proposed by Yao and Brockwell (2006) for the case of $d = 2$ dimensions, works for any process with an AR(∞) representation and not just the ARMA, as the innovations algorithm allows at each observation to be expressed as a linear function of all the other observations in the sample from its 'past' only, until it reaches the AR(∞) representation. Again, unless the original process is an auto-regression of finite order, that method does not distinguish between the ARMA and any other process with an AR(∞) representation.

For observations from an ARMA process, we minimize a quantity, which produces consistent, asymptotically unbiased and normal estimators. For Gaussian processes, the estimators are efficient too. Based on arguments from the previous section, this quantity is the quadratic form of a modified Gaussian likelihood, and, thus, we consider our estimators to be Gaussian likelihood estimators. This is the same idea as the quasi-maximum likelihood estimators introduced by White (1982). Consequently, that same idea was also used by Guyon (1982), Yao and Brockwell (2006) and all our maximum likelihood suggestions for estimation so far, refer to Gaussian likelihoods too. Thus, our

proposed method for the ARMA generalizes the case of modified likelihood estimation for the parameters of a moving-average, which was presented in the previous section. The main idea is finding a finite filter, which, if applied on the ARMA process, produces a moving-average process. Moreover, as we are going to see in Section 3.5.2, there can be three different moving-averages produced, but only one of them is appropriate to be used in the quantity; this is, in order to avoid the edge-effect and not to come up with the original ARMA process and its auto-covariance function, after finding the derivatives of the quantity with respect to the parameters. Thus, we select this moving-average process, for which the derivatives with respect to all the parameters are moving-average processes. The auto-regression that corresponds to the moving-average of interest is again of great importance for the derivation of the results, though it is only possible to produce it after applying an infinite filter on the ARMA, and, so, not possible with a finite sample available.

3.5.2 Definitions

For $\mathbf{0} < \mathbf{i}_1 < \cdots < \mathbf{i}_p$ and $\mathbf{0} < \mathbf{j}_1 < \cdots < \mathbf{j}_q$, we consider the causal and invertible ARMA process $\{Z(\mathbf{v}),\ \mathbf{v}^\tau \in \mathbb{Z}^d\}$, such that it satisfies

$$Z(\mathbf{v}) - \sum_{n=1}^{p} b_{\mathbf{i}_n} Z(\mathbf{v} - \mathbf{i}_n) \equiv \varepsilon(\mathbf{v}) + \sum_{m=1}^{q} a_{\mathbf{j}_m} \varepsilon(\mathbf{v} - \mathbf{j}_m),\ \{\varepsilon(\mathbf{v})\} \sim WN(0, \sigma^2). \quad (3.5.1)$$

We define the polynomials

$$b(\mathbf{z}) \equiv 1 - \sum_{n=1}^{p} b_{\mathbf{i}_n}\ \mathbf{z}^{\mathbf{i}_n} \quad (3.5.2)$$

$$a(\mathbf{z}) \equiv 1 + \sum_{m=1}^{q} a_{\mathbf{j}_m}\ \mathbf{z}^{\mathbf{j}_m} \quad (3.5.3)$$

and the polynomial

$$d(\mathbf{z}) \equiv \{b(\mathbf{z})a(\mathbf{z})\ b(\mathbf{z}^{-1})a(\mathbf{z}^{-1})\}^{-1} \equiv \sum_{\mathbf{j}^\tau \in \mathbb{Z}^d} d_{\mathbf{j}}\ \mathbf{z}^{\mathbf{j}}. \quad (3.5.4)$$

Consequently from (3.5.1), (3.5.2) and (3.5.3), we may write the following equations

$$b(\mathbf{B})Z(\mathbf{v}) = a(\mathbf{B})\varepsilon(\mathbf{v}) = a(\mathbf{B}^{-1})\varepsilon^*(\mathbf{v}) \quad (3.5.5)$$

and

$$b(\mathbf{B}^{-1})Z(\mathbf{v}) = a(\mathbf{B}^{-1})u(\mathbf{v}) = a(\mathbf{B})u^*(\mathbf{v}), \quad (3.5.6)$$

where it holds that

$$\{\varepsilon(\mathbf{v})\}, \{\varepsilon^*(\mathbf{v})\}, \{u(\mathbf{v})\}, \{u^*(\mathbf{v})\} \sim WN(0, \sigma^2)$$

and, so, the processes $\{\varepsilon^*(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$, $\{u(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ and $\{u^*(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ are sequences of uncorrelated random variables with variance σ^2, too. The process $\{u(\mathbf{v})\}$ is the unilateral counterpart of $\{\varepsilon(\mathbf{v})\}$ in the AR(∞) representation of $\{Z(\mathbf{v})\}$. Then $\{\varepsilon^*(\mathbf{v})\}$ and $\{u^*(\mathbf{v})\}$ are the unilateral counterparts of $\{\varepsilon(\mathbf{v})\}$ and $\{u(\mathbf{v})\}$ for the moving-average processes $\{b(\mathbf{B})Z(\mathbf{v})\}$ and $\{b(\mathbf{B}^{-1})Z(\mathbf{v})\}$, respectively.

We define the new process $\{M(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ by the equations

$$\begin{aligned} M(\mathbf{v}) \equiv b(\mathbf{B})b(\mathbf{B}^{-1})Z(\mathbf{v}) &= b(\mathbf{B}^{-1})a(\mathbf{B}^{-1})\varepsilon^*(\mathbf{v}),\ \{\varepsilon^*(\mathbf{v})\} \sim WN(0,\sigma^2) \quad (3.5.7) \\ &= b(\mathbf{B})a(\mathbf{B})u^*(\mathbf{v}), \qquad \{u^*(\mathbf{v})\} \sim WN(0,\sigma^2), \quad (3.5.8) \end{aligned}$$

which has been expressed in (3.5.8) as an invertible moving-average. Its spectral density can be written as

$$g_M(\boldsymbol{\omega}) = \frac{\sigma^2}{(2\pi)^d}\, b(e^{i\boldsymbol{\omega}})a(e^{i\boldsymbol{\omega}})\, b(e^{-i\boldsymbol{\omega}})a(e^{-i\boldsymbol{\omega}}) = \frac{\sigma^2}{(2\pi)^d}\, \frac{1}{d(e^{i\boldsymbol{\omega}})},\ \boldsymbol{\omega}^\tau \in [-\pi,\pi]^d. \quad (3.5.9)$$

We also define the process $\{A(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ by the equations

$$A(\mathbf{v}) \equiv a(\mathbf{B})^{-1}a(\mathbf{B}^{-1})^{-1}Z(\mathbf{v}) = b(\mathbf{B})^{-1}a(\mathbf{B})^{-1}\varepsilon^*(\mathbf{v}) = b(\mathbf{B}^{-1})^{-1}a(\mathbf{B}^{-1})^{-1}u^*(\mathbf{v}), \quad (3.5.10)$$

such that

$$b(\mathbf{B})a(\mathbf{B})A(\mathbf{v}) = \varepsilon^*(\mathbf{v}),\ \{\varepsilon^*(\mathbf{v})\} \sim WN(0,\sigma^2) \quad (3.5.11)$$

and

$$b(\mathbf{B}^{-1})a(\mathbf{B}^{-1})A(\mathbf{v}) = u^*(\mathbf{v}),\ \{u^*(\mathbf{v})\} \sim WN(0,\sigma^2). \quad (3.5.12)$$

In (3.5.11), the process $\{A(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ has been expressed as a causal auto-regression. We can see immediately from (3.5.7), (3.5.8) and (3.5.10) that $M(\mathbf{v})$ and $A(\mathbf{v}-\mathbf{j})$ are two uncorrelated random variables for any $\mathbf{j} \neq \mathbf{0}$.

Apart from the process $\{M(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$, we also define the moving-average processes

$$\begin{aligned} Y_1(\mathbf{v}) \equiv b(\mathbf{B})Z(\mathbf{v}) &= a(\mathbf{B})\varepsilon(\mathbf{v}), \qquad \{\varepsilon(\mathbf{v})\} \sim WN(0,\sigma^2) \quad (3.5.13) \\ &= a(\mathbf{B}^{-1})\varepsilon^*(\mathbf{v}),\ \{\varepsilon^*(\mathbf{v})\} \sim WN(0,\sigma^2) \quad (3.5.14) \end{aligned}$$

and

$$Y_2(\mathbf{v}) \equiv b(\mathbf{B}^{-1})Z(\mathbf{v}) = a(\mathbf{B}^{-1})u(\mathbf{v}),\ \{u(\mathbf{v})\} \sim WN(0,\sigma^2), \quad (3.5.15)$$

$$= a(\mathbf{B})u^*(\mathbf{v}), \quad \{u(\mathbf{v})\} \sim WN(0,\sigma^2). \quad (3.5.16)$$

We consider $\mathcal{F}$ to be the minimal set, such that for any $\mathbf{j}^\tau \notin \mathcal{F}$, it holds that

$$E\{Y_1(\mathbf{v})\, M(\mathbf{v}+\mathbf{j})\} = E\{Y_2(\mathbf{v})\, M(\mathbf{v}+\mathbf{j})\} = E\{M(\mathbf{v})\, Y_1(\mathbf{v}+\mathbf{j})\} = E\{M(\mathbf{v})\, Y_2(\mathbf{v}+\mathbf{j})\} = 0$$

and that

$$E\{Y_1(\mathbf{v}+\mathbf{i}_n)\, M(\mathbf{v}+\mathbf{j})\} = 0, \quad E\{Y_2(\mathbf{v}-\mathbf{i}_n)\, M(\mathbf{v}+\mathbf{j})\} = 0, \quad (3.5.17)$$

$$E\{M(\mathbf{v})\, Y_1(\mathbf{v}+\mathbf{i}_n+\mathbf{j})\} = 0, \quad E\{M(\mathbf{v})\, Y_2(\mathbf{v}-\mathbf{i}_n+\mathbf{j})\} = 0, \quad (3.5.18)$$

for all $n = 1, \cdots, p$. Since it holds that

$$M(\mathbf{v}) = b(\mathbf{B})Y_2(\mathbf{v}) = Y_2(\mathbf{v}) - \sum_{n=1}^{p} b_{\mathbf{i}_n} Y_2(\mathbf{v}-\mathbf{i}_n) \quad (3.5.19)$$

$$= b(\mathbf{B}^{-1})Y_1(\mathbf{v}) = Y_1(\mathbf{v}) - \sum_{n=1}^{p} b_{\mathbf{i}_n} Y_1(\mathbf{v}+\mathbf{i}_n), \quad (3.5.20)$$

we can see immediately that if $\mathbf{j}^\tau \notin \mathcal{F}$, then $E\{M(\mathbf{v})M(\mathbf{v}+\mathbf{j})\} = 0$.

We consider $\mathcal{F}_1$ the maximal set of lags $\mathbf{j}^\tau \in \mathcal{F}_1$, such that

$$E\{Y_1(\mathbf{v})\, M(\mathbf{v}+\mathbf{j})\} \neq 0,$$

and $\mathcal{F}_2$ the maximal set of lags $\mathbf{j}^\tau \in \mathcal{F}_2$, such that

$$E\{Y_2(\mathbf{v})\, M(\mathbf{v}+\mathbf{j})\} \neq 0,$$

and we go on the same way with the sets $\mathcal{F}_3, \mathcal{F}_4, \mathcal{F}_{4+1}, \cdots, \mathcal{F}_{4+p}, \cdots \mathcal{F}_{4+4p-1}$, to finish with the maximal set $\mathcal{F}_{4+4p}$ of lags $\mathbf{j}^\tau \in \mathcal{F}_{4+4p}$, such that

$$E\{M(\mathbf{v})\, Y_2(\mathbf{v}-\mathbf{i}_p+\mathbf{j})\} \neq 0.$$

We can see immediately that all the sets $\mathcal{F}_1, \cdots, \mathcal{F}_{4+4p}$ are of finite cardinality. This is because all processes $\{M(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$, $\{Y_1(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ and $\{Y_2(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ are moving-averages. Moreover, it holds that

$$\mathcal{F} = \mathcal{F}_1 \cup \cdots \cup \mathcal{F}_{4+4p}. \quad (3.5.21)$$

Thus, $\mathcal{F}$ is a set of finite cardinality too. In general, we may consider $\mathcal{F}$ to be the set of lags, where the process $\{M(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ has non-zero auto-covariances.

3.5.3 Estimators

We observe $\{Z(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{I}\}$ and wish to estimate the unknown parameter vector

$$\boldsymbol{\theta}_0 = [b_{\mathbf{i}_1,0}, \cdots, b_{\mathbf{i}_p,0}, a_{\mathbf{j}_1,0}, \cdots, a_{\mathbf{j}_q,0}]^\tau = [\mathbf{b}_0^\tau, \mathbf{a}_0^\tau]^\tau. \tag{3.5.22}$$

We assume the following condition holds.

(C4) The parameter space $\Theta = \Theta_1 \times \Theta_2 \subset \mathcal{R}^{p+q}$ is a compact set containing the true value $\boldsymbol{\theta}_0$ as an inner point. Further, for any $\boldsymbol{\theta} = [\mathbf{b}^\tau, \mathbf{a}^\tau]^\tau \in \Theta$, the ARMA process (3.5.1) is causal and invertible.

For a set $\mathcal{I} \subset \mathcal{Z}^d$ and for observations $\{Z(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{I}\}$ we consider the maximal set $\mathcal{S}$, such that $\mathbf{v}^\tau \in \mathcal{S}$ if $\mathbf{v}^\tau + \mathbf{i}_n^\tau, \mathbf{v}^\tau - \mathbf{i}_n^\tau, \mathbf{v}^\tau + \mathbf{i}_n^\tau - \mathbf{i}_m^\tau \in \mathcal{I}$ for all $n, m = 1, \cdots, p$. Then we consider the maximal set $\mathcal{S}^*$, such that $\mathbf{v}^\tau \in \mathcal{S}^*$ if $\mathbf{v}^\tau - \mathbf{j}^\tau \in \mathcal{S}$ for all $\mathbf{j}^\tau \in \mathcal{F}$. We assume that $\mathcal{I}$ is a large enough set, so that $\mathcal{S}^*$ is not the empty set. We denote with $N_\mathcal{I}, N, N^*$ the cardinalities of $\mathcal{I}, \mathcal{S}$ and $\mathcal{S}^*$, respectively. Finally, for any $\mathbf{v}^\tau \in \mathcal{Z}^d$, we define the set $\mathcal{F}_\mathbf{v} \subset \mathcal{Z}^d$ to be such that $\mathbf{j}^\tau \in \mathcal{F}_\mathbf{v}$ if $\mathbf{v}^\tau - \mathbf{j}^\tau \in \mathcal{S}$. We can see immediately that for any $\mathbf{v}^\tau \in \mathcal{S}^*$, it holds that $\mathcal{F} \subseteq \mathcal{F}_\mathbf{v}$.

For a better understanding, we present the following example. We define the ARMA process $\{Z_t,\ t \in \mathcal{Z}\}$ by the equation

$$Z_t - 0.6\ Z_{t-1} + 0.09\ Z_{t-2} \equiv \varepsilon_t + 0.1\ \varepsilon_{t-1},\ \{\varepsilon_t\} \sim WN(0, \sigma^2).$$

For the set $\mathcal{I} = \{1, \cdots, 20\}$, the new set $\mathcal{S} = \{3, \cdots, 18\}$, is such that for all $t \in \mathcal{S}$, it holds that $t-2, \cdots, t+2 \in \mathcal{I}$. Before we define the set $\mathcal{S}^*$, we will need to find $\mathcal{F}$. Indeed, if we imagine the moving-average process $\{M_t,\ t \in \mathcal{Z}\}$ defined by

$$\begin{aligned} M_t &\equiv [u_t^* + 0.1\ u_{t-1}^*] - 0.6\ [u_{t-1}^* + 0.1\ u_{t-2}^*] + 0.09\ [u_{t-2}^* + 0.1\ u_{t-3}^*] \\ &= u_t^* - 0.5\ u_{t-1}^* + 0.03\ u_{t-2}^* + 0.009\ u_{t-3}^*,\ \{u_t^*\} \sim WN(0, \sigma^2), \end{aligned}$$

then it holds that $E\{M_t\ M_{t+j}\} \neq 0$ when $j \in \mathcal{F} = \{0, \pm 1, \pm 2, \pm 3\}$. Thus, we consider the set $\mathcal{S}^* = \{6, \cdots, 15\}$ with cardinality $N^* = 10$, instead of the cardinality of the original set $N_\mathcal{I} = 20$.

Back to our general case, for any $\mathbf{v}^\tau \in \mathcal{Z}^d$ and $\mathbf{b} \in \Theta_1$, we define the random variables

$$\begin{aligned} M(\mathbf{v}, \mathbf{b}) &\equiv b(\mathbf{B})b(\mathbf{B}^{-1})Z(\mathbf{v}) \\ &= Z(\mathbf{v}) - \sum_{n=1}^{p} b_{\mathbf{i}_n}[Z(\mathbf{v} + \mathbf{i}_n) + Z(\mathbf{v} - \mathbf{i}_n)] + \sum_{n,m=1}^{p} b_{\mathbf{i}_n} b_{\mathbf{i}_m} Z(\mathbf{v} + \mathbf{i}_n - \mathbf{i}_m). \end{aligned} \tag{3.5.23}$$

and

$$M(\mathbf{v}, \mathbf{b}_0) \equiv M(\mathbf{v}). \tag{3.5.24}$$

For any $\mathbf{v}^\tau \in \mathcal{Z}^d$ and $\mathbf{b} \in \Theta_1$, we also define

$$Y_1(\mathbf{v}, \mathbf{b}) \equiv b(\mathbf{B})Z(\mathbf{v}) = Z(\mathbf{v}) - \sum_{n=1}^{p} b_{\mathbf{i}_n} Z(\mathbf{v} - \mathbf{i}_n) \tag{3.5.25}$$

$$Y_2(\mathbf{v}, \mathbf{b}) \equiv b(\mathbf{B}^{-1})Z(\mathbf{v}) = Z(\mathbf{v}) - \sum_{n=1}^{p} b_{\mathbf{i}_n} Z(\mathbf{v} + \mathbf{i}_n) \tag{3.5.26}$$

and

$$Y_1(\mathbf{v}, \mathbf{b}_0) \equiv Y_1(\mathbf{v}) \tag{3.5.27}$$

$$Y_2(\mathbf{v}, \mathbf{b}_0) \equiv Y_2(\mathbf{v}). \tag{3.5.28}$$

We first define the quantity

$$Q^*(\boldsymbol{\theta}) \equiv \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v}, \mathbf{b}) \sum_{\mathbf{j}^\tau \in \mathcal{F}_\mathbf{v}} d_\mathbf{j} \, M(\mathbf{v} - \mathbf{j}, \mathbf{b}), \ \boldsymbol{\theta} \in \Theta \tag{3.5.29}$$

and then the estimators

$$\widehat{\boldsymbol{\theta}} \equiv \text{argmin}_{\boldsymbol{\theta} \in \Theta} Q^*(\boldsymbol{\theta}). \tag{3.5.30}$$

If

$$\{\varepsilon(\mathbf{v})\} \sim NID(0, \sigma^2)$$

and we had observed $\{M(\mathbf{v}), \ \mathbf{v}^\tau \in \mathcal{S}\}$, we saw in Section 3.4.3 why

$$L_M(\boldsymbol{\theta}, \sigma^2) \propto \frac{|\text{Var}\{\mathbf{A}\}|^{1/2}}{(\sigma^4)^{N/2}} \exp\{-\frac{1}{2\sigma^2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v}) \sum_{\mathbf{j}^\tau \in \mathcal{F}_\mathbf{v}} d_\mathbf{j} \, M(\mathbf{v} - \mathbf{j})\}, \ \boldsymbol{\theta} \in \Theta, \ \sigma^2 > 0, \tag{3.5.31}$$

is a conditional likelihood. These are the same reasons that make (3.4.73) a conditional likelihood. Again $\mathbf{A}$ is the vector of random variables $\{A(\mathbf{v}), \ \mathbf{v}^\tau \in \mathcal{S}\}$ from the auto-regression $\{A(\mathbf{v}), \ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ as it was defined in (3.5.10). Following the same arguments as to derive (3.4.70), we may also write the modified version of the Gaussian likelihood

$$L_M^*(\boldsymbol{\theta}, \sigma^2) \propto \frac{1}{(\sigma^2)^{N^*/2}} \exp\{-\frac{1}{2\sigma^2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v}) \sum_{\mathbf{j}^\tau \in \mathcal{F}_\mathbf{v}} d_\mathbf{j} \, M(\mathbf{v} - \mathbf{j})\}, \ \boldsymbol{\theta} \in \Theta, \ \sigma^2 > 0. \tag{3.5.32}$$

If we now consider the random variables $\{Z(\mathbf{v}), \ \mathbf{v}^\tau \in \mathcal{S}\}$ and using (3.5.23), we can write

$$\mathbf{M} = \mathbf{B} \cdot \mathbf{Z} + \mathbf{B}_0 \cdot \mathbf{Z}_0 \tag{3.5.33}$$

where $\mathbf{M} = [M(\mathbf{v}_1), \cdots, M(\mathbf{v}_N)]^\tau$, $\mathbf{Z} = [Z(\mathbf{v}_1), \cdots, Z(\mathbf{v}_N)]^\tau$ and $\mathbf{v}_1 < \cdots < \mathbf{v}_N$ are all the elements of $\mathcal{S}$. The vector $\mathbf{Z}_0$ has elements the random variables $\{Z(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{I} - \mathcal{S}\}$. If we write

$$b(\mathbf{z})\ b(\mathbf{z}^{-1}) \equiv \sum_{\mathbf{j}^\tau \in \mathcal{F}_b} \beta_{\mathbf{j}}\ \mathbf{z}^{\mathbf{j}} \equiv \sum_{\mathbf{j}^\tau \in \mathcal{Z}^d} \beta_{\mathbf{j}}\ \mathbf{z}^{\mathbf{j}}, \tag{3.5.34}$$

with $\beta_{\mathbf{j}} = 0, \mathbf{j}^\tau \notin \mathcal{F}_b$, and

$$\mathbf{B} \equiv [\beta^*_{r,l}]_{r,l=1}^N, \tag{3.5.35}$$

it holds that

$$\beta^*_{r,l} = \beta_{\mathbf{v}_r - \mathbf{v}_l} = \beta_{\mathbf{v}_l - \mathbf{v}_r},\ r, l = 1, \cdots, N. \tag{3.5.36}$$

If we denote with $f_{\mathbf{M}|\mathbf{Z}_0}$ and $f_{\mathbf{Z}|\mathbf{Z}_0}$ the conditional densities of the random vectors $\mathbf{M}$ and $\mathbf{Z}$, respectively, given the values of the random vector $\mathbf{Z}_0$, then it holds, according to (3.5.33), that

$$f_{\mathbf{Z}|\mathbf{Z}_0} = |\mathbf{B}| \cdot f_{\mathbf{M}|\mathbf{Z}_0}. \tag{3.5.37}$$

Moreover, if we consider the moving-average process $\{Y_H(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ defined by the equation

$$Y_H(\mathbf{v}) \equiv b(\mathbf{B})e(\mathbf{v}),\ \{e(\mathbf{v})\} \sim WN(0,1), \tag{3.5.38}$$

then it holds that

$$\mathrm{Var}\{\mathbf{Y}_H\} = \mathbf{B} \tag{3.5.39}$$

for the random vector $\mathbf{Y}_H = [Y_H(\mathbf{v}_1), \cdots, Y_H(\mathbf{v}_N)]^\tau$. The innovations algorithm teaches us how to factorize the determinant of a variance matrix

$$|\mathbf{B}| = \prod_{\mathbf{v}^\tau \in \mathcal{S}} r_b(\mathbf{v}), \tag{3.5.40}$$

where

$$r_b(\mathbf{v}) \equiv E\{Y_H(\mathbf{v}) - \widetilde{Y}_H(\mathbf{v})\}^2,\ \mathbf{v}^\tau \in \mathcal{S}, \tag{3.5.41}$$

and $\widetilde{Y}_H(\mathbf{v})$ is the best linear predictor of $Y_H(\mathbf{v})$ based on all random variables $Y_H(\mathbf{v} - \mathbf{i}), \mathbf{i} > \mathbf{0}, \mathbf{v}^\tau - \mathbf{i}^\tau \in \mathcal{S}$.

We may also see that since

$$Z(\mathbf{v}) = a(\mathbf{B})a(\mathbf{B}^{-1})A(\mathbf{v}),$$

it holds that

$$\mathrm{Cov}\{M(\mathbf{v} - \mathbf{j}), Z(\mathbf{v})\} = \mathrm{Cov}\{M(\mathbf{v} - \mathbf{j}), a(\mathbf{B})a(\mathbf{B}^{-1})A(\mathbf{v})\} \propto \alpha_{\mathbf{j}},\ \mathbf{j}^\tau \in \mathcal{Z}^d, \tag{3.5.42}$$

where we consider

$$a(\mathbf{z})\ a(\mathbf{z}^{-1}) \equiv \sum_{\mathbf{j}^\tau \in \mathcal{F}_a} \alpha_{\mathbf{j}}\ \mathbf{z}^{\mathbf{j}} \equiv \sum_{\mathbf{j}^\tau \in \mathcal{Z}^d} \alpha_{\mathbf{j}}\ \mathbf{z}^{\mathbf{j}}, \tag{3.5.43}$$

with $\alpha_{\mathbf{j}} = 0$ for $\mathbf{j}^\tau \notin \mathcal{F}_a$. Thus, for Gaussian random variables, the random variables $\{M(\mathbf{v}), \mathbf{v}^\tau \in \mathcal{S}^*\}$ are independent of the random vector $\mathbf{Z}_0$. Using this last argument and also (3.5.32), (3.5.37) and (3.5.40), we may come up with the modified conditional likelihood of the observations $\{Z(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{S}^*\}$ given the realization of the random vector $\mathbf{Z}_0$

$$L_Z^*(\boldsymbol{\theta}, \sigma^2) \propto \frac{\prod_{\mathbf{v}^\tau \in \mathcal{S}^*} r_b(\mathbf{v})}{(\sigma^2)^{N^*/2}} \exp\{-\frac{1}{2\sigma^2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v}, \mathbf{b}) \sum_{\mathbf{j}^\tau \in \mathcal{F}_\mathbf{v}} d_{\mathbf{j}}\ M(\mathbf{v}-\mathbf{j}, \mathbf{b})\},\ \boldsymbol{\theta} \in \Theta,\ \sigma^2 > 0, \tag{3.5.44}$$

where $M(\mathbf{v}, \mathbf{b})$ was defined in (3.5.23) to be a function of the data and the parameters. For convenience, we will ignore the factor

$$\prod_{\mathbf{v}^\tau \in \mathcal{S}^*} r_b(\mathbf{v})$$

and we will finally consider the modified version of conditional Gaussian likelihood to be

$$L^*(\boldsymbol{\theta}, \sigma^2) \propto \frac{1}{(\sigma^2)^{N^*/2}} \exp\{-\frac{1}{2\sigma^2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v}, \mathbf{b}) \sum_{\mathbf{j}^\tau \in \mathcal{F}_\mathbf{v}} d_{\mathbf{j}}\ M(\mathbf{v}-\mathbf{j}, \mathbf{b})\},\ \boldsymbol{\theta} \in \Theta,\ \sigma^2 > 0. \tag{3.5.45}$$

The factor

$$\prod_{\mathbf{v}^\tau \in \mathcal{S}^*} r_b(\mathbf{v})$$

involves the parameters $\mathbf{b} \in \Theta_1$ and not the data $\{Z(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{I}\}$ and it is part of the deterministic, not the random part of the likelihood. For more information on the properties of the prediction variances

$$r_b(\mathbf{v}, \mathbf{b}) \equiv r_b(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{S},$$

we refer to Yao and Brockwell (2006). For example, for any $\mathbf{b} \in \Theta_1$, an invertible moving-average $\{Y_H(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ is defined from (3.5.38), and for any fixed element $\mathbf{v}^\tau \in \mathcal{S}$, it holds that

$$r_b(\mathbf{v}, \mathbf{b}) \to \mathrm{Var}\{e(\mathbf{v})\} = 1, \tag{3.5.46}$$

as $N \to \infty$ and (C1)(i) holds. The assumption for the causality of the original ARMA process or invertibility of the moving-average process defined by (3.5.38) is essential for (3.5.46) to hold and for the factor $\prod_{\mathbf{v}^\tau \in \mathcal{S}^*} r_b(\mathbf{v})$ to be omitted. Otherwise, as we are going

to see in Section 3.6, a modification needs to be taken into account in the deterministic part of the likelihood.

Finally, if we re-write (3.5.45) as

$$L^*(\boldsymbol{\theta}, \sigma^2) \propto \frac{1}{(\sigma^2)^{N^*/2}} \exp\{-\frac{1}{2\sigma^2}\, Q^*(\boldsymbol{\theta})\},\ \boldsymbol{\theta} \in \Theta,\ \sigma^2 > 0, \tag{3.5.47}$$

we can see that $\widehat{\boldsymbol{\theta}}$, as it was defined in (3.5.30), and

$$\widehat{\sigma^2} \equiv Q^*(\widehat{\boldsymbol{\theta}})/N^* \tag{3.5.48}$$

are maximum modified Gaussian likelihood estimators.

3.5.4 Properties of estimators

Theorem 3.5 (Consistency). If $\{\varepsilon^*(\mathbf{v})\} \sim IID(0, \sigma^2)$, then under conditions (C1)(i) and (C4), it holds that

$$\widehat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}_0$$

and

$$\widehat{\sigma^2} \xrightarrow{P} \sigma^2$$

as $N \to \infty$.

Proof. Similar arguments like the ones for Theorem 3.3 will be used. For any $\boldsymbol{\theta} \in \Theta$, we can write

$$\begin{aligned}
& \frac{1}{N} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} [b(\mathbf{B})b(\mathbf{B}^{-1})Z(\mathbf{v})] \sum_{\mathbf{i}^\tau \in \mathcal{F}_\mathbf{v}} d_\mathbf{i}\, [b(\mathbf{B})b(\mathbf{B}^{-1})Z(\mathbf{v}-\mathbf{i})] \\
= & \frac{1}{N} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} [b(\mathbf{B})b_0(\mathbf{B})^{-1}b(\mathbf{B}^{-1})b_0(\mathbf{B}^{-1})^{-1}M(\mathbf{v})] \\
& \qquad \sum_{\mathbf{i}^\tau \in \mathcal{F}_\mathbf{v}} d_\mathbf{i}\, [b(\mathbf{B})b_0(\mathbf{B})^{-1}\, b(\mathbf{B}^{-1})b_0(\mathbf{B}^{-1})^{-1}M(\mathbf{v}-\mathbf{i})] \\
\xrightarrow{P} & \frac{1}{N} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} [b(\mathbf{B})b_0(\mathbf{B})^{-1}b(\mathbf{B}^{-1})b_0(\mathbf{B}^{-1})^{-1}M(\mathbf{v})] \\
& \qquad \sum_{\mathbf{i}^\tau \in \mathcal{Z}^d} d_\mathbf{i}\, [b(\mathbf{B})b_0(\mathbf{B})^{-1}\, b(\mathbf{B}^{-1})b_0(\mathbf{B}^{-1})^{-1}M(\mathbf{v}-\mathbf{i})] \\
= & \frac{1}{N} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} [b(\mathbf{B})b_0(\mathbf{B})^{-1}\, b(\mathbf{B}^{-1})b_0(\mathbf{B}^{-1})^{-1}M(\mathbf{v})] \\
& \qquad [d(\mathbf{B})d_0(\mathbf{B})^{-1}b(\mathbf{B})b_0(\mathbf{B})^{-1}b(\mathbf{B}^{-1})b_0(\mathbf{B}^{-1})^{-1}A(\mathbf{v})] \\
= & \frac{1}{N} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} [b(\mathbf{B})b_0(\mathbf{B})^{-1}b(\mathbf{B}^{-1})b_0(\mathbf{B}^{-1})^{-1}M(\mathbf{v})] \\
& \qquad [a(\mathbf{B})^{-1}a_0(\mathbf{B})a(\mathbf{B}^{-1})^{-1}a_0(\mathbf{B}^{-1})A(\mathbf{v})],
\end{aligned} \tag{3.5.49}$$

where the convergence in probability holds since (C1)(i) holds. Moreover, since

$$\{\varepsilon^*(\mathbf{v})\} \sim IID(0, \sigma^2),$$

it holds that (3.5.49) tends in probability to

$$E\{[b(\mathbf{B})b_0(\mathbf{B})^{-1}b(\mathbf{B}^{-1})b_0(\mathbf{B}^{-1})^{-1}M(\mathbf{v})]\ [a(\mathbf{B})^{-1}a_0(\mathbf{B})a(\mathbf{B}^{-1})^{-1}a_0(\mathbf{B}^{-1})A(\mathbf{v})]\}, \tag{3.5.50}$$

as $N \to \infty$. If we define, for any $\mathbf{b} \in \Theta_1$, the polynomial

$$b(\mathbf{z})b_0(\mathbf{z})^{-1}b(\mathbf{z}^{-1})b_0(\mathbf{z}^{-1})^{-1} \equiv \sum_{\mathbf{j}^\tau \in \mathcal{Z}^d} \beta_{\mathbf{j}}(\mathbf{b})\ \mathbf{z}^{\mathbf{j}}, \tag{3.5.51}$$

and, for any $\mathbf{a} \in \Theta_2$, the polynomial

$$a(\mathbf{z})^{-1}a_0(\mathbf{z})\ a(\mathbf{z}^{-1})^{-1}a_0(\mathbf{z}^{-1}) \equiv \sum_{\mathbf{j}^\tau \in \mathcal{Z}^d} \alpha_{\mathbf{j}}(\mathbf{a})\ \mathbf{z}^{\mathbf{j}}, \tag{3.5.52}$$

then (3.5.50) is equal to

$$\sigma^2\{\sum_{\mathbf{j}^\tau \in \mathcal{Z}^d} \beta_{\mathbf{j}}(\mathbf{b}) \cdot \alpha_{\mathbf{j}}(\mathbf{a})\}. \tag{3.5.53}$$

On the other hand, the polynomial

$$\begin{aligned} & \sigma^2[b(\mathbf{z})b_0(\mathbf{z})^{-1}b(\mathbf{z}^{-1})b_0(\mathbf{z}^{-1})^{-1}]\ [a(\mathbf{z})^{-1}a_0(\mathbf{z})\ a(\mathbf{z}^{-1})^{-1}a_0(\mathbf{z}^{-1})] \\ \equiv\ & \sigma^2 \sum_{\mathbf{j}_1^\tau, \mathbf{j}_2^\tau \in \mathcal{Z}^d} \beta_{\mathbf{j}_1}(\mathbf{b}) \cdot \alpha_{\mathbf{j}_2}(\mathbf{a})\ \mathbf{z}^{\mathbf{j}_1 - \mathbf{j}_2} \end{aligned} \tag{3.5.54}$$

generates (3.5.53) to be the coefficient of $\mathbf{z}^{\mathbf{0}}$. Thus, we may write

$$\begin{aligned} & E\{[b(\mathbf{B})b_0(\mathbf{B})^{-1}b(\mathbf{B}^{-1})b_0(\mathbf{B}^{-1})^{-1}M(\mathbf{v})]\ [a(\mathbf{B})^{-1}a_0(\mathbf{B})a(\mathbf{B}^{-1})^{-1}a_0(\mathbf{B}^{-1})A(\mathbf{v})]\} \\ =\ & E\{\varepsilon^*(\mathbf{v})\ [b(\mathbf{z})b_0(\mathbf{z})^{-1}b(\mathbf{z}^{-1})b_0(\mathbf{z}^{-1})^{-1}\ a(\mathbf{z})^{-1}a_0(\mathbf{z})a(\mathbf{z}^{-1})^{-1}a_0(\mathbf{z}^{-1})\ \varepsilon^*(\mathbf{v})]\} \\ =\ & E\{\varepsilon^*(\mathbf{v})\ [\{b(\mathbf{z})a(\mathbf{z})^{-1}\}\{b_0(\mathbf{z})a_0(\mathbf{z})^{-1}\}^{-1}\ \{b(\mathbf{z}^{-1})a(\mathbf{z}^{-1})^{-1}\} \cdot \\ & \{b_0(\mathbf{z}^{-1})a_0(\mathbf{z}^{-1})^{-1}\}^{-1}\varepsilon^*(\mathbf{v})]\} \geq E\{\varepsilon^*(\mathbf{v})^2\} = \sigma^2, \end{aligned} \tag{3.5.55}$$

and the equality holds if and only if $\mathbf{b} \equiv \mathbf{b}_0$ and $\mathbf{a} \equiv \mathbf{a}_0$. At this final stage and for the inequality to hold, the assumptions of causality and invertibility are essential, so that all the polynomials $b(\mathbf{z})^{-1}, b_0(\mathbf{z})^{-1}$ and $a(\mathbf{z})^{-1}, a_0(\mathbf{z})^{-1}$, can generate unity to be the coefficient of $\mathbf{z}^{\mathbf{0}}$ and they can extend over one side only $\mathbf{z}^{\mathbf{i}}, \mathbf{i} \geq \mathbf{0}$ or $\mathbf{z}^{\mathbf{i}}, \mathbf{i} \leq \mathbf{0}$. The rest of the proof is identical to the one for Theorem 3.3. ■

To prove the asymptotic normality, we will define the following variables. First, we might write for any $n = 1, \cdots, p$, and any $\mathbf{v}^\tau \in \mathcal{Z}^d$,

$$-\frac{\partial}{\partial b_{\mathbf{i}_n}} M(\mathbf{v}, \mathbf{b}) = Y_1(\mathbf{v} + \mathbf{i}_n, \mathbf{b}) + Y_2(\mathbf{v} - \mathbf{i}_n, \mathbf{b}). \tag{3.5.56}$$

Depending on the set $\mathcal{S}$ and, consequently, on the sampling set $\mathcal{I}$, we define the random variables

$$H_M(\mathbf{v}, \mathbf{b}) \equiv \begin{cases} M(\mathbf{v}, \mathbf{b}), & \mathbf{v}^\tau \in \mathcal{S} \\ 0, & \mathbf{v}^\tau \notin \mathcal{S} \end{cases}, \tag{3.5.57}$$

and

$$H_M(\mathbf{v}, \mathbf{b}_0) \equiv H_M(\mathbf{v}), \ \mathbf{v}^\tau \in \mathcal{Z}^d. \tag{3.5.58}$$

Equation (3.5.57) also implies that for any $n = 1, \cdots, p$,

$$-\frac{\partial}{\partial b_{\mathbf{i}_n}} H_M(\mathbf{v}, \mathbf{b}) = \begin{cases} -\frac{\partial}{\partial b_{\mathbf{i}_n}} M(\mathbf{v}, \mathbf{b}), & \mathbf{v}^\tau \in \mathcal{S} \\ 0, & \mathbf{v}^\tau \notin \mathcal{S} \end{cases}. \tag{3.5.59}$$

As a result, we may define for any $n = 1, \cdots, p$, the variables

$$H_{Y_1}^{(n)}(\mathbf{v}, \mathbf{b}) \equiv \begin{cases} Y_1(\mathbf{v}, \mathbf{b}), & \mathbf{v}^\tau - \mathbf{i}_n^\tau \in \mathcal{S} \\ 0, & \mathbf{v}^\tau - \mathbf{i}_n^\tau \notin \mathcal{S} \end{cases} \tag{3.5.60}$$

and

$$H_{Y_2}^{(n)}(\mathbf{v}, \mathbf{b}) \equiv \begin{cases} Y_2(\mathbf{v}, \mathbf{b}), & \mathbf{v}^\tau + \mathbf{i}_n^\tau \in \mathcal{S} \\ 0, & \mathbf{v}^\tau + \mathbf{i}_n^\tau \notin \mathcal{S} \end{cases}. \tag{3.5.61}$$

If we combine (3.5.56), (3.5.59), (3.5.60) and (3.5.61), we may then write that

$$-\frac{\partial}{\partial b_{\mathbf{i}_n}} H_M(\mathbf{v}, \mathbf{b}) = H_{Y_1}^{(n)}(\mathbf{v} + \mathbf{i}_n, \mathbf{b}) + H_{Y_2}^{(n)}(\mathbf{v} - \mathbf{i}_n, \mathbf{b}), \ n = 1, \cdots, p, \ \mathbf{v}^\tau \in \mathcal{Z}^d. \tag{3.5.62}$$

We finally consider that

$$H_{Y_1}^{(n)}(\mathbf{v}, \mathbf{b}_0) \equiv H_{Y_1}^{(n)}(\mathbf{v}), \tag{3.5.63}$$

$$H_{Y_2}^{(n)}(\mathbf{v}, \mathbf{b}_0) \equiv H_{Y_2}^{(n)}(\mathbf{v}), \tag{3.5.64}$$

for any $\mathbf{v}^\tau \in \mathcal{Z}^d$ and $n = 1, \cdots, p$.

We may now re-write (3.5.29), as

$$Q^*(\boldsymbol{\theta}) \equiv \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v}, \mathbf{b}) \ [d(\mathbf{B}) \ H_M(\mathbf{v}, \mathbf{b})], \ \boldsymbol{\theta} \in \Theta. \tag{3.5.65}$$

We write the derivatives of (3.5.65) with respect to the auto-regressive parameters $n = 1, \cdots, p$,

$$\begin{aligned} D_n(\boldsymbol{\theta}) \equiv -\frac{\partial}{\partial b_{\mathbf{i}_n}} Q^*(\boldsymbol{\theta}) &= \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_1(\mathbf{v}+\mathbf{i}_n, \mathbf{b}) \; [d(\mathbf{B}) \; H_M(\mathbf{v}, \mathbf{b})] \\ &+ \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_2(\mathbf{v}-\mathbf{i}_n, \mathbf{b}) \; [d(\mathbf{B}) \; H_M(\mathbf{v}, \mathbf{b})] \\ &- \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v}, \mathbf{b}) \; [b(\mathbf{B})^{-1} d(\mathbf{B}) \; H_M(\mathbf{v}-\mathbf{i}_n, \mathbf{b})] \\ &- \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v}, \mathbf{b}) \; [b(\mathbf{B}^{-1})^{-1} d(\mathbf{B}) \; H_M(\mathbf{v}+\mathbf{i}_n, \mathbf{b})] \\ &+ \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v}, \mathbf{b}) \; [d(\mathbf{B}) \; H_{Y_1}^{(n)}(\mathbf{v}+\mathbf{i}_n, \mathbf{b})] \\ &+ \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v}, \mathbf{b}) \; [d(\mathbf{B}) \; H_{Y_2}^{(n)}(\mathbf{v}-\mathbf{i}_n, \mathbf{b})], \end{aligned} \tag{3.5.66}$$

and the derivatives with respect to the moving-average parameters $m = 1, \cdots, q$,

$$\begin{aligned} D_{p+m}(\boldsymbol{\theta}) \equiv -\frac{\partial}{\partial a_{\mathbf{j}_m}} Q^*(\boldsymbol{\theta}) &= \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v}, \mathbf{b}) \; [a(\mathbf{B})^{-1} d(\mathbf{B}) \; H_M(\mathbf{v}-\mathbf{j}_m, \mathbf{b})] \\ &+ \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v}, \mathbf{b}) \; [a(\mathbf{B}^{-1})^{-1} d(\mathbf{B}) \; H_M(\mathbf{v}+\mathbf{j}_m, \mathbf{b})]. \end{aligned} \tag{3.5.67}$$

For any $n = 1, \cdots, p$, we re-write (3.5.66) as

$$D_n(\boldsymbol{\theta}) = D_n(\boldsymbol{\theta}_0) - \mathbf{J}_n^\tau(\boldsymbol{\theta})[\boldsymbol{\theta} - \boldsymbol{\theta}_0], \tag{3.5.68}$$

where

$$\mathbf{J}_n^\tau(\boldsymbol{\theta}) \equiv [J_{n,1}(\boldsymbol{\theta}), \cdots, J_{n,p+q}(\boldsymbol{\theta})] \tag{3.5.69}$$

with

$$\begin{aligned} J_{n,r}(\boldsymbol{\theta}) &\equiv \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Z(\mathbf{v}+\mathbf{i}_n-\mathbf{i}_r) \; [d_0(\mathbf{B}) H_M(\mathbf{v})] \\ &- \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_1(\mathbf{v}+\mathbf{i}_n) \; [b_0(\mathbf{B})^{-1} d_0(\mathbf{B}) \; H_M(\mathbf{v}-\mathbf{i}_r)] \\ &- \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_1(\mathbf{v}+\mathbf{i}_n) \; [b_0(\mathbf{B}^{-1})^{-1} d_0(\mathbf{B}) \; H_M(\mathbf{v}+\mathbf{i}_r)] \\ &+ \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_1(\mathbf{v}+\mathbf{i}_n) \; [d_0(\mathbf{B}) \; H_{Y_1}^{(r)}(\mathbf{v}+\mathbf{i}_r)] \\ &+ \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_1(\mathbf{v}+\mathbf{i}_n) \; [d_0(\mathbf{B}) \; H_{Y_2}^{(r)}(\mathbf{v}-\mathbf{i}_r)] \\ &+ \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Z(\mathbf{v}-\mathbf{i}_n+\mathbf{i}_r) \; [d_0(\mathbf{B}) H_M(\mathbf{v})] \end{aligned}$$

$$
\begin{aligned}
&- \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_2(\mathbf{v}-\mathbf{i}_n)\ [b_0(\mathbf{B})^{-1} d_0(\mathbf{B})\ H_M(\mathbf{v}-\mathbf{i}_r)] \\
&- \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_2(\mathbf{v}-\mathbf{i}_n)\ [b_0(\mathbf{B}^{-1})^{-1} d_0(\mathbf{B})\ H_M(\mathbf{v}+\mathbf{i}_r)] \\
&+ \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_2(\mathbf{v}-\mathbf{i}_n)\ [d_0(\mathbf{B})\ H^{(r)}_{Y_1}(\mathbf{v}+\mathbf{i}_r)] \\
&+ \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_2(\mathbf{v}-\mathbf{i}_n)\ [d_0(\mathbf{B})\ H^{(r)}_{Y_2}(\mathbf{v}-\mathbf{i}_r)] \\
&- \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_1(\mathbf{v}+\mathbf{i}_r)\ [b_0(\mathbf{B})^{-1} d_0(\mathbf{B})\ H_M(\mathbf{v}-\mathbf{i}_n)] \\
&- \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_2(\mathbf{v}-\mathbf{i}_r)\ [b_0(\mathbf{B})^{-1} d_0(\mathbf{B})\ H_M(\mathbf{v}-\mathbf{i}_n)] \\
&+ 2 \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v})\ [b_0(\mathbf{B})^{-2} d_0(\mathbf{B})\ H_M(\mathbf{v}-\mathbf{i}_n-\mathbf{i}_r)] \\
&+ \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v})\ [b_0(\mathbf{B})^{-1} b_0(\mathbf{B}^{-1})^{-1} d_0(\mathbf{B})\ H_M(\mathbf{v}-\mathbf{i}_n+\mathbf{i}_r)] \\
&- \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v})\ [b_0(\mathbf{B})^{-1} d_0(\mathbf{B})\ H^{(r)}_{Y_1}(\mathbf{v}-\mathbf{i}_n+\mathbf{i}_r)] \\
&- \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v})\ [b_0(\mathbf{B})^{-1} d_0(\mathbf{B})\ H^{(r)}_{Y_2}(\mathbf{v}-\mathbf{i}_n-\mathbf{i}_r)] \\
&- \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_1(\mathbf{v}+\mathbf{i}_r)\ [b_0(\mathbf{B}^{-1})^{-1} d_0(\mathbf{B})\ H_M(\mathbf{v}+\mathbf{i}_n)] \\
&- \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_2(\mathbf{v}-\mathbf{i}_r)\ [b_0(\mathbf{B}^{-1})^{-1} d_0(\mathbf{B})\ H_M(\mathbf{v}+\mathbf{i}_n)] \\
&+ \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v})\ [b_0(\mathbf{B})^{-1} b_0(\mathbf{B}^{-1})^{-1} d_0(\mathbf{B})\ H_M(\mathbf{v}+\mathbf{i}_n-\mathbf{i}_r)] \\
&+ 2 \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v})\ [b_0(\mathbf{B}^{-1})^{-2} d_0(\mathbf{B})\ H_M(\mathbf{v}+\mathbf{i}_n+\mathbf{i}_r) \\
&- \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v})\ [b_0(\mathbf{B}^{-1})^{-1} d_0(\mathbf{B})\ H^{(r)}_{Y_1}(\mathbf{v}+\mathbf{i}_n+\mathbf{i}_r)] \\
&- \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v})\ [b_0(\mathbf{B}^{-1})^{-1} d_0(\mathbf{B})\ H^{(r)}_{Y_2}(\mathbf{v}+\mathbf{i}_n-\mathbf{i}_r)] \\
&+ \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_1(\mathbf{v}+\mathbf{i}_r)\ [d_0(\mathbf{B})\ H^{(n)}_{Y_1}(\mathbf{v}+\mathbf{i}_n)] \\
&+ \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_2(\mathbf{v}-\mathbf{i}_r)\ [d_0(\mathbf{B})\ H^{(n)}_{Y_1}(\mathbf{v}+\mathbf{i}_n)] \\
&- \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v})\ [b_0(\mathbf{B})^{-1} d_0(\mathbf{B})\ H^{(n)}_{Y_1}(\mathbf{v}+\mathbf{i}_n-\mathbf{i}_r)] \\
&- \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v})\ [b_0(\mathbf{B}^{-1})^{-1} d_0(\mathbf{B})\ H^{(n)}_{Y_1}(\mathbf{v}+\mathbf{i}_n+\mathbf{i}_r)]
\end{aligned}
$$

$$
\begin{aligned}
&+ \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v}) \sum_{\mathbf{j}^\tau \in \mathcal{F}_\mathbf{v}} d_{\mathbf{j},0}\, Z(\mathbf{v}+\mathbf{i}_n-\mathbf{i}_r-\mathbf{j}) \\
&+ \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v}) \sum_{\mathbf{j}^\tau \in \mathcal{F}_\mathbf{v}} d_{\mathbf{j},0}\, Z(\mathbf{v}-\mathbf{i}_n+\mathbf{i}_r-\mathbf{j}) \\
&+ \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_1(\mathbf{v}+\mathbf{i}_r)\, [d_0(\mathbf{B})\, H^{(n)}_{Y_2}(\mathbf{v}-\mathbf{i}_n)] \\
&+ \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_2(\mathbf{v}-\mathbf{i}_r)\, [d_0(\mathbf{B})\, H^{(n)}_{Y_2}(\mathbf{v}-\mathbf{i}_n)] \\
&- \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v})\, [b_0(\mathbf{B})^{-1} d_0(\mathbf{B})\, H^{(n)}_{Y_2}(\mathbf{v}-\mathbf{i}_n-\mathbf{i}_r)] \\
&- \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v})\, [b_0(\mathbf{B}^{-1})^{-1} d_0(\mathbf{B})\, H^{(n)}_{Y_2}(\mathbf{v}-\mathbf{i}_n+\mathbf{i}_r)] + O_P(N||\boldsymbol{\theta}-\boldsymbol{\theta}_0||),
\end{aligned}
\tag{3.5.70}
$$

for $r = 1, \cdots, p$, and

$$
\begin{aligned}
J_{n,r}(\boldsymbol{\theta}) \equiv J_{n,p+l}(\boldsymbol{\theta}) \;\equiv\; & \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_1(\mathbf{v}+\mathbf{i}_n)\, [a_0(\mathbf{B})^{-1} d_0(\mathbf{B})\, H_M(\mathbf{v}-\mathbf{j}_l)] \\
&+ \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_1(\mathbf{v}+\mathbf{i}_n)\, [a_0(\mathbf{B}^{-1})^{-1} d_0(\mathbf{B})\, H_M(\mathbf{v}+\mathbf{j}_l)] \\
&+ \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_2(\mathbf{v}-\mathbf{i}_n)\, [a_0(\mathbf{B})^{-1} d_0(\mathbf{B})\, H_M(\mathbf{v}-\mathbf{j}_l)] \\
&+ \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_2(\mathbf{v}-\mathbf{i}_n)\, [a_0(\mathbf{B}^{-1})^{-1} d_0(\mathbf{B})\, H_M(\mathbf{v}+\mathbf{j}_l)] \\
&- \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v})\, [b_0(\mathbf{B})^{-1} a_0(\mathbf{B})^{-1} d_0(\mathbf{B})\, H_M(\mathbf{v}-\mathbf{i}_n-\mathbf{j}_l)] \\
&- \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v})\, [b_0(\mathbf{B})^{-1} a_0(\mathbf{B}^{-1})^{-1} d_0(\mathbf{B})\, H_M(\mathbf{v}-\mathbf{i}_n+\mathbf{j}_l)] \\
&- \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v})\, [a_0(\mathbf{B})^{-1} b_0(\mathbf{B}^{-1})^{-1} d_0(\mathbf{B})\, H_M(\mathbf{v}+\mathbf{i}_n-\mathbf{j}_l)] \\
&- \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v})\, [a_0(\mathbf{B}^{-1})^{-1} b_0(\mathbf{B}^{-1})^{-1} d_0(\mathbf{B})\, H_M(\mathbf{v}+\mathbf{i}_n+\mathbf{j}_l)] \\
&+ \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v})\, [a_0(\mathbf{B})^{-1} d_0(\mathbf{B})\, H^{(n)}_{Y_1}(\mathbf{v}+\mathbf{i}_n-\mathbf{j}_l)] \\
&+ \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v})\, [a_0(\mathbf{B}^{-1})^{-1} d_0(\mathbf{B})\, H^{(n)}_{Y_1}(\mathbf{v}+\mathbf{i}_n+\mathbf{j}_l)] \\
&+ \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v})\, [a_0(\mathbf{B})^{-1} d_0(\mathbf{B})\, H^{(n)}_{Y_2}(\mathbf{v}-\mathbf{i}_n-\mathbf{j}_l)] \\
&+ \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v})\, [a_0(\mathbf{B}^{-1})^{-1} d_0(\mathbf{B})\, H^{(n)}_{Y_2}(\mathbf{v}-\mathbf{i}_n+\mathbf{j}_l)] \\
&+ O_P(N||\boldsymbol{\theta}-\boldsymbol{\theta}_0||),
\end{aligned}
\tag{3.5.71}
$$

for $l = r - p = 1, \cdots, q$.

On the other hand, for any $m = 1, \cdots, q$, we re-write (3.5.67) as

$$D_{p+m}(\boldsymbol{\theta}) = D_{p+m}(\boldsymbol{\theta}_0) - \mathbf{J}^{\tau}_{p+m}(\boldsymbol{\theta})[\boldsymbol{\theta} - \boldsymbol{\theta}_0], \tag{3.5.72}$$

where

$$\mathbf{J}^{\tau}_{p+m}(\boldsymbol{\theta}) = [J_{p+m,1}(\boldsymbol{\theta}), \cdots, J_{p+m,p+q}(\boldsymbol{\theta})] \tag{3.5.73}$$

with

$$\begin{aligned}
J_{p+m,r}(\boldsymbol{\theta}) \equiv & \sum_{\mathbf{v}^{\tau} \in \mathcal{S}^*} Y_1(\mathbf{v} + \mathbf{i}_r)\ [a_0(\mathbf{B})^{-1} d_0(\mathbf{B})\ H_M(\mathbf{v} - \mathbf{j}_m)] \\
& + \sum_{\mathbf{v}^{\tau} \in \mathcal{S}^*} Y_2(\mathbf{v} - \mathbf{i}_r)\ [a_0(\mathbf{B})^{-1} d_0(\mathbf{B})\ H_M(\mathbf{v} - \mathbf{j}_m)] \\
& - \sum_{\mathbf{v}^{\tau} \in \mathcal{S}^*} M(\mathbf{v})\ [a_0(\mathbf{B})^{-1} b_0(\mathbf{B})^{-1} d_0(\mathbf{B})\ H_M(\mathbf{v} - \mathbf{j}_m - \mathbf{i}_r)] \\
& - \sum_{\mathbf{v}^{\tau} \in \mathcal{S}^*} M(\mathbf{v})\ [a_0(\mathbf{B})^{-1} b_0(\mathbf{B}^{-1})^{-1} d_0(\mathbf{B})\ H_M(\mathbf{v} - \mathbf{j}_m + \mathbf{i}_r)] \\
& + \sum_{\mathbf{v}^{\tau} \in \mathcal{S}^*} M(\mathbf{v})\ [a_0(\mathbf{B})^{-1} d_0(\mathbf{B})\ H^{(r)}_{Y_1}(\mathbf{v} - \mathbf{j}_m + \mathbf{i}_r)] \\
& + \sum_{\mathbf{v}^{\tau} \in \mathcal{S}^*} M(\mathbf{v})\ [a_0(\mathbf{B})^{-1} d_0(\mathbf{B})\ H^{(r)}_{Y_2}(\mathbf{v} - \mathbf{j}_m - \mathbf{i}_r)] \\
& + \sum_{\mathbf{v}^{\tau} \in \mathcal{S}^*} Y_1(\mathbf{v} + \mathbf{i}_r)\ [a_0(\mathbf{B}^{-1})^{-1} d_0(\mathbf{B})\ H_M(\mathbf{v} + \mathbf{j}_m)] \\
& + \sum_{\mathbf{v}^{\tau} \in \mathcal{S}^*} Y_2(\mathbf{v} - \mathbf{i}_r)\ [a_0(\mathbf{B}^{-1})^{-1} d_0(\mathbf{B})\ H_M(\mathbf{v} + \mathbf{j}_m)] \\
& - \sum_{\mathbf{v}^{\tau} \in \mathcal{S}^*} M(\mathbf{v})\ [a_0(\mathbf{B}^{-1})^{-1} b_0(\mathbf{B})^{-1} d_0(\mathbf{B})\ H_M(\mathbf{v} + \mathbf{j}_m - \mathbf{i}_r)] \\
& - \sum_{\mathbf{v}^{\tau} \in \mathcal{S}^*} M(\mathbf{v})\ [a_0(\mathbf{B}^{-1})^{-1} b_0(\mathbf{B}^{-1})^{-1} d_0(\mathbf{B})\ H_M(\mathbf{v} + \mathbf{j}_m + \mathbf{i}_r)] \\
& + \sum_{\mathbf{v}^{\tau} \in \mathcal{S}^*} M(\mathbf{v})\ [a_0(\mathbf{B}^{-1})^{-1} d_0(\mathbf{B})\ H^{(r)}_{Y_1}(\mathbf{v} + \mathbf{j}_m + \mathbf{i}_r)] \\
& + \sum_{\mathbf{v}^{\tau} \in \mathcal{S}^*} M(\mathbf{v})\ [a_0(\mathbf{B}^{-1})^{-1} d_0(\mathbf{B})\ H^{(r)}_{Y_2}(\mathbf{v} + \mathbf{j}_m - \mathbf{i}_r)] \\
& + O_P(N||\boldsymbol{\theta} - \boldsymbol{\theta}_0||),
\end{aligned} \tag{3.5.74}$$

for $r = 1, \cdots, p$, and

$$\begin{aligned}
J_{p+m,r}(\boldsymbol{\theta}) \equiv J_{p+m,p+l}(\boldsymbol{\theta}) \equiv & \ 2 \sum_{\mathbf{v}^{\tau} \in \mathcal{S}^*} M(\mathbf{v})\ [a_0(\mathbf{B})^{-2} d_0(\mathbf{B}) H_M(\mathbf{v} - \mathbf{j}_m - \mathbf{j}_l)] \\
& + \sum_{\mathbf{v}^{\tau} \in \mathcal{S}^*} M(\mathbf{v}) [a_0(\mathbf{B})^{-1} a_0(\mathbf{B}^{-1})^{-1} d_0(\mathbf{B})\ H_M(\mathbf{v} - \mathbf{j}_m + \mathbf{j}_l)] \\
& + \sum_{\mathbf{v}^{\tau} \in \mathcal{S}^*} M(\mathbf{v}) [a_0(\mathbf{B}^{-1})^{-1} a_0(\mathbf{B})^{-1} d_0(\mathbf{B})\ H_M(\mathbf{v} + \mathbf{j}_m - \mathbf{j}_l)] \\
& + 2 \sum_{\mathbf{v}^{\tau} \in \mathcal{S}^*} M(\mathbf{v})\ [a_0(\mathbf{B}^{-1})^{-2} d_0(\mathbf{B})\ H_M(\mathbf{v} + \mathbf{j}_m + \mathbf{j}_l)] \\
& + O_P(N||\boldsymbol{\theta} - \boldsymbol{\theta}_0||),
\end{aligned} \tag{3.5.75}$$

for $l = r - p = 1, \cdots, q$.

Finally, for any $\boldsymbol{\theta} \in \Theta$, we define the matrix

$$\mathbf{J}(\boldsymbol{\theta}) \equiv \begin{bmatrix} \mathbf{J}_1^\tau(\boldsymbol{\theta}) \\ \vdots \\ \mathbf{J}_{p+q}^\tau(\boldsymbol{\theta}) \end{bmatrix} \tag{3.5.76}$$

and

$$\mathbf{J}(\widehat{\boldsymbol{\theta}}) \equiv \mathbf{J}. \tag{3.5.77}$$

We also define the random vector

$$\mathbf{D} \equiv [D_1(\boldsymbol{\theta}_0), \cdots, D_{p+q}(\boldsymbol{\theta}_0)]^\tau. \tag{3.5.78}$$

From (3.5.68) and (3.5.72), we can conclude that

$$\mathbf{J}[\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0] = \mathbf{D} \tag{3.5.79}$$

and

$$N^{1/2}[\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0] = \{\mathbf{J}/N\}^{-1}\{N^{-1/2}\mathbf{D}\}. \tag{3.5.80}$$

Before we move to the next proposition and theorem, for a process $\{W(\mathbf{v})\} \sim WN(0,1)$, we define the auto-regression $\{\xi(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ by the equation

$$b_0(\mathbf{B})\xi(\mathbf{v}) \equiv W(\mathbf{v}) \tag{3.5.81}$$

and the auto-regression $\{\eta(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ by the equation

$$a_0(\mathbf{B})\eta(\mathbf{v}) \equiv W(\mathbf{v}). \tag{3.5.82}$$

We let the random vector

$$\boldsymbol{\xi} \equiv [\xi(-\mathbf{i}_1), \cdots, \xi(-\mathbf{i}_p), \eta(-\mathbf{j}_1), \cdots, \eta(-\mathbf{j}_q)]^\tau \tag{3.5.83}$$

and the variance matrix

$$\mathbf{W}_{p+q} \equiv \mathrm{Var}\{\boldsymbol{\xi}\}. \tag{3.5.84}$$

Proposition 3.4. If $\{\varepsilon^*(\mathbf{v})\} \sim IID(0, \sigma^2)$, then under conditions (C1)(i) and (C4), it holds that

$$\mathbf{J}/N \xrightarrow{P} 2 \cdot \sigma^2 \mathbf{W}_{p+q} \tag{3.5.85}$$

as $N \to \infty$.

Proof. Since

$$\{\varepsilon^*(\mathbf{v})\} \sim IID(0, \sigma^2)$$

and (C1)(i) holds, we may use the same arguments as in Proposition 3.3. In (3.5.70), we can find 14 zero terms, *i.e.*

$$\begin{aligned}
&\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_1(\mathbf{v}+\mathbf{i}_n)[b_0(\mathbf{B})^{-1}d_0(\mathbf{B})H_M(\mathbf{v}-\mathbf{i}_r)]/N \\
\xrightarrow{P}\ & E\{[b_0(\mathbf{B}^{-1})^{-1}M(\mathbf{v}+\mathbf{i}_n)][b_0(\mathbf{B})^{-1}A(\mathbf{v}-\mathbf{i}_r)]\} = 0 \\
&\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_1(\mathbf{v}+\mathbf{i}_n)[d_0(\mathbf{B})H^{(r)}_{Y_2}(\mathbf{v}-\mathbf{i}_r)]/N \\
\xrightarrow{P}\ & E\{[b_0(\mathbf{B}^{-1})^{-1}M(\mathbf{v}+\mathbf{i}_n)][b_0(\mathbf{B})^{-1}A(\mathbf{v}-\mathbf{i}_r)]\} = 0 \\
&\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_2(\mathbf{v}-\mathbf{i}_n)[b_0(\mathbf{B}^{-1})^{-1}d_0(\mathbf{B})H_M(\mathbf{v}+\mathbf{i}_r)]/N \\
\xrightarrow{P}\ & E\{[b_0(\mathbf{B})^{-1}M(\mathbf{v}-\mathbf{i}_n)][b_0(\mathbf{B}^{-1})^{-1}A(\mathbf{v}+\mathbf{i}_r)]\} = 0 \\
&\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_2(\mathbf{v}-\mathbf{i}_n)[d_0(\mathbf{B})H^{(r)}_{Y_1}(\mathbf{v}+\mathbf{i}_r)]/N \\
\xrightarrow{P}\ & E\{[b_0(\mathbf{B})^{-1}M(\mathbf{v}-\mathbf{i}_n)][b_0(\mathbf{B}^{-1})^{-1}A(\mathbf{v}+\mathbf{i}_n)]\} = 0 \\
&\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_1(\mathbf{v}+\mathbf{i}_r)[b_0(\mathbf{B})^{-1}d_0(\mathbf{B})H_M(\mathbf{v}-\mathbf{i}_n)]/N \\
\xrightarrow{P}\ & E\{[b_0(\mathbf{B}^{-1})^{-1}M(\mathbf{v}+\mathbf{i}_r)][b_0(\mathbf{B})^{-1}A(\mathbf{v}-\mathbf{i}_n)]\} = 0 \\
&\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v})[b_0(\mathbf{B})^{-2}d_0(\mathbf{B})H_M(\mathbf{v}-\mathbf{i}_n-\mathbf{i}_r)]/N \\
\xrightarrow{P}\ & E\{M(\mathbf{v})[b_0(\mathbf{B})^{-2}A(\mathbf{v}-\mathbf{i}_n-\mathbf{i}_r)]\} = 0 \\
&\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v})[b_0(\mathbf{B})^{-1}d_0(\mathbf{B})H^{(r)}_{Y_2}(\mathbf{v}-\mathbf{i}_n-\mathbf{i}_r)]/N \\
\xrightarrow{P}\ & E\{M(\mathbf{v})[b_0(\mathbf{B})^{-2}A(\mathbf{v}-\mathbf{i}_n-\mathbf{i}_r)]\} = 0 \qquad (3.5.86)
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_2(\mathbf{v}-\mathbf{i}_r)[b_0(\mathbf{B}^{-1})^{-1} d_0(\mathbf{B}) H_M(\mathbf{v}+\mathbf{i}_n)]/N \\
\xrightarrow{P}\ & E\{[b_0(\mathbf{B})^{-1} M(\mathbf{v}-\mathbf{i}_r)][b_0(\mathbf{B}^{-1})^{-1} A(\mathbf{v}+\mathbf{i}_n)]\} = 0 \\
&\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v})[b_0(\mathbf{B}^{-1})^{-2} d_0(\mathbf{B}) H_M(\mathbf{v}+\mathbf{i}_n+\mathbf{i}_r)]/N \\
\xrightarrow{P}\ & E\{M(\mathbf{v})[b_0(\mathbf{B}^{-1})^{-2} A(\mathbf{v}+\mathbf{i}_n+\mathbf{i}_r)]\} = 0 \\
&\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v})[b_0(\mathbf{B}^{-1})^{-1} d_0(\mathbf{B}) H_{Y_1}^{(r)}(\mathbf{v}+\mathbf{i}_n+\mathbf{i}_r)]/N \\
\xrightarrow{P}\ & E\{M(\mathbf{v})[b_0(\mathbf{B}^{-1})^{-2} A(\mathbf{v}+\mathbf{i}_n+\mathbf{i}_r)]\} = 0 \\
&\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_2(\mathbf{v}-\mathbf{i}_r)[d_0(\mathbf{B}) H_{Y_1}^{(n)}(\mathbf{v}+\mathbf{i}_n)]/N \\
\xrightarrow{P}\ & E\{[b_0(\mathbf{B})^{-2} M(\mathbf{v}-\mathbf{i}_r)][b_0(\mathbf{B}^{-1})^{-1} A(\mathbf{v}+\mathbf{i}_n)]\} = 0 \\
&\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v})[b_0(\mathbf{B}^{-1})^{-1} d_0(\mathbf{B}) H_{Y_1}^{(n)}(\mathbf{v}+\mathbf{i}_n+\mathbf{i}_r)]/N \\
\xrightarrow{P}\ & E\{M(\mathbf{v})[b_0(\mathbf{B}^{-1})^{-2} A(\mathbf{v}+\mathbf{i}_n+\mathbf{i}_r)]\} = 0 \\
&\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_1(\mathbf{v}+\mathbf{i}_r)[d_0(\mathbf{B}) H_{Y_2}^{(n)}(\mathbf{v}-\mathbf{i}_n)]/N \\
\xrightarrow{P}\ & E\{[b_0(\mathbf{B}^{-1})^{-1} M(\mathbf{v}+\mathbf{i}_r)][b_0(\mathbf{B})^{-1} A(\mathbf{v}-\mathbf{i}_n)]\} = 0 \\
&\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v})[b_0(\mathbf{B})^{-1} d_0(\mathbf{B}) H_{Y_2}(\mathbf{v}-\mathbf{i}_n-\mathbf{i}_r)]/N \\
\xrightarrow{P}\ & E\{M(\mathbf{v})[b_0(\mathbf{B})^{-2} A(\mathbf{v}-\mathbf{i}_n-\mathbf{i}_r)]\} = 0. \qquad (3.5.87)
\end{aligned}$$

In (3.5.70), there can be also obtained 10 positive and equal terms, *i.e.*

$$
\begin{aligned}
&\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Z(\mathbf{v}+\mathbf{i}_n-\mathbf{i}_r)[d_0(\mathbf{B})H_M(\mathbf{v})]/N \\
\xrightarrow{P}\ & E\{[b_0(\mathbf{B})^{-1}b_0(\mathbf{B}^{-1})^{-1}M(\mathbf{v}+\mathbf{i}_n-\mathbf{i}_r)]A(\mathbf{v})\} \\
&\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_1(\mathbf{v}+\mathbf{i}_n)[d_0(\mathbf{B})H^{(r)}_{Y_1}(\mathbf{v}+\mathbf{i}_r)]/N \\
\xrightarrow{P}\ & E\{[b_0(\mathbf{B}^{-1})^{-1}M(\mathbf{v}+\mathbf{i}_n)][b_0(\mathbf{B}^{-1})^{-1}A(\mathbf{v}+\mathbf{i}_r)]\} \\
&\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Z(\mathbf{v}-\mathbf{i}_n+\mathbf{i}_r)[d_0(\mathbf{B})H_M(\mathbf{v})]/N \\
\xrightarrow{P}\ & E\{[b_0(\mathbf{B})^{-1}b_0(\mathbf{B}^{-1})^{-1}M(\mathbf{v}-\mathbf{i}_n+\mathbf{i}_r)]A(\mathbf{v})\} \\
&\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_2(\mathbf{v}-\mathbf{i}_n)[d_0(\mathbf{B})H^{(r)}_{Y_2}(\mathbf{v}-\mathbf{i}_r)]/N \\
\xrightarrow{P}\ & E\{[b_0(\mathbf{B})^{-1}M(\mathbf{v}-\mathbf{i}_n)][b_0(\mathbf{B})^{-1}A(\mathbf{v}-\mathbf{i}_r)]\} \\
&\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v})[b_0(\mathbf{B})^{-1}b_0(\mathbf{B}^{-1})^{-1}d_0(\mathbf{B})H_M(\mathbf{v}-\mathbf{i}_n+\mathbf{i}_r)]/N \\
\xrightarrow{P}\ & E\{M(\mathbf{v})[b_0(\mathbf{B})^{-1}b_0(\mathbf{B}^{-1})^{-1}A(\mathbf{v}-\mathbf{i}_n+\mathbf{i}_r)]\}
\end{aligned}
\tag{3.5.88}
$$

and

$$
\begin{aligned}
&\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v})[b_0(\mathbf{B})^{-1}b_0(\mathbf{B}^{-1})^{-1}d_0(\mathbf{B})H_M(\mathbf{v}+\mathbf{i}_n-\mathbf{i}_r)]/N \\
\xrightarrow{P}\ & E\{M(\mathbf{v})[b_0(\mathbf{B})^{-1}b_0(\mathbf{B}^{-1})^{-1}A(\mathbf{v}+\mathbf{i}_n-\mathbf{i}_r)]\} \\
&\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_1(\mathbf{v}+\mathbf{i}_r)[d_0(\mathbf{B})H^{(n)}_{Y_1}(\mathbf{v}+\mathbf{i}_n)]/N \\
\xrightarrow{P}\ & E\{[b_0(\mathbf{B}^{-1})^{-1}M(\mathbf{v}+\mathbf{i}_r)][b_0(\mathbf{B}^{-1})^{-1}A(\mathbf{v}+\mathbf{i}_n)]\} \\
&\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v}) \sum_{\mathbf{j}^\tau \in \mathcal{F}_\mathbf{v}} d_{\mathbf{j},0}Z(\mathbf{v}+\mathbf{i}_n-\mathbf{i}_r-\mathbf{j})/N \\
\xrightarrow{P}\ & E\{M(\mathbf{v})[b_0(\mathbf{B})^{-1}b_0(\mathbf{B}^{-1})^{-1}A(\mathbf{v}+\mathbf{i}_n-\mathbf{i}_r)]\} \\
&\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v}) \sum_{\mathbf{j}^\tau \in \mathcal{F}_\mathbf{v}} d_{\mathbf{j},0}Z(\mathbf{v}-\mathbf{i}_n+\mathbf{i}_r-\mathbf{j})/N \\
\xrightarrow{P}\ & E\{M(\mathbf{v})[b_0(\mathbf{B})^{-1}b_0(\mathbf{B}^{-1})^{-1}A(\mathbf{v}-\mathbf{i}_n+\mathbf{i}_r)]\} \\
&\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_2(\mathbf{v}-\mathbf{i}_r)[d_0(\mathbf{B})H^{(n)}_{Y_2}(\mathbf{v}-\mathbf{i}_n)]/N \\
\xrightarrow{P}\ & E\{[b_0(\mathbf{B})^{-1}M(\mathbf{v}-\mathbf{i}_r)][b_0(\mathbf{B})^{-1}A(\mathbf{v}-\mathbf{i}_n)]\}.
\end{aligned}
\tag{3.5.89}
$$

Finally, in (3.5.70) there are also 8 terms of opposite sign, *i.e.*

$$\begin{aligned}
&\sum_{\mathbf{v}^\tau\in\mathcal{S}^*} Y_1(\mathbf{v}+\mathbf{i}_n)[b_0(\mathbf{B}^{-1})^{-1}d_0(\mathbf{B})H_M(\mathbf{v}+\mathbf{i}_r)]/N \\
\xrightarrow{P}\ & E\{[b_0(\mathbf{B}^{-1})^{-1}M(\mathbf{v}+\mathbf{i}_n)][b_0(\mathbf{B}^{-1})^{-1}A(\mathbf{v}+\mathbf{i}_r)]\} \\
&\sum_{\mathbf{v}^\tau\in\mathcal{S}^*} Y_2(\mathbf{v}-\mathbf{i}_n)[b_0(\mathbf{B})^{-1}d_0(\mathbf{B})H_M(\mathbf{v}-\mathbf{i}_r)]/N \\
\xrightarrow{P}\ & E\{[b_0(\mathbf{B})^{-1}M(\mathbf{v}-\mathbf{i}_n)][b_0(\mathbf{B})^{-1}A(\mathbf{v}-\mathbf{i}_r)]\} \\
&\sum_{\mathbf{v}^\tau\in\mathcal{S}^*} Y_2(\mathbf{v}-\mathbf{i}_r)[b_0(\mathbf{B})^{-1}d_0(\mathbf{B})H_M(\mathbf{v}-\mathbf{i}_n)]/N \\
\xrightarrow{P}\ & E\{[b_0(\mathbf{B})^{-1}M(\mathbf{v}-\mathbf{i}_r)][b_0(\mathbf{B})^{-1}A(\mathbf{v}-\mathbf{i}_n)]\} \\
&\sum_{\mathbf{v}^\tau\in\mathcal{S}^*} M(\mathbf{v})[b_0(\mathbf{B})^{-1}d_0(\mathbf{B})H^{(r)}_{Y_1}(\mathbf{v}-\mathbf{i}_n+\mathbf{i}_r)]/N \\
\xrightarrow{P}\ & E\{M(\mathbf{v})[b_0(\mathbf{B})^{-1}b_0(\mathbf{B}^{-1})^{-1}A(\mathbf{v}-\mathbf{i}_n+\mathbf{i}_r)]\} \\
&\sum_{\mathbf{v}^\tau\in\mathcal{S}^*} Y_1(\mathbf{v}+\mathbf{i}_r)[b_0(\mathbf{B}^{-1})^{-1}d_0(\mathbf{B})H_M(\mathbf{v}+\mathbf{i}_n)]/N \\
\xrightarrow{P}\ & E\{[b_0(\mathbf{B}^{-1})^{-1}M(\mathbf{v}+\mathbf{i}_n)][b_0(\mathbf{B}^{-1})^{-1}A(\mathbf{v}+\mathbf{i}_n)]\} \\
&\sum_{\mathbf{v}^\tau\in\mathcal{S}^*} M(\mathbf{v})[b_0(\mathbf{B}^{-1})^{-1}d_0(\mathbf{B})H^{(r)}_{Y_2}(\mathbf{v}+\mathbf{i}_n-\mathbf{i}_r)]/N \\
\xrightarrow{P}\ & E\{M(\mathbf{v})[b_0(\mathbf{B})^{-1}b_0(\mathbf{B}^{-1})^{-1}A(\mathbf{v}+\mathbf{i}_n-\mathbf{i}_r)]\} \\
&\sum_{\mathbf{v}^\tau\in\mathcal{S}^*} M(\mathbf{v})[b_0(\mathbf{B})^{-1}d_0(\mathbf{B})H^{(n)}_{Y_1}(\mathbf{v}+\mathbf{i}_n-\mathbf{i}_r)]/N \\
\xrightarrow{P}\ & E\{M(\mathbf{v})[b_0(\mathbf{B})^{-1}b_0(\mathbf{B}^{-1})^{-1}A(\mathbf{v}+\mathbf{i}_n-\mathbf{i}_r)]\} \\
&\sum_{\mathbf{v}^\tau\in\mathcal{S}^*} M(\mathbf{v})[b_0(\mathbf{B}^{-1})^{-1}d_0(\mathbf{B})H^{(n)}_{Y_2}(\mathbf{v}-\mathbf{i}_n+\mathbf{i}_r)]/N \\
\xrightarrow{P}\ & E\{M(\mathbf{v})[b_0(\mathbf{B})^{-1}b_0(\mathbf{B}^{-1})^{-1}A(\mathbf{v}-\mathbf{i}_n+\mathbf{i}_r)]\}. \qquad (3.5.90)
\end{aligned}$$

Thus, from the sum of all 32 terms, we may come up with 2 terms, *i.e.*

$$E\{[b_0(\mathbf{B})^{-1}b_0(\mathbf{B}^{-1})^{-1}M(\mathbf{v}+\mathbf{i}_n-\mathbf{i}_r)]A(\mathbf{v})\}+E\{[b_0(\mathbf{B})^{-1}b_0(\mathbf{B}^{-1})^{-1}M(\mathbf{v}-\mathbf{i}_n+\mathbf{i}_r)]A(\mathbf{v})\},$$

and use similar arguments like in Proposition 3.3, to explain why, for any $n, r = 1, \cdots, p$, the (n,r)-th element of $\mathbf{J}/N$ tends in probability to the (n,r)-th element of $2\cdot\sigma^2\mathbf{W}_{p+q}$. Similarly, we can obtain all the other elements of $\mathbf{J}/N$. ■

Theorem 3.6. If $\{\varepsilon^*(\mathbf{v})\} \sim IID(0,\sigma^2)$ and $E\{\varepsilon^*(\mathbf{v})^4\} < \infty$, then under conditions (C1) and (C4), it holds that

$$N^{1/2}[\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0] \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Delta}) \qquad (3.5.91)$$

as $N \to \infty$. Otherwise, if $\{\varepsilon^*(\mathbf{v})\}, \{u^*(\mathbf{v})\} \sim IID(0, \sigma^2)$ and $|E\{\varepsilon^*(\mathbf{v})^3\}| < \infty$, then under conditions (C1) and (C4), it holds that

$$N^{1/2}[\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0] \xrightarrow{D} N(\mathbf{0}, \mathbf{W}_{p+q}^{-1}) \tag{3.5.92}$$

as $N \to \infty$.

Proof. The proof will come immediately from (3.5.80) and Proposition 3.4, after we prove that

$$N^{-1/2}\mathbf{D} \xrightarrow{D} N(\mathbf{0}, \mathbf{M}) \tag{3.5.93}$$

or that for any $\boldsymbol{\lambda} \in \mathcal{R}^{p+q}$, it holds that

$$N^{-1/2}\boldsymbol{\lambda}^\tau \mathbf{D} \xrightarrow{D} N(0, \boldsymbol{\lambda}^\tau \mathbf{M} \boldsymbol{\lambda}). \tag{3.5.94}$$

For $n = 1, \cdots, p$, it holds that

$$\begin{aligned}
N^{-1/2} D_n(\boldsymbol{\theta}_0) &= N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_1(\mathbf{v} + \mathbf{i}_n) \, [d_0(\mathbf{B}) H_M(\mathbf{v})] \\
&+ N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_2(\mathbf{v} - \mathbf{i}_n) \, [d_0(\mathbf{B}) H_M(\mathbf{v})] \\
&- N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v}) \, [b_0(\mathbf{B})^{-1} d_0(\mathbf{B}) H_M(\mathbf{v} - \mathbf{i}_n)] \\
&- N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v}) \, [b_0(\mathbf{B}^{-1})^{-1} d_0(\mathbf{B}) H_M(\mathbf{v} + \mathbf{i}_n)] \\
&+ N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v}) \, [d_0(\mathbf{B}) H_{Y_1}^{(n)}(\mathbf{v} + \mathbf{i}_n)] \\
&+ N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v}) \, [d_0(\mathbf{B}) H_{Y_2}^{(n)}(\mathbf{v} - \mathbf{i}_n)] \\
&= N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_1(\mathbf{v} + \mathbf{i}_n) \, [d_0(\mathbf{B}) M(\mathbf{v})] \\
&+ N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_2(\mathbf{v} - \mathbf{i}_n) \, [d_0(\mathbf{B}) M(\mathbf{v})] \\
&- N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v}) \, [b_0(\mathbf{B})^{-1} d_0(\mathbf{B}) M(\mathbf{v} - \mathbf{i}_n)] \\
&- N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v}) \, [b_0(\mathbf{B}^{-1})^{-1} d_0(\mathbf{B}) M(\mathbf{v} + \mathbf{i}_n)] \\
&+ N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v}) \, [d_0(\mathbf{B}) Y_1(\mathbf{v} + \mathbf{i}_n)] \\
&+ N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v}) \, [d_0(\mathbf{B}) Y_2(\mathbf{v} - \mathbf{i}_n)] + o_P(1),
\end{aligned} \tag{3.5.95}$$

where the equality holds thanks to (3.5.17) and (3.5.18). Also, the condition $E\{\varepsilon^*(\mathbf{v})^4\} < \infty$ has been used there. This is because, when

$$\{\varepsilon^*(\mathbf{v})\} \sim IID(0, \sigma^2),$$

we may consider that $\{M(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ and $\{Y_1(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ are K_1 and K_2-dependent processes for positive and finite integers K_1, K_2. Thus for any $\mathbf{j}^\tau \notin \mathcal{F}$, we can claim the independence between pairs of random variables, such as $Y_1(\mathbf{v}+\mathbf{i}_n), M(\mathbf{v}+\mathbf{j})$ or $M(\mathbf{v}), M(\mathbf{v}+\mathbf{j})$ or $M(\mathbf{v}), Y_1(\mathbf{v}+\mathbf{i}_n+\mathbf{j})$. Unfortunately, the same cannot be said for the process $\{Y_2(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$, as

$$Y_2(\mathbf{v}) = b(\mathbf{B}^{-1})b(\mathbf{B})^{-1}a(\mathbf{B}^{-1})\varepsilon^*(\mathbf{v})$$

and an infinite linear filter needs to be applied on the one and only sequence of independent and identically distributed random variables $\{\varepsilon^*(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$.

Unless it holds that

$$\{u^*(\mathbf{v})\} \sim IID(0, \sigma^2),$$

we may proceed as follows to show that

$$N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_2(\mathbf{v}-\mathbf{i}_n)\ [d_0(\mathbf{B})H_M(\mathbf{v})] = N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_2(\mathbf{v}-\mathbf{i}_n)\ [d_0(\mathbf{B})M(\mathbf{v})] + o_P(1)$$

and that

$$N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v})\ [d_0(\mathbf{B})H^{(n)}_{Y_2}(\mathbf{v}-\mathbf{i}_n)] = N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v})\ [d_0(\mathbf{B})Y_2(\mathbf{v}-\mathbf{i}_n)] + o_P(1).$$

We demonstrate the first case only. It holds that

$$\begin{aligned} & N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_2(\mathbf{v}-\mathbf{i}_n)\ [d_0(\mathbf{B})M(\mathbf{v})] - N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_2(\mathbf{v}-\mathbf{i}_n)\ [d_0(\mathbf{B})H_M(\mathbf{v})] \\ = \ & N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_2(\mathbf{v}-\mathbf{i}_n) \sum_{\mathbf{j}^\tau \notin \mathcal{F}_\mathbf{v}} d_{\mathbf{j},0}\ M(\mathbf{v}-\mathbf{j}) \end{aligned}$$

with expected value

$$E\{N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_2(\mathbf{v}-\mathbf{i}_n) \sum_{\mathbf{j}^\tau \notin \mathcal{F}_\mathbf{v}} d_{\mathbf{j},0}\ M(\mathbf{v}-\mathbf{j})\} = 0$$

and variance

$$N^{-1}\mathrm{Var}\{ \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_2(\mathbf{v}-\mathbf{i}_n) \sum_{\mathbf{j}^\tau \notin \mathcal{F}_\mathbf{v}} d_{\mathbf{j},0}\ M(\mathbf{v}-\mathbf{j})\},$$

which involves the sum of variances

$$N^{-1} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \mathrm{Var}\{Y_2(\mathbf{v}-\mathbf{i}_n) \sum_{\mathbf{j}^\tau \notin \mathcal{F}_\mathbf{v}} d_{\mathbf{j},0}\ M(\mathbf{v}-\mathbf{j})\},$$

as well as a sum of cross-terms. We can see immediately that the sum of variances will involve

$$\mathrm{Var}\{\varepsilon^*(\mathbf{v})^2\} = E\{\varepsilon^*(\mathbf{v})^4\} - \sigma^4$$

and it will also involve

$$\mathrm{Var}\{\varepsilon^*(\mathbf{v})\varepsilon^*(\mathbf{v}-\mathbf{j})\} = E\{\varepsilon^*(\mathbf{v})^2\}\ E\{\varepsilon^*(\mathbf{v}-\mathbf{j})^2\} = \sigma^4,$$

for $\mathbf{j} \neq \mathbf{0}$. Thus, a condition on the fourth moment is required to move on.

We may re-write (3.5.95) as

$$\begin{aligned}
N^{-1/2}D_n(\boldsymbol{\theta}_0) &= N^{-1/2}\sum_{\mathbf{v}^\tau\in\mathcal{S}^*}[b_0(\mathbf{B}^{-1})^{-1}M(\mathbf{v}+\mathbf{i}_n)]\ A(\mathbf{v}) \\
&+ N^{-1/2}\sum_{\mathbf{v}^\tau\in\mathcal{S}^*}[b_0(\mathbf{B})^{-1}M(\mathbf{v}-\mathbf{i}_n)]\ A(\mathbf{v}) \\
&- N^{-1/2}\sum_{\mathbf{v}^\tau\in\mathcal{S}^*}M(\mathbf{v})\ [b_0(\mathbf{B})^{-1}A(\mathbf{v}-\mathbf{i}_n)] \\
&- N^{-1/2}\sum_{\mathbf{v}^\tau\in\mathcal{S}^*}M(\mathbf{v})\ [b_0(\mathbf{B}^{-1})^{-1}A(\mathbf{v}+\mathbf{i}_n)] \\
&+ N^{-1/2}\sum_{\mathbf{v}^\tau\in\mathcal{S}^*}M(\mathbf{v})\ [b_0(\mathbf{B}^{-1})^{-1}A(\mathbf{v}+\mathbf{i}_n)] \\
&+ N^{-1/2}\sum_{\mathbf{v}^\tau\in\mathcal{S}^*}M(\mathbf{v})\ [b_0(\mathbf{B})^{-1}A(\mathbf{v}-\mathbf{i}_n)] + o_P(1) \\
&= N^{-1/2}\sum_{\mathbf{v}^\tau\in\mathcal{S}^*}[b_0(\mathbf{B}^{-1})^{-1}M(\mathbf{v}+\mathbf{i}_n)]\ A(\mathbf{v}) \\
&+ N^{-1/2}\sum_{\mathbf{v}^\tau\in\mathcal{S}^*}[b_0(\mathbf{B})^{-1}M(\mathbf{v}-\mathbf{i}_n)]\ A(\mathbf{v}) + o_P(1). \qquad (3.5.96)
\end{aligned}$$

On the other hand, for $m = 1, \cdots, q$, it holds that

$$\begin{aligned}
N^{-1/2}D_{p+m}(\boldsymbol{\theta}_0) &= N^{-1/2}\sum_{\mathbf{v}^\tau\in\mathcal{S}^*}M(\mathbf{v})\ [a_0(\mathbf{B})^{-1}d_0(\mathbf{B})H_M(\mathbf{v}-\mathbf{j}_m)] \\
&+ N^{-1/2}\sum_{\mathbf{v}^\tau\in\mathcal{S}^*}M(\mathbf{v})\ [a_0(\mathbf{B}^{-1})^{-1}d_0(\mathbf{B})H_M(\mathbf{v}+\mathbf{j}_m)] \\
&= N^{-1/2}\sum_{\mathbf{v}^\tau\in\mathcal{S}^*}M(\mathbf{v})\ [a_0(\mathbf{B})^{-1}d_0(\mathbf{B})M(\mathbf{v}-\mathbf{j}_m)] \\
&+ N^{-1/2}\sum_{\mathbf{v}^\tau\in\mathcal{S}^*}M(\mathbf{v})\ [a_0(\mathbf{B}^{-1})^{-1}d_0(\mathbf{B})M(\mathbf{v}+\mathbf{j}_m)] + o_P(1) \\
&= N^{-1/2}\sum_{\mathbf{v}^\tau\in\mathcal{S}^*}M(\mathbf{v})\ [a_0(\mathbf{B})^{-1}A(\mathbf{v}-\mathbf{j}_m)] \\
&+ N^{-1/2}\sum_{\mathbf{v}^\tau\in\mathcal{S}^*}M(\mathbf{v})\ [a_0(\mathbf{B}^{-1})^{-1}A(\mathbf{v}+\mathbf{j}_m)] + o_P(1), \qquad (3.5.97)
\end{aligned}$$

using the same arguments. Thus, if we define, for any $\mathbf{v}^\tau \in \mathcal{Z}^d$, the random variables

$$U_n(\mathbf{v}) \equiv [b_0(\mathbf{B})^{-1}M(\mathbf{v}-\mathbf{i}_n)]\ A(\mathbf{v}) + [b_0(\mathbf{B}^{-1})^{-1}M(\mathbf{v}+\mathbf{i}_n)]\ A(\mathbf{v}),\ n = 1, \cdots, p, \qquad (3.5.98)$$

$$U_{p+m}(\mathbf{v}) \equiv [a_0(\mathbf{B})^{-1}A(\mathbf{v}-\mathbf{j}_m)]\ M(\mathbf{v}) + [a_0(\mathbf{B}^{-1})^{-1}A(\mathbf{v}+\mathbf{j}_m)]\ M(\mathbf{v}),\ m = 1, \cdots, q, \qquad (3.5.99)$$

and the random vectors

$$\mathbf{U}(\mathbf{v}) \equiv [U_1(\mathbf{v}), \cdots, U_{p+q}(\mathbf{v})]^\tau, \tag{3.5.100}$$

then it holds that

$$N^{-1/2}\boldsymbol{\lambda}^\tau \mathbf{D} = N^{-1/2}\boldsymbol{\lambda}^\tau \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \mathbf{U}(\mathbf{v}) + o_P(1). \tag{3.5.101}$$

For convenience, we may re-write for any $\mathbf{v}^\tau \in \mathcal{Z}^d$

$$\begin{aligned} U_n(\mathbf{v}) &= Y_2(\mathbf{v}-\mathbf{i}_n)\ [d_0(\mathbf{B})M(\mathbf{v})] + Y_1(\mathbf{v}+\mathbf{i}_n)\ [d_0(\mathbf{B})M(\mathbf{v})] \\ &= [b_0(\mathbf{B})^{-1}M(\mathbf{v}-\mathbf{i}_n)]\ [d_0(\mathbf{B})M(\mathbf{v})] + Y_1(\mathbf{v}+\mathbf{i}_n)\ [d_0(\mathbf{B})M(\mathbf{v})],\ n=1,\cdots,p, \end{aligned} \tag{3.5.102}$$

$$U_{p+m}(\mathbf{v}) = [h_0(\mathbf{B})M(\mathbf{v}-\mathbf{j}_m)]\ M(\mathbf{v}) + [h_0(\mathbf{B}^{-1})M(\mathbf{v}+\mathbf{j}_m)]\ M(\mathbf{v}),\ m=1,\cdots,q, \tag{3.5.103}$$

where

$$h_0(\mathbf{z}) \equiv a_0(\mathbf{z})^{-1}d_0(\mathbf{z}) \equiv \sum_{\mathbf{j}^\tau \in \mathcal{Z}^d} h_{\mathbf{j},0}\ \mathbf{z}^{\mathbf{j}}. \tag{3.5.104}$$

For any positive integer K, we define

$$A^{(K)}(\mathbf{v}) \equiv \sum_{\mathbf{j}^\tau \in \mathcal{F}_K} d_{\mathbf{j},0}\ M(\mathbf{v}-\mathbf{j}), \tag{3.5.105}$$

where the set $\mathcal{F}_K$ was defined in (2.5.4). We define the polynomial

$$b_0(\mathbf{z})^{-1} \equiv \sum_{\mathbf{j}\geq \mathbf{0}} \Phi_{\mathbf{j},0}\ \mathbf{z}^{\mathbf{j}},\ \Phi_{\mathbf{0},0}=1. \tag{3.5.106}$$

Then we define for any $\mathbf{v}^\tau \in \mathcal{Z}^d$

$$Y_2^{(K)}(\mathbf{v}) \equiv \begin{cases} Y_2(\mathbf{v}), & \text{if } \{u(\mathbf{v})\} \sim IID(0,\sigma^2) \\ \sum_{\mathbf{j}^\tau \in \mathcal{M}_K} \Phi_{\mathbf{j},0}\ M(\mathbf{v}-\mathbf{j}), & \text{if } E\{\varepsilon^*(\mathbf{v})^4\} < \infty \end{cases}, \tag{3.5.107}$$

where the set $\mathcal{M}_K$ was defined in (2.3.6). Then, we define for any $\mathbf{v}^\tau \in \mathcal{Z}^d$ the random variables

$$U_n^{(K)}(\mathbf{v}) \equiv [Y_2^{(K)}(\mathbf{v}-\mathbf{i}_n)\ A^{(K)}(\mathbf{v}) + Y_1(\mathbf{v}+\mathbf{i}_n)\ A^{(K)}(\mathbf{v})], \tag{3.5.108}$$

for $n=1,\cdots,p$, and

$$U_{p+m}^{(K)}(\mathbf{v}) \equiv [\sum_{\mathbf{j}^\tau \in \mathcal{F}_K} h_{\mathbf{j},0}\ M(\mathbf{v}-\mathbf{j}_m-\mathbf{j})]\ M(\mathbf{v}) + [\sum_{\mathbf{j}^\tau \in \mathcal{F}_K} h_{\mathbf{j},0}\ M(\mathbf{v}+\mathbf{j}_m+\mathbf{j})]\ M(\mathbf{v}), \tag{3.5.109}$$

for $m = 1, \cdots, q$. For any $\mathbf{v}^\tau \in \mathcal{Z}^d$, we also define the random vector

$$\mathbf{U}^{(K)}(\mathbf{v}) \equiv [U_1^{(K)}(\mathbf{v}), \cdots, U_{p+q}^{(K)}(\mathbf{v})]^\tau. \tag{3.5.110}$$

We can see immediately that when

$$\{\varepsilon^*(\mathbf{v})\} \sim IID(0, \sigma^2)$$

the process $\{\boldsymbol{\lambda}^\tau \mathbf{U}^{(K)}(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ is a K^*-dependent process for some positive integer K^*. Then since (C1) holds, we may write

$$N^{-1/2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} \boldsymbol{\lambda}^\tau \mathbf{U}^{(K)}(\mathbf{v}) \xrightarrow{D} N(0, \boldsymbol{\lambda}^\tau \mathbf{M}_K \boldsymbol{\lambda}), \tag{3.5.111}$$

as $N \to \infty$. Next and similarly to Theorem 3.4, we may show that

$$\boldsymbol{\lambda}^\tau \mathbf{M}_K \boldsymbol{\lambda} \to \boldsymbol{\lambda}^\tau \mathbf{M}\ \boldsymbol{\lambda}, \tag{3.5.112}$$

where

$$\mathbf{M} \equiv \sum_{\mathbf{i}^\tau \in \mathcal{Z}^d} \boldsymbol{\Gamma}(\mathbf{i}) \tag{3.5.113}$$

and

$$\boldsymbol{\Gamma}(\mathbf{i}) \equiv E\{\mathbf{U}(\mathbf{v})\mathbf{U}^\tau(\mathbf{v} - \mathbf{i})\}. \tag{3.5.114}$$

Of course, we will need again the assumption

$$E\{\varepsilon^*(\mathbf{v})^4\} < \infty,$$

unless it holds that

$$\{u^*(\mathbf{v})\} \sim IID(0, \sigma^2).$$

In both cases, we may conclude that, for any $\boldsymbol{\lambda} \in \mathcal{R}^{p+q}$, (3.5.94) holds and $\mathbf{M}$ comes from (3.5.113).

The next step will be to show that if

$$\{u^*(\mathbf{v})\} \sim IID(0, \sigma^2),$$

then

$$\mathbf{M} \equiv 4 \cdot \sigma^4 \mathbf{W}_{p+q}. \tag{3.5.115}$$

In Theorem 3.4, we have shown why (3.5.115) holds for the elements $(n, r), n, r = 1, \cdots, p$, and for the elements $(p + m, p + l), m, l = 1, \cdots, q$, of the two matrices.

We may proceed the same way for the elements $(n, p+m), n = 1, \cdots, p, m = 1, \cdots, q$, of the two matrices. First, let us write the $(n, p+m)$-th element of $\mathbf{\Gamma}(\mathbf{i})$ to be equal to

$$\begin{aligned}
& E\{[[b_0(\mathbf{B})^{-1}M(\mathbf{v}-\mathbf{i}_n)]A(\mathbf{v}) + [b_0(\mathbf{B}^{-1})^{-1}M(\mathbf{v}+\mathbf{i}_n)]A(\mathbf{v})] \\
& [[a_0(\mathbf{B})^{-1}A(\mathbf{v}-\mathbf{j}_m-\mathbf{i})]M(\mathbf{v}-\mathbf{i}) + [a_0(\mathbf{B}^{-1})^{-1}A(\mathbf{v}+\mathbf{j}_m-\mathbf{i})]M(\mathbf{v}-\mathbf{i})]\} \\
= \; & E\{[b_0(\mathbf{B})^{-1}M(\mathbf{v}-\mathbf{i}_n)][a_0(\mathbf{B})^{-1}A(\mathbf{v}-\mathbf{j}_m-\mathbf{i})]A(\mathbf{v})M(\mathbf{v}-\mathbf{i})\} \qquad (3.5.116) \\
+ \; & E\{[b_0(\mathbf{B})^{-1}M(\mathbf{v}-\mathbf{i}_n)][a_0(\mathbf{B}^{-1})^{-1}A(\mathbf{v}+\mathbf{j}_m-\mathbf{i})]A(\mathbf{v})M(\mathbf{v}-\mathbf{i})\} \qquad (3.5.117) \\
+ \; & E\{[b_0(\mathbf{B}^{-1})^{-1}M(\mathbf{v}+\mathbf{i}_n)][a_0(\mathbf{B})^{-1}A(\mathbf{v}-\mathbf{j}_m-\mathbf{i})]A(\mathbf{v})M(\mathbf{v}-\mathbf{i})\} \qquad (3.5.118) \\
+ \; & E\{[b_0(\mathbf{B}^{-1})^{-1}M(\mathbf{v}+\mathbf{i}_n)][a_0(\mathbf{B}^{-1})^{-1}A(\mathbf{v}+\mathbf{j}_m-\mathbf{i})]A(\mathbf{v})M(\mathbf{v}-\mathbf{i})\}. \qquad (3.5.119)
\end{aligned}$$

We define the polynomial

$$a_0(\mathbf{z})^{-1} \equiv \sum_{\mathbf{j}\geq\mathbf{0}} \Theta_{\mathbf{j},0}\, \mathbf{z}^{\mathbf{j}}, \;\; \Theta_{\mathbf{0},0} = 1. \qquad (3.5.120)$$

From the first term (3.5.116) and in the $(n, p+m)$-th element of $\mathbf{M}$, we will come up with

$$\begin{aligned}
& \sum_{\mathbf{i}^\tau\in\mathcal{Z}^d} \sum_{\mathbf{j},\mathbf{j}^*\geq\mathbf{0}} \Phi_{\mathbf{j},0}\cdot\Theta_{\mathbf{j}^*,0}\; E\{M(\mathbf{v}-\mathbf{i}_n-\mathbf{j})A(\mathbf{v}-\mathbf{j}_m-\mathbf{i}-\mathbf{j}^*)A(\mathbf{v})M(\mathbf{v}-\mathbf{i})\} \\
= \; & \sum_{\mathbf{i}^\tau\in\mathcal{Z}^d} \sum_{\mathbf{j},\mathbf{j}^*\geq\mathbf{0}} \Phi_{\mathbf{j},0}\cdot\Theta_{\mathbf{j}^*,0}\; E\{M(\mathbf{v}-\mathbf{i}_n-\mathbf{j})M(\mathbf{v}-\mathbf{i})\}\; E\{A(\mathbf{v}-\mathbf{j}_m-\mathbf{i}-\mathbf{j}^*)A(\mathbf{v})\} \\
+ \; & \sigma^4 \sum_{\substack{\mathbf{j},\mathbf{j}^*\geq\mathbf{0},\\ \mathbf{j}+\mathbf{i}_n=\mathbf{j}^*+\mathbf{j}_m}} \Phi_{\mathbf{j},0}\cdot\Theta_{\mathbf{j}^*,0}, \qquad (3.5.121)
\end{aligned}$$

where the equality holds, according to Proposition 2.6, since for any $\mathbf{j} \geq \mathbf{0}$ and any $n = 1, \cdots, p$, it cannot be that $\mathbf{v} - \mathbf{i}_n - \mathbf{j} = \mathbf{v}$, but it can be that

$$\mathbf{v} - \mathbf{i} = \mathbf{v}$$

when $\mathbf{i} = \mathbf{0}$ and that

$$\mathbf{v} - \mathbf{i}_n - \mathbf{j} = \mathbf{v} - \mathbf{j}_m - \mathbf{i} - \mathbf{j}^* = \mathbf{v} - \mathbf{j}_m - \mathbf{j}^*,$$

when $\mathbf{i}_n + \mathbf{j} = \mathbf{j}_m + \mathbf{j}^*$. Of course, it is then that it cannot be $\mathbf{v} = \mathbf{v} - \mathbf{i}_n - \mathbf{j}$ to risk that all four locations are the same. Thus, we may write

$$\sum_{\substack{\mathbf{j},\mathbf{j}^*\geq\mathbf{0},\\ \mathbf{j}+\mathbf{i}_n=\mathbf{j}^*+\mathbf{j}_m}} \Phi_{\mathbf{j},0}\cdot\Theta_{\mathbf{j}^*,0} = E\{\xi(-\mathbf{i}_n)\eta(-\mathbf{j}_m)\}. \qquad (3.5.122)$$

For the rest of (3.5.121), *i.e.* the sum

$$\sum_{\mathbf{i}^\tau \in \mathcal{Z}^d} \sum_{\mathbf{j},\mathbf{j}^* \geq \mathbf{0}} \Phi_{\mathbf{j},0} \cdot \Theta_{\mathbf{j}^*,0} \; E\{M(\mathbf{v}-\mathbf{i}_n-\mathbf{j})M(\mathbf{v}-\mathbf{i})\} \; E\{A(\mathbf{v}-\mathbf{j}_m-\mathbf{i}-\mathbf{j}^*)A(\mathbf{v})\}, \quad (3.5.123)$$

we write

$$E\{M(\mathbf{v})M(\mathbf{v}-\mathbf{i})\} \equiv \sigma^2 \cdot \gamma_M(\mathbf{i}) \quad (3.5.124)$$

and

$$E\{A(\mathbf{v})A(\mathbf{v}-\mathbf{i})\} \equiv \sigma^2 \cdot c_A(\mathbf{i}). \quad (3.5.125)$$

Then we may re-write (3.5.123) as

$$\sigma^4 \sum_{\mathbf{j},\mathbf{j}^* \geq \mathbf{0}} \Phi_{\mathbf{j},0} \cdot \Theta_{\mathbf{j}^*,0} \sum_{\mathbf{i}^\tau \in \mathcal{Z}^d} \gamma_M(\mathbf{i}-\mathbf{i}_n-\mathbf{j}) \; c_A(\mathbf{j}_m+\mathbf{j}^*+\mathbf{i}). \quad (3.5.126)$$

From the general Yule-Walker equations, it holds that

$$\sum_{\mathbf{i}^\tau \in \mathcal{Z}^d} \gamma_M(\mathbf{i}-\mathbf{i}_n-\mathbf{j}) \; c_A(\mathbf{j}_m+\mathbf{j}^*+\mathbf{i}) = 0. \quad (3.5.127)$$

As a result, from the first term (3.5.116) and in the $(n, p+m)$-th element of $\mathbf{M}$, we will come up with

$$\sigma^4 \; E\{\xi(-\mathbf{i}_n)\eta(-\mathbf{j}_m)\}. \quad (3.5.128)$$

Following the same way for the three terms (3.5.117), (3.5.118) and (3.5.119), we may show that, for any $n = 1, \cdots, p$, $m = 1, \cdots, q$, the $(n, p+m)$-th element of $\mathbf{M}$ is equal to

$$4 \cdot \sigma^4 \; E\{\xi(-\mathbf{i}_n)\eta(-\mathbf{j}_m)\}.$$

Similarly for the $(p+m, n)$-th element of $\mathbf{M}$. Thus, (3.5.115) holds and the proof of the theorem has been completed. ■

Remark 3.4. (i) In Theorem 3.5, Proposition 3.4 and Theorem 3.6, we have used the condition

$$\{\varepsilon^*(\mathbf{v})\} \sim IID(0, \sigma^2). \quad (3.5.129)$$

This is because both the moving-average $\{M(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ and the auto-regression $\{A(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ can be expressed as unilateral functions of this error sequence. Nevertheless, the condition

$$\{\varepsilon(\mathbf{v})\} \sim IID(0, \sigma^2) \quad (3.5.130)$$

would still imply that all the processes of interest are linear functions of independent and identically distributed random variables and, thus, it could replace (3.5.129) in

Theorem 3.5 and Proposition 3.4. As for the first part of Theorem 3.6, which claims the asymptotic normality of the estimators, since we can write

$$M(\mathbf{v}) = b(\mathbf{B}^{-1})a(\mathbf{B})\varepsilon(\mathbf{v}) \tag{3.5.131}$$

and

$$Y_1(\mathbf{v}) = a(\mathbf{B})\varepsilon(\mathbf{v}), \tag{3.5.132}$$

then two finite linear filters, though one is bilateral, can be applied on $\{\varepsilon(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ to derive $\{M(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ and $\{Y_1(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$, and condition (3.5.130) can be used to consider those two as K_1 and K_2-dependent processes, for some finite positive integers K_1, K_2. Condition

$$E\{\varepsilon(\mathbf{v})^4\} < \infty$$

would then be used to fix the fact that $\{Y_2(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ is not a K-dependent process.

(ii) For the second part of Theorem 3.6 that proves the form of the variance matrix $\mathbf{\Delta}$, we have used the extra assumption that

$$\{u^*(\mathbf{v})\} \sim IID(0, \sigma^2), \tag{3.5.133}$$

which together with (3.5.129) implies that for any $\mathbf{v}^\tau \in \mathcal{Z}^d$, it holds that $M(\mathbf{v})$ and $A(\mathbf{v} - \mathbf{j})$ are two independent random variables for any $\mathbf{j} \neq \mathbf{0}$. Then, Proposition 2.6 might be used.

(iii) According to (ii), conditions (3.5.129) and (3.5.133) have been used for the proof of the second part of Theorem 3.6, while it is only the process

$$\{\varepsilon(\mathbf{v})\} \sim WN(0, \sigma^2),$$

which is such that the ARMA process $\{Z(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$, from which we have obtained the observations, might be written as a causal function of this process. Thus, it would be a question of interest if (3.5.130) could hold, in addition to (3.5.129) and (3.5.133) for non-Gaussian processes. Even if we only require (3.5.129) and (3.5.133) to hold, it is interesting to see if those conditions could be obtained for non-Gaussian random variables. This is also the special case of Theorem 3.4. For Gaussian processes though, the estimators $\widehat{\boldsymbol{\theta}}$ of either Theorem 3.4 or 3.6 are efficient anyway.

3.6 Bilateral ARMA models

For $\mathbf{0} < \mathbf{i}_1 < \cdots < \mathbf{i}_p$ and $\mathbf{0} < \mathbf{j}_1 < \cdots < \mathbf{j}_q$, we define the (weakly) stationary ARMA process $\{Z(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ by the equation

$$b(\mathbf{B})Z(\mathbf{v}) \equiv a(\mathbf{B})W(\mathbf{v}),\ \{W(\mathbf{v})\} \sim WN(0, \sigma_W^2). \tag{3.6.1}$$

We consider that

$$b(\mathbf{z}) \equiv 1 - \sum_{n=1}^{p} b_{\mathbf{i}_n}, \tag{3.6.2}$$

$$a(\mathbf{z}) \equiv 1 + \sum_{m=1}^{q} a_{\mathbf{j}_m}, \tag{3.6.3}$$

and $\{Z(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ has not necessarily been expressed as a causal and invertible ARMA process in (3.6.1). We then write

$$b(\mathbf{z})\ b(\mathbf{z}^{-1}) \equiv c_b\ \phi(\mathbf{z})\ \phi(\mathbf{z}^{-1}), \tag{3.6.4}$$

where

$$\phi(\mathbf{z}) \equiv 1 - \sum_{\mathbf{j} \in \mathcal{I}_p} \varphi_{\mathbf{j}}\ \mathbf{z}^{\mathbf{j}},\ \sum_{\mathbf{j} \in \mathcal{I}_p} |\varphi_{\mathbf{j}}| < \infty, \tag{3.6.5}$$

with the set $\mathcal{I}_p \subseteq \{\mathbf{j} > \mathbf{0}\}$ and with

$$\Phi(\mathbf{z}) \equiv \phi(\mathbf{z})^{-1} \equiv 1 + \sum_{\mathbf{j} > \mathbf{0}} \Phi_{\mathbf{j}}\ \mathbf{z}^{\mathbf{j}},\ \sum_{\mathbf{j} > \mathbf{0}} |\Phi_{\mathbf{j}}| < \infty. \tag{3.6.6}$$

Similarly, we write

$$a(\mathbf{z})\ a(\mathbf{z}^{-1}) \equiv c_a\ \theta(\mathbf{z})\ \theta(\mathbf{z}^{-1}), \tag{3.6.7}$$

where

$$\theta(\mathbf{z}) \equiv 1 + \sum_{\mathbf{j} \in \mathcal{J}_q} \theta_{\mathbf{j}}\ \mathbf{z}^{\mathbf{j}},\ \sum_{\mathbf{j} \in \mathcal{J}_q} |\theta_{\mathbf{j}}| < \infty, \tag{3.6.8}$$

with the set $\mathcal{J}_q \subseteq \{\mathbf{j} > \mathbf{0}\}$ and with

$$\Theta(\mathbf{z}) \equiv \theta(\mathbf{z})^{-1} \equiv 1 + \sum_{\mathbf{j} > \mathbf{0}} \Theta_{\mathbf{j}}\ \mathbf{z}^{\mathbf{j}},\ \sum_{\mathbf{j} > \mathbf{0}} |\Theta_{\mathbf{j}}| < \infty. \tag{3.6.9}$$

We can write the spectral density of $\{Z(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ as

$$g_Z(\boldsymbol{\omega}) \equiv \sigma_W^2\ \frac{a(e^{i\boldsymbol{\omega}})\ a(e^{-i\boldsymbol{\omega}})}{b(e^{i\boldsymbol{\omega}})\ b(e^{-i\boldsymbol{\omega}})} \equiv \sigma^2\ \frac{\theta(e^{i\boldsymbol{\omega}})\ \theta(e^{-i\boldsymbol{\omega}})}{\phi(e^{i\boldsymbol{\omega}})\ \phi(e^{-i\boldsymbol{\omega}})} = \sigma^2\ \frac{\Phi(e^{i\boldsymbol{\omega}})\ \Phi(e^{-i\boldsymbol{\omega}})}{\Theta(e^{i\boldsymbol{\omega}})\ \Theta(e^{-i\boldsymbol{\omega}})},\ \boldsymbol{\omega}^\tau \in [-\pi, \pi]^d, \tag{3.6.10}$$

where

$$\sigma^2 \equiv \sigma_W^2 \frac{c_a}{c_b}. \tag{3.6.11}$$

Thus, there exist processes

$$\begin{aligned} \{\varepsilon(\mathbf{v})\} &\sim WN(0,\sigma^2) \\ \{u(\mathbf{v})\} &\sim WN(0,\sigma^2), \end{aligned}$$

such that we can write

$$\phi(\mathbf{B})Z(\mathbf{v}) = \theta(\mathbf{B})\varepsilon(\mathbf{v}) \tag{3.6.12}$$

and

$$\phi(\mathbf{B}^{-1})Z(\mathbf{v}) = \theta(\mathbf{B}^{-1})u(\mathbf{v}), \tag{3.6.13}$$

but $\mathcal{I}_p$ and $\mathcal{J}_q$ are not necessarily sets of finite cardinality when $d \geq 2$ and the process cannot necessarily be expressed as a causal and invertible ARMA process of finite order. A description of this problem has been attempted before by Whittle (1954, p.439) for the case of two-dimensional auto-regressions. We will also refer to it again in Section 4.4.

We may now define for any $\mathbf{v}^\tau \in \mathcal{Z}^d$ the random variables

$$M(\mathbf{v}) \equiv \phi(\mathbf{B})\phi(\mathbf{B}^{-1})Z(\mathbf{v}) \tag{3.6.14}$$

and

$$M^*(\mathbf{v}) \equiv b(\mathbf{B})b(\mathbf{B}^{-1})Z(\mathbf{v}) = c_b\, M(\mathbf{v}). \tag{3.6.15}$$

We also define for any $\mathbf{v}^\tau \in \mathcal{Z}^d$

$$A(\mathbf{v}) \equiv \theta(\mathbf{B})^{-1}\theta(\mathbf{B}^{-1})^{-1}Z(\mathbf{v}) = \Theta(\mathbf{B})\Theta(\mathbf{B}^{-1})Z(\mathbf{v}) \tag{3.6.16}$$

and

$$A^*(\mathbf{v}) \equiv a(\mathbf{B})^{-1}a(\mathbf{B}^{-1})^{-1}Z(\mathbf{v}) = \frac{1}{c_a}\, A(\mathbf{v}). \tag{3.6.17}$$

We established in Section 3.5.2 why the two random variables $M(\mathbf{v})$ and $A(\mathbf{v}-\mathbf{j})$ are uncorrelated for any $\mathbf{v}^\tau \in \mathcal{Z}^d$ and any $\mathbf{j} \neq \mathbf{0}$. This is because, for the two sequences of uncorrelated random variables $\{\varepsilon^*(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ and $\{u^*(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ we could write

$$\begin{aligned} M(\mathbf{v}) &= \phi(\mathbf{B}^{-1})\theta(\mathbf{B}^{-1})\varepsilon^*(\mathbf{v}) \\ &= \phi(\mathbf{B})\theta(\mathbf{B})u^*(\mathbf{v}) \end{aligned}$$

and

$$\begin{aligned} A(\mathbf{v}) &= \phi(\mathbf{B})^{-1}\theta(\mathbf{B})^{-1}\varepsilon^*(\mathbf{v}) = \Phi(\mathbf{B})\Theta(\mathbf{B})\varepsilon^*(\mathbf{v}) \\ &= \phi(\mathbf{B}^{-1})^{-1}\theta(\mathbf{B}^{-1})^{-1}u^*(\mathbf{v}) = \Phi(\mathbf{B}^{-1})\Theta(\mathbf{B}^{-1})u^*(\mathbf{v}). \end{aligned}$$

Consequently, the two random variables $M^*(\mathbf{v})$ and $A^*(\mathbf{v}-\mathbf{j})$ are uncorrelated for any $\mathbf{v}^\tau \in \mathcal{Z}^d$ and any $\mathbf{j} \neq \mathbf{0}$.

Moreover, the original process $\{Z(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ and the two new processes of interest $\{M^*(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ and $\{A^*(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ are all linear functions of $\{W(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$. Especially, for $\{M^*(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ it holds that

$$M^*(\mathbf{v}) = b(\mathbf{B}^{-1})a(\mathbf{B})W(\mathbf{v}), \tag{3.6.18}$$

and $b(\mathbf{z}^{-1})a(\mathbf{z})$ is a finite filter. Thus, if

$$\{W(\mathbf{v})\} \sim IID(0, \sigma_W^2) \tag{3.6.19}$$

then it holds that $\{M^*(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ is K-dependent process for some positive and finite integer number K.

All these arguments are mentioned here to convince for the equivalence of the two random quantities

$$\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M(\mathbf{v}, \varphi) \sum_{\mathbf{j}^\tau \in \mathcal{F}_\mathbf{v}} d_\mathbf{j}\ M(\mathbf{v}-\mathbf{j}, \varphi) \tag{3.6.20}$$

and

$$\sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M^*(\mathbf{v}, \mathbf{b}) \sum_{\mathbf{j}^\tau \in \mathcal{F}_\mathbf{v}} d_\mathbf{j}^*\ M^*(\mathbf{v}-\mathbf{j}, \mathbf{b}), \tag{3.6.21}$$

where φ, $\mathbf{b}$ are the auto-regressive parameters of the two representations and we define the polynomials

$$d(\mathbf{z}) \equiv \{\varphi(\mathbf{z})\varphi(\mathbf{z}^{-1})\ \theta(\mathbf{z})\theta(\mathbf{z}^{-1})\}^{-1} \tag{3.6.22}$$

and

$$d^*(\mathbf{z}) \equiv \{b(\mathbf{z})b(\mathbf{z}^{-1})\ a(\mathbf{z})a(\mathbf{z}^{-1})\}^{-1} = \frac{d(\mathbf{z})}{c_b\ c_a}. \tag{3.6.23}$$

The set $\mathcal{S}^* \subset \mathcal{Z}^d$ has been created exactly like before from an original set $\mathcal{I}$; it depends on the set $\mathcal{F}$, which is related to the auto-covariance function of any of the two processes $\{M(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ or $\{M^*(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$, since both functions are equal to zero at exactly the same vector lags. Similarly for the sets $\mathcal{F}_\mathbf{v}$ and any $\mathbf{v}^\tau \in \mathcal{Z}^d$.

Equivalence of the two random quantities implies that the asymptotic normality of the estimators can be achieved for bilateral ARMA processes. The variances of the

Gaussian likelihood estimators are expected to be proportional to the variances achieved by efficient estimators. Whittle (1954) performed a similar transition from a sum of squares of errors of a bilateral and two-dimensional auto-regression to the sum of squares of errors of its unique causal second-order equivalent. This quantity closely resembled our conditional likelihood maximized in Section 3.3, for the case of an auto-regression. Thus, if one wishes to apply the results of Section 3.3 to bilateral auto-regressions, they will have to follow Whittle (1954).

The equivalence of (3.6.21) and (3.6.20) follows immediately, since we have defined our quantities in a way that treats the two sides of the polynomials equally. In other words, we come up with processes, such as the moving-averages M or M^* or the auto-regressions A or A^*, and we do not depend on causal formulations. In the contrary, it was not that obvious when Whittle (1954) transformed a function of the errors of a bilateral auto-regression to the same function of the errors of its unique causal representation. Let us also not forget that Whittle (1954, p.441), after ignoring the edge-effect and any candidate corrections required to proceed without it, he provided the Gaussian likelihood in four different forms, two of which only are those already mentioned to compute sum of squares of uncorrelated variables. The fourth form is not other than the spectral domain quantity, which Guyon (1982) used next to correct the edge-effect. The third quantity closely resembled our suggestion, as it involved the auto-covariance function of an auto-regression and a moving-average; one in terms of data and the other in terms of parameters.

The deterministic part can have an effect on the bias of the estimators. Indeed, we demonstrate this next. If we consider again

$$Y_1(\mathbf{v}) \equiv \varphi(\mathbf{B})Z(\mathbf{v}) = \theta(\mathbf{B})\varepsilon(\mathbf{v}), \tag{3.6.24}$$

$$Y_2(\mathbf{v}) \equiv \varphi(\mathbf{B}^{-1})Z(\mathbf{v}) = \theta(\mathbf{B}^{-1})u(\mathbf{v}), \tag{3.6.25}$$

and if we define

$$Y_1^*(\mathbf{v}) \equiv b(\mathbf{B})Z(\mathbf{v}) = a(\mathbf{B})W(\mathbf{v}), \tag{3.6.26}$$

$$Y_2^*(\mathbf{v}) \equiv b(\mathbf{B}^{-1})Z(\mathbf{v}), \tag{3.6.27}$$

then $\{Y_1^*(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ and $\{Y_2^*(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ have been expressed as finite moving-averages, though not necessarily invertible. Those two processes share non-zero auto-covariances with the process $\{M^*(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$, for finite sets of vector lags. Thus, we

may deal with the edge-effect and, after we take the derivatives of (3.6.21) with respect to the parameters and set the true parameter vector, we may come up (with opposite signs)

$$\begin{aligned} D_n^* \equiv & \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_1^*(\mathbf{v}+\mathbf{i}_n)\ [d_0^*(\mathbf{B})M^*(\mathbf{v})] \\ & + \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} Y_2^*(\mathbf{v}-\mathbf{i}_n)\ [d_0^*(\mathbf{B})M^*(\mathbf{v})] \\ & - \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M^*(\mathbf{v})\ [b_0(\mathbf{B})^{-1}d_0^*(\mathbf{B})M^*(\mathbf{v}-\mathbf{i}_n)] \\ & - \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M^*(\mathbf{v})\ [b_0(\mathbf{B}^{-1})^{-1}d_0^*(\mathbf{B})M^*(\mathbf{v}+\mathbf{i}_n)] \\ & + \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M^*(\mathbf{v})\ [d_0^*(\mathbf{B})Y_1^*(\mathbf{v}+\mathbf{i}_n)] \\ & + \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M^*(\mathbf{v})\ [d_0^*(\mathbf{B})Y_2^*(\mathbf{v}-\mathbf{i}_n)], \end{aligned} \tag{3.6.28}$$

for $n = 1, \cdots, p$, or

$$\begin{aligned} D_n^* = & \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} [b_0(\mathbf{B}^{-1})^{-1}M^*(\mathbf{v}+\mathbf{i}_n)]\ A^*(\mathbf{v}) \\ & + \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} [b_0(\mathbf{B})^{-1}M^*(\mathbf{v}-\mathbf{i}_n)]\ A^*(\mathbf{v}) \\ & - \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M^*(\mathbf{v})\ [b_0(\mathbf{B})^{-1}A(\mathbf{v}-\mathbf{i}_n)] \\ & - \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M^*(\mathbf{v})\ [b_0(\mathbf{B}^{-1})^{-1}A^*(\mathbf{v}+\mathbf{i}_n)] \\ & + \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M^*(\mathbf{v})\ [b_0(\mathbf{B}^{-1})^{-1}A^*(\mathbf{v}+\mathbf{i}_n)] \\ & + \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M^*(\mathbf{v})\ [b_0(\mathbf{B})^{-1}A^*(\mathbf{v}-\mathbf{i}_n)] \end{aligned} \tag{3.6.29}$$

$$\begin{aligned} = & \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} [b_0(\mathbf{B}^{-1})^{-1}M^*(\mathbf{v}+\mathbf{i}_n)]\ A^*(\mathbf{v}) \\ & + \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} [b_0(\mathbf{B})^{-1}M^*(\mathbf{v}-\mathbf{i}_n)]\ A^*(\mathbf{v}). \end{aligned} \tag{3.6.30}$$

Similarly, for $m = 1, \cdots, q$, we will come up (with opposite signs)

$$\begin{aligned} D_{p+m}^* \equiv & \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M^*(\mathbf{v})\ [a_0(\mathbf{B})^{-1}d_0^*(\mathbf{B})M^*(\mathbf{v}-\mathbf{j}_m)] \\ & + \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M^*(\mathbf{v})\ [a_0(\mathbf{B}^{-1})^{-1}d_0(\mathbf{B})M^*(\mathbf{v}+\mathbf{j}_m)] \end{aligned} \tag{3.6.31}$$

$$\begin{aligned} = & \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M^*(\mathbf{v})\ [a_0(\mathbf{B})^{-1}A^*(\mathbf{v}-\mathbf{j}_m)] \\ & + \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M^*(\mathbf{v})\ [a_0(\mathbf{B}^{-1})^{-1}A^*(\mathbf{v}+\mathbf{j}_m)]. \end{aligned} \tag{3.6.32}$$

Now, both filters $b_0(\mathbf{z})$ and $a_0(\mathbf{z})$ in (3.6.30) and (3.6.32) might not be used to define a causal and invertible process. Similarly, the filters $b_0(\mathbf{z})^{-1}$ and $a_0(\mathbf{z})^{-1}$ might extend over both sides, and we may define

$$b_0(\mathbf{z})^{-1} \equiv \sum_{\mathbf{j}^\tau \in \mathbb{Z}^d} r_{\mathbf{j},0}^{(b)}\, \mathbf{z}^{\mathbf{j}}, \sum_{\mathbf{j}^\tau \in \mathbb{Z}^d} |r_{\mathbf{j},0}^{(b)}| < \infty, \tag{3.6.33}$$

$$a_0(\mathbf{z})^{-1} \equiv \sum_{\mathbf{j}^\tau \in \mathbb{Z}^d} r_{\mathbf{j},0}^{(a)}\, \mathbf{z}^{\mathbf{j}}, \sum_{\mathbf{j}^\tau \in \mathbb{Z}^d} |r_{\mathbf{j},0}^{(a)}| < \infty. \tag{3.6.34}$$

Since $M^*(\mathbf{v})$ and $A^*(\mathbf{v}+\mathbf{j})$ are two uncorrelated random variables for any $\mathbf{j} \neq \mathbf{0}$, in (3.6.30), we will come up with the expectations

$$E\{[b_0(\mathbf{B}^{-1})^{-1}M^*(\mathbf{v}+\mathbf{i}_n)]\ A^*(\mathbf{v})\} = E\{[b_0(\mathbf{B})^{-1}M^*(\mathbf{v}-\mathbf{i}_n)]\ A^*(\mathbf{v})\} = \sigma_W^2 \cdot r_{-\mathbf{i}_n,0}^{(b)}, \tag{3.6.35}$$

for $n = 1, \cdots, p$ and, in (3.6.32), we will come up with

$$E\{M^*(\mathbf{v})\ [a_0(\mathbf{B})^{-1}A^*(\mathbf{v}-\mathbf{j}_m)]\} = E\{M^*(\mathbf{v})\ [a_0(\mathbf{B}^{-1})^{-1}A^*(\mathbf{v}+\mathbf{j}_m)]\} = \sigma_W^2 \cdot r_{-\mathbf{j}_m,0}^{(a)\cdot}, \tag{3.6.36}$$

for $m = 1, \cdots, q$. According (3.6.35) and (3.6.36), unless the ARMA defined by (3.6.1) is causal and invertible, our estimators will be biased. In the end of the proof of Theorem 3.5, we have seen why the estimators we have defined, also risk not to be consistent when the ARMA is not causal or invertible.

Following the same sequel as Whittle (1954), the next step is to find the modified Gaussian likelihood which, if maximized, produces asymptotically unbiased and normal estimators for the parameters of a bilateral ARMA process. For this, we will have to compute the constants c_b and c_a. From (3.6.4), we may write for any $\boldsymbol{\omega}^\tau \in [-\pi, \pi]^d$

$$b(e^{i\boldsymbol{\omega}})\ b(e^{-i\boldsymbol{\omega}}) = c_b\ \varphi(e^{i\boldsymbol{\omega}})\ \varphi(e^{-i\boldsymbol{\omega}}), \tag{3.6.37}$$

or

$$\log\{b(e^{i\boldsymbol{\omega}})\ b(e^{-i\boldsymbol{\omega}})\} = \log c_b + \log\{\varphi(e^{i\boldsymbol{\omega}})\ \varphi(e^{-i\boldsymbol{\omega}})\} \tag{3.6.38}$$

and, consequently,

$$\int_{[-\pi,\pi]^d} \log\{b(e^{i\boldsymbol{\omega}})\ b(e^{i\boldsymbol{\omega}})\} d\boldsymbol{\omega} = (2\pi)^d\ \log c_b + \int_{[-\pi,\pi]^d} \log\{\varphi(e^{i\boldsymbol{\omega}})\ \varphi(e^{-i\boldsymbol{\omega}})\} d\boldsymbol{\omega}. \tag{3.6.39}$$

Then, according to Whittle (1954), for a causal polynomial like $\varphi(\mathbf{z})$, it holds that

$$\int_{[-\pi,\pi]^d} \log\{\varphi(e^{i\boldsymbol{\omega}})\ \varphi(e^{-i\boldsymbol{\omega}})\} d\boldsymbol{\omega} = 0 \tag{3.6.40}$$

and, so,

$$\log c_b = \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} \log\{b(e^{i\omega})\ b(e^{-i\omega})\} d\omega. \tag{3.6.41}$$

In exactly the same way, we may derive that

$$\log c_a = \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} \log\{a(e^{i\omega})\ a(e^{-i\omega})\} d\omega. \tag{3.6.42}$$

Using (3.6.11), (3.6.15) and (3.6.23), we find our estimators by maximizing the modified Gaussian likelihood

$$L^* \propto (\sigma_W^2)^{-N^*/2} \left(\frac{c_b}{c_a}\right)^{N^*/2} \exp\{-\frac{1}{2\sigma_W^2} \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M^*(\mathbf{v},\mathbf{b}) \sum_{\mathbf{j}^\tau \in \mathcal{F}_\mathbf{v}} d_\mathbf{j}^*\ M^*(\mathbf{v}-\mathbf{j},\mathbf{b})\} \tag{3.6.43}$$

or minimizing the natural logarithm

$$l^* = \log \sigma_W^2 + \log\{c_a/c_b\} + \frac{Q^*}{N^*\ \sigma_W^2}, \tag{3.6.44}$$

for all the values $[\mathbf{b}^\tau, \mathbf{a}^\tau]^\tau \in \Theta$ and $\sigma_W^2 > 0$, where $\Theta \subset \mathcal{R}^{p+q}$ is the parameter space and

$$Q^* \equiv \sum_{\mathbf{v}^\tau \in \mathcal{S}^*} M^*(\mathbf{v},\mathbf{b}) \sum_{\mathbf{j}^\tau \in \mathcal{F}_\mathbf{v}} d_\mathbf{j}^*\ M^*(\mathbf{v}-\mathbf{j},\mathbf{b}). \tag{3.6.45}$$

Regarding the parameter space Θ, the assumption that the ARMA process of interest is causal and invertible is not only necessary for the derivation of results in Sections 3.4 and 3.5, as well as in Section 3.3, but also it guarantees the uniqueness of the coefficients that correspond to the second-order properties of the process of interest. In other words, while every ARMA process has a unique auto-covariance function, the same auto-covariance function might be computed from more than one processes. Only for the simplest case of an one-dimensional ARMA(p,q), there are 2^{p+q} ARMA processes that might share exactly the same second-order properties. One of them only is causal and invertible. When $d \geq 2$, the equivalence of different ARMA processes, in terms of their second-order properties, becomes much more complicated as, not only the variety of bilateral representations is huge, but also it is doubtful whether there can be found a finite number of coefficients to express each representation. Thus, our parameter space must make sure that there cannot be two different elements $[\mathbf{b}_1^\tau, \mathbf{a}_1^\tau]^\tau$, $[\mathbf{b}_2^\tau, \mathbf{a}_2^\tau]^\tau \in \Theta$, which generate exactly the same auto-covariance function.

The log-likelihood (3.6.44) closely resembles the quantity suggested by Whittle (1954, p.441) for the case of two-dimensional auto-regressions. We generalize this result for the case of any ARMA(p,q) process on $\mathcal{Z}^d$. Of course, as we have explained before,

the random part Q^* of the likelihood has not been computed via a sum of squares of uncorrelated random variables. This is because, unless the process of interest is a unilateral or bilateral auto-regression of finite order, one has to prewhiten the data in terms of the innovations algorithm, like Yao and Brockwell (2006).

The changes in the proposed Gaussian likelihood, because of the lack of causality or invertibility, might be attributed to the following two reasons. On the one hand, for the moving-average polynomial $a(\mathbf{z})$, we noted in the end of Section 2.4.1 that, although conditional variance matrices of zero-mean Gaussian random variables involve their second-order properties only, still we could only derive the conditional variance matrix like this, if the polynomial $a(\mathbf{z})$ was expressing the process as an invertible moving-average process. For example, for the bilateral moving-average $\{Y_W(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}\}$, defined by

$$Y_W(\mathbf{v}) \equiv a(\mathbf{B})W_1(\mathbf{v}),\ \{W_1(\mathbf{v})\} \sim WN(0,1),$$

the error sequence $\{W_1(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ would not express the process as a unilateral MA, but we would have to find another MA representation of the process, possibly of infinite order. In Section 3.4.3, instead of factorizing the determinant of this conditional variance matrix, we factorized the determinant of its inverse, *i.e.* a quantity proportional to the determinant of the variance matrix of random variables from the bilateral auto-regression

$$a(\mathbf{B})X_W(\mathbf{v}) = X_W(\mathbf{v}) - \sum_{m=1}^{q} a_{\mathbf{j}_m}\ X_W(\mathbf{v}-\mathbf{j}_m) \equiv W_2(\mathbf{v}),\ \{W_2(\mathbf{v})\} \sim WN(0,1).$$

Since the auto-regression $\{X_W(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$ is bilateral, under no means, would the random variable

$$\sum_{m=1}^{q} a_{\mathbf{j}_m}\ X_W(\mathbf{v}-\mathbf{j}_m)$$

be the best linear predictor of $X_W(\mathbf{v})$ based on $X_W(\mathbf{v}-\mathbf{j}), \mathbf{j} > \mathbf{0}$, and would

$$1 = E\{X_W(\mathbf{v}) - \sum_{m=1}^{q} a_{\mathbf{j}_m} X_W(\mathbf{v}-\mathbf{j}_m)\}^2$$

be the prediction variance. In the contrary, the prediction variance would be equal to the variance of the error sequence of the unilateral representations. Moreover, since the bilateral auto-regression cannot necessarily be expressed as a finite unilateral auto-regression, the selection of prediction variances on locations $\mathbf{v}^\tau \in \mathcal{S}^*$, would not necessarily be computed as a product of the same N^* prediction variances, but we could rest that the limit of this product would be equal to this number under (C1)(i).

On the other hand, this last argument is the same as the one used in Section 3.5.3 for the auto-regressive polynomial $b(\mathbf{z})$ and the causality of the ARMA. Indeed, we referred there to a moving-average process $\{Y_H(\mathbf{v}),\ \mathbf{v}^\tau \in \mathcal{Z}^d\}$, which, if bilateral, is defined now by the equation

$$Y_H(\mathbf{v}) = b(\mathbf{B})W_3(\mathbf{v}),\ \{W_3(\mathbf{v})\} \sim WN(0,1).$$

We wrote $|\mathbf{B}|$ to be the Jacobian of transformation between two conditional Gaussian densities, where $\mathbf{B}$ is a variance matrix of random variables from the bilateral moving-average. Since a moving-average always has an AR(∞) representation, using the same arguments as for the polynomial $a(\mathbf{z})$ and the invertibility of the ARMA process, we may claim that writing $|\mathbf{B}|$ as a product of N or N^* prediction variances, on the original or selected locations, respectively, does not guarantee that the determinant will tend to unity, but the limit will be a different number then.

Regarding the bias of the estimators, we may see, for example, from the moving-average equations $m = 1, \cdots, q$, that

$$E\{D^*_{p+m}\} = 2\ N^*\sigma^2_W \cdot r^{(a)}_{-\mathbf{j}_m,0}, \tag{3.6.46}$$

thanks to (3.6.32) and (3.6.36). This implies that

$$E\{\frac{\partial}{\partial_{a_{\mathbf{j}_m}}}\ [-Q^*/N^*\sigma^2_W]\Big|_{\mathbf{b}=\mathbf{b}_0,\mathbf{a}=\mathbf{a}_0}\} = 2\ r^{(a)}_{-\mathbf{j}_m,0},\ m = 1, \cdots, q. \tag{3.6.47}$$

Equation (3.6.47) would reveal the bias of the moving-average estimators, unless the quantity $(\log c_a)$ had been included in the log-likelihood l^*. Indeed, it holds for $m = 1, \cdots, q$, that

$$\begin{aligned}
\partial \log c_a/\partial a_{\mathbf{j}_m} &= \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} \partial \log\{a(e^{i\omega})\ a(e^{-i\omega})\}/\partial a_{\mathbf{j}_m} d\omega \\
&= \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} \partial \log\{a(e^{i\omega})\}/\partial a_{\mathbf{j}_m} d\omega \\
&+ \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} \partial \log\{a(e^{-i\omega})\}/\partial a_{\mathbf{j}_m} d\omega \\
&= \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} \frac{e^{i\omega\mathbf{j}^\tau_m}}{a(e^{i\omega})}\ d\omega + \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} \frac{e^{-i\omega\mathbf{j}^\tau_m}}{a(e^{-i\omega})}\ d\omega
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial}{\partial a_{\mathbf{j}_m}} \log c_a \Big|_{\mathbf{a}=\mathbf{a}_0} &= \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} \frac{e^{i\omega \mathbf{j}_m^\tau}}{a_0(e^{i\omega})}\, d\omega + \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} \frac{e^{-i\omega \mathbf{j}_m^\tau}}{a_0(e^{-i\omega})}\, d\omega \\
&= \frac{1}{(2\pi)^d} \sum_{\mathbf{j}^\tau \in \mathbb{Z}^d} r_{\mathbf{j},0}^{(a)} \int_{[-\pi,\pi]^d} e^{i\omega(\mathbf{j}_m^\tau + \mathbf{j}^\tau)}\, d\omega \\
&+ \frac{1}{(2\pi)^d} \sum_{\mathbf{j}^\tau \in \mathbb{Z}^d} r_{\mathbf{j},0}^{(a)} \int_{[-\pi,\pi]^d} e^{-i\omega(\mathbf{j}_m^\tau + \mathbf{j}^\tau)}\, d\omega \\
&= 2\, r_{-\mathbf{j}_m,0}^{(a)}, \qquad (3.6.48)
\end{aligned}$$

where the last equality follows from the same argument as in Remark 2.3 and proves the asymptotic unbiasedness of the estimators. We trust that we may use similar arguments to Theorem 3.6 to find the asymptotic covariance matrix of the estimators, which is expected to generalize formula (42) of Whittle (1954, p.441).

3.7 Spatio-temporal auto-regressions

Including time as one of the d dimensions in the analysis is a privilege that usually gives meaningful interpretations to causal formulations, as we are going to explain next. Either we choose a parametrization in terms of the second-order properties of the process of interest or we assume that a causal relationship is taking place there, we can always find ways to come up with estimators of the parameters, which have the desired properties. We have tried to establish this so far in Chapter 3, and we will also do it in Chapter 4. When we have observations available from a stationary process with a d-dimensional index, but we do not know what these dimensions represent, the first thing we do is to compute the sample auto-covariances, in order to have some knowledge on the second-order properties of the process. Moreover, if we know that these are spatial dimensions, we still like to proceed the same way, so that we can avoid ordering the dimensions and the locations of each dimension in a nonsensical way. On the other hand, if we do have the time axis in our analysis, we prefer to think that there is a causal formulation, one that is taking place over time. This is a generalization of causality from the one-dimensional time series to the spatio-temporal processes, which use at least two dimensions. Previously, in Sections 3.3-3.5, we discussed different ways of estimating the parameters of auto-regressions, moving-averages and ARMA processes. All the results there referred to causal and invertible ARMA processes. In this section we will use the results of Section 3.3 further for two reasons; the first reason is that it is allowed or it is

even mandatory to resort to causal schemes when a spatio-temporal process is studied, and the second reason is that it is only for causal auto-regressions that Section 3.3 is accurate. Of course, we should not forget that there are also time series, for which contemporaneous associations are very important, such as some weather series; and there are spatial processes that require the assumption of a causal relationship, such as the example given by Whittle (1954, p.434).

We study the specific form of spatio-temporal auto-regression

$$\begin{aligned} X_t(u,v) &= \sum_{i=1}^{p}[\varphi_{i,1}\, X_{t-i}(u+1,v) + \varphi_{i,2}\, X_{t-i}(u,v+1) + \varphi_{i,3}\, X_{t-i}(u,v) \\ &+ \varphi_{i,4}\, X_{t-i}(u,v-1) + \varphi_{i,5}\, X_{t-i}(u-1,v)] + \varepsilon_t(u,v), \\ &\{\varepsilon_t(u,v)\} \sim IID(0,\sigma^2), \end{aligned} \tag{3.7.1}$$

where $t \in Z$ is a time index and $[u,v]^\tau \in \mathcal{Z}^2$ are space indexes. In (3.7.1), we are writing $X_t(u,v)$ as a linear function of the values of the same process, which come from up to p lags from the past, *i.e.* $t-1,\cdots,t-p$. From each one of this p points on the time axis, we are using the values of the process on five different locations. Thus, we may refer to models such as (3.7.1) as five-nearest neighbours models.

We let $\boldsymbol{\varphi}$ to be the vector of autoregressive parameters. We consider the parameter space $\Theta \subset \mathcal{R}^{5p}$, such that for every $\boldsymbol{\varphi} \in \Theta$ a causal auto-regression is defined. Indeed, the concept of a causal formulation for (3.7.1) closely resembles now the notion of a causal time series. Looking at (3.7.1), one can see that its MA(∞) representation runs everywhere over the $\mathcal{Z}^2$ space and backwards in time only.

We collect N observations $\{X_t(u,v),\ t=1-p,\cdots,T,\ u=1,\cdots,N_1,\ v=1,\cdots,N_2\}$, based on which we wish to make inference on the true parameter vector $\boldsymbol{\varphi}_0 \in \Theta$. We select the N^* observations $\{X_t(u,v),\ t=1,\cdots,T,\ u=2,\cdots,N_1-1,\ v=2,\cdots,N_3-1\}$ and consider the vector $\mathbf{Y}^*$ of the selected observations with locations in the ascending order and the vector, say, $\boldsymbol{\varepsilon}^*$ of the errors in the same order, as described in Section 3.3.2. We write the linear model

$$\mathbf{Y}^* = \mathbf{X}^*\boldsymbol{\varphi}_0 + \boldsymbol{\varepsilon}^*, \tag{3.7.2}$$

where $\mathbf{X}^*$ is a $N^* \times 5p$ matrix of available observations from the sample as it has also been described in Section 3.3.2. For the least squares estimator

$$\boldsymbol{\varphi}^* = (\mathbf{X}^{*\tau}\mathbf{X}^*)^{-1}(\mathbf{X}^{*\tau}\mathbf{Y}^*)$$

of the parameter vector φ_0 and for the variance matrix

$$\mathbf{W}(\varphi_0) = \frac{1}{\sigma^2}\text{Var}\{[X_{-1}(1,0), \cdots, X_{-p}(-1,0)]^\tau\},$$

it holds that

$$N^{*1/2}[\varphi^* - \varphi_0] \xrightarrow{D} N(\mathbf{0}, \mathbf{W}(\varphi_0)^{-1}) \tag{3.7.3}$$

and

$$N^* \, [\varphi^* - \varphi_0]^\tau \mathbf{W}(\varphi_0)[\varphi^* - \varphi_0] \xrightarrow{D} \chi^2_{5p}, \tag{3.7.4}$$

as $\min\{T, N_1, N_2\} \to \infty$ and (C1) holds, according to Theorem 3.1.

3.7.1 Order selection

For pure autoregressive models of the form (3.7.1), we may use similar results to the ones for time series to select the unique number p, which specifies the order for our model. We select the model of the form (3.7.1) with this order p, such that the Final Prediction Error (FPE) is minimized.

First for fixed p, we record the realization
$\{X_t(u,v),\ t = 1-p, \cdots, T,\ u = 1, \cdots, N_1,\ v = 1, \cdots, N_2\}$ of the process (3.7.1), with $T > 0,\ N_1, N_2 > 2$, from which we obtain our estimators φ^*. For any $[t,u,v]^\tau \in \mathcal{Z}$, we also record independently the realization $\{Y_{t-i}(u-k, v-l),\ i = 1, \cdots, p,\ k,l = 0, \pm 1\}$ from the same process. We define the linear predictor of $Y_t(u,v)$ based on the available observations as

$$\begin{aligned} \widetilde{Y}_t(u,v) \ \equiv \ & \sum_{i=1}^{p} [\varphi^*_{i,1}\, Y_{t-i}(u+1,v) + \varphi^*_{i,2}\, Y_{t-i}(u,v+1) + \varphi^*_{i,3}\, Y_{t-i}(u,v) \\ + \ & \varphi^*_{i,4}\, Y_{t-i}(u,v-1) + \varphi^*_{i,5}\, Y_{t-i}(u-1,v)]. \end{aligned}$$

The one step prediction mean squared error is

$$\begin{aligned} FPE = E\{(Y_t(u,v) - \widetilde{Y}_t(u,v))^2\} \ &= \ \sigma^2 + \frac{\sigma^2}{N^*}\, E\{N^*[\varphi^* - \varphi_0]^\tau \mathbf{W}(\varphi_0)[\varphi^* - \varphi_0]\} \\ &= \ \sigma^2 \left(1 + \frac{5p}{N^*}\right), \end{aligned} \tag{3.7.5}$$

since the two realizations used for prediction and for estimation are independent. Using standard results of the general linear model, we may also show that if model (3.7.1) is correct, then

$$\frac{(\mathbf{Y}^* - \mathbf{X}^*\varphi^*)^\tau (\mathbf{Y}^* - \mathbf{X}^*\varphi^*)}{\sigma^2} \xrightarrow{D} \chi^2_{N^*-5p}. \tag{3.7.6}$$

The random vector $\mathbf{Y}^*$ still refers to the original observations used for estimation, but on the selected locations $\{t = 1, \cdots, T,\ u = 2, \cdots, N_1 - 1,\ v = 2, \cdots, N_2 - 1\}$, as this was described in the end of Section 3.7. We may estimate the Final Prediction Error as

$$\widehat{FPE} = \widehat{\sigma^2}\left(1 + \frac{5p}{N^*}\right), \tag{3.7.7}$$

by plugging-in the unbiased estimator of the error variance, $\widehat{\sigma^2} = (\mathbf{Y}^* - \mathbf{X}^*\boldsymbol{\varphi}^*)^\tau(\mathbf{Y}^* - \mathbf{X}^*\boldsymbol{\varphi}^*)/(N^* - 5p)$. We often call $\widehat{\sigma^2}$ the Mean Square Error (MSE) of the model of interest.

3.7.2 Tests for linear models

Suppose that we have a set of competing spatio-temporal auto-regressions of the form (3.7.1) and we want to determine which one is more appropriate to fit our data, taking into account the two main statistical requirements, parsimony and goodness of fit. The goodness of fit of any linear model to the observed data is encapsulated by its deviance.

We want to test

$$\begin{aligned} H_0 &: \boldsymbol{\varphi} \in \Theta_0 \\ H_1 &: \boldsymbol{\varphi} \in \Theta_1, \end{aligned}$$

where $\Theta_0,\ \Theta_1 \subset \Theta$ and $\Theta_0 \cap \Theta_1 = \emptyset$. The model under H_0 assumes that

$$\mathbf{Y}^* = \mathbf{X}_0^*\,\boldsymbol{\varphi} + \boldsymbol{\varepsilon}^*$$

with df_0 unknown parameters and it is, of course, nested in the model for which $\boldsymbol{\varphi} \in \Theta_0 \cup \Theta_1$, which can be written in the form

$$\mathbf{Y}^* = \mathbf{X}_1^*\,\boldsymbol{\varphi} + \boldsymbol{\varepsilon}^*$$

and allows for df_1 parameters. We define the deviances

$$\begin{aligned} D_0 &\equiv (\mathbf{Y}^* - \mathbf{X}_0^*\,\boldsymbol{\varphi}_0^*)^\tau(\mathbf{Y}^* - \mathbf{X}_0^*\,\boldsymbol{\varphi}_0^*) \\ D_1 &\equiv (\mathbf{Y}^* - \mathbf{X}_1^*\,\boldsymbol{\varphi}_1^*)^\tau(\mathbf{Y}^* - \mathbf{X}_1^*\,\boldsymbol{\varphi}_1^*), \end{aligned}$$

where $\mathbf{Y}^*$ is the maximal selection of random variables on N^* locations both for H_0 and H_1,

$$\boldsymbol{\varphi}_0^* = \arg\min_{\boldsymbol{\varphi} \in \Theta_0} [(\mathbf{Y}^* - \mathbf{X}_0^*\,\boldsymbol{\varphi})^\tau(\mathbf{Y}^* - \mathbf{X}_0^*\,\boldsymbol{\varphi})]$$

and, similarly,

$$\varphi_1^* = \arg \min_{\varphi \in \Theta_0 \cup \Theta_1} \left[(\mathbf{Y}^* - \mathbf{X}_1^* \, \varphi)^\tau (\mathbf{Y}^* - \mathbf{X}_1^* \, \varphi) \right].$$

Under H_0, it holds for both deviances that

$$\begin{aligned} \frac{D_0}{\sigma^2} &\xrightarrow{D} \mathcal{X}^2_{N^*-df_0} \\ \frac{D_1}{\sigma^2} &\xrightarrow{D} \mathcal{X}^2_{N^*-df_1} \end{aligned}$$

and the random variables $D_0 - D_1$, D_1 are asymptotically independent. Thus,

$$F = \frac{D_0 - D_1}{D_1} \xrightarrow{D} F_{df_1 - df_0, N^* - df_1}. \tag{3.7.8}$$

For the model (3.7.1) with fixed order p, some examples of null hypotheses of interest are now in order. In all these examples, we consider that $\Theta_1 = \Theta - \Theta_0$.

Tests for white noise

We test

$$H_0 : \varphi = \mathbf{0},$$

with $D_0 = \mathbf{Y}^{*\tau}\mathbf{Y}^*$ and the statistic (3.7.8) is $F \xrightarrow{D} F_{5p, N^*-5p}$.

Tests for equal coefficients over time

We test

$$H_0 : \varphi_{(1,j)} = \varphi_{(2,j)} = \cdots = \varphi_{(p,j)} = \varphi_j, \; j = 1, \cdots, 5,$$

where the assumed model, under H_0, is

$$\begin{aligned} X_t(u,v) &= \varphi_1 \sum_{i=1}^{p} X_{t-i}(u+1,v) + \varphi_2 \sum_{i=1}^{p} X_{t-i}(u,v+1) + \varphi_3 \sum_{i=1}^{p} X_t(u,v) \\ &+ \varphi_4 \sum_{i=1}^{p} X_{t-i}(u,v-1) + \varphi_5 \sum_{i=1}^{p} X_{t-i}(u-1,v) + \varepsilon_t(u,v), \end{aligned}$$

where $\{\varepsilon_t(u,v)\} \sim IID(0, \sigma^2)$ and the statistic (3.7.8) is $F \xrightarrow{D} F_{5(p-1), N^*-5p}$.

Tests for symmetry in space

We test

$$H_0 : \varphi_{(i,1)} = \varphi_{(i,5)}, \; \varphi_{(i,2)} = \varphi_{(i,4)}, \; i = 1, \cdots, p,$$

where the assumed model under H_0 is

$$\begin{aligned} X_t(u,v) &= \sum_{i=1}^{p}[\varphi_{(i,1)}\ (X_{t-i}(u+1,v)+X_{t-i}(u-1,v)) \\ &+ \varphi_{(i,2)}\ (X_{t-i}(u,v+1)+X_{t-i}(u,v-1)) \\ &+ \varphi_{(i,3)}\ X_{t-i}(u,v)] + \varepsilon_t(u,v),\ \{\varepsilon_t(u,v)\} \sim IID(0,\sigma^2) \end{aligned}$$

and the statistic (3.7.8) is $F \xrightarrow{D} F_{2p,N^*-5p}$.

3.7.3 An application on data recorded regularly in time and space

The data we are using in this section, have been provided by the National Centers for Environmental Prediction- National Center for Atmospheric Research (NCEP - NCAR). A small area of sea level pressure, measured in units of Pascal, in the North Sea has been chosen. The longitudes are from 20 degrees West to 20 degrees East and the latitudes are from 50 to 60 degrees North. The grid points are of size 2.5×2.5 degrees^2 with total number of $17 \times 5 = 85$ spatial locations. The time period is winter 2001-2002 with 100 daily observations starting from December 1st 2001.

First, we write our observations $\{X_t(u,v),\ t = 1,\cdots,100,\ u = 1,\cdots,5,\ v = 1,\cdots,17\}$. The 5 labels refer to the dimension 'South-North', starting from 50 degrees North for the first label, 52.5 degrees for the second label, up to 60 degrees North for the last label. On the other hand, for the dimension 'West-East', we start with the first label for 20 degrees West, the second label for 17.5 degrees West and we go on the line transect up to the last label, *i.e.* -20 degrees West. We are interested in fitting an auto-regression of the type (3.7.1) in the centralized series; the mean has been estimated from the data as $\bar{X} = 101,131.9$. We try the values $p = 1,\cdots,10$. While we have started the analysis with $N = 100 \times 5 \times 17 = 8,500$ observations available, we come up with $N^* = 90 \times 3 \times 15 = 4,050$ observations, instead. The first 10 recordings in time as well as the extreme locations 50 and 60 degrees North and 20 and -20 degrees West have been omitted.

The fact that we have to reduce our sample size from $N = 8,500$ to $N^* = 4,050$, reveals a great weakness of all the methods proposed in this chapter. We have used the locations on a selected set $\mathcal{S}^*$ of cardinality N^*, rather than the original set $\mathcal{S}$ of cardinality N. We know that as $N \to \infty$, it holds that $N^*/N \to 1$. Though our original sample size is very big, we can see that more than half of the locations have

been rejected, which brings to mind the following question. How large does the sample size have to be in practise, in order to exclude a relatively small number of locations? The answer is closely related to the number of dimensions d. The more the dimensions we have available, the slower we obtain this small percentage of excluded locations. In our example, we have $d = 3$ dimensions, and this should be seen as a reason why the percentage is close to 50%.

What is it possible to do, in order to make the most of our observations? To answer this question, we need to make a choice. Either we prefer to use estimators, for which there are not any theoretical results available and we use all the observations available, or we stick to our established results and we pay the price of excluding many locations. As an example, we may imagine the classical case of time series and a causal AR(10), which has produced 100 consecutive observations. If we find the least squares estimators or conditional Gaussian likelihood estimators of the parameters, we need to exclude 10 timings and fully use 90 observations. Alternatively, we may find the exact Gaussian likelihood estimators using all the 100 observations. That would usually imply computing the best linear predictor coefficients of a random variable from the process, based on the random variables from its previous $1, 2, \cdots, 9$ lags; the innovations algorithm or the Kalman filter could easily derive those coefficients. Nevertheless, for the simple case that $d = 1$, we know that both conditional and exact Gaussian likelihood estimators possess similar statistical properties, as the edge-effect is hidden then. When $d \geq 2$, the edge-effect causes a dilemma. We have chosen to follow the route of the estimators with the established statistical properties, which are going to be necessary, in order to perform statistical tests and make decisions.

Since we have tried to fit an auto-regression to the centralized process, it means that we expect our process $\{X_t(u, v),\ t, u, v \in \mathcal{Z}\}$ to be stationary on $\mathcal{Z}^3$. For example, that would also imply that if we fix any $t, u, v \in \mathcal{Z}$, then the processes on $[u, v]^\tau, [t, v]^\tau, [t, u]^\tau \in \mathcal{Z}^2$, respectively, are also stationary. Similarly, if we fix any $[u, v]^\tau, [t, v]^\tau, [t, u]^\tau \in \mathcal{Z}^2$, then the processes on $t, u, v \in \mathcal{Z}$, respectively, must be stationary. In Figure 3.1, we can see a realization of 100 consecutive observations of the process, which takes place over $t \in \mathcal{Z}$, when we have fixed the location to be 60 degrees North and 0 degrees West. It does not look that there is a systematic trend there. Maybe if the observations covered the whole of the year, we would see clearly bigger values of sea pressure during the winter compared to these of the summer. The observations are moving upside down, but it does

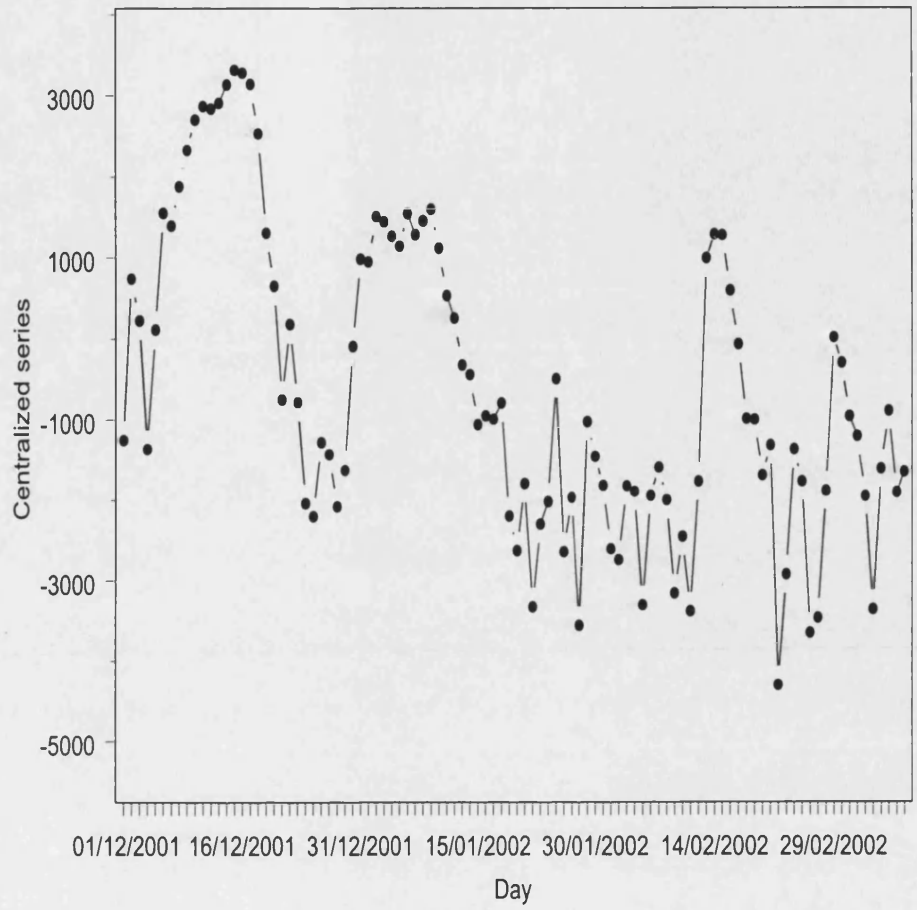


Figure 3.1: The daily centralized series on the location 60 degrees North and 0 degrees West over the period 01/12/2001 - 08/03/2002.

not look that there are cycles, for which it takes a specific number of days to complete them. A more analytical study is required here, in order to make sure that the trend and periodical components, if any, have been extracted from the original series and the remaining series is stationary indeed. We have treated the series $\{X_t(u,v),\ t,u,v \in \mathcal{Z}\}$ as a stationary series, as it has not been a purpose of this chapter to deal with non-stationary processes on $\mathcal{Z}^d$.

We have selected the order of the model according to the FPE and the BIC. As we

Table 3.2: Estimated Final Prediction Error for the five-nearest neighbours model of order p.

Order p	Mean Square Error (MSE)	Estimated FPE = MSE $\left(1+\frac{5p}{N^*}\right)$	% change in the FPE
1	686,608.917	687,456.582	
2	596,084.673	597,556.487	−13.077%
3	584,426.819	586,591.363	−1.835%
4	577,933.248	580,787.239	−0.989%
5	567,124.639	570,625.408	−1.750%
6	561,137.809	565,294.385	−0.934%
7	541,642.671	546,323.534	−3.356%
8	539,802.990	545,134.378	−0.218%
9	531.354.104	537,258.038	−1.445%
10	524,440.621	530,915.197	−1.181%

explained in Section 3.7.1, the FPE can be used for pure auto-regressive models like the ones of interest. We have also extended the definition of BIC for the case of our models. Nevertheless, we have not searched for the properties of the estimated FPE or BIC, as the number of observations increases towards at least one of the three directions. We do trust, though, that since we deal with finite auto-regressions, even when the dimensionality reaches the number $d = 3$, there should not be many differences from the simple case when $d = 1$. What would be interesting to be investigated further, would be what happens to the FPE or BIC as a function of p and as one of the spatial dimensions only increases to infinity. We do not have such results available from time series.

In Table 3.2, we can see that as we increase the order p of the model, the estimated Final Prediction Error keeps becoming smaller and smaller. This is because of the large sample size $N^* = 4,050$, which makes the effect of adding 5 parameters in the model every time unimportant. Still, if we look at the relative decrease in the Final Prediction Error, it looks like this is not bigger than 2% for $p \geq 8$. This means that all orders $p \geq 7$ give estimated prediction errors that are really close.

For the selection of order p, we may also compute the BIC, as this is given by the formula (9.3.5) of Brockwell and Davis (1991, p.304) for the case of causal, invertible

and one-dimensional ARMA(p,q) models. There, we replace $(p+q)$ by the number of parameters of our model, *i.e.* $5p$ and write

$$\begin{aligned} BIC &\equiv 5p\ \ln[N^*MSE/(N^*-5p)] + N^*(1+\ln\sqrt{2\pi}) \\ &+ 5p\ \ln[(\sum_{[t,u,v]^\tau\in\mathcal{S}^*} X_t(u,v)^2 - N^*MSE)/5p]. \end{aligned} \tag{3.7.9}$$

We summarize the results in the next table. Similar conclusions might be made from

Table 3.3: BIC for the five-nearest neighbours model of order p.

Order p	BIC	Order p	BIC	Order p	BIC	Order p	BIC
1	$6.22\cdot10^4$	2	$6.17\cdot10^4$	3	$6.17\cdot10^4$	4	$6.17\cdot10^4$
5	$6.16\cdot10^4$	6	$6.16\cdot10^4$	7	$6.15\cdot10^4$	8	$6.15\cdot10^4$
9	$6.15\cdot10^4$	10	$6.15\cdot10^4$				

Tables 3.2 and 3.3, as it seems that the BIC has been computed the same for all orders $p \geq 7$. Thus, we select the model of order $p=7$ and we write

$$\begin{aligned} X_t(u,v)-\bar{X} &= \sum_{i=1}^{7}\{\varphi_i\ [X_{t-i}(u,v)-\bar{X}] \\ &+ \varphi_{i,S}\ [X_{t-i}(u-1,v)-\bar{X}] + \varphi_{i,N}\ [X_{t-i}(u+1,v)-\bar{X}] \\ &+ \varphi_{i,W}\ [X_{t-i}(u,v-1)-\bar{X}] + \varphi_{i,E}\ [X_{t-i}(u,v+1)-\bar{X}] \\ &+ \varepsilon_t(u,v), \quad \{\varepsilon_t(u,v)\} \sim IID(0,\sigma^2), \end{aligned} \tag{3.7.10}$$

where we use 'S','N','W','E' for the dependence from the north, south, west and east, respectively.

When $p=7$, we estimate 35 parameters altogether and come up with a significant F-statistic, where $F \sim F_{35,4015}$ tests the goodness of fit of the model. We have estimated the parameters, according to Table 3.4. We can see immediately that for $i=2,5,6$, there seems to be one estimate that is very small, and it would not be a surprise if this was insignificant. For $i=2$ and $i=6$, this expresses the dependence from the north neighbour, while for $i=5$ it expresses the dependence from the south neighbour. Both these neighbours refer to the same dimension of space. Indeed, it is not surprising that these estimates give insignificant results for the corresponding parameters, as shown in Table 3.5.

Table 3.5 reveals that a new model should be considered. The usual tactic is to exclude the independent variables that relate to the insignificant coefficients, but we

Table 3.4: Estimates of the parameters $\varphi_i, \varphi_{i,S}, \varphi_{i,N}, \varphi_{i,W}, \varphi_{i,E}$, $i = 1, \cdots, 7$, for the five-nearest neighbours model of order 7.

i	φ_i^*	$\varphi_{i,S}^*$	$\varphi_{i,N}^*$	$\varphi_{i,W}^*$	$\varphi_{i,E}^*$
1	-5.959	0.212	0.750	2.023	4.003
2	4.928	0.491	$1.216 \cdot 10^{-2}$	-2.176	-3.287
3	-1.321	-0.140	0.109	0.843	0.529
4	2.478	0.271	0.371	-1.265	-1.771
5	-0.692	$-1.5 \cdot 10^{-2}$	-0352	0.700	0.340
6	0.895	0.118	$3.061 \cdot 10^{-2}$	-0.676	-0.456
7	-1.123	-0.734	-0.589	0.913	1.471

Table 3.5: Insignificant results for the five-nearest neighbours model of order 7.

Estimated parameter	Observed significance level (p-value)
$\varphi_{2,N}^* = 1.216 \cdot 10^{-2}$	86.6%
$\varphi_{3,S}^* = -0.140$	10.3%
$\varphi_{3,N}^* = 0.109$	12.8%
$\varphi_5^* = -0.692$	17.5%
$\varphi_{5,S}^* = -1.50 \cdot 10^{-2}$	86.2%
$\varphi_{5,E}^* = 0.340$	19.3%
$\varphi_6^* = 0.895$	7.5%
$\varphi_{6,S}^* = 0.118$	16.6%
$\varphi_{6,N}^* = 3.061 \cdot 10^{-2}$	67.2%

prefer to take into account that we are analyzing a spatio-temporal dataset. Thus, we

write the model

$$\begin{aligned} X_t(u,v) - \bar{X} &= \sum_{i=1}^{4}\{\varphi_i\,[X_{t-i}(u,v) - \bar{X}]\} + \varphi_7\,[X_{t-7}(u,v) - \bar{X}] \\ &+ \sum_{i=1,4,7}\{\varphi_{i,S}\,[X_{t-i}(u-1,v) - \bar{X}] + \varphi_{i,N}\,[X_{t-i}(u+1,v) - \bar{X}]\} \\ &+ \sum_{i=2,5}\{\varphi_{i,SN}\,([X_{t-i}(u-1,v) - \bar{X}] + [X_{t-i}(u+1,v) - \bar{X}])\} \\ &+ \sum_{\substack{i=1,\\ i\neq 5}}^{7}\{\varphi_{i,W}\,[X_{t-i}(u,v-1) - \bar{X}] + \varphi_{i,E}\,[X_{t-i}(u,v+1) - \bar{X}]\} \\ &+ \varphi_{5,WE}\,([X_{t-5}(u,v-1) - \bar{X}] + [X_{t-5}(u,v+1) - \bar{X}]) \\ &+ \varepsilon_t(u,v), \quad \{\varepsilon_t(u,v)\} \sim IID(0,\sigma^2), \end{aligned} \tag{3.7.11}$$

which, for the insignificant results as indicated by Table 3.5 and for each $i = 1, \cdots, 7$, either excludes both the parameters that refer to the two sides of the same spatial dimension or uses one parameter for both the sides together. This time, the results for only two parameters were insignificant, *i.e.* $\varphi^*_{6,W} = -2.90 \cdot 10^{-2}$, $\varphi^*_{6,E} = 1.849 \cdot 10^{-2}$, with observed significance levels equal to 53.2% and 70.9%, respectively. Indeed, for the updated model

$$\begin{aligned} X_t(u,v) - \bar{X} &= \sum_{i=1}^{4}\{\varphi_i\,[X_{t-i}(u,v) - \bar{X}]\} + \varphi_7\,[X_{t-7}(u,v) - \bar{X}] \\ &+ \sum_{i=1,4,7}\{\varphi_{i,S}\,[X_{t-i}(u-1,v) - \bar{X}] + \varphi_{i,N}\,[X_{t-i}(u+1,v) - \bar{X}]\} \\ &+ \sum_{i=2,5}\{\varphi_{i,SN}\,([X_{t-i}(u-1,v) - \bar{X}] + [X_{t-i}(u+1,v) - \bar{X}])\} \\ &+ \sum_{\substack{i=1,\\ i\neq 5,6}}^{7}\{\varphi_{i,W}\,[X_{t-i}(u,v-1) - \bar{X}] + \varphi_{i,E}\,[X_{t-i}(u,v+1) - \bar{X}]\} \\ &+ \varphi_{5,WE}\,([X_{t-5}(u,v-1) - \bar{X}] + [X_{t-5}(u,v+1) - \bar{X}]) \\ &+ \varepsilon_t(u,v), \quad \{\varepsilon_t(u,v)\} \sim IID(0,\sigma^2), \end{aligned} \tag{3.7.12}$$

all the coefficients were estimated to be significant. The estimates of all the parameters that could be taken as different than 0 are given in Table 3.6. It is interesting to see there the coordination of the estimated coefficients for the two sides of the same spatial dimension, in terms of both the absolute values and the signs. For $i = 1$ and $i = 4$, it is true that $\varphi^*_{i,S}$ and $\varphi^*_{i,N}$ are both positive and very close to each other, while $\varphi^*_{7,S}$ and $\varphi^*_{7,N}$ are negative and close to each other. The same conclusion can be made for the West and South neighbours. This might imply that the model distributes equally

Table 3.6: Estimates of parameters for the model (3.7.12).

$\varphi_1^* = -5.72$	$\varphi_{1,S}^* = 0.481,\ \varphi_{1,N}^* = 0.761$	$\varphi_{1,W}^* = 1.757,\ \varphi_{1,E}^* = 3.695$
$\varphi_2^* = 5.441$	$\varphi_{2,SN}^* = 0.127$	$\varphi_{2,W}^* = -2.352,\ \varphi_{2,E}^* = -3.440$
$\varphi_3^* = -1.788$		$\varphi_{3,W}^* = 1.232,\ \varphi_{3,E}^* = 0.682$
$\varphi_4^* = 2.484$	$\varphi_{4,S}^* = 0.426,\ \varphi_{4,N}^* = 0.340$	$\varphi_{4,W}^* = -1.405,\ \varphi_{4,E}^* = -1.832$
	$\varphi_{5,SN}^* = -0.243$	$\varphi_{5,WE}^* = 0.207$
$\varphi_7^* = -1.068$	$\varphi_{7,S}^* = -0.636,\ \varphi_{7,N}^* = -0.660$	$\varphi_{7,W}^* = 0.879,\ \varphi_{7,E}^* = 1.395$

the dependence within each spatial dimension to its two sides and this makes sense. If a reduction of the number of parameters of the model should be achieved, we could put these coefficients together.

But before we consider to reduce the parameters of the model even more, we should see whether it is really worth it to move from the original nearest five-neighbours model of order 7 to (3.7.12). For the first model, we have computed the sum of squares of the error as

$$SSE_o = 2.17 \cdot 10^9 \tag{3.7.13}$$

and the degrees of freedom of the error are

$$df_o = 4015. \tag{3.7.14}$$

Similarly, for the model (3.7.12)

$$SSE_f = 2.26 \cdot 10^9 \tag{3.7.15}$$

and

$$df_f = 4026. \tag{3.7.16}$$

Then for the value

$$\frac{(2.26 \cdot 10^9 - 2.17 \cdot 10^9)/(4026 - 4015)}{2.17 \cdot 10^9/4015} = 15.13825,$$

it holds that

$$P(F \geq 15.13825) < 0.05\%,$$

where $F \sim F_{11,4015}$. This means that, although we proceeded by eliminating the variables with insignificant coefficients only, still those variables altogether were significantly

contributing in interpreting the variability of the data. The original model is preferred than model (3.7.12), which has 11 parameters less.

The result is mainly because of the big df_0 and the huge number of observations we are dealing with here. This should allow for a model with many parameters, otherwise most of the degrees freedom will be left to estimate the variance of the model. If more parameters are to be included in the model, more neighbouring values could be used to offer information on the value of interest. Thus, we may extend our results to spatio-temporal auto-regressions with more than five-neighbours.

Finally, the elimination of coefficients from the original model (3.7.10), mostly concerned the parameters of the dimension 'South-North'. This was very apparent in the first elimination, according to Table 3.5. Even though the original model was tested to be overall more significant than the other model and the coefficients did not have to be excluded, we could see that the sea level pressure in the North Sea is probably moving in the 'West-East' dimension, rather than in the 'South-North' dimension. Later in Section 4.7, we will have the opportunity to analyze a part of the same dataset and to verify how the dimension 'West-East' dominates over the dimension 'South-North', indeed, regarding the sea level pressure.

Chapter 4

Statistical inference for spatial auto-linear schemes

4.1 Introduction

In Chapter 3, we were concerned with the simultaneous equation ARMA models defined on any positive integer d number of dimensions. The use of causal models was of primary importance and naturally justified when including the time axis as one of the dimensions. Therefore, when applying these methods to spatial processes, one would have to artificially introduce a unilateral order which, typically, lacks of any practical meaning. To avoid this, an alternative broad strategy for the analysis of spatial models was introduced by Besag (1974). The model is now defined in terms of a conditional distribution of the observation on one location given the values on all other locations. Thus, the way we approach the problem of estimation of the parameters of a (weakly) stationary process on $\mathcal{Z}^d$ in this chapter, is based on a parametrization used, different than the one of Chapter 3; the motivation for that stems from the next step after estimation, *i.e.* prediction. In problems of spatial prediction, we usually have available information about locations all around the location of interest. Similarly, in time series we often have to deal with the problem of missing a value in a set of data from a (weakly) stationary process, for which we have available values not only from its past, but also from its future. The suggestions of this chapter might also be seen as the basic ingredient that can be used for the remedy of this problem, which is known as smoothing.

We consider a (weakly) stationary process $\{X(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ and without loss of

generality, we denote by $\mathbf{s}^\tau \in \mathcal{Z}^d$ the spatial index vector of dimensions d. Below we consider some models, for which the conditional distribution on each site is a function of its finite number of neighbour sites and we call them spatial auto-schemes. We examine the cases where the dependence structure is reflected in the conditional expectation, which has a linear form. The conditional variance is assumed to be a constant. When the random variables are Gaussian, Besag (1974) referred to these as spatial auto-normal schemes. We will use the term auto-linear schemes, as we will not be dealing with Gaussian random variables only.

For observations from an auto-scheme, two methods of estimation have been proposed by Besag (1974); the coding and the pseudo-likelihood techniques. The coding technique produces estimators of reduced efficiency, as they do not make direct use of all the observations at the same time. As a result, there is an issue of finding an optimal coding for the specific scheme of interest. Examples of this for auto-normal schemes can be found in Besag and Moran (1975), Besag (1977) and Besag (1974) followed by a comment by Clifford. The pseudo-likelihood method, on the other hand, simplifies, for auto-linear schemes, to minimizing a sum of squares of random variables that are correlated and ignores their dependence. This has an effect on the variance matrices of the estimators. Minimizing a sum of squares of random variables is usually preferred for the estimation of the parameters of a simultaneous auto-regression rather than an auto-linear scheme, as we are going to see in Section 4.5.3. The important contribution to the study of bilateral auto-regressions made by Whittle (1954), though, has shown how important the existence of a causal representation is, as minimizing the sum of squares of errors for a bilateral auto-regression cannot produce consistent estimators.

For observations from an auto-linear scheme, we propose a method of moments estimation after we point out the difficulties of reaching the estimators derived by maximizing a conditional likelihood, a modified form of which has been proposed by Besag (1974). Besag suggested the specific form of the conditional likelihood but did not manage to obtain the properties of the estimators because of two obstacles; the computation of the determinant involved and the fact that the quadratic form is a linear function of the parameters. We show that the proposed method of moments estimators are consistent and asymptotically normal, with a variance matrix of an easy form and we derive tests for the unknown coefficients of interest. For observations lying on a rectangle of a two-dimensional spatial process, we conclude and apply our results, by estimating the

coefficients and fitting the first-order model.

4.2 Best linear predictors

For a (weakly) stationary process $\{X(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ with an absolutely summable autocovariance function, if we manage to write its spectral density in the form

$$g_X(\boldsymbol{\omega}) = \frac{1}{(2\pi)^d}\ \frac{\sigma^2}{\theta(e^{i\boldsymbol{\omega}})\cdot\theta(e^{-i\boldsymbol{\omega}})},\ \boldsymbol{\omega}^\tau \in [-\pi,\pi]^d, \tag{4.2.1}$$

with polynomial

$$\theta(\mathbf{z}) \equiv 1 - \sum_{\mathbf{j}>\mathbf{0}} \theta_{\mathbf{j}}\ \mathbf{z}^{\mathbf{j}},\ \sum_{\mathbf{j}>\mathbf{0}} |\theta_{\mathbf{j}}| < \infty, \tag{4.2.2}$$

and if we define the random variables

$$\varepsilon(\mathbf{s}) \equiv \theta(\mathbf{B})X(\mathbf{s}) = X(\mathbf{s}) - \sum_{\mathbf{j}>\mathbf{0}} \theta_{\mathbf{j}}\ X(\mathbf{s}-\mathbf{j}),\ \mathbf{s}^\tau \in \mathcal{Z}^d, \tag{4.2.3}$$

then it holds for the (weakly) stationary process $\{\varepsilon(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ that

$$\{\varepsilon(\mathbf{s})\} \sim WN(0,\sigma^2). \tag{4.2.4}$$

This is because we have applied a linear filter and, according to Theorem 4.4.1 of Brockwell and Davis (1991, p.122), we can immediately write its spectral density as

$$g_\varepsilon(\boldsymbol{\omega}) = \frac{\sigma^2}{(2\pi)^d},\ \boldsymbol{\omega}^\tau \in [-\pi,\pi]^d. \tag{4.2.5}$$

Similarly, for

$$u(\mathbf{s}) \equiv \theta(\mathbf{B}^{-1})X(\mathbf{s}) = X(\mathbf{s}) - \sum_{\mathbf{j}>\mathbf{0}} \theta_{\mathbf{j}}\ X(\mathbf{s}+\mathbf{j}), \tag{4.2.6}$$

it holds that

$$\{u(\mathbf{s})\} \sim WN(0,\sigma^2). \tag{4.2.7}$$

The representation (4.2.3) with (4.2.4) does not directly answer the question which one is the best linear predictor of $X(\mathbf{s})$ based on the values $X(\mathbf{s}-\mathbf{j}), \mathbf{j} > \mathbf{0}$, or what we would call its 'past' values. Since the process is stationary, it holds that

$$\{\theta(\mathbf{z})\theta(\mathbf{z}^{-1})\}^{-1} \propto [1 + \sum_{\mathbf{j}^\tau \in \mathcal{Z}^d - \{\mathbf{0}^\tau\}} \Theta_{\mathbf{j}}\ \mathbf{z}^{\mathbf{j}}][1 + \sum_{\mathbf{j}^\tau \in \mathcal{Z}^d - \{\mathbf{0}^\tau\}} \Theta_{\mathbf{j}}\ \mathbf{z}^{-\mathbf{j}}] \tag{4.2.8}$$

with squared summable coefficients

$$\sum_{\mathbf{j}^\tau \in \mathcal{Z}^d - \{\mathbf{0}^\tau\}} \Theta_{\mathbf{j}}^2 < \infty, \tag{4.2.9}$$

according to the Wold decomposition. Moreover, if it holds

$$\theta(\mathbf{z})^{-1} \equiv 1 + \sum_{\mathbf{j}>\mathbf{0}} \Theta_{\mathbf{j}}\, \mathbf{z}^{\mathbf{j}},\ \sum_{\mathbf{j}>\mathbf{0}} |\Theta_{\mathbf{j}}| < \infty, \tag{4.2.10}$$

we may write

$$X(\mathbf{s}) = \varepsilon(\mathbf{s}) + \sum_{\mathbf{j}>\mathbf{0}} \Theta_{\mathbf{j}}\, \varepsilon(\mathbf{s}-\mathbf{j}), \tag{4.2.11}$$

which implies that $X(\mathbf{s})$ is a linear function of $\varepsilon(\mathbf{s}-\mathbf{j}), \mathbf{j} \geq \mathbf{0}$, and it is uncorrelated with any 'future' value $\varepsilon(\mathbf{s}+\mathbf{j}), \mathbf{j} > \mathbf{0}$. As a result,

$$E\{\varepsilon(\mathbf{s})\ X(\mathbf{s}-\mathbf{i})\} = 0,\ \mathbf{i} > \mathbf{0}, \tag{4.2.12}$$

or

$$E\{[X(\mathbf{s}) - \sum_{\mathbf{j}>\mathbf{0}} \theta_{\mathbf{j}}\ X(\mathbf{s}-\mathbf{j})]\ X(\mathbf{s}-\mathbf{i})\} = 0,\ \mathbf{i} > \mathbf{0}, \tag{4.2.13}$$

which guarantees that

$$\widetilde{X}(\mathbf{s}) \equiv \sum_{\mathbf{j}>\mathbf{0}} \theta_{\mathbf{j}}\ X(\mathbf{s}-\mathbf{j}) \tag{4.2.14}$$

is the best linear predictor of $X(\mathbf{s})$ based on all its 'past' values. The fact that this predictor is unique comes immediately from Theorem 2.3.1 of Brockwell and Davis (1991, p.51). The fact that the decomposition

$$X(\mathbf{s}) = \widetilde{X}(\mathbf{s}) + \varepsilon(\mathbf{s}) \tag{4.2.15}$$

and the coefficients $\{\theta_{\mathbf{j}},\ \mathbf{j} > \mathbf{0}\}$ are unique comes from Guyon (1982, p.96), since the spectral density is bounded away from 0 and ∞.

Next, we try to extend this concept to the more general case of best linear predictors based on all other locations and not just the ones that refer to the 'past' or 'future'. This can be more meaningful and useful for spatial statistics, when there is no particular reason to assume that one side of a spatial dimension offers more information or should be preferred over the other side. It becomes more obvious in prediction rather than in estimation. Given that the parameters of a process have been somehow estimated, in time series it is usually necessary to predict the value of the process in the future, given what has happened in the present and past, while in spatial statistics all the values of the process around the value that should be predicted are usually known and used for this.

For a real-valued zero-mean process $\{X(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ with $E\{X(\mathbf{s})^2\} < \infty$ on any $\mathbf{s}^\tau \in \mathcal{Z}^d$, we consider $\widehat{X}(\mathbf{s})$ to be the best linear predictor of $X(\mathbf{s})$ based on all $X(\mathbf{s}-\mathbf{j})$, $\mathbf{j}^\tau \in \mathcal{Z}^d$, $\mathbf{j} \neq \mathbf{0}$, and we write for any $\mathbf{s}^\tau \in \mathcal{Z}^d$

$$\widehat{X}(\mathbf{s}) \equiv \sum_{\substack{\mathbf{j}^\tau \in \mathcal{Z}^d, \\ \mathbf{j} \neq \mathbf{0}}} \beta_{\mathbf{s},\mathbf{j}}\, X(\mathbf{s}-\mathbf{j}), \tag{4.2.16}$$

in the sense that

$$E\{X(\mathbf{s}) - \widehat{X}(\mathbf{s})\}^2 = \min_{\{\psi_{\mathbf{s},\mathbf{j}}\}} E\{X(\mathbf{s}) - \sum_{\substack{\mathbf{j}^\tau \in \mathcal{Z}^d, \\ \mathbf{j} \neq \mathbf{0}}} \psi_{\mathbf{s},\mathbf{j}}\, X(\mathbf{s}-\mathbf{j})\}^2. \tag{4.2.17}$$

Due to the least squares property, it holds that

$$E\{(X(\mathbf{s}) - \widehat{X}(\mathbf{s}))X(\mathbf{s}-\mathbf{i})\} = 0,\ \mathbf{s}^\tau, \mathbf{i}^\tau \in \mathcal{Z}^d,\ \mathbf{i} \neq \mathbf{0}. \tag{4.2.18}$$

Equation (4.2.18) can be re-written as

$$E\{X(\mathbf{s})X(\mathbf{s}-\mathbf{i})\} - \sum_{\substack{\mathbf{j}^\tau \in \mathcal{Z}^d, \\ \mathbf{j} \neq \mathbf{0}}} \beta_{\mathbf{s},\mathbf{j}}\, E\{X(\mathbf{s}-\mathbf{j})X(\mathbf{s}-\mathbf{i})\} = 0,\ \mathbf{s}^\tau, \mathbf{i}^\tau \in \mathcal{Z}^d,\ \mathbf{i} \neq \mathbf{0}. \tag{4.2.19}$$

Similarly, for any $\mathbf{s}^\tau \in \mathcal{Z}^d$, we may define the prediction variances

$$\begin{aligned} \nu_{\mathbf{s}} &\equiv E\{X(\mathbf{s}) - \widehat{X}(\mathbf{s})\}^2 = E\{X(\mathbf{s})(X(\mathbf{s}) - \widehat{X}(\mathbf{s}))\} \\ &= E\{X(\mathbf{s})(X(\mathbf{s}) - \sum_{\substack{\mathbf{j}^\tau \in \mathcal{Z}^d, \\ \mathbf{j} \neq \mathbf{0}}} \beta_{\mathbf{s},\mathbf{j}}\, X(\mathbf{s}-\mathbf{j}))\} \\ &= E\{X(\mathbf{s})^2\} - \sum_{\substack{\mathbf{j}^\tau \in \mathcal{Z}^d, \\ \mathbf{j} \neq \mathbf{0}}} \beta_{\mathbf{s},\mathbf{j}}\, E\{X(\mathbf{s}-\mathbf{j})X(\mathbf{s})\} < \infty, \end{aligned} \tag{4.2.20}$$

where the inequality holds since $E\{X(\mathbf{s}) - \widehat{X}(\mathbf{s})\}^2 < E\{X(\mathbf{s})^2\} < \infty$, for any $\mathbf{s}^\tau \in \mathcal{Z}^d$.

If $\{X(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ is a (weakly) stationary process, then $\gamma_X(\mathbf{j}) = E\{X(\mathbf{s})X(\mathbf{s}-\mathbf{j})\}$ and for any $\mathbf{s}^\tau \in \mathcal{Z}^d$, we can simplify $\beta_{\mathbf{s},\mathbf{j}} \equiv \beta_{\mathbf{j}}$ and $\nu_{\mathbf{s}} \equiv \nu$, in (4.2.19) and (4.2.20). The two equations can then be summarized as

$$\gamma_X(\mathbf{i}) - \sum_{\substack{\mathbf{j}^\tau \in \mathcal{Z}^d, \\ \mathbf{j} \neq \mathbf{0}}} \beta_{\mathbf{j}}\, \gamma_X(\mathbf{j}-\mathbf{i}) = 0,\ \mathbf{i} \neq \mathbf{0}, \tag{4.2.21}$$

$$\gamma_X(\mathbf{0}) - \sum_{\substack{\mathbf{j}^\tau \in \mathcal{Z}^d, \\ \mathbf{j} \neq \mathbf{0}}} \beta_{\mathbf{j}}\, \gamma_X(\mathbf{j}) = \nu \in (0, \infty). \tag{4.2.22}$$

Although the uniqueness of the predictor $\widehat{X}(\mathbf{s})$ is guaranteed by Theorem 2.3.1 of Brockwell and Davis (1991, p.51), so far we cannot say that the equations (4.2.21) generate

unique coefficients $\{\beta_{\mathbf{j}},\ \mathbf{j}^\tau \in \mathcal{Z}^d - \{\mathbf{0}^\tau\}\}$. Moreover, for a simple example such as that of the following auto-regression $\{X_t,\ t \in \mathcal{Z}\}$, which satisfies the equation

$$X_t - \theta\ X_{t-1} = \varepsilon_t,\ |\theta| < 1,\ \{\varepsilon_t\} \sim WN(0, \sigma^2), \tag{4.2.23}$$

we would like to be able to find these coefficients $\{\beta_i,\ i \in \mathcal{Z},\ i \neq 0\}$. From (4.2.23), we may derive

$$-\theta\ X_{t+1} + \theta^2\ X_t = -\theta\ \varepsilon_{t+1}. \tag{4.2.24}$$

If we add (4.2.23) and (4.2.24), we may define

$$Y_t \equiv -\theta\ X_{t-1} + (1+\theta^2)\ X_t - \theta\ X_{t+1} = \varepsilon_t - \theta\ \varepsilon_{t+1}. \tag{4.2.25}$$

We can see immediately that $\{Y_t,\ t \in \mathcal{Z}\}$ is an invertible moving-average that corresponds to the same polynomial as the causal auto-regression $\{X_t,\ t \in \mathcal{Z}\}$, *i.e.* they are both generated by $\{\varepsilon_t,\ t \in \mathcal{Z}\}$, one from its past and the other from its future and present values. Thanks to the same arguments as the ones used in Section 2.4.1, it holds that Y_t is uncorrelated with X_{t-i}, for any $i \neq 0$. We may re-write (4.2.25) as

$$X_t - \frac{\theta}{1+\theta^2}\ (X_{t-1} + X_{t+1}) = \frac{1}{1+\theta^2}\ Y_t, \tag{4.2.26}$$

from which we may see that

$$\widehat{X}_t \equiv \frac{\theta}{1+\theta^2}\ (X_{t-1} + X_{t+1}) \tag{4.2.27}$$

is the unique best linear predictor of X_t based on all other $X_{t-i}, i \neq 0$, since $X_t - \widehat{X}_t$ is proportional to Y_t and, thus, uncorrelated with these random variables. Also, it holds that

$$E\{X_t - \widehat{X}_t\}^2 = \frac{1}{(1+\theta^2)^2}\mathrm{Var}\{Y_t\} = \frac{\sigma^2}{1+\theta^2} \tag{4.2.28}$$

is the variance of the prediction error. In (4.2.27), we can see that for any $i > 0$, the coefficient of X_{t-i} is the same as the coefficients of X_{t+i}. Indeed, it holds that

$$\beta_i = \begin{cases} \theta/1+\theta^2, & i = 1 \\ 0, & i > 1 \\ \beta_{-i}, & i < 0 \end{cases}. \tag{4.2.29}$$

In (4.2.29), we can see that

$$\left|\frac{\theta}{1+\theta^2}\right| = \frac{|\theta|}{1+\theta^2} < 1.$$

In general, the fact that the coefficients $\{\beta_{\mathbf{j}},\ \mathbf{j}^\tau \in \mathcal{Z}^d - \{\mathbf{0}^\tau\}\}$ are symmetric and that their absolute value is not greater than one, as well as a condition for their uniqueness are established in the next proposition.

Proposition 4.1. If $\{X(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ is a (weakly) stationary process with absolutely summable auto-covariance function $\gamma_X(\cdot)$ and there are coefficients $\{\beta_{\mathbf{j}},\ \mathbf{j}^\tau \in \mathcal{Z}^d - \{\mathbf{0}^\tau\}\}$ satisfying (4.2.21),(4.2.22) and such that $\sum_{\substack{\mathbf{j}^\tau \in \mathcal{Z}^d,\\ \mathbf{j}\neq\mathbf{0}}} |\beta_{\mathbf{j}}| < \infty$, then

1. $\beta_{\mathbf{j}} = \beta_{-\mathbf{j}}$, for all $\mathbf{j}^\tau \in \mathcal{Z}^d - \{\mathbf{0}^\tau\}$,
2. $|\beta_{\mathbf{j}}| \leq 1$, for all $\mathbf{j}^\tau \in \mathcal{Z}^d - \{\mathbf{0}^\tau\}$,
3. The coefficients are the unique solutions of (4.2.21).

Proof. For the polynomial

$$\beta(\mathbf{z}) \equiv 1 - \sum_{\substack{\mathbf{j}^\tau \in \mathcal{Z}^d,\\ \mathbf{j}\neq\mathbf{0}}} \beta_{\mathbf{j}}\ \mathbf{z}^{\mathbf{j}}, \tag{4.2.30}$$

we define the new process

$$Y(\mathbf{s}) \equiv \beta(\mathbf{B})X(\mathbf{s}) = X(\mathbf{s}) - \sum_{\substack{\mathbf{j}^\tau \in \mathcal{Z}^d,\\ \mathbf{j}\neq\mathbf{0}}} \beta_{\mathbf{j}}\ X(\mathbf{s}-\mathbf{j}), \tag{4.2.31}$$

which is also (weakly) stationary. Its auto-covariance function is equal to

$$\begin{aligned}
&\mathrm{Cov}\{Y(\mathbf{s}), Y(\mathbf{s}-\mathbf{i})\} = \mathrm{Cov}\{X(\mathbf{s}) - \sum_{\substack{\mathbf{j}^\tau \in \mathcal{Z}^d,\\ \mathbf{j}\neq\mathbf{0}}} \beta_{\mathbf{j}}\ X(\mathbf{s}-\mathbf{j}), X(\mathbf{s}-\mathbf{i}) - \sum_{\substack{\mathbf{j}^{*\tau} \in \mathcal{Z}^d,\\ \mathbf{j}^*\neq\mathbf{0}}} \beta_{\mathbf{j}^*}\ X(\mathbf{s}-\mathbf{i}-\mathbf{j}^*)\} \\
= \ &\gamma_X(\mathbf{i}) - \sum_{\substack{\mathbf{j}^\tau \in \mathcal{Z}^d,\\ \mathbf{j}\neq\mathbf{0}}} \beta_{\mathbf{j}}\ \gamma_X(\mathbf{j}-\mathbf{i}) - \sum_{\substack{\mathbf{j}^{*\tau} \in \mathcal{Z}^d,\\ \mathbf{j}^*\neq\mathbf{0}}} \beta_{\mathbf{j}^*}[\gamma_X(\mathbf{i}+\mathbf{j}^*) - \sum_{\substack{\mathbf{j}^\tau \in \mathcal{Z}^d,\\ \mathbf{j}\neq\mathbf{0}}} \beta_{\mathbf{j}}\ \gamma_X(\mathbf{j}-\mathbf{i}-\mathbf{j}^*)]
\end{aligned}$$

By equations (4.2.21) and (4.2.22), we can write

$$\gamma_Y(\mathbf{i}) \equiv \mathrm{Cov}\{Y(\mathbf{s}), Y(\mathbf{s}-\mathbf{i})\} = \begin{cases} \nu, & \mathbf{i} = \mathbf{0} \\ -\beta_{-\mathbf{i}} \cdot \nu, & \mathbf{i} \neq \mathbf{0} \end{cases}, \tag{4.2.32}$$

which verifies the first two statements of the proposition, since $\beta_{\mathbf{j}}$ are auto-correlations at lag $\mathbf{j}$.

For the third statement of the proposition, we define the polynomial

$$\gamma_X(\mathbf{z}) \equiv \sum_{\mathbf{j}^\tau \in \mathcal{Z}^d} \gamma_X(\mathbf{j})\ \mathbf{z}^{\mathbf{j}}, \tag{4.2.33}$$

and we can see immediately from (4.2.21) and (4.2.22) that

$$\gamma_X(\mathbf{z}) \cdot \frac{\beta(\mathbf{z})}{\nu} = 1. \tag{4.2.34}$$

The spectral density of the original process can be expressed as

$$g_X(\omega) = \frac{1}{(2\pi)^d}\,\gamma_X(e^{i\omega}) = \frac{1}{(2\pi)^d}\frac{\nu}{\beta(e^{i\omega})},\ \omega^\tau \in [-\pi,\pi]^d, \tag{4.2.35}$$

and it is bounded away from 0 and ∞, since $\sum_{\mathbf{j}^\tau\in\mathcal{Z}^d}|\gamma_X(\mathbf{j})| < \infty$ and $\sum_{\mathbf{j}>\mathbf{0}}|\beta_\mathbf{j}| < \infty$. According to Guyon (1982, p.96), $\{X(\mathbf{j}),\ \mathbf{j}^\tau \in \mathcal{Z}^d\}$ is a basis of the Hilbert space generated by $\{X(\mathbf{j}),\ \mathbf{j}^\tau \in \mathcal{Z}^d\}$. A direct consequence of that is that for any $\mathbf{s}^\tau \in \mathcal{Z}^d$, it holds that $\{X(\mathbf{s}-\mathbf{j})+X(\mathbf{s}+\mathbf{j}),\ \mathbf{j}>\mathbf{0}\}$ is also a basis of the Hilbert space generated by $\{X(\mathbf{s}-\mathbf{j})+X(\mathbf{s}+\mathbf{j}),\ \mathbf{j}^\tau \in \mathcal{Z}^d - \{\mathbf{0}^\tau\}\}$. Then the decomposition

$$X(\mathbf{s}) = \sum_{\mathbf{j}>\mathbf{0}}\beta_\mathbf{j}\,[X(\mathbf{s}-\mathbf{j})+X(\mathbf{s}+\mathbf{j})] + Y(\mathbf{s}), \tag{4.2.36}$$

with $E\{Y(\mathbf{s})X(\mathbf{s}-\mathbf{j})\} = 0,\ \mathbf{j}\neq\mathbf{0}$, is unique. This verifies the final statement of the proposition.

We can see that, since $\sum_{\mathbf{j}>\mathbf{0}}|\beta_\mathbf{j}| < \infty$, then the spectral density of $\{Y(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ exists and is equal to

$$g_Y(\omega) \equiv \frac{\nu}{(2\pi)^d}\,\beta(e^{i\omega}),\ \omega^\tau \in [-\pi,\pi]^d. \tag{4.2.37}$$

■

Remark 4.1. An interesting question would be to find a necessary condition for

$$\sum_{\mathbf{j}^\tau\in\mathcal{Z}^d}|\gamma_X(\mathbf{j})| < \infty \tag{4.2.38}$$

in terms of the prediction coefficients $\beta_\mathbf{j}$. A sufficient condition for (4.2.38) is that

$$\sum_{\mathbf{j}^\tau\in\mathcal{Z}^d-\{\mathbf{0}^\tau\}}|\beta_\mathbf{j}| < 1 \Leftrightarrow \sum_{\mathbf{j}>\mathbf{0}}|\beta_\mathbf{j}| < 1/2. \tag{4.2.39}$$

This has been mentioned, by Besag (1974, p.232) and Besag (1977, p.74), by Besag and Moran (1975, p.558) and by Moran (1973, p.58) for the case when $d = 2$. Indeed, it holds for any $\omega^\tau \in [-\pi,\pi]^d$ that

$$\begin{aligned}\sum_{\mathbf{j}^\tau\in\mathcal{Z}^d-\{\mathbf{0}^\tau\}}\beta_\mathbf{j}\,\cos\{\omega\mathbf{j}^\tau\} &\leq \Big|\sum_{\mathbf{j}^\tau\in\mathcal{Z}^d-\{\mathbf{0}^\tau\}}\beta_\mathbf{j}\,\cos\{\omega\mathbf{j}^\tau\}\Big| \\ &\leq \sum_{\mathbf{j}^\tau\in\mathcal{Z}^d-\{\mathbf{0}^\tau\}}|\beta_\mathbf{j}|\,|\cos\{\omega\mathbf{j}^\tau\}| \leq \sum_{\mathbf{j}^\tau\in\mathcal{Z}^d-\{\mathbf{0}^\tau\}}|\beta_\mathbf{j}|.\end{aligned}$$

Thus, if condition (4.2.39) holds, we can write

$$\beta(e^{i\omega}) = 1 - \sum_{\mathbf{j}^\tau\in\mathcal{Z}^d-\{\mathbf{0}^\tau\}}\beta_\mathbf{j}\,\cos\{\omega\mathbf{j}^\tau\} > 0, \tag{4.2.40}$$

for any $\omega^\tau \in [-\pi,\pi]^d$.

4.3 Spatial auto-linear schemes

Definition 4.1. For a set of finite cardinality $\mathcal{F}_0 \subset \mathcal{Z}^d$, such that $\mathbf{0}^\tau \notin \mathcal{F}_0$, we consider the (weakly) stationary zero-mean process $\{X(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ to form a spatial auto-linear scheme if it has an absolutely summable auto-covariance function and for every $\mathbf{s}^\tau \in \mathcal{Z}^d$, it holds that

$$\begin{aligned} E\{X(\mathbf{s})|X(\mathbf{s}-\mathbf{j}) = x(\mathbf{s}-\mathbf{j}),\ \mathbf{j}^\tau \in \mathcal{S} \supseteq \mathcal{F}_0,\ \mathbf{j} \neq \mathbf{0}\} &= \\ E\{X(\mathbf{s})|X(\mathbf{s}-\mathbf{j}) = x(\mathbf{s}-\mathbf{j}),\ \mathbf{j}^\tau \in \mathcal{F}_0\} &= \sum_{\mathbf{j}^\tau \in \mathcal{F}_0} \beta_{\mathbf{j}}\, x(\mathbf{s}-\mathbf{j}) \end{aligned} \quad (4.3.1)$$

and

$$\begin{aligned} \mathrm{Var}\{X(\mathbf{s})|X(\mathbf{s}-\mathbf{j}) = x(\mathbf{s}-\mathbf{j}),\ \mathbf{j}^\tau \in \mathcal{S} \supseteq \mathcal{F}_0,\ \mathbf{j} \neq \mathbf{0}\} &= \\ \mathrm{Var}\{X(\mathbf{s})|X(\mathbf{s}-\mathbf{j}) = x(\mathbf{s}-\mathbf{j}),\ \mathbf{j}^\tau \in \mathcal{F}_0\} &= \nu \in (0,\infty). \end{aligned} \quad (4.3.2)$$

According to Proposition 4.1, now that the best linear predictors take the form of conditional expectations, the former scheme is valid only when two conditions are imposed. The first is a symmetry condition, that

$$\beta_{\mathbf{j}} = \beta_{-\mathbf{j}},\ \mathbf{j}^\tau \in \mathcal{F}_0. \quad (4.3.3)$$

For the finite set $\mathcal{F}_0$ this also means that $\mathbf{j}^\tau \in \mathcal{F}_0$ implies that $-\mathbf{j}^\tau \in \mathcal{F}_0$. The second condition is that

$$|\beta_{\mathbf{j}}| < 1,\ \mathbf{j}^\tau \in \mathcal{F}_0. \quad (4.3.4)$$

If we define the polynomial

$$\beta(\mathbf{z}) \equiv 1 - \sum_{\mathbf{j}^\tau \in \mathcal{F}_0} \beta_{\mathbf{j}}\, \mathbf{z}^{\mathbf{j}}, \quad (4.3.5)$$

and the spectral density exists, it can be expressed as

$$g_X(\boldsymbol{\omega}) = \frac{1}{(2\pi)^d} \cdot \frac{\nu}{\beta(e^{i\boldsymbol{\omega}})},\ \boldsymbol{\omega}^\tau \in [-\pi,\pi]^d. \quad (4.3.6)$$

The definition of an auto-linear scheme is, for the special case where $d = 2$, equivalent to the definition of an auto-normal scheme given by Besag and Moran (1975, p.555), with the main exception that we have not made any specific assumptions on the distribution of the process. Of course, for stationary Gaussian processes, the best linear predictors are conditional expectations and the prediction variances are conditional variances.

In the general case that we do not study Gaussian random variables only, the construction of a stationary auto-scheme is the same as the construction of a valid conditional distribution on one location given the values on all other locations. For example, apart from the auto-normal scheme, Besag (1974) also introduced the auto-logistic scheme which for a stationary process $\{X(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ takes the form

$$\begin{aligned} & P\{X(\mathbf{s}) = k | X(\mathbf{s}-\mathbf{j}) = x(\mathbf{s}-\mathbf{j}),\ \mathbf{j}^\tau \in \mathcal{S} \supseteq \mathcal{F}_0,\ \mathbf{j} \neq \mathbf{0}\} \\ = \ & P\{X(\mathbf{s}) = k | X(\mathbf{s}-\mathbf{j}) = x(\mathbf{s}-\mathbf{j}),\ \mathbf{j}^\tau \in \mathcal{F}_0\} \\ = \ & \begin{cases} \frac{\exp\{\alpha + \sum_{\mathbf{j}^\tau \in \mathcal{F}_0} \beta_{\mathbf{j}}\ x(\mathbf{s}-\mathbf{j})\}}{1+\exp\{\alpha + \sum_{\mathbf{j}^\tau \in \mathcal{F}_0} \beta_{\mathbf{j}}\ x(\mathbf{s}-\mathbf{j})\}}, & k = 1 \\ \frac{1}{1+\exp\{\alpha + \sum_{\mathbf{j}^\tau \in \mathcal{F}_0} \beta_{\mathbf{j}}\ x(\mathbf{s}-\mathbf{j})\}}, & k = 0 \end{cases} . \end{aligned} \tag{4.3.7}$$

Like the auto-normal scheme, Besag showed that the former representation is valid when there is a symmetry in the neighbourhood structure and the coefficients, *i.e.* if $\mathbf{j}^\tau \in \mathcal{F}_0$ implies that $-\mathbf{j}^\tau \in \mathcal{F}_0$ and if it holds that $\beta_{\mathbf{j}} = \beta_{-\mathbf{j}}$.

4.4 Unilateral and some bilateral auto-regressions

For any unilateral or bilateral (weakly) stationary auto-regression of finite order $\{X(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$, only a look at its spectral density together with Proposition 4.1 shows that the best linear predictor of $X(\mathbf{s})$ based on all $X(\mathbf{s}-\mathbf{j}), \mathbf{j} \neq \mathbf{0}$, is equal to a finite sum of random variables. This, of course, does not imply that the auto-regression is an auto-linear scheme as the predictors are not necessarily conditional expectations. In this section, we examine under which conditions is it possible to form an auto-regression that is an auto-linear scheme, even though it might not be an auto-normal scheme.

From the sequence of uncorrelated random variables

$$\{W(\mathbf{s})\} \sim WN(0, \sigma^2), \tag{4.4.1}$$

we may define the auto-regression $\{X(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$, such that

$$\theta_p(\mathbf{B})\ \theta_f(\mathbf{B}^{-1}) X(\mathbf{s}) \equiv W(\mathbf{s}), \tag{4.4.2}$$

where

$$\theta_p(\mathbf{z}) \equiv 1 + \sum_{\mathbf{j} \in \mathcal{U}_p} \theta_{\mathbf{j}}^{(p)} \mathbf{z}^{\mathbf{j}}, \tag{4.4.3}$$

$$\theta_f(\mathbf{z}) \equiv 1 + \sum_{\mathbf{j} \in \mathcal{U}_f} \theta_{\mathbf{j}}^{(f)} \mathbf{z}^{\mathbf{j}}, \tag{4.4.4}$$

and $\mathcal{U}_p, \mathcal{U}_f \subset \{\mathbf{j} > \mathbf{0}\}$ are two sets of finite cardinalities. Also, we assume we can write

$$\theta_p(\mathbf{z})^{-1} \equiv 1 + \sum_{\mathbf{j}>\mathbf{0}} p_{\mathbf{j}}\ \mathbf{z}^{\mathbf{j}},\ \sum_{\mathbf{j}>\mathbf{0}} |p_{\mathbf{j}}| < \infty, \tag{4.4.5}$$

$$\theta_f(\mathbf{z})^{-1} \equiv 1 + \sum_{\mathbf{j}>\mathbf{0}} f_{\mathbf{j}}\ \mathbf{z}^{\mathbf{j}},\ \sum_{\mathbf{j}>\mathbf{0}} |f_{\mathbf{j}}| < \infty. \tag{4.4.6}$$

Since we have assumed that $\{X(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ is an auto-regression, we consider that it cannot be $\theta_{\mathbf{j}}^{(p)} = 0,\ \mathbf{j} \in \mathcal{U}_p$ and $\theta_{\mathbf{j}}^{(f)} = 0,\ \mathbf{j} \in \mathcal{U}_f$ at the same time, as this would imply that

$$\{X(\mathbf{s})\} \sim WN(0, \sigma^2).$$

If $\theta_{\mathbf{j}}^{(p)} = 0,\ \mathbf{j} \in \mathcal{U}_p$ or $\theta_{\mathbf{j}}^{(f)} = 0,\ \mathbf{j} \in \mathcal{U}_f$, then the auto-regression is unilateral. Otherwise, it is a bilateral auto-regression. Not all bilateral auto-regressions can be written in the form (4.4.2) when $d \geq 2$. In (4.4.2), the auto-regressive polynomial can be factorized into two distinct polynomials, one expressing the dependence from the 'past' and the other from the 'future'.

The spectral density of $\{X(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ can be expressed as

$$g_X(\boldsymbol{\omega}) = \frac{1}{(2\pi)^d} \cdot \frac{\sigma^2}{\theta_p(e^{i\boldsymbol{\omega}})\ \theta_f(e^{i\boldsymbol{\omega}})\ \theta_p(e^{-i\boldsymbol{\omega}})\ \theta_f(e^{-i\boldsymbol{\omega}})},\ \boldsymbol{\omega}^\tau \in [-\pi, \pi]^d. \tag{4.4.7}$$

If $\theta_p(\mathbf{z}) \neq 1$ and $\theta_f(\mathbf{z}) \neq 1$ and the auto-regression is bilateral, then from the spectral density we can see immediately that there is another process $\{u(\mathbf{s})\} \sim WN(0, \sigma^2)$, such that we can write

$$\theta_p(\mathbf{B})\ \theta_f(\mathbf{B}) X(\mathbf{s}) \equiv u(\mathbf{s}) \tag{4.4.8}$$

and express the same process $\{X(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ as a unilateral auto-regression. We can write then

$$X(\mathbf{s}) \equiv u(\mathbf{s}) + \sum_{\mathbf{j}>\mathbf{0}} \psi_{\mathbf{j}}\ u(\mathbf{s} - \mathbf{j}), \tag{4.4.9}$$

where

$$1 + \sum_{\mathbf{j}>\mathbf{0}} \psi_{\mathbf{j}}\ \mathbf{z}^{\mathbf{j}} = (1 + \sum_{\mathbf{j}>\mathbf{0}} p_{\mathbf{j}}\ \mathbf{z}^{\mathbf{j}}) \cdot (1 + \sum_{\mathbf{j}>\mathbf{0}} f_{\mathbf{j}}\ \mathbf{z}^{\mathbf{j}}),\ \sum_{\mathbf{j}>\mathbf{0}} |\psi_{\mathbf{j}}| < \infty. \tag{4.4.10}$$

Next, we define the polynomial

$$\sum_{\mathbf{j}^\tau \in \mathcal{F}} b_{\mathbf{j}}\ \mathbf{z}^{\mathbf{j}} \equiv \theta_p(\mathbf{z})\ \theta_f(\mathbf{z})\ \theta_p(\mathbf{z}^{-1})\ \theta_f(\mathbf{z}^{-1}) = (1 + \sum_{\mathbf{j}>\mathbf{0}} \psi_{\mathbf{j}}\ \mathbf{z}^{\mathbf{j}})^{-1} \cdot (1 + \sum_{\mathbf{j}<\mathbf{0}} \psi_{-\mathbf{j}}\ \mathbf{z}^{\mathbf{j}})^{-1}, \tag{4.4.11}$$

where $\mathcal{F}$ is a set of finite cardinality, such that

1. $\mathbf{0}^\tau \in \mathcal{F}$,

2. $\mathbf{j}^\tau \in \mathcal{F}$ if and only if $-\mathbf{j}^\tau \in \mathcal{F}$,

and the coefficients $b_\mathbf{j}$, are such that

1. $b_\mathbf{j} = b_{-\mathbf{j}}$, $\mathbf{j}^\tau \in \mathcal{F}$,

2. $|b_\mathbf{j}| \leq b_\mathbf{0}$, $\mathbf{j}^\tau \in \mathcal{F}$.

While the first three parts derive immediately from the way the new polynomial has been defined, the fourth part might be justified by saying that the spectral density

$$\begin{aligned} g^*(\omega) &\equiv \frac{1}{(2\pi)^d} \cdot \frac{1}{(1+\sum_{\mathbf{j}>\mathbf{0}} \psi_\mathbf{j}\, e^{i\omega\mathbf{j}^\tau}) \cdot (1+\sum_{\mathbf{j}<\mathbf{0}} \psi_{-\mathbf{j}}\, e^{i\omega\mathbf{j}^\tau})} \\ &= \frac{1}{(2\pi)^d} \cdot \sum_{\mathbf{j}^\tau \in \mathcal{F}} b_\mathbf{j}\, e^{i\omega\mathbf{j}^\tau},\ \omega^\tau \in [-\pi,\pi]^d, \end{aligned}$$

generates the auto-covariance function $\{b_\mathbf{j},\ \mathbf{j}^\tau \in \mathcal{F}\}$.

From the unilateral representation (4.4.8), we can define the polynomial

$$\theta(\mathbf{z}) \equiv \theta_p(\mathbf{z})\ \theta_f(\mathbf{z}) \equiv 1 + \sum_{\mathbf{j}\in\mathcal{U}} \theta_\mathbf{j}\, \mathbf{z}^\mathbf{j}, \tag{4.4.12}$$

where again $\mathcal{U}$ is a set of finite cardinality and $\mathcal{U}_p \cup \mathcal{U}_f \subseteq \mathcal{U} \subset \{\mathbf{j} > \mathbf{0}\}$. Then it holds that

$$b_\mathbf{0} = 1 + \sum_{\mathbf{j}\in\mathcal{U}} \theta_\mathbf{j}^2. \tag{4.4.13}$$

If we define the set

$$\mathcal{F}_0 \equiv \mathcal{F} - \{\mathbf{0}^\tau\}, \tag{4.4.14}$$

the coefficients

$$\beta_\mathbf{j} \equiv -b_\mathbf{j}/b_\mathbf{0},\ \mathbf{j}^\tau \in \mathcal{F}_0, \tag{4.4.15}$$

the polynomial

$$\beta(\mathbf{z}) \equiv 1 - \sum_{\mathbf{j}^\tau\in\mathcal{F}_0} \beta_\mathbf{j}\, \mathbf{z}^\mathbf{j} \tag{4.4.16}$$

and the constant

$$\nu \equiv \sigma^2/b_\mathbf{0}, \tag{4.4.17}$$

then we can re-write the spectral density of $\{X(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ as

$$g_X(\omega) = \frac{1}{(2\pi)^d} \cdot \frac{\nu}{\beta(e^{i\omega})},\ \omega^\tau \in [-\pi,\pi]^d. \tag{4.4.18}$$

The former representation concerns the second-order properties of $\{X(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ and does not imply yet that they form an auto-linear scheme. To make such a conclusion, we need further assumptions. If we know that for any $\mathbf{j} \neq \mathbf{0}$ the random variable $u(\mathbf{s})$ is independent of $u(\mathbf{s}-\mathbf{j})$, then from (4.4.8) and (4.4.12), we can write

$$\theta(\mathbf{B}^{-1})\ \theta(\mathbf{B})X(\mathbf{s}) = \theta(\mathbf{B}^{-1})u(\mathbf{s}) \tag{4.4.19}$$

or

$$b_0\ X(\mathbf{s}) + \sum_{\mathbf{j}^\tau \in \mathcal{F}_0} b_\mathbf{j}\ X(\mathbf{s}-\mathbf{j}) = u(\mathbf{s}) + \sum_{\mathbf{j} \in \mathcal{U}} \theta_\mathbf{j}\ u(\mathbf{s}+\mathbf{j})$$

or

$$X(\mathbf{s}) = \sum_{\mathbf{j}^\tau \in \mathcal{F}_0} \beta_\mathbf{j}\ X(\mathbf{s}-\mathbf{j}) + \frac{1}{b_0}[u(\mathbf{s}) + \sum_{\mathbf{j} \in \mathcal{U}} \theta_\mathbf{j}\ u(\mathbf{s}+\mathbf{j})]. \tag{4.4.20}$$

It holds that $X(\mathbf{s}-\mathbf{j}), \mathbf{j} > \mathbf{0}$ are independent of $u(\mathbf{s}), u(\mathbf{s}+\mathbf{j}), \mathbf{j} \in \mathcal{U}$. Now for the other error sequence

$$\{\varepsilon(\mathbf{s})\} \sim WN(0, \sigma^2), \tag{4.4.21}$$

for which we can write

$$\theta(\mathbf{B}^{-1})X(\mathbf{s}) \equiv \varepsilon(\mathbf{s}), \tag{4.4.22}$$

we need to assume that for any $\mathbf{j} \neq \mathbf{0}$, the random variable $\varepsilon(\mathbf{s})$ is independent of $\varepsilon(\mathbf{s}-\mathbf{j})$. We can then reverse and write

$$\theta(\mathbf{B})\ \theta(\mathbf{B}^{-1})X(\mathbf{s}) = \theta(\mathbf{B})\varepsilon(\mathbf{s}) \tag{4.4.23}$$

and now $X(\mathbf{s}+\mathbf{j}), \mathbf{j} > \mathbf{0}$ are independent of $\varepsilon(\mathbf{s}), \varepsilon(\mathbf{s}-\mathbf{j}), \mathbf{j} \in \mathcal{U}$. We may then combine the two and conclude that $X(\mathbf{s}-\mathbf{j})$ for any $\mathbf{j} \neq \mathbf{0}$ are independent of the random quantity $\theta(\mathbf{B}^{-1})u(\mathbf{s}) = \theta(\mathbf{B})\varepsilon(\mathbf{s})$. As a result, from (4.4.20), we can write

$$E\{X(\mathbf{s})|X(\mathbf{s}-\mathbf{j}) = x(\mathbf{s}-\mathbf{j}),\ \mathbf{j}^\tau \in \mathcal{S} \supseteq \mathcal{F}_0,\ \mathbf{j} \neq \mathbf{0}\} = \sum_{\mathbf{j}^\tau \in \mathcal{F}_0} \beta_\mathbf{j}\ x(\mathbf{s}-\mathbf{j}) \tag{4.4.24}$$

and

$$\begin{aligned}
\mathrm{Var}\{X(\mathbf{s})|X(\mathbf{s}-\mathbf{j}) = x(\mathbf{s}-\mathbf{j}),\ \mathbf{j}^\tau \in \mathcal{S} \supseteq \mathcal{F}_0,\ \mathbf{j} \neq \mathbf{0}\} &= \frac{1}{b_0^2} \cdot \mathrm{Var}\{u(\mathbf{s}) + \sum_{\mathbf{j} \in \mathcal{U}} \theta_\mathbf{j}\ u(\mathbf{s}+\mathbf{j})\} \\
&= \frac{1}{b_0^2} \cdot (1 + \sum_{\mathbf{j} \in \mathcal{U}} \theta_\mathbf{j}^2) \cdot \sigma^2 \\
&= \frac{1}{b_0^2} \cdot b_0 \cdot \sigma^2 \\
&= \sigma^2/b_0 = \nu.
\end{aligned} \tag{4.4.25}$$

Thus, it takes the two error sequences of the auto-regressive unilateral representations, to be sequences of independent random variables, in order to come up with an auto-linear scheme.

According to Proposition 2.5, we have seen how $\{\varepsilon(\mathbf{s})\} \sim IID(0, \sigma^2)$ implies that $\{u(\mathbf{s})\} \sim IID(0, \sigma^2)$, if meanwhile it has produced a reverse strictly stationary process $\{X(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$. If, for example, we consider

$$\theta_p(\mathbf{z}) = \theta_f(\mathbf{z}) \tag{4.4.26}$$

in the original representation (4.4.2), then according to Remark 2.1, the symmetric filter

$$\theta_p(\mathbf{B})\ \theta_p(\mathbf{B}^{-1})X(\mathbf{s}) = W(\mathbf{s}) \tag{4.4.27}$$

could generate a reverse strictly stationary process, if $\{W(\mathbf{s})\} \sim IID(0, \sigma^2)$. But again, that would be no use, unless we could assume that either $\{\varepsilon(\mathbf{s})\} \sim IID(0, \sigma^2)$ or $\{u(\mathbf{s})\} \sim IID(0, \sigma^2)$, in order to come up with an auto-linear scheme. Thus, there does not seem to exist a safe way to consider that a process $\{X(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ satisfying (4.4.2) can also satisfy (4.3.1) and (4.3.2), if one sequence only of independent and identically distributed random errors can be assumed. Of course, for Gaussian random variables, one sequence of Gaussian random errors immediately implies the auto-normal formulation.

The conditional moments of an auto-linear scheme relate to the assumptions on the dependence between the random variables, rather than the second-order properties only. Regarding the second-order properties, we have seen in this section how from a unilateral or bilateral auto-regression of the form (4.4.2) we may recover the coefficients of the best linear predictors. But what if we are trying to recover the causal auto-regressive representation of an auto-linear scheme? Looking at the spectral densities, this might be viewed as a problem similar to that of matching a (weakly) stationary sequence of random variables that have non-zero auto-covariances at a finite number of lags only, to an invertible moving-average sharing the same second-order properties. For $d = 1$, one can always find an invertible moving-average of finite order, equal to the number of non-zero auto-covariances at the positive lags. This has been established by Proposition 3.2.1 of Brockwell and Davis (1991, p.89). For $d \geq 2$, this might not be possible though. For example, for the simple case $d = 2$, one might have the best linear predictor

$$\widehat{X}(u, v) = \beta_1\ [X(u-1, v) + X(u+1, v)] + \beta_2\ [X(u, v-1) + X(u, v+1)], \tag{4.4.28}$$

with $|\beta_1|, |\beta_2| < 1$. Then we need to write the polynomial

$$b(z_1, z_2) \propto 1 - \beta_1 \, (z_1 + z_1^{-1}) - \beta_2 \, (z_2 + z_2^{-1}) \tag{4.4.29}$$

as a product of, say,

$$\begin{aligned} b(z_1, z_2) &\propto (1 + \theta_1 \, z_1 + \theta_2 \, z_2) \cdot (1 + \theta_1 \, z_1^{-1} + \theta_2 \, z_2^{-1}) \\ &= (1 + \theta_1^2 + \theta_2^2) + \theta_1 \, (z_1 + z_1^{-1}) + \theta_2 \, (z_2 + z_2^{-1}) + \theta_1 \cdot \theta_2 \, (z_1 z_2^{-1} + z_1^{-1} z_2). \end{aligned}$$

It should be

$$\theta_i \propto -\beta_i \neq 0, \; i = 1, 2,$$

and

$$\theta_1 \cdot \theta_2 = 0,$$

which cannot happen at the same time and we will have to look harder than that. A problem of a similar nature occurs when we try to match a bilateral auto-regression to a unilateral auto-regression sharing the same second-order properties, according to Whittle (1954, p.439).

In conclusion, in this section we have seen how any auto-regression that shares the same second-order properties as a finite unilateral auto-regression can be expressed as an auto-linear scheme, if the error sequences of the two finite unilateral auto-regressive representations are sequences of independent random variables. Moreover, we have seen how to perform the re-parametrization from the coefficients of unilateral representation to the coefficients of the auto-linear scheme. That is a generalization of the example of the causal AR(1), which was presented in Section 4.2. Bilateral auto-regressions can also be auto-linear schemes. We have selected a convenient class of bilateral auto-regressions to demonstrate that. As a result, the results achieved next in this chapter, regarding the estimators of the coefficients of an auto-linear scheme and their properties, are totally unconnected to any assumption of a causal representation of the process of interest.

4.5 Estimation

We collect observations $\{X(\mathbf{s}), \; \mathbf{s}^\tau \in \mathcal{S}\}$ from a spatial auto-linear scheme. We consider $\mathcal{S}^*$ to be the maximal set such that for every $\mathbf{s}^\tau \in \mathcal{S}^*$, it holds that $\mathbf{s}^\tau - \mathbf{j}^\tau \in \mathcal{S}$ for all $\mathbf{j}^\tau \in \mathcal{F}$. Then we may define the maximal set $\mathcal{S}^{**}$, such that for every $\mathbf{s}^\tau \in \mathcal{S}^{**}$, it holds

that $\mathbf{s}^\tau - \mathbf{j}^\tau \in \mathcal{S}^*$ for all $\mathbf{j}^\tau \in \mathcal{F} \cap \{\mathbf{j}^\tau : \mathbf{j} \geq \mathbf{0}\}$. We assume that $\mathcal{S}^{**}$ is not the empty set. We write N, N^*, N^{**} for the cardinalities of $\mathcal{S}$, $\mathcal{S}^*$ and $\mathcal{S}^{**}$, respectively.

We will consider $\mathbf{j}_1^\tau < \cdots < \mathbf{j}_p^\tau$ the elements of the set $\mathcal{F}_0 \cap \{\mathbf{j}^\tau : \mathbf{j} > \mathbf{0}\}$, the vector

$$\boldsymbol{\beta} \equiv [\beta_{\mathbf{j}_1}, \cdots, \beta_{\mathbf{j}_p}]^\tau \tag{4.5.1}$$

and the parameter space $\mathcal{B} \subset \mathcal{R}^p$. We will use the following condition.

(C5) The parameter space $\mathcal{B}$ is a compact set containing the true value $\boldsymbol{\beta}_0$ as an inner point. Further, for any $\boldsymbol{\beta} \in \mathcal{B}$, it holds that $|\beta_{\mathbf{j}_n}| < 1$, $n = 1, \cdots, p$, and that the polynomial $[1 - \sum_{n=1}^p \beta_{\mathbf{j}_n}(\mathbf{z}^{\mathbf{j}_n} + \mathbf{z}^{-\mathbf{j}_n})]^{-1}$ has absolutely summable coefficients.

4.5.1 Conditions

In order to prove the properties of the different estimators of the parameters, we will use some extra conditions apart from (C5). These conditions involve the new (weakly) stationary process $\{Y(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$. Like in Proposition 4.1, we may define, for any $\mathbf{s}^\tau \in \mathcal{Z}^d$, the error of best linear prediction as

$$\begin{aligned} Y(\mathbf{s}) &\equiv \beta(\mathbf{B})X(\mathbf{s}) = X(\mathbf{s}) - \sum_{\mathbf{j}^\tau \in \mathcal{F}_0} \beta_{\mathbf{j}}\, X(\mathbf{s}-\mathbf{j}) \\ &= X(\mathbf{s}) - E\{X(\mathbf{s}) | X(\mathbf{s}-\mathbf{j}) = x(\mathbf{s}-\mathbf{j}),\ \mathbf{j}^\tau \in \mathcal{F}_0\}. \end{aligned} \tag{4.5.2}$$

The second equality holds only when the best linear predictors are also best predictors. According to (4.2.32), it holds that

$$\mathrm{Cov}\{Y(\mathbf{s}), Y(\mathbf{s}-\mathbf{i})\} = E\{Y(\mathbf{s})Y(\mathbf{s}-\mathbf{i})\} = \begin{cases} \nu, & \mathbf{i} = \mathbf{0} \\ -\beta_{\mathbf{i}} \cdot \nu, & \mathbf{i}^\tau \in \mathcal{F}_0 \\ 0, & \mathbf{i}^\tau \notin \mathcal{F} \end{cases} \tag{4.5.3}$$

and the spectral density function of $\{Y(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ is defined as in (4.2.37). For the set of finite cardinality $\mathcal{F}_0$ as considered before, equations (4.2.21) and (4.2.22) can be re-written as

$$\gamma_X(\mathbf{i}) - \sum_{\mathbf{j}^\tau \in \mathcal{F}_0} \beta_{\mathbf{j}}\, \gamma_X(\mathbf{j}-\mathbf{i}) = 0,\ \mathbf{i} \neq \mathbf{0}, \tag{4.5.4}$$

$$\gamma_X(\mathbf{0}) - \sum_{\mathbf{j}^\tau \in \mathcal{F}_0} \beta_{\mathbf{j}}\, \gamma_X(\mathbf{j}) = \nu, \tag{4.5.5}$$

and they are general Yule-Walker equations. Thus for $\mathbf{i} \neq \mathbf{0}$,

$$E\{Y(\mathbf{s})X(\mathbf{s}-\mathbf{i})\} = E\{(X(\mathbf{s}) - \sum_{\mathbf{j}^\tau \in \mathcal{F}_0} \beta_{\mathbf{j}}\, X(\mathbf{s}-\mathbf{j}))X(\mathbf{s}-\mathbf{i})\} = \gamma_X(\mathbf{i}) - \sum_{\mathbf{j}^\tau \in \mathcal{F}_0} \beta_{\mathbf{j}}\, \gamma_X(\mathbf{i}-\mathbf{j}) = 0 \tag{4.5.6}$$

and the random variables $Y(\mathbf{s}), X(\mathbf{s}-\mathbf{i})$ are uncorrelated. This is not surprising, since $Y(\mathbf{s})$ was defined as a prediction error. Also,

$$E\{Y(\mathbf{s})X(\mathbf{s})\} = E\{(X(\mathbf{s}) - \sum_{\mathbf{j}^\tau \in \mathcal{F}_0} \beta_{\mathbf{j}}\, X(\mathbf{s}-\mathbf{j}))X(\mathbf{s})\} = \gamma_X(\mathbf{0}) - \sum_{\mathbf{j}^\tau \in \mathcal{F}_0} \beta_{\mathbf{j}}\, \gamma_X(\mathbf{j}) = \nu. \tag{4.5.7}$$

We write the following conditions:

(C6) For any $\mathbf{s}^\tau \in \mathcal{Z}^d$, the conditional distribution of $Y(\mathbf{s})$ given all $X(\mathbf{s}-\mathbf{j}), \mathbf{j} \neq \mathbf{0}$, is the same as the distribution of $Y(\mathbf{s})$.

(C7) It holds that $\{X(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ is a strictly stationary sequence of random variables.

Condition (C6) generalizes the two properties that

$$\begin{aligned} & E\{Y(\mathbf{s})|X(\mathbf{s}-\mathbf{j}) = x(\mathbf{s}-\mathbf{j}),\ \mathbf{j} \neq \mathbf{0}\} \\ = \ & E\{X(\mathbf{s})|X(\mathbf{s}-\mathbf{j}) = x(\mathbf{s}-\mathbf{j}),\ \mathbf{j} \neq \mathbf{0}\} - \sum_{\mathbf{j}^\tau \in \mathcal{F}_0} \beta_{\mathbf{j}}\, x(\mathbf{s}-\mathbf{j}) = 0 \end{aligned} \tag{4.5.8}$$

and

$$\mathrm{Var}\{Y(\mathbf{s})|X(\mathbf{s}-\mathbf{j}) = x(\mathbf{s}-\mathbf{j}),\ \mathbf{j} \neq \mathbf{0}\} = \mathrm{Var}\{X(\mathbf{s})|X(\mathbf{s}-\mathbf{j}) = x(\mathbf{s}-\mathbf{j}),\ \mathbf{j} \neq \mathbf{0}\} = \nu, \tag{4.5.9}$$

i.e. that the conditional expectation and variance of $Y(\mathbf{s})$ given $X(\mathbf{s}-\mathbf{j}), \mathbf{j} \neq \mathbf{0}$, do not depend on the values $x(\mathbf{s}-\mathbf{j}), \mathbf{j} \neq \mathbf{0}$. We may separate (C6) into two different parts, as we will often be needing only one of them:

(C6)(i) For any $\mathbf{s}^\tau \in \mathcal{Z}^d$, $\mathbf{j} > \mathbf{0}$, the random variables $Y(\mathbf{s})$ and $X(\mathbf{s}+\mathbf{j})$ are independent.

(C6)(ii) For any $\mathbf{s}^\tau \in \mathcal{Z}^d$, $\mathbf{j} > \mathbf{0}$, the random variables $Y(\mathbf{s})$ and $X(\mathbf{s}-\mathbf{j})$ are independent.

In other words, (C6) says that, for any $\mathbf{s}^\tau \in \mathcal{Z}^d$, the two random variables $X(\mathbf{s})$ and $Y(\mathbf{s}-\mathbf{j}), \mathbf{j} \neq \mathbf{0}$, are independent. As a result, for

$$Y(\mathbf{s}) = X(\mathbf{s}) - \sum_{\mathbf{j}^\tau \in \mathcal{F}_0} \beta_{\mathbf{j}}\, X(\mathbf{s}-\mathbf{j})$$

it holds that

$$\begin{aligned} 0 = E\{Y(\mathbf{s})\} &= E\{Y(\mathbf{s})|X(\mathbf{s}-\mathbf{j}) = x(\mathbf{s}-\mathbf{j}),\ \mathbf{j} \neq \mathbf{0}\} \\ &= E\{X(\mathbf{s}) - \sum_{\mathbf{j}^\tau \in \mathcal{F}_0} \beta_{\mathbf{j}}\ X(\mathbf{s}-\mathbf{j})|X(\mathbf{s}-\mathbf{j}) = x(\mathbf{s}-\mathbf{j}),\ \mathbf{j} \neq \mathbf{0}\} \\ &= E\{X(\mathbf{s})|X(\mathbf{s}-\mathbf{j}) = x(\mathbf{s}-\mathbf{j}),\ \mathbf{j} \neq \mathbf{0}\} - \sum_{\mathbf{j}^\tau \in \mathcal{F}_0} \beta_{\mathbf{j}}\ x(\mathbf{s}-\mathbf{j}), \end{aligned}$$

from which we may immediately derive (4.3.1). From (C6), it also holds that

$$\nu = \mathrm{Var}\{Y(\mathbf{s})\} = \mathrm{Var}\{Y(\mathbf{s})|X(\mathbf{s}-\mathbf{j}) = x(\mathbf{s}-\mathbf{j}),\ \mathbf{j} \neq \mathbf{0}\},$$

according to (4.5.3), which implies (4.3.2). We may conclude that condition (C6) also implies that the (weakly) stationary process $\{X(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ form an auto-linear scheme.

If we combine the conditions (C6) and (C7) with (4.5.2), then it holds that $\{Y(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ is a strictly stationary but also K-dependent process, for some positive and finite integer K. Actually, if we keep (C6)(i) or (C6)(ii) only and combine it with (C7), then we may come up with the same conclusion, as we are going to see later in the proofs of the theorems for the properties of estimators.

Whether it is possible to satisfy condition (C7) for an auto-linear scheme is a question of interest, as (C7) is very strong given that we have first required for (4.3.1) and (4.3.2) to hold. Since (C6) implies the auto-linear formulation, we may understand that imposing both (C6) and (C7) is a very strong requirement, indeed. When $\{X(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ form a (weakly) stationary auto-normal scheme, by definition, the equations (4.3.1) and (4.3.2) and, by consequence, conditions (C6) and (C7) are satisfied. Conditions (C6) and (C7) were hinted when Besag (1975, p.192) proved the consistency of the least squares estimators for the parameters of an auto-normal scheme.

From now on, we will consider that our observations $\{X(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{S}\}$ have been collected from the (weakly) stationary process $\{X(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$, such that, for any $\mathbf{s}^\tau \in \mathcal{Z}^d$, the best linear predictor of $X(\mathbf{s})$ based on all other $X(\mathbf{s}-\mathbf{j}), \mathbf{j} \neq \mathbf{0}$, is equal to a finite sum

$$\widehat{X}(\mathbf{s}) \equiv \sum_{\mathbf{j}^\tau \in \mathcal{F}_0} \beta_{\mathbf{j}}\ X(\mathbf{s}-\mathbf{j}),$$

and it is not necessarily the best predictor of $X(\mathbf{s})$ based on $X(\mathbf{s}-\mathbf{j}), \mathbf{j} \neq \mathbf{0}$. The linear prediction error is defined by

$$Y(\mathbf{s}) \equiv X(\mathbf{s}) - \widehat{X}(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d.$$

Thus, we will propose estimators for the parameters of a class of (weakly) stationary processes, which includes all the processes with an auto-linear formulation. For example, any unilateral or bilateral auto-regression is included in that class, while we saw in Section 4.4 how hard it is to find an auto-regression that is an auto-linear scheme, unless it is a Gaussian process.

4.5.2 Conditional likelihood estimators

We write a conditional Gaussian likelihood, *i.e.* a likelihood of the observations, as if they have been generated from a Gaussian process. Since the likelihood is conditional, it uses conditional moments, and since it is Gaussian, for the second-order properties this simplifies to the auto-covariance function of the process. If (C6) holds, then these are the conditional expectation and conditional variance matrix of the vector of observations indeed, though the process might not be Gaussian.

A very similar form of the likelihood was proposed by Besag (1974) for the estimation of the parameters of an auto-normal scheme, but the properties of the estimators were not found, as the likelihood included a determinant that was not easy to compute. Here, we propose a computational solution for that determinant using the innovations algorithm. Moreover, in Besag's likelihood, the quadratic form was a linear function of the parameters of interest. We show that this likelihood could be useful, if a different parametrization has been achieved and the process has been expressed as a finite auto-regression.

First from the processes $\{X(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ and $\{Y(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ as defined in this chapter, and the sets of locations $\mathcal{S}$, $\mathcal{S}^*$ of Section 4.5, we define the random vectors $\mathbf{Y}^*, \mathbf{X}^*$ and $\mathbf{X}_0$ on the same locations like in Section 2.4.1 and the subsection on the conditional variance matrix of random variables from an auto-regression. When condition (C6) holds, we may directly from (2.4.55) and (2.4.56), write the conditional Gaussian likelihood of $\mathbf{X}^*$ given the realization $\mathbf{x}_0$ of the random vector $\mathbf{X}_0$ as

$$L(\boldsymbol{\beta}, \nu) \propto \frac{|\mathrm{Var}\{\mathbf{Y}^*\}|^{1/2}}{\nu^{N^*}} \cdot \exp\{-\frac{1}{2\nu^2}[(\mathbf{X}^* - E\{\mathbf{X}^*|\mathbf{X}_0 = \mathbf{x}_0\})^\tau \mathrm{Var}\{\mathbf{Y}^*\}\ (\mathbf{X}^* - E\{\mathbf{X}^*|\mathbf{X}_0 = \mathbf{x}_0\})]\}, \tag{4.5.10}$$

for any $\beta \in \mathcal{B}$ and $\nu \in (0, \infty)$. We use ν and ν^2 in (4.5.10), as it holds that

$$\mathrm{Var}\{\mathbf{X}^*|\mathbf{X}_0 = \mathbf{x}_0\}^{-1} = \nu^{-2} \cdot \mathrm{Var}\{\mathbf{Y}^*\}.$$

We may see that, using the same sequel as in Section 2.4.1. The general Yule-Walker equations have allowed us to use the results of this subsection, though the second-order properties of the process might not necessarily be described by a finite auto-regression.

Besag (1974, p.213) claimed that the right form for the conditional likelihood is

$$L^*(\beta, \nu) \propto \frac{|\mathrm{Var}\{\mathbf{Y}^*\}|^{1/2}}{\nu^{N^*}} \exp\{-\frac{1}{2\nu^2}[\mathbf{X}^{*\tau}\mathrm{Var}\{\mathbf{Y}^*\}\ \mathbf{X}^*]\}, \tag{4.5.11}$$

instead, and omitted the term $E\{\mathbf{X}^*|\mathbf{X}_0 = \mathbf{x}_0\} = \mathrm{Cov}\{\mathbf{X}^*, \mathbf{X}_0\}\ \mathrm{Var}\{\mathbf{X}_0\}^{-1}\mathbf{x}_0$. As a result, the quadratic form in (4.5.11) has been expressed as a linear function of the unknown parameters and all the derivatives of first order are not functions of the parameters any more. One has to resort to the determinant $|\mathrm{Var}\{\mathbf{Y}^*\}|$, for which the innovations algorithm might be used. This is because the inverse of the original conditional variance matrix is a variance matrix of random variables from a process, for which we know its second-order properties, according to (4.5.3). When we write the spectral density of $\{Y(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ as

$$g_Y(\boldsymbol{\omega}) = \frac{\nu}{(2\pi)^d}\ \beta(e^{i\boldsymbol{\omega}}) \equiv \frac{1}{(2\pi)^d}\ \frac{\sigma^2}{\varphi(e^{i\boldsymbol{\omega}}) \cdot \varphi(e^{-i\boldsymbol{\omega}})},\ \boldsymbol{\omega}^\tau \in [-\pi, \pi]^d, \tag{4.5.12}$$

where

$$\varphi(\mathbf{z}) \equiv 1 - \sum_{\mathbf{i}>\mathbf{0}} \varphi_{\mathbf{i}}\ \mathbf{z}^{\mathbf{i}},\ \sum_{\mathbf{i}>\mathbf{0}} |\varphi_{\mathbf{i}}| < \infty, \tag{4.5.13}$$

then according to the innovations algorithm, it holds that

$$|\mathrm{Var}\{\mathbf{Y}^*\}| \equiv \prod_{\mathbf{s}^\tau \in \mathcal{S}^*} r(\mathbf{s}, \beta, \nu), \tag{4.5.14}$$

where $r(\mathbf{s}, \beta, \nu)$ is the variance of $Y(\mathbf{s}) - \widetilde{Y}(\mathbf{s})$ and $\widetilde{Y}(\mathbf{s})$ is the best linear predictor of $Y(\mathbf{s})$ based on the random variables $Y(\mathbf{s} - \mathbf{i}), \mathbf{i} > \mathbf{0}, \mathbf{s}^\tau - \mathbf{i}^\tau \in \mathcal{S}^*$, only. If

$$\varphi(\mathbf{z})^{-1} \equiv 1 + \sum_{\mathbf{j}>\mathbf{0}} \Phi_{\mathbf{j}}\ \mathbf{z}^{\mathbf{j}},\ \sum_{\mathbf{j}>\mathbf{0}} |\Phi_{\mathbf{j}}| < \infty,$$

then using the same arguments as Yao and Brockwell (2006) for the case $d = 2$, it holds for a fixed element $\mathbf{s}^\tau \in \mathcal{S}^*$ that

$$r(\mathbf{s}, \beta, \nu) \to \sigma^2 \tag{4.5.15}$$

as $N \to \infty$ and (C1)(i) holds.

Let us now consider the special case when we have observations from a bilateral auto-regression $\{X(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ defined by

$$X(\mathbf{s}) + \sum_{n=1}^{q} \theta_{\mathbf{i}_n}\ X(\mathbf{s} - \mathbf{i}_n) \equiv W(\mathbf{s}),\ \{W(\mathbf{s})\} \sim WN(0, \sigma_W^2) \tag{4.5.16}$$

with $\mathbf{0} < \mathbf{i}_1 < \cdots < \mathbf{i}_q$, and the process also has a causal AR(∞) representation. We may define the polynomial

$$\theta(\mathbf{z}) \equiv 1 + \sum_{n=1}^{q} \theta_{\mathbf{i}_n}\ \mathbf{z}^{\mathbf{i}_n} \tag{4.5.17}$$

and the bilateral moving-average process $\{Y(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ by the equation

$$Y(\mathbf{s}) \equiv \theta(\mathbf{B})\theta(\mathbf{B}^{-1})X(\mathbf{s}) = \theta(\mathbf{B}^{-1})W(\mathbf{s}),\ \{W(\mathbf{s})\} \sim WN(0, \sigma_W^2). \tag{4.5.18}$$

From observations $\{X(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{S}\}$, we will find our modified likelihood estimators for $\boldsymbol{\theta}_0 \equiv [\theta_{\mathbf{i}_1,0}, \cdots, \theta_{\mathbf{i}_q,0}] \in \Theta$, by minimizing the quantity

$$l_X^* = \log \sigma_W^2 - \log c_\theta + \frac{Q_X^*}{N^*\ \sigma_W^2},\ \boldsymbol{\theta} \in \Theta,\ \sigma_W^2 > 0, \tag{4.5.19}$$

where

$$\log c_\theta \equiv \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} \log\{\theta(e^{i\boldsymbol{\omega}})\ \theta(e^{-i\boldsymbol{\omega}})\} d\boldsymbol{\omega} \tag{4.5.20}$$

and

$$Q_X^* \equiv \sum_{\mathbf{s}^\tau \in \mathcal{S}^*} [\theta(\mathbf{B})\theta(\mathbf{B}^{-1})X(\mathbf{s})]\ X(\mathbf{s}) \tag{4.5.21}$$

or

$$Q_X^* \equiv \sum_{\mathbf{s}} [\theta(\mathbf{B})\theta(\mathbf{B}^{-1})X(\mathbf{s})]\ [\sum_{\mathbf{i}^\tau \in \mathcal{F}_\mathbf{s}} c(\mathbf{i})\ (\theta(\mathbf{B})\theta(\mathbf{B}^{-1})X(\mathbf{s} - \mathbf{i}))], \tag{4.5.22}$$

for the polynomial

$$c(\mathbf{z}) \equiv \sum_{\mathbf{i}^\tau \in \mathcal{Z}^d} c(\mathbf{i})\ \mathbf{z}^{\mathbf{i}} \equiv \{\theta(\mathbf{z})\ \theta(\mathbf{z}^{-1})\}^{-1} \tag{4.5.23}$$

and the set $\mathcal{F}_\mathbf{s}$, as it was defined in Section 3.5.3, for any $\mathbf{s}^\tau \in \mathcal{Z}^d$. The form (4.5.21) refers to the quadratic form

$$\mathbf{X}^{*\tau}[\mathrm{Var}\{\mathbf{Y}^*\}\mathbf{X}^*]$$

in (4.5.11). If we re-write the quadratic form as

$$[\mathrm{Var}\{\mathbf{Y}^*\}\mathbf{X}^*]^\tau \mathrm{Var}\{\mathbf{Y}^*\}^{-1}[\mathrm{Var}\{\mathbf{Y}^*\}\mathbf{X}^*],$$

then we may see the relevance with (4.5.22).

Quantities (4.5.19) and (4.5.21) or (4.5.22) closely resemble

$$l_Y^* = \log \sigma_W^2 + \log c_\theta + \frac{Q_Y^*}{N^* \sigma_W^2}, \ \boldsymbol{\theta} \in \Theta, \ \sigma_W^2 > 0, \tag{4.5.24}$$

and

$$Q_Y^* \equiv \sum_{\mathbf{s}^\tau \in \mathcal{S}^*} Y(\mathbf{s}) \, [\sum_{\mathbf{i}^\tau \in \mathcal{F}_\mathbf{s}} c(\mathbf{i}) \, Y(\mathbf{s}-\mathbf{i})], \tag{4.5.25}$$

respectively. The negative log-likelihood (4.5.24) will be appropriate to minimize, in order to derive the estimators for the same parameters, if we have observed $\{Y(\mathbf{s}), \ \mathbf{s}^\tau \in \mathcal{S}\}$ instead. If $\theta(\mathbf{z})$ is a causal polynomial and the moving-average of interest is invertible, then

$$\log c_\theta = 0$$

and we would generate the same estimators $\widehat{\boldsymbol{\theta}}$, which are also obtained by maximizing the Gaussian likelihood (3.4.70). As a result, (4.5.19) is the quantity to be minimized in the case of observations from a unilateral or bilateral auto-regression and is a special case of the results presented in Sections 3.5 and 3.6, if (4.5.22) has been used instead of (4.5.21).

4.5.3 Pseudo-likelihood and least squares estimators

The Gaussian pseudo-likelihood was introduced by Besag (1975, p.190) as an alternative technique for the statistical analysis of sets of random variables, for which the Gaussian likelihood was intractable. The pseudo-likelihood estimators are very easy to compute; in this section, we also discover some of their statistical properties for the case when they are used to estimate the parameters of an auto-linear formulation.

The representation (4.5.2) often becomes a source of confusion as it strongly resembles with the standard linear model. The fact that $\{\beta_\mathbf{j}, \ \mathbf{j}^\tau \in \mathcal{F}_0\}$ in the model are set to be the coefficients of the best linear predictor and that $\mathcal{F}_0$ is not a subset of $\{\mathbf{j}^\tau : \mathbf{j} > \mathbf{0}\}$, should be taken together as the only two reasons why $\{Y(\mathbf{s}), \ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ is not a sequence of uncorrelated random variables. Treating $\{Y(\mathbf{s}), \ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ as a sequence of uncorrelated random variables is mistaken and can mislead the analysis as, for example, when the pseudo-likelihood replaces the likelihood of the observations. Maximizing the Gaussian pseudo-likelihood is the same as minimizing the sum of squares $\sum_\mathbf{s} Y(\mathbf{s})^2$, but the random variables are not uncorrelated with each other. Equation (4.5.3) has been derived instead. On the other hand, the bilateral auto-regression $\{X^*(\mathbf{s}), \ \mathbf{s}^\tau \in \mathcal{Z}^d\}$

defined by the equation

$$\beta(\mathbf{B})X^*(\mathbf{s}) \equiv W(\mathbf{s}),\ \{W(\mathbf{s})\} \sim WN(0, \sigma_W^2), \tag{4.5.26}$$

does not share the same second-order properties as $\{X(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$, for which we can write

$$\beta(\mathbf{B})X(\mathbf{s}) = Y(\mathbf{s}),$$

i.e. it does not have the same best linear predictor coefficients. Regarding the finite bilateral auto-regressions, such as the one defined in (4.5.26), Whittle (1954) studied whether the least squares estimators of the parameters of interest are consistent. For (4.5.26), those estimators would be obtained after minimizing $\sum_{\mathbf{s}} W(\mathbf{s})^2$. More specifically, exactly as we described in Section 3.2.1, Whittle (1954) would re-write (4.5.26) as

$$X^*(\mathbf{s}) + \sum_{n=1}^{p-1} \frac{\beta_{\mathbf{j}_n}}{\beta_{\mathbf{j}_p}} X^*(\mathbf{s}+\mathbf{j}_n-\mathbf{j}_p) - \frac{1}{\beta_{\mathbf{j}_p}} X^*(\mathbf{s}-\mathbf{j}_p) + \sum_{n=1}^{p} \frac{\beta_{\mathbf{j}_n}}{\beta_{\mathbf{j}_p}} X^*(\mathbf{s}-\mathbf{j}_n-\mathbf{j}_p) = W^*(\mathbf{s}), \tag{4.5.27}$$

with

$$W^*(\mathbf{s}) \equiv -\frac{1}{\beta_{\mathbf{j}_p}} W(\mathbf{s}-\mathbf{j}_p) \tag{4.5.28}$$

and

$$\{W^*(\mathbf{s})\} \sim WN(0, \frac{\sigma_W^2}{\beta_{\mathbf{j}_p}^2}). \tag{4.5.29}$$

Then he would prove that the estimators of the auto-regressive parameters derived by minimizing $\sum_{\mathbf{s}} W^*(\mathbf{s})^2$ are not consistent. Although $\{W(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ and $\{W^*(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ form sequences of uncorrelated random variables now, it is the property that $W(\mathbf{s})$ is, in general, correlated to $X^*(\mathbf{s}-\mathbf{i}), \mathbf{i} \neq \mathbf{0}$ that is the problem for the estimation, since the auto-regression is not unilateral.

While (4.5.2) and (4.5.6) make it clear that $\{\beta_{\mathbf{j}},\ \mathbf{j}^\tau \in \mathcal{F}_0\}$ are the best linear predictor coefficients for the original process $\{X(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$, one has to write down the spectral density

$$g_{X^*}(\boldsymbol{\omega}) \equiv \frac{1}{(2\pi)^d} \frac{\sigma_W^2}{\beta(e^{i\boldsymbol{\omega}})\ \beta(e^{-i\boldsymbol{\omega}})},\ \boldsymbol{\omega}^\tau \in [-\pi, \pi]^d, \tag{4.5.30}$$

to discover there the best linear predictor coefficients of the auto-regression (4.5.26), according to Proposition 4.1. In other words, it is the bilateral moving-average defined as

$$Y^*(\mathbf{s}) \equiv \beta(\mathbf{B}^{-1})\beta(\mathbf{B})X^*(\mathbf{s}) = \beta(\mathbf{B})^2 X(\mathbf{s}) = \beta(\mathbf{B})W(\mathbf{s}), \tag{4.5.31}$$

that it is such that $Y^*(\mathbf{s})$ and $X^*(\mathbf{s}-\mathbf{i})$ are two uncorrelated random variables for any $\mathbf{i} \neq \mathbf{0}$.

Remark 4.2. When the process of interest has an absolutely summable auto-covariance function, the solution of the equations (4.5.6) is unique, according to Proposition 4.1, since $\mathcal{F}_0$ is now a set of finite cardinality. This is equivalent to Remark 2.3 for a (weakly) stationary auto-regression of finite order. As a result, we may rest that the variance matrix

$$\mathbf{\Gamma}_p^* \equiv [\mathrm{Cov}\{X(\mathbf{s}-\mathbf{j})+X(\mathbf{s}+\mathbf{j}), X(\mathbf{s}-\mathbf{i})+X(\mathbf{s}+\mathbf{i})\}]_{\mathbf{j}^\tau,\ \mathbf{i}^\tau \in \mathcal{F}_0,\ \mathbf{j},\mathbf{i}>\mathbf{0}} \tag{4.5.32}$$

defined for any $\mathbf{s}^\tau \in \mathcal{Z}^d$, is non-singular.

Properties of the pseudo-likelihood estimators

For any $\boldsymbol{\beta} \in \mathcal{B}$, we define

$$Y(\mathbf{s},\boldsymbol{\beta}) \equiv X(\mathbf{s}) - \sum_{\mathbf{j}^\tau \in \mathcal{F}_0} \beta_{\mathbf{j}}\, X(\mathbf{s}-\mathbf{j}) = X(\mathbf{s}) - \sum_{n=1}^{p} \beta_{\mathbf{j}_n}\, [X(\mathbf{s}-\mathbf{j}_n)+X(\mathbf{s}+\mathbf{j}_n)] \tag{4.5.33}$$

and we write

$$Y(\mathbf{s},\boldsymbol{\beta}_0) \equiv Y(\mathbf{s}). \tag{4.5.34}$$

For

$$\mathbf{X}(\mathbf{s}) \equiv [X(\mathbf{s}-\mathbf{j}_1)+X(\mathbf{s}+\mathbf{j}_1), \cdots, X(\mathbf{s}-\mathbf{j}_p)+X(\mathbf{s}+\mathbf{j}_p)]^\tau \tag{4.5.35}$$

and any $\boldsymbol{\beta} \in \mathcal{B}$, it holds that

$$Y(\mathbf{s},\boldsymbol{\beta}) = Y(\mathbf{s}) - \mathbf{X}^\tau(\mathbf{s})[\boldsymbol{\beta}-\boldsymbol{\beta}_0]. \tag{4.5.36}$$

We define the Gaussian pseudo-likelihood estimators

$$\boldsymbol{\beta}^* \equiv \mathrm{argmin}_{\boldsymbol{\beta}\in\mathcal{B}} \sum_{\mathbf{s}^\tau \in \mathcal{S}^*} Y(\mathbf{s},\boldsymbol{\beta})^2 \tag{4.5.37}$$

and we can see immediately that it holds that

$$\sum_{\mathbf{s}^\tau \in \mathcal{S}^*} Y(\mathbf{s},\boldsymbol{\beta}^*)\, [X(\mathbf{s}-\mathbf{j}_n)+X(\mathbf{s}+\mathbf{j}_n)] = 0,\ n = 1, \cdots, p. \tag{4.5.38}$$

Indeed, one can write the solution

$$\boldsymbol{\beta}^* = \{\sum_{\mathbf{s}^\tau \in \mathcal{S}^*} \mathbf{X}(\mathbf{s})\mathbf{X}^\tau(\mathbf{s})\}^{-1} \sum_{\mathbf{s}^\tau \in \mathcal{S}^*} X(\mathbf{s})\mathbf{X}(\mathbf{s}), \tag{4.5.39}$$

which imitates the equation

$$\mathbf{\Gamma}_p^* \cdot \boldsymbol{\beta}_0 = E\{X(\mathbf{s})\mathbf{X}(\mathbf{s})\}, \tag{4.5.40}$$

for the true parameter vector $\boldsymbol{\beta}_0 = [\beta_{\mathbf{j}_1,0}, \cdots, \beta_{\mathbf{j}_p,0}]^\tau \in \mathcal{B}$. Indeed, (4.5.40) holds, since from the equation

$$\sum_{n=1}^{p} \beta_{\mathbf{j}_n,0} \left[X(\mathbf{s}-\mathbf{j}_n) + X(\mathbf{s}+\mathbf{j}_n)\right] = X(\mathbf{s}) - Y(\mathbf{s}) \tag{4.5.41}$$

we can multiply by $X(\mathbf{s}-\mathbf{j}_m) + X(\mathbf{s}+\mathbf{j}_m), m = 1, \cdots, p$, and find the expected values.

As we are going to justify later in Theorem 4.2, if (C6)(i) (or (C6)(ii)) and (C7) hold, then for any $\mathbf{i}^\tau \in \mathcal{Z}^d$, we can write

$$\sum_{\mathbf{s}^\tau \in \mathcal{S}^*} X(\mathbf{s})X(\mathbf{s}-\mathbf{i})/N \xrightarrow{P} E\{X(\mathbf{s})X(\mathbf{s}-\mathbf{i})\}$$

as $N \to \infty$. The consistency of the estimators comes as an immediate consequence of that.

Moreover, we can come up with

$$N^{1/2}[\boldsymbol{\beta}^* - \boldsymbol{\beta}_0] = \left(\frac{N}{N^*}\right)^{1/2} \{\frac{1}{N^*} \sum_{\mathbf{s}^\tau \in \mathcal{S}^*} \mathbf{X}(\mathbf{s})\mathbf{X}^\tau(\mathbf{s})\}^{-1} \{N^{*-1/2} \sum_{\mathbf{s}^\tau \in \mathcal{S}^*} \mathbf{X}(\mathbf{s})Y(\mathbf{s})\}. \tag{4.5.42}$$

Again if (C6)(i) (or (C6)(ii)) and (C7) hold, we can write that

$$\frac{1}{N^*} \sum_{\mathbf{s}^\tau \in \mathcal{S}^*} \mathbf{X}(\mathbf{s})\mathbf{X}^\tau(\mathbf{s}) \xrightarrow{P} \mathbf{\Gamma}_p^*, \tag{4.5.43}$$

as $N \to \infty$. On the other hand, provided that $|E\{X(\mathbf{s})^3\}| < \infty$, as $N \to \infty$ and (C1),(C5)-(C7) hold

$$N^{*-1/2} \sum_{\mathbf{s}^\tau \in \mathcal{S}^*} \mathbf{X}(\mathbf{s})Y(\mathbf{s}) \xrightarrow{D} N(\mathbf{0}, 4\nu^2 \cdot \mathbf{I}_p), \tag{4.5.44}$$

with similar arguments like in Theorem 3.2. In (4.5.44) we come up with 4, since in the

(n,m)-th element of the variance matrix, for $n,m=1,\cdots,p$, there are four terms

$$\begin{aligned}
&\sum_{\mathbf{i}^\tau\in\mathcal{Z}^d}\mathrm{Cov}\{[X(\mathbf{s}-\mathbf{j}_n)+X(\mathbf{s}+\mathbf{j}_n)]Y(\mathbf{s}),[X(\mathbf{s}-\mathbf{j}_m+\mathbf{i})+X(\mathbf{s}+\mathbf{j}_m+\mathbf{i})]Y(\mathbf{s}+\mathbf{i})\}\\
=&\sum_{\mathbf{i}^\tau\in\mathcal{Z}^d}\mathrm{Cov}\{X(\mathbf{s}-\mathbf{j}_n)Y(\mathbf{s}),X(\mathbf{s}-\mathbf{j}_m+\mathbf{i})Y(\mathbf{s}+\mathbf{i})\}\\
+&\sum_{\mathbf{i}^\tau\in\mathcal{Z}^d}\mathrm{Cov}\{X(\mathbf{s}-\mathbf{j}_n)Y(\mathbf{s}),X(\mathbf{s}+\mathbf{j}_m+\mathbf{i})Y(\mathbf{s}+\mathbf{i})\}\\
+&\sum_{\mathbf{i}^\tau\in\mathcal{Z}^d}\mathrm{Cov}\{X(\mathbf{s}+\mathbf{j}_n)Y(\mathbf{s}),X(\mathbf{s}-\mathbf{j}_m+\mathbf{i})Y(\mathbf{s}+\mathbf{i})\}\\
+&\sum_{\mathbf{i}^\tau\in\mathcal{Z}^d}\mathrm{Cov}\{X(\mathbf{s}+\mathbf{j}_n)Y(\mathbf{s}),X(\mathbf{s}+\mathbf{j}_m+\mathbf{i})Y(\mathbf{s}+\mathbf{i})\}\\
=&\;2\sum_{\mathbf{i}^\tau\in\mathcal{Z}^d}\mathrm{Cov}\{X(\mathbf{s}-\mathbf{j}_n)Y(\mathbf{s}),X(\mathbf{s}-\mathbf{j}_m+\mathbf{i})Y(\mathbf{s}+\mathbf{i})\}\\
+&\;2\sum_{\mathbf{i}^\tau\in\mathcal{Z}^d}\mathrm{Cov}\{X(\mathbf{s}-\mathbf{j}_n)Y(\mathbf{s}),X(\mathbf{s}+\mathbf{j}_m+\mathbf{i})Y(\mathbf{s}+\mathbf{i})\}
\end{aligned}$$

and the proof for the identity matrix $\mathbf{I}_p$ in (4.5.44) is similar to the one in Theorem 3.2. Combining everything together, we can conclude

$$N^{1/2}[\boldsymbol{\beta}^*-\boldsymbol{\beta}_0]\xrightarrow{D}N(\mathbf{0},4\nu^2\,\boldsymbol{\Gamma}_p^{*-2}). \tag{4.5.45}$$

Thus, we can state the next theorem.

Theorem 4.1 (Asymptotic normality). Let the Gaussian and stationary process $\{\xi(\mathbf{s}),\ \mathbf{s}^\tau\in\mathcal{Z}^d\}$, such that

$$E\{\xi(\mathbf{s})|\xi(\mathbf{s}-\mathbf{j}),\ \mathbf{j}\neq\mathbf{0}\}=\sum_{n=1}^{p}\beta_{\mathbf{j}_n}[\xi(\mathbf{s}-\mathbf{j}_n)+\xi(\mathbf{s}+\mathbf{j}_n)]$$

and

$$\mathrm{Var}\{\xi(\mathbf{s})|\xi(\mathbf{s}-\mathbf{j}),\ \mathbf{j}\neq\mathbf{0}\}=1$$

and let the variance matrix

$$\mathbf{W}_p^*\equiv\mathrm{Var}\{[\xi(\mathbf{s}-\mathbf{j}_1)+\xi(\mathbf{s}+\mathbf{j}_1),\cdots,\xi(\mathbf{s}-\mathbf{j}_p)+\xi(\mathbf{s}+\mathbf{j}_p)]^\tau\}. \tag{4.5.46}$$

Then if $E\{X(\mathbf{s})^4\}<\infty$, under conditions (C1),(C5), (C6)(i) and (C7), it holds that

$$N^{1/2}[\boldsymbol{\beta}^*-\boldsymbol{\beta}_0]\xrightarrow{D}N(\mathbf{0},\boldsymbol{\Delta}) \tag{4.5.47}$$

as $N\to\infty$. Otherwise if $|E\{X(\mathbf{s})^3\}|<\infty$, under conditions (C1),(C5),(C6) and (C7), it holds that

$$N^{1/2}[\boldsymbol{\beta}^*-\boldsymbol{\beta}_0]\xrightarrow{D}N(\mathbf{0},4\,\mathbf{W}_p^{*-2}). \tag{4.5.48}$$

We are not going to prove this theorem, as the parts which are not obvious due to the previous derivations are similar to Theorem 4.3, which is proven in the next section. Since, it holds for $n, m = 1, \cdots, p$, that

$$\begin{aligned} & \mathrm{Cov}\{\xi(\mathbf{s}-\mathbf{j}_n)+\xi(\mathbf{s}+\mathbf{j}_n), \xi(\mathbf{s}-\mathbf{j}_m)+\xi(\mathbf{s}+\mathbf{j}_m)\} \\ = \; & \mathrm{Cov}\{\xi(\mathbf{s}-\mathbf{j}_n), \xi(\mathbf{s}-\mathbf{j}_m)\} + \mathrm{Cov}\{\xi(\mathbf{s}-\mathbf{j}_n), \xi(\mathbf{s}+\mathbf{j}_m)\} \\ + \; & \mathrm{Cov}\{\xi(\mathbf{s}+\mathbf{j}_n), \xi(\mathbf{s}-\mathbf{j}_m)\} + \mathrm{Cov}\{\xi(\mathbf{s}+\mathbf{j}_n), \xi(\mathbf{s}+\mathbf{j}_m)\} \\ = \; & 2\ \mathrm{Cov}\{\xi(\mathbf{s}-\mathbf{j}_n), \xi(\mathbf{s}-\mathbf{j}_m)\} + 2\ \mathrm{Cov}\{\xi(\mathbf{s}+\mathbf{j}_n), \xi(\mathbf{s}-\mathbf{j}_m)\} \\ = \; & 2\ \mathrm{Cov}\{\xi(\mathbf{s}-\mathbf{j}_n)+\xi(\mathbf{s}+\mathbf{j}_n), \xi(\mathbf{s}-\mathbf{j}_m)\}, \end{aligned}$$

then if we write

$$\mathbf{C}_p^* \equiv \mathrm{Cov}\{[\xi(\mathbf{s}-\mathbf{j}_1)+\xi(\mathbf{s}+\mathbf{j}_1), \cdots, \xi(\mathbf{s}-\mathbf{j}_p)+\xi(\mathbf{s}+\mathbf{j}_p)]^\tau, [\xi(\mathbf{s}-\mathbf{j}_1), \cdots, \xi(\mathbf{s}-\mathbf{j}_p)]\}, \quad (4.5.49)$$

it holds that

$$\mathbf{W}_p^* = 2\ \mathbf{C}_p^* \quad (4.5.50)$$

and that

$$\mathbf{\Delta} = \mathbf{C}_p^{*-2}. \quad (4.5.51)$$

4.5.4 Method of moments estimators

We write the method of moments estimators

$$\widehat{\boldsymbol{\beta}} = [\widehat{\beta}_{\mathbf{j}_1}, \cdots, \widehat{\beta}_{\mathbf{j}_p}]^\tau \quad (4.5.52)$$

and define them such that

$$-\widehat{\beta}_{\mathbf{j}_n} \cdot \nu \equiv \frac{1}{N^{**}} \sum_{\mathbf{s}^\tau \in \mathcal{S}^{**}} Y(\mathbf{s}-\mathbf{j}_n, \widehat{\boldsymbol{\beta}}) Y(\mathbf{s}, \widehat{\boldsymbol{\beta}}),\ n = 1, \cdots, p. \quad (4.5.53)$$

The equations above are used when we know the variance ν. If ν is also unknown, then from the polynomial

$$b(\mathbf{z}) \equiv \frac{\beta(\mathbf{z})}{\nu} \equiv \sum_{\mathbf{j}^\tau \in \mathcal{F}} b_{\mathbf{j}}\ \mathbf{z}^{\mathbf{j}} \quad (4.5.54)$$

we may define the process

$$Y_b(\mathbf{s}) \equiv b(\mathbf{B}) X(\mathbf{s}) \quad (4.5.55)$$

and proceed exactly like before to derive the estimators $\widehat{b}_0, \widehat{b}_{\mathbf{j}_n},\ n = 1, \cdots, p$, from the equations

$$\widehat{b}_0 \equiv \frac{1}{N^{**}} \sum_{\mathbf{s}^\tau \in \mathcal{S}^{**}} Y_b^2(\mathbf{s}, \widehat{\mathbf{b}}) \quad (4.5.56)$$

and

$$\widehat{b}_{\mathbf{j}_n} \equiv \frac{1}{N^{**}} \sum_{\mathbf{s}^\tau \in \mathcal{S}^{**}} Y_b(\mathbf{s}-\mathbf{j}_n, \widehat{\mathbf{b}}) Y_b(\mathbf{s}, \widehat{\mathbf{b}}),\ n = 1, \cdots, p. \tag{4.5.57}$$

From (4.5.54), we may then define the estimators of the coefficients

$$\widehat{\beta}_{\mathbf{j}_n} \equiv -\widehat{b}_{\mathbf{j}_n} \cdot \widehat{\nu}, \tag{4.5.58}$$

where the estimated variance is

$$\widehat{\nu} \equiv \frac{1}{\widehat{b}_{\mathbf{0}}}. \tag{4.5.59}$$

Next, we will prove the properties of the method of moments estimators when we do know in advance the variance ν. Otherwise, one may read Remark 4.3(ii).

Theorem 4.2 (Consistency). Under conditions (C5),(C6)(i) (or (C6)(ii)) and (C7), it holds that

$$\widehat{\boldsymbol{\beta}} \xrightarrow{P} \boldsymbol{\beta}_0$$

as $N \to \infty$.

Proof. For any $\boldsymbol{\beta} \in \mathcal{B}$ and any $n = 1, \cdots, p$, we define the quantity

$$\begin{aligned} Q_n(\boldsymbol{\beta}) &\equiv \beta_{\mathbf{j}_n} \cdot \nu + \frac{1}{N^{**}} \sum_{\mathbf{s}^\tau \in \mathcal{S}^{**}} Y(\mathbf{s}-\mathbf{j}_n, \boldsymbol{\beta}) Y(\mathbf{s}, \boldsymbol{\beta}) \\ &= \frac{1}{N^{**}} \sum_{\mathbf{s}^\tau \in \mathcal{S}^{**}} Y(\mathbf{s}) Y(\mathbf{s}-\mathbf{j}_n) \\ &- \frac{1}{N^{**}} \sum_{\mathbf{s}^\tau \in \mathcal{S}^{**}} [Y(\mathbf{s}-\mathbf{j}_n)\mathbf{X}^\tau(\mathbf{s}) + Y(\mathbf{s})\mathbf{X}^\tau(\mathbf{s}-\mathbf{j}_n)]\ [\boldsymbol{\beta} - \boldsymbol{\beta}_0] \\ &+ [\boldsymbol{\beta} - \boldsymbol{\beta}_0]^\tau [\frac{1}{N^{**}} \sum_{\mathbf{s}^\tau \in \mathcal{S}^{**}} \mathbf{X}(\mathbf{s}-\mathbf{j}_n)\mathbf{X}^\tau(\mathbf{s})]\ [\boldsymbol{\beta} - \boldsymbol{\beta}_0] + \beta_{\mathbf{j}_n} \cdot \nu. \end{aligned} \tag{4.5.60}$$

The estimators we have defined set

$$Q_n(\widehat{\boldsymbol{\beta}}) \equiv 0,\ n = 1, \cdots, p. \tag{4.5.61}$$

Thanks to the same argument used back in Proposition 2.6, the assumption of independence of $Y(\mathbf{s})$ with $X(\mathbf{s}+\mathbf{j}), \mathbf{j} > \mathbf{0}$ under (C6)(i) together with (C7), allows us to write both $\{X(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ and $\{Y(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ as linear combinations of the same process, say

$$\{\varepsilon(\mathbf{s})\} \sim IID(0, \sigma^2).$$

Then as $N \to \infty$, it holds that

$$\frac{1}{N^{**}} \sum_{\mathbf{s}^\tau \in \mathcal{S}^{**}} Y(\mathbf{s})Y(\mathbf{s}-\mathbf{j}_n) \xrightarrow{P} E\{Y(\mathbf{s})Y(\mathbf{s}-\mathbf{j}_n)\} = -\beta_{\mathbf{j}_n,0} \cdot \nu, \tag{4.5.62}$$

for any $n = 1, \cdots, p$. From (4.5.60), we can write immediately that

$$Q_n(\boldsymbol{\beta}_0) = \frac{1}{N^{**}} \sum_{\mathbf{s}^\tau \in \mathcal{S}^{**}} Y(\mathbf{s})Y(\mathbf{s}-\mathbf{j}_n) + \beta_{\mathbf{j}_n,0} \cdot \nu \xrightarrow{P} 0, \tag{4.5.63}$$

according to (4.5.62). Since $Q_n(\boldsymbol{\beta})$, $n = 1, \cdots, p$, are smooth functions of $\boldsymbol{\beta}$, equations (4.5.61) and (4.5.63) imply that

$$\boldsymbol{\beta} \xrightarrow{P} \boldsymbol{\beta}_0 \tag{4.5.64}$$

as $N \to \infty$. ■

Theorem 4.3 (Asymptotic normality). Under conditions (C1),(C5),(C6)(i) and (C7), it holds that

$$N^{1/2}[\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0] \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Delta})$$

as $N \to \infty$. Moreover if (C6)(ii) holds and $|E\{X(\mathbf{s})^3\}| < \infty$, the variance matrix $\boldsymbol{\Delta} = [\delta_{n,m}]_{n,m=1}^p$, is such that $\delta_{n,m}$ is equal to

$$\begin{cases} 1 + \sum_{\mathbf{j}\neq\mathbf{0}} \beta_{\mathbf{j}}^2 - 2\,\beta_{2\mathbf{j}_n} + \sum_{\mathbf{j}\neq\pm\mathbf{j}_n} \beta_{\mathbf{j}+\mathbf{j}_n} \cdot \beta_{\mathbf{j}-\mathbf{j}_n}, & n = m \\ \sum_{\mathbf{j}\neq-\mathbf{j}_n,\mathbf{j}_m} \beta_{\mathbf{j}+\mathbf{j}_n} \cdot \beta_{\mathbf{j}-\mathbf{j}_m} - 2(\beta_{\mathbf{j}_n-\mathbf{j}_m} + \beta_{\mathbf{j}_n+\mathbf{j}_m}) + \sum_{\mathbf{j}\neq\mathbf{0},\mathbf{j}_m-\mathbf{j}_n} \beta_{\mathbf{j}} \cdot \beta_{\mathbf{j}+\mathbf{j}_n-\mathbf{j}_m}, & n \neq m \end{cases}$$

Proof. From (4.5.60) and (4.5.61), we can write for $n = 1, \cdots, p$,

$$-[\widehat{\beta}_{\mathbf{j}_n} - \beta_{\mathbf{j}_n,0}] \cdot \nu - \frac{1}{N} \cdot \mathbf{D}_n[\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0] = \frac{1}{N} \sum_{\mathbf{s}^\tau \in \mathcal{S}^{**}} [Y(\mathbf{s}-\mathbf{j}_n)Y(\mathbf{s}) + \beta_{\mathbf{j}_n,0} \cdot \nu], \tag{4.5.65}$$

where

$$\mathbf{D}_n \equiv [D_{n,1}, \cdots, D_{n,p}] \tag{4.5.66}$$

and for any $n, m = 1, \cdots, p$, we define

$$\begin{aligned} D_{n,m} \equiv & \sum_{\mathbf{s}^\tau \in \mathcal{S}^{**}} \{[X(\mathbf{s}-\mathbf{j}_n-\mathbf{j}_m) + X(\mathbf{s}-\mathbf{j}_n+\mathbf{j}_m)]Y(\mathbf{s}) \\ & + [X(\mathbf{s}-\mathbf{j}_m) + X(\mathbf{s}+\mathbf{j}_m)]Y(\mathbf{s}-\mathbf{j}_n)\} + O_p(N||\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0||). \end{aligned} \tag{4.5.67}$$

Using the same argument as in Theorem 4.2, we can write as $N \to \infty$

$$\frac{D_{n,m}}{N} \xrightarrow{P} \begin{cases} -2 \cdot \nu, & n = m \\ 0, & n \neq m \end{cases} \tag{4.5.68}$$

and, so,

$$\frac{1}{N} \cdot \mathbf{D} \xrightarrow{P} -2 \cdot \nu \cdot \mathbf{I}_p, \tag{4.5.69}$$

where $\mathbf{I}_p$ is the identity matrix and

$$\mathbf{D}^\tau \equiv [\mathbf{D}_1^\tau, \cdots, \mathbf{D}_p^\tau]. \tag{4.5.70}$$

Stacking all the equations together, we can write them as

$$N^{1/2}[\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0] = \{-\nu \cdot \mathbf{I}_p - \frac{1}{N} \cdot \mathbf{D}\}^{-1} \left(\frac{N^{**}}{N}\right)^{1/2} N^{**-1/2} \sum_{\mathbf{s}^\tau \in \mathcal{S}^{**}} \mathbf{M}(\mathbf{s}), \tag{4.5.71}$$

where we define

$$\mathbf{M}(\mathbf{s}) \equiv \begin{bmatrix} Y(\mathbf{s} - \mathbf{j}_1) \\ \vdots \\ Y(\mathbf{s} - \mathbf{j}_p) \end{bmatrix} Y(\mathbf{s}) - \boldsymbol{\beta}_0 \cdot \nu. \tag{4.5.72}$$

It holds that

$$E\{\mathbf{M}(\mathbf{s})\} = \mathbf{0}. \tag{4.5.73}$$

Since $\{Y(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ is strictly stationary and K-dependent process, for any $\boldsymbol{\lambda} \in \mathcal{R}^p$, the process $\{\boldsymbol{\lambda}^\tau \mathbf{M}(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ is also a zero mean, strictly stationary and K^*-dependent process for some finite positive integer K^*. As a result, we can write as $N \to \infty$ and (C1) holds

$$N^{**-1/2} \sum_{\mathbf{s}^\tau \in \mathcal{S}^{**}} \boldsymbol{\lambda}^\tau \mathbf{M}(\mathbf{s}) \xrightarrow{D} N(0, \nu^2 \cdot \boldsymbol{\lambda}^\tau \boldsymbol{\Delta}\ \boldsymbol{\lambda}), \tag{4.5.74}$$

where the exact form of the variance matrix $\boldsymbol{\Delta}$ is given later.

Finally, from the Cramer-Wold device and (4.5.74), it holds that

$$N^{**-1/2} \sum_{\mathbf{s}^\tau \in \mathcal{S}^{**}} \mathbf{M}(\mathbf{s}) \xrightarrow{D} N(\mathbf{0}, \nu^2 \cdot \boldsymbol{\Delta}) \tag{4.5.75}$$

From (4.5.69), we can also write as $N \to \infty$

$$-\nu \cdot \mathbf{I}_p - \frac{1}{N} \cdot \mathbf{D} \xrightarrow{P} \nu \cdot \mathbf{I}_p. \tag{4.5.76}$$

Combining (4.5.71), (4.5.75),(4.5.76) and under (C1), it holds that

$$N^{1/2}[\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0] \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Delta}) \tag{4.5.77}$$

as $N \to \infty$.

The variance matrix of the estimators

First, we define the new processes

$$Y^*(\mathbf{s}) \equiv \frac{Y(\mathbf{s})}{\sqrt{\nu}}, \tag{4.5.78}$$

$$X^*(\mathbf{s}) \equiv \beta(\mathbf{B})^{-1}Y^*(\mathbf{s}), \tag{4.5.79}$$

and

$$\mathbf{M}^*(\mathbf{s}) \equiv \begin{bmatrix} Y^*(\mathbf{s}-\mathbf{j}_1) \\ \vdots \\ Y^*(\mathbf{s}-\mathbf{j}_p) \end{bmatrix} Y^*(\mathbf{s}) - \beta_0. \tag{4.5.80}$$

Then, we consider the minimal set of finite cardinality, say $\mathcal{F}^*$, such that $\mathcal{F} \subseteq \mathcal{F}^* \subset \mathcal{Z}^d$, and we can write

$$\mathrm{Cov}\{\mathbf{M}^*(\mathbf{s}), \mathbf{M}^*(\mathbf{s}+\mathbf{j})\} = \mathbf{O}_{p\times p},\ \mathbf{j}^\tau \notin \mathcal{F}^*. \tag{4.5.81}$$

It holds that

$$\mathbf{\Delta} = \sum_{\mathbf{j}^\tau \in \mathcal{F}^*} \mathrm{Cov}\{\mathbf{M}^*(\mathbf{s}), \mathbf{M}^*(\mathbf{s}+\mathbf{j})\}. \tag{4.5.82}$$

Similarly, we can define

$$\mathbf{M}_c^*(\mathbf{s}) \equiv \begin{bmatrix} Y^*(\mathbf{s}-\mathbf{j}_1) \\ \vdots \\ Y^*(\mathbf{s}-\mathbf{j}_p) \end{bmatrix} Y^*(\mathbf{s}) \tag{4.5.83}$$

and

$$\mathbf{\Delta} = \mathrm{Var}\{\mathbf{M}_c^*(\mathbf{s})\} + \sum_{\mathbf{j}>\mathbf{0}} [\mathrm{Cov}\{\mathbf{M}_c^*(\mathbf{s}), \mathbf{M}_c^*(\mathbf{s}+\mathbf{j})\} + \mathrm{Cov}^\tau\{\mathbf{M}_c^*(\mathbf{s}), \mathbf{M}_c^*(\mathbf{s}+\mathbf{j})\}]. \tag{4.5.84}$$

For convenience, we consider

$$\beta_{\mathbf{j}} = \begin{cases} \beta_{\mathbf{j}}, & \mathbf{j}^\tau \in \mathcal{F}_0 \\ 0, & \mathbf{j}^\tau \notin \mathcal{F} \end{cases}. \tag{4.5.85}$$

When $\mathbf{j} = \mathbf{0}$, the (n,m)-th element of $\mathrm{Var}\{\mathbf{M}_c^*(\mathbf{s})\}$ is equal to

$$\begin{aligned}
& \mathrm{Cov}\{Y^*(\mathbf{s}-\mathbf{j}_n)\ Y^*(\mathbf{s}), Y^*(\mathbf{s}-\mathbf{j}_m)\ Y^*(\mathbf{s})\} \\
= \ & \mathrm{Cov}\{Y^*(\mathbf{s}-\mathbf{j}_n)\ Y^*(\mathbf{s}), X^*(\mathbf{s}-\mathbf{j}_m)\ X^*(\mathbf{s})\} \\
- \ & \sum_{\mathbf{i}\neq\mathbf{0}} \beta_{\mathbf{i}}\ \mathrm{Cov}\{Y^*(\mathbf{s}-\mathbf{j}_n)\ Y^*(\mathbf{s}), X^*(\mathbf{s}-\mathbf{j}_m)\ X^*(\mathbf{s}-\mathbf{i})\} \\
- \ & \sum_{\mathbf{i}^*\neq\mathbf{0}} \beta_{\mathbf{i}^*}\ \mathrm{Cov}\{Y^*(\mathbf{s}-\mathbf{j}_n)\ Y^*(\mathbf{s}), X^*(\mathbf{s}-\mathbf{j}_m-\mathbf{i}^*)\ X^*(\mathbf{s})\} \\
+ \ & \sum_{\mathbf{i},\mathbf{i}^*\neq\mathbf{0}} \beta_{\mathbf{i}}\cdot\beta_{\mathbf{i}^*}\ \mathrm{Cov}\{Y^*(\mathbf{s}-\mathbf{j}_n)\ Y^*(\mathbf{s}), X^*(\mathbf{s}-\mathbf{j}_m-\mathbf{i}^*)\ X^*(\mathbf{s}-\mathbf{i})\}.
\end{aligned}$$

Now, since both (C6) and (C7) hold and thanks to Proposition 2.6, we can write, in general, that

$$\mathrm{Cov}\{Y^*(\mathbf{s}-\mathbf{j}_n)\ Y^*(\mathbf{s}), X^*(\mathbf{s}-\mathbf{j}_m)\ X^*(\mathbf{s})\} = \begin{cases} 1, & n=m \\ 0, & n\neq m \end{cases}$$

that

$$\sum_{\mathbf{i}\neq\mathbf{0}} \beta_{\mathbf{i}}\ \mathrm{Cov}\{Y^*(\mathbf{s}-\mathbf{j}_n)\ Y^*(\mathbf{s}), X^*(\mathbf{s}-\mathbf{j}_m)\ X^*(\mathbf{s}-\mathbf{i})\} = 0,$$

that

$$\sum_{\mathbf{i}^*\neq\mathbf{0}} \beta_{\mathbf{i}^*}\ \mathrm{Cov}\{Y^*(\mathbf{s}-\mathbf{j}_n)\ Y^*(\mathbf{s}), X^*(\mathbf{s}-\mathbf{j}_m-\mathbf{i}^*)\ X^*(\mathbf{s})\} = \begin{cases} 0, & n=m \\ \beta_{\mathbf{j}_n-\mathbf{j}_m}, & n\neq m \end{cases},$$

and finally that

$$\sum_{\mathbf{i},\mathbf{i}^*\neq\mathbf{0}} \beta_{\mathbf{i}}\cdot\beta_{\mathbf{i}^*}\ \mathrm{Cov}\{Y^*(\mathbf{s}-\mathbf{j}_n)\ Y^*(\mathbf{s}), X^*(\mathbf{s}-\mathbf{j}_m-\mathbf{i}^*)\ X^*(\mathbf{s}-\mathbf{i})\} = \beta_{\mathbf{j}_n}\cdot\beta_{-\mathbf{j}_m} = \beta_{\mathbf{j}_n}\cdot\beta_{\mathbf{j}_m}.$$

The same pattern can be used for the lags $\mathbf{j}>\mathbf{0}$, when we need look at the (n,m)-th element of the covariance matrix $\mathrm{Cov}\{\mathbf{M}_c^*(\mathbf{s}), \mathbf{M}_c^*(\mathbf{s}+\mathbf{j})\}$. This is equal to

$$\begin{aligned}
& \mathrm{Cov}\{Y^*(\mathbf{s}-\mathbf{j}_n)\ Y^*(\mathbf{s}), Y^*(\mathbf{s}+\mathbf{j}-\mathbf{j}_m)\ Y^*(\mathbf{s}+\mathbf{j})\} \\
= & \ \mathrm{Cov}\{Y^*(\mathbf{s}-\mathbf{j}_n)\ Y^*(\mathbf{s}), X^*(\mathbf{s}+\mathbf{j}-\mathbf{j}_m)\ X^*(\mathbf{s}+\mathbf{j})\} \\
- & \sum_{\mathbf{i}\neq\mathbf{0}} \beta_{\mathbf{i}}\ \mathrm{Cov}\{Y^*(\mathbf{s}-\mathbf{j}_n)\ Y^*(\mathbf{s}), X^*(\mathbf{s}+\mathbf{j}-\mathbf{j}_m)\ X^*(\mathbf{s}+\mathbf{j}-\mathbf{i})\} \\
- & \sum_{\mathbf{i}^*\neq\mathbf{0}} \beta_{\mathbf{i}^*}\ \mathrm{Cov}\{Y^*(\mathbf{s}-\mathbf{j}_n)\ Y^*(\mathbf{s}), X^*(\mathbf{s}+\mathbf{j}-\mathbf{j}_m-\mathbf{i}^*)\ X^*(\mathbf{s}+\mathbf{j})\} \\
+ & \sum_{\mathbf{i},\mathbf{i}^*\neq\mathbf{0}} \beta_{\mathbf{i}}\cdot\beta_{\mathbf{i}^*}\mathrm{Cov}\{Y^*(\mathbf{s}-\mathbf{j}_n)\ Y^*(\mathbf{s}), X^*(\mathbf{s}+\mathbf{j}-\mathbf{j}_m-\mathbf{i}^*)\ X^*(\mathbf{s}+\mathbf{j}-\mathbf{i})\}.
\end{aligned}$$

Again, it holds for $\mathbf{j}>\mathbf{0}$, that

$$\mathrm{Cov}\{Y^*(\mathbf{s}-\mathbf{j}_n)\ Y^*(\mathbf{s}), X^*(\mathbf{s}+\mathbf{j}-\mathbf{j}_m)\ X^*(\mathbf{s}+\mathbf{j})\} = 0,$$

that

$$\begin{aligned}
& \sum_{\mathbf{j}>\mathbf{0}}\sum_{\mathbf{i}\neq\mathbf{0}} \beta_{\mathbf{i}}\ \mathrm{Cov}\{Y^*(\mathbf{s}-\mathbf{j}_n)\ Y^*(\mathbf{s}), X^*(\mathbf{s}+\mathbf{j}-\mathbf{j}_m)\ X^*(\mathbf{s}+\mathbf{j}-\mathbf{i})\} \\
= & \begin{cases} \beta_{\mathbf{j}_n+\mathbf{j}_m}, & n\geq m \\ \beta_{\mathbf{j}_n+\mathbf{j}_m}+\beta_{\mathbf{j}_m-\mathbf{j}_n}, & n<m \end{cases},
\end{aligned}$$

that

$$\sum_{\mathbf{i}^*\neq\mathbf{0}} \beta_{\mathbf{i}^*}\ \mathrm{Cov}\{Y^*(\mathbf{s}-\mathbf{j}_n)\ Y^*(\mathbf{s}), X^*(\mathbf{s}+\mathbf{j}-\mathbf{j}_m-\mathbf{i}^*)\ X^*(\mathbf{s}+\mathbf{j})\} = 0,$$

and finally that

$$\sum_{\mathbf{j}>\mathbf{0}} \sum_{\mathbf{i},\mathbf{i}^* \neq \mathbf{0}} \beta_{\mathbf{i}} \cdot \beta_{\mathbf{i}^*} \operatorname{Cov}\{Y^*(\mathbf{s}-\mathbf{j}_n)\, Y^*(\mathbf{s}), X^*(\mathbf{s}+\mathbf{j}-\mathbf{j}_m-\mathbf{i}^*)\, X^*(\mathbf{s}+\mathbf{j}-\mathbf{i})\}$$

$$= \begin{cases} \sum_{\mathbf{j}>\mathbf{0}} \beta_{\mathbf{j}} \cdot \beta_{\mathbf{j}+\mathbf{j}_n-\mathbf{j}_m} + \sum_{\substack{\mathbf{j}>\mathbf{0}, \\ \mathbf{j}\neq\mathbf{j}_m}} \beta_{\mathbf{j}+\mathbf{j}_n} \cdot \beta_{\mathbf{j}-\mathbf{j}_m}, & n \geq m \\ \sum_{\substack{\mathbf{j}>\mathbf{0}, \\ \mathbf{j}\neq\mathbf{j}_m-\mathbf{j}_n}} \beta_{\mathbf{j}} \cdot \beta_{\mathbf{j}+\mathbf{j}_n-\mathbf{j}_m} + \sum_{\substack{\mathbf{j}>\mathbf{0}, \\ \mathbf{j}\neq\mathbf{j}_m}} \beta_{\mathbf{j}+\mathbf{j}_n} \cdot \beta_{\mathbf{j}-\mathbf{j}_m}, & n < m \end{cases}$$

As a result, if we write $\delta_{n,m}$ for the (n,m)-th element of $\mathbf{\Delta}$, then it holds that it is equal to

$$\begin{cases} 1 + \sum_{\mathbf{j}\neq\mathbf{0}} \beta_{\mathbf{j}}^2 - 2\,\beta_{2\mathbf{j}_n} + \sum_{\mathbf{j}\neq\pm\mathbf{j}_n} \beta_{\mathbf{j}+\mathbf{j}_n} \cdot \beta_{\mathbf{j}-\mathbf{j}_n}, & n = m \\ \sum_{\mathbf{j}\neq-\mathbf{j}_n,\mathbf{j}_m} \beta_{\mathbf{j}+\mathbf{j}_n} \cdot \beta_{\mathbf{j}-\mathbf{j}_m} - 2(\beta_{\mathbf{j}_n-\mathbf{j}_m} + \beta_{\mathbf{j}_n+\mathbf{j}_m}) + \sum_{\mathbf{j}\neq\mathbf{0},\mathbf{j}_m-\mathbf{j}_n} \beta_{\mathbf{j}} \cdot \beta_{\mathbf{j}+\mathbf{j}_n-\mathbf{j}_m}, & n \neq m \end{cases} \tag{4.5.86}$$

■

Remark 4.3. (i) After (4.5.85), we may define

$$\rho_{\mathbf{j}} \equiv \begin{cases} -\beta_{\mathbf{j}}, & \mathbf{j} \neq \mathbf{0} \\ 1, & \mathbf{j} = \mathbf{0} \end{cases}. \tag{4.5.87}$$

and then re-write

$$\delta_{n,m} = \sum_{\mathbf{j}^\tau \in \mathcal{Z}^d} \rho_{\mathbf{j}} \cdot \rho_{\mathbf{j}+\mathbf{j}_n-\mathbf{j}_m} + \sum_{\mathbf{j}^\tau \in \mathcal{Z}^d} \rho_{\mathbf{j}+\mathbf{j}_n} \cdot \rho_{\mathbf{j}-\mathbf{j}_m}, \tag{4.5.88}$$

for all $n, m = 1, \cdots, p$. This formula has been obtained under the condition that $Y(\mathbf{s})$ and $X(\mathbf{s}-\mathbf{j})$ are independent for any $\mathbf{j} \neq \mathbf{0}$. The formula (4.5.88) strongly resembles formula (7.3.13) of Proposition 7.3.4 of Brockwell and Davis (1991, p.230) which gives the (n,m)-th element of a variance matrix equal to

$$(\eta - 3)\rho_{\mathbf{j}_n} \cdot \rho_{\mathbf{j}_m} + \sum_{\mathbf{j}^\tau \in \mathcal{Z}^d} \rho_{\mathbf{j}} \cdot \rho_{\mathbf{j}+\mathbf{j}_n-\mathbf{j}_m} + \sum_{\mathbf{j}^\tau \in \mathcal{Z}^d} \rho_{\mathbf{j}+\mathbf{j}_n} \cdot \rho_{\mathbf{j}-\mathbf{j}_m} \tag{4.5.89}$$

and η is such that

$$E\{\varepsilon(\mathbf{s})^4\} = \eta \cdot \sigma^4 < \infty. \tag{4.5.90}$$

The variance matrix there refers to the estimators of the theoretical auto-covariances of an invertible, one-dimensional moving-average process. The moving-average might have a finite or infinite order and $\{\varepsilon(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ is the sequence of uncorrelated random variables of interest.

The same proposition requires (4.5.90) as a condition, *i.e.* that the fourth moment of the sequence $\{\varepsilon(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ is finite, but it does not require condition (C6) like we did,

i.e. that the other error sequence, say, $\{u(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ is also a sequence of independent random variables. Of course, for Gaussian random variables we know that $\eta = 3$ and the two formulas give the same result. If the sequences are not Gaussian, one may always estimate the coefficients of best linear predictor and derive consistent and asymptotically normal estimators if (4.5.90) holds.

(ii) When we do not know the variance ν, we will need to proceed as follows. First, we will estimate $(p+1)$ unknown parameters according to (4.5.56) and (4.5.57). Then we will follow the same sequel as Brockwell and Davis (1991, p.221) for the Theorem 7.2.1. A condition on a finite fourth rather than third moment would be necessary this time, and we would expect to come up with Bartlett's formula for the elements of the variance matrix of the estimators.

4.6 Tests for spatial auto-linear schemes

4.6.1 Goodness of fit test

For a stationary Gaussian process $\{X(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ and a given set $\mathcal{F}_0$, which satisfies the restrictions mentioned before, we want to test whether we can write

$$E\{X(\mathbf{s})|X(\mathbf{s}-\mathbf{j}) = x(\mathbf{s}-\mathbf{j}),\ \mathbf{j}^\tau \in \mathcal{F}_0\} = \sum_{\mathbf{j}^\tau \in \mathcal{F}_0} \beta_{\mathbf{j}}\ x(\mathbf{s}-\mathbf{j}) \tag{4.6.1}$$

and

$$\mathrm{Var}\{X(\mathbf{s})|X(\mathbf{s}-\mathbf{j}) = x(\mathbf{s}-\mathbf{j}),\ \mathbf{j}^\tau \in \mathcal{F}_0\} = \nu, \tag{4.6.2}$$

where ν is a known constant and $\{\beta_{\mathbf{j}},\ \mathbf{j}^\tau \in \mathcal{F}_0\}$ are non-zero coefficients.

We write

$$\begin{aligned} H_0 &: \beta_{\mathbf{j},0} = 0,\ \mathbf{j}^\tau \in \mathcal{F}_0 \\ H_1 &: \text{otherwise} \end{aligned}$$

Under the null hypothesis, it holds that $\mathcal{F}_0 = \emptyset$ and we can write under (C1),

$$N^{1/2}\widehat{\boldsymbol{\beta}} \xrightarrow{D} N(\mathbf{0}, \mathbf{I}_p) \tag{4.6.3}$$

and

$$N \cdot \widehat{\boldsymbol{\beta}}^\tau \widehat{\boldsymbol{\beta}} \xrightarrow{D} \chi^2_p \tag{4.6.4}$$

as $N \to \infty$.

The test presented above is usually known as the Portmanteau test as it tests whether the auto-correlations of the transformed sequence $\{Y(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ are all zero. If $\beta_{\mathbf{j},0} = 0$ for all $\mathbf{j}^\tau \in \mathcal{Z}^d$, then $\{X(\mathbf{s}),\ \mathbf{s}^\tau \in \mathcal{Z}^d\}$ is a sequence of independent random variables. We expect our estimators $\widehat{\beta}_\mathbf{j}$, $\mathbf{j}^\tau \in \mathcal{F}_0$, to be close enough to 0 and the value

$$Y(\mathbf{s}, \widehat{\boldsymbol{\beta}}) \equiv X(\mathbf{s}) - \sum_{\mathbf{j}^\tau \in \mathcal{F}_0} \widehat{\beta}_\mathbf{j}\, X(\mathbf{s}-\mathbf{j}) \tag{4.6.5}$$

to be close enough to $X(\mathbf{s})$. Under the null hypothesis, the original and transformed series are the same and they share the same auto-correlation function.

4.6.2 Test for zero coefficients

We consider the set $\mathcal{F}_1 \subset \{\mathcal{F}_0 \cap \{\mathbf{j}^\tau : \mathbf{j} > \mathbf{0}\}$ with cardinality p_1 and we are interested in testing

$$\begin{aligned} H_0 &: \beta_{\mathbf{j},0} = 0,\ \mathbf{j}^\tau \in \mathcal{F}_1 \\ H_1 &: \text{otherwise} \end{aligned}$$

We may write the estimators $\widehat{\boldsymbol{\beta}} = [\widehat{\boldsymbol{\beta}}_1^\tau, \widehat{\boldsymbol{\beta}}_2^\tau]^\tau$, where $\widehat{\boldsymbol{\beta}}_1$ refer to the lags $\mathbf{j}^\tau \in \mathcal{F}_1$ and $\widehat{\boldsymbol{\beta}}_2$ to the remaining lags. Under the null hypothesis and as $N \to \infty$, it holds that

$$N^{1/2}\widehat{\boldsymbol{\beta}}_1 \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Delta}_1), \tag{4.6.6}$$

where

$$\boldsymbol{\Delta}_1 = \left[\ \mathbf{I}_{p_1} \quad \mathbf{O}_{p_1 \times (p-p_1)}\ \right] \cdot \boldsymbol{\Delta} \cdot \begin{bmatrix} \mathbf{I}_{p_1} \\ \mathbf{O}_{(p-p_1)\times p_1} \end{bmatrix} \tag{4.6.7}$$

and $\boldsymbol{\Delta}$ is the covariance matrix of the estimators as defined before. And, of course, (4.6.6) can be turned to

$$N \cdot \widehat{\boldsymbol{\beta}}_1^\tau\ \boldsymbol{\Delta}_1^{-1}\ \widehat{\boldsymbol{\beta}}_1 \xrightarrow{D} \mathcal{X}^2_{p_1}. \tag{4.6.8}$$

Since the variance matrix $\boldsymbol{\Delta}_1$ involves the true but unknown coefficients $\beta_\mathbf{j}$, we may replace it by its consistent estimator $\widehat{\boldsymbol{\Delta}}_1$ by setting $\widehat{\beta}_\mathbf{j}$, instead. As a result, under (C1)

$$N \cdot \widehat{\boldsymbol{\beta}}_1^\tau\ \widehat{\boldsymbol{\Delta}}_1^{-1}\ \widehat{\boldsymbol{\beta}}_1 \xrightarrow{D} \mathcal{X}^2_{p_1} \tag{4.6.9}$$

as $N \to \infty$.

4.7 Application on the climate data

We have the observations $\{X(u,v),\ u = 1,\cdots,5,\ v = 1,\cdots,17\}$ used back in Section 3.7.3 for the last time recording $t = 100$ only. In other words, for the 8th of March 2002, we have available the measurements of sea level pressure on 85 regular spatial locations. Again, the longitudes are from 20 degrees West to 20 degrees East and the latitudes are from 50 to 60 degrees North.

Originally, we consider a stationary spatial process $\{X(u,v),\ u,v,\in \mathcal{Z}\}$. The fact that we have used the same dataset as in the previous chapter should not be confused with the fact that we are using two instead of three dimensions. In other words, we do not analyze the observations at a single time point, thinking that there are other time points as well. That would imply that we are using three dimensions and we need an appropriate setting for a stationary process on $\mathcal{Z}^3$. Instead, we are using two dimensions and the process of interest now takes place on $\mathcal{Z}^2$.

Figure 4.1 presents all the 85 realizations of $X(u,v)$, from which the overall mean has been subtracted. Although we have treated our series as a stationary process over $\mathcal{Z}^2$, the figure reveals that there are serious reasons to consider that this is not true. An additive main effect seems to be taking place over the dimension 'North-South' and it is probably affecting the expected values of the process. Similarly, a periodical component seems to be related to the first-order properties of the process with respect to the dimension 'East-West'. In such cases, a more analytical study is required to come up with a stationary process from the original one, and then to apply the methods introduced.

There are $N = 5 \times 17 = 85$ observations in $\mathcal{S}$ which are reduced to $N^* = 3 \times 15 = 45$ elements of the set $\mathcal{S}^*$, when we exclude the extreme cases 60 and 50 degrees 'North' from the latitude vector and -20 and 20 degrees 'West' from the longitude vector. This is because we will fit the first-order scheme, as it will be described next. We consider the primary dimension to be the y-axis, *i.e.* the 'North-South' axis and the secondary dimension the 'East-West'. The directions 'South' and 'West' are considered as the 'past' of each dimension, respectively. This choice is not essential as the auto-linear scheme formulation is not dependent on any ordering on $\mathcal{Z}^2$. Still, we have mentioned this, as the final set of selected indexes $\mathcal{S}^{**}$ is comprised of $N^{**} = 2 \times 14 = 28$ elements excluding one more location from the 'past' of each dimension, *i.e.* the 52.5 degrees 'North' and 17.5 degrees 'West'.

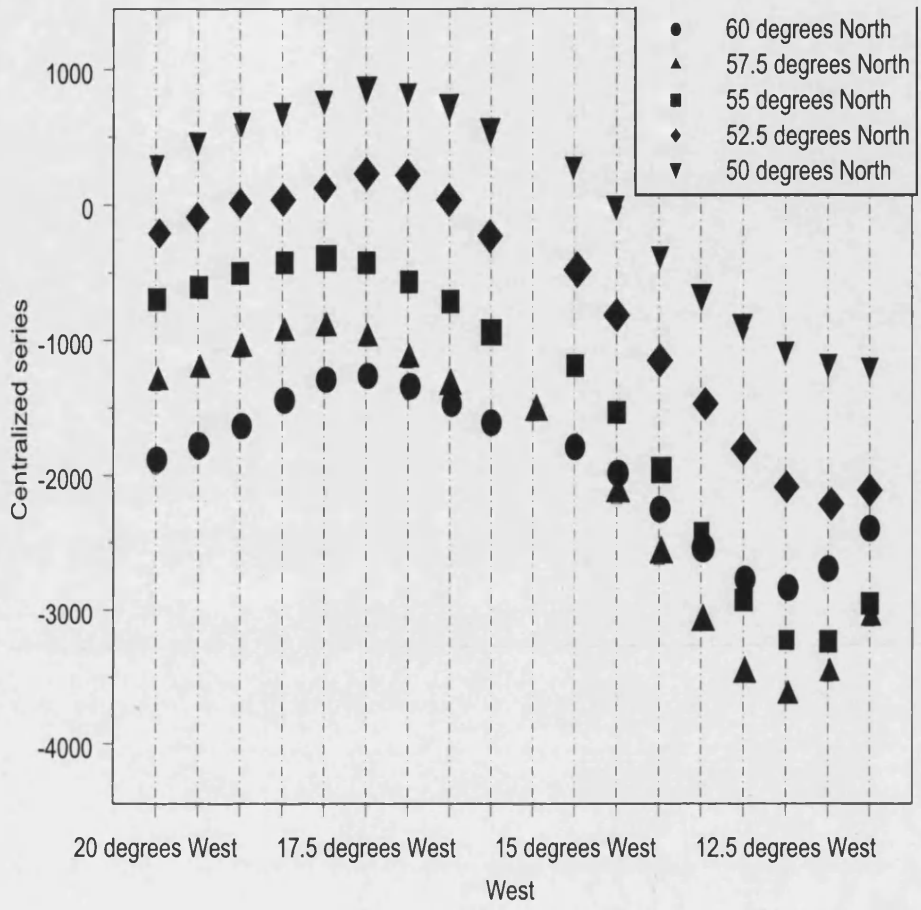


Figure 4.1: The centralized series on the 8th of March 2002 versus the 'East-West' axis for the different latitudes of the 'North-South' axis.

The 85 original observations had a sample mean $\bar{X} = 99,892$ and a sample variance $\widehat{\sigma}_X^2 = 1,352,616$. We created the centered version

$$X_c(u,v) = \frac{X(u,v) - \bar{X}}{100}.$$

We would like to see whether the first-order filter fits the data and whether we can write

$$\widehat{X}_c(u,v) = \beta_1 \left[X_c(u,v-1) + X_c(u,v+1)\right] + \beta_2 \left[X_c(u-1,v) + X_c(u+1,v)\right], \quad (4.7.1)$$

where $\widehat{X}_c(u,v)$ is the best linear predictor of $X_c(u,v)$ based on all $X_c(u-k,v-l), k,l \neq 0$ and the prediction variance of the original series is

$$\nu = 10,000 \cdot \text{Var}\{X_c(u,v) - \widehat{X}_c(u,v)\}. \tag{4.7.2}$$

We also require that the following condition is true

$$|\beta_1| + |\beta_2| < 0.5. \tag{4.7.3}$$

We saw back in Section 4.5.4, that when we do not know the conditional variance ν, we define our estimators according to (4.5.54) to (4.5.59). In order to find our estimators $\widehat{\beta}_1, \widehat{\beta}_2$ and $\widehat{\nu}$, for every $\beta_1, \beta_2 = -0.49, \cdots, 0.49$, such that $|\beta_1| + |\beta_2| < 0.5$ and every

$$\nu^* \equiv \nu/10,000 = 1, 2, \cdots, 200, \tag{4.7.4}$$

we follow the steps numbered next. The upper bound 200 for the conditional variance that we set in the parameter space, might be justified by the argument that ν is smaller than the actual variance which was estimated by the sample as $1,352,616$.

1. We set $b_0 \equiv 1/\nu^*$ and $b_j \equiv -\beta_j/\nu^*,\ j = 1,2$.

2. We compute $Y_b(u,v) \equiv b_0\ X_c(u,v) + b_1\ [X_c(u,v-1) + X_c(u,v+1)] + b_2\ [X_c(u-1,v) + X_c(u+1,v)]$ for all $[u,v]^\tau \in \mathcal{S}^*$.

3. We compute the sample auto-covariances

$$acvf_0 \equiv \sum_{[u,v]^\tau \in \mathcal{S}^{**}} Y_b^2(u,v)/N^{**}$$

and

$$\begin{aligned} acvf_1 &= \sum_{[u,v]^\tau \in \mathcal{S}^{**}} Y_b(u,v)Y_b(u,v-1)/N^{**} \\ acvf_2 &= \sum_{[u,v]^\tau \in \mathcal{S}^{**}} Y_b(u,v)Y_b(u-1,v)/N^{**}. \end{aligned}$$

4. We compute the quantity

$$S(\nu,\beta_1,\beta_2) \equiv |acvf_0 - b_0| + |acvf_1 - b_1| + |acvf_2 - b_2|. \tag{4.7.5}$$

Now, if we would select our estimators by finding the values that would minimize (4.7.5) over all the candidate values, we would be more likely to estimate ν as big as possible,

as the distances $|acvf_j - b_j|$, $j = 0, 1, 2$, depend on the variances of the original and transformed processes. That would distort the results. Indeed, for the transformed series $\{Y_b(u,v),\ u,v \in \mathcal{Z}\}$, the bigger the ν the smaller the actual variance of the process. As a result, we may replace (4.7.5) in Step 4 by

$$S^*(\nu, \beta_1, \beta_2) \equiv \left|\frac{acvf_0}{b_0} - 1\right| + \left|\frac{acvf_1}{b_1} - 1\right| + \left|\frac{acvf_2}{b_2} - 1\right| \tag{4.7.6}$$

and select our estimators $\widehat{\nu}, \widehat{\beta}_1, \widehat{\beta}_2$ that minimize (4.7.6).

We estimated the prediction variance as $\widehat{\nu} = 1,220,000$. The estimates for the coefficients β_1, β_2 might be found in Table 4.1. Table 4.1 demonstrates that there is no

Table 4.1: Estimated Prediction Coefficients and goodness of fit test.

Estimated coefficient $\widehat{\beta}$	$N^{**1/2}\ \widehat{\beta}$	Significance at 5% level
$\widehat{\beta}_1 = 0.48$	$28^{1/2} \cdot 0.48 = 2.54$	> 1.96
$\widehat{\beta}_2 = -0.01$	$28^{1/2} \cdot (-0.01) = -0.053$	> -1.96

need to move to the second-order model by including interaction terms and assuming that

$$\begin{aligned}
\widehat{X}_c(u,v) &= \beta_1 \left[X_c(u, v-1) + X_c(u, v+1)\right] + \beta_2 \left[X_c(u-1, v) + X_c(u+1, v)\right] \\
&+ \gamma_1 \left[X_c(u-1, v-1) + X_c(u+1, v+1)\right] \\
&+ \gamma_2 \left[X_c(u+1, v-1) + X_c(u-1, v+1)\right].
\end{aligned}$$

In the contrary, one might exclude the parameter β_2 from the model and ignore the influence of the neighbours that are located in the dimension 'South-North'. Still, the goodness of fit test generates

$$\mathcal{X}^2 = 2.54^2 + (-0.053)^2 = 6.454409$$

with observed significance level 3.97% using 2 degrees of freedom, which allows us to stop here and consider representation (4.7.1) appropriate. Let us not forget that under condition $|\beta_1| + |\beta_2| < 0.5$, it is not surprising that one of the two parameters will have to absorb and reflect all the discrepancies from the independence assumption, if we are expecting any significant results at all.

In conclusion, it seems that the sea level pressure follows a dependence over both the two dimensions of space. It is mainly for the dimension 'East-West' that the values of the process are associated; it could be considered that any dependence that is taking place over the 'South-North' axis is so weak, that it could be omitted.

Chapter 5

Modeling data observed irregularly over space and regularly in time

5.1 Introduction

When the data has been collected regularly over time and irregularly over space, it is difficult to impose an explicit auto-regressive structure over the space. In this chapter, we assume that we study a phenomenon on a number of fixed locations, and for each location the process forms an auto-regressive time series. The dependence over space is reflected by the covariance matrix of the noise process, which is 'white' in time but not over the space. We consider the asymptotic properties of our inference methods when the number of observations in time only tends to infinity.

We observe over time a multi-dimensional process $\{Y_t(\mathbf{s}_j),\ j = 1, \cdots, N,\ t \in \mathcal{Z}\}$, which takes place on any locations $\mathbf{s}_j^\tau \in \mathcal{R}^d,\ j = 1, \cdots, N$, for a positive integer number d. In order to model the serial dependence, for each fixed location $j = 1, \cdots, N$, we consider that $\{Y_t(\mathbf{s}_j),\ t \in \mathcal{Z}\}$ is an auto-regression caused by a white noise sequence $\{\varepsilon_t(\mathbf{s}_j),\ t \in \mathcal{Z}\}$. The N auto-regressions take place simultaneously on the different sites. For each fixed $t \in \mathcal{Z}$, we parameterize the spatial interdependence between the N sites, via the variance matrix of the random variables $\varepsilon_t(\mathbf{s}_j),\ j = 1, \cdots, N$, which is not a function of t. In Section 5.3.3, we use a Gaussian likelihood to estimate both the auto-regressive coefficients and the elements of the inverse covariance matrix. When

$\{[\varepsilon_t(\mathbf{s}_1), \cdots, \varepsilon_t(\mathbf{s}_N)]^\tau,\ t \in \mathcal{Z}\}$ form an N-variate sequence of independent and identically distributed random vectors, we may prove the consistency and asymptotic normality of estimators, as the number of observations increases over time.

The estimate of the inverse covariance matrix provides estimates for the coefficients of best linear predictors for a location $j = 1, \cdots, N$, based on all other locations $k \neq j,\ k = 1, \cdots, N$, and estimates for the prediction variances. As a result, the parametrization in terms of the inverse variance matrix is not only useful when writing down the Gaussian likelihood of observations, but also a meaningful interpretation regarding the dependence between the different sites, since it does not imply any unnatural ordering of the sites, as we are going to see in detail in Section 5.2.1. This is a similar idea to this used in Chapter 4, which allowed for naturally justified statistical analysis over space. While in Chapter 4 stationary spatial processes on the d-dimensional lattice $\mathcal{Z}^d$ were considered, here we only deal with random variables on N fixed sites on $\mathcal{R}^d$, which do not have to follow any pattern in their second-order properties. Finally, in Section 5.4 we proceed with hypotheses testing for the auto-regressive parameters, as well as for the coefficients of best linear predictors and for the prediction variances.

Other statistical analysis for data distributed regularly in time only, has been proposed before. Hjellvik and Tjøstheim (1999) also assumed n different auto-regressions to be taking place over time. Nevertheless, they embedded the interaction between the n sites, only into the first-order properties of the processes and, thus, they used additive models with n nuisance parameters, expressing the spatial dependence, and with sequences of uncorrelated random variables over space. As they focused on the estimation of the auto-regressive coefficients only, they studied the asymptotic properties of the estimators, as both the number of observations in time tends to infinity and the number of sites under observation $n \to \infty$. With their work, Zhang, Yao, Tong and Stenseth (2003) have escaped the auto-regressive structure at each post over time and they have assumed models, which are not necessarily linear. They have also clothed the temporal models at each post with spatially dependent noise, similarly to our approach. A non-parametric method of spatial smoothing, based on kernel functions, has been proposed there, and the results have been applied on the mink and muskrat Canadian fur sales. We have used the same dataset in Section 5.6, in order to apply our methods. There are two time series at each of 82 locations over a period of 25 years; via a bivariate time series at each post we have considered both of them to be dependent series with independent variables

some of their lagged values.

5.2 Spatial modeling

Before including time in our analysis, it would be interesting to focus on the form of spatial interdependence. For a given set of N different locations, say $\{\mathbf{s}_j^\tau \in \mathcal{R}^d,\ j = 1, \cdots, N\}$, we consider the zero-mean random vector

$$\mathbf{Y} = [Y(\mathbf{s}_1), \cdots, Y(\mathbf{s}_N)]^\tau. \tag{5.2.1}$$

We let

$$\gamma_{j,k} = E\{Y(\mathbf{s}_j)Y(\mathbf{s}_k)\} \tag{5.2.2}$$

and assume that $\gamma_{j,j} < \infty,\ j = 1, \cdots, N$, and that the matrix $\mathbf{V} \equiv \mathrm{Var}\{\mathbf{Y}\}$ is positive-definite. We do not make any other assumption on the form of dependence between the different sites. In the general case that the sites are not scattered on the d-dimensional lattice, we would expect the variance matrix to include up to $q = N(N+1)/2$ different elements. We write down the inverse variance matrix

$$\mathbf{V}^{-1} \equiv \begin{bmatrix} a_{1,1} & -a_{1,2} & \cdots & -a_{1,N} \\ -a_{1,2} & a_{2,2} & \cdots & -a_{2,N} \\ \vdots & & \ddots & \\ -a_{1,N} & -a_{2,N} & & a_{N,N} \end{bmatrix} \tag{5.2.3}$$

and let the two $(q \times 1)$ parameter vectors

$$\mathbf{a} = [a_{1,2}, \cdots, a_{1,N}, a_{2,3}, \cdots, a_{2,N}, \cdots, a_{N-1,N}, a_{1,1}, \cdots, a_{N,N}]^\tau \tag{5.2.4}$$

and

$$\boldsymbol{\gamma} = [\gamma_{1,2}, \cdots, \gamma_{1,N}, \gamma_{2,3}, \cdots, \gamma_{2,N}, \cdots, \gamma_{N-1,N}, \gamma_{1,1}, \cdots, \gamma_{N,N}]^\tau. \tag{5.2.5}$$

We consider $\mathcal{A} \subset \mathcal{R}^q$, such that the following condition is true:

(C8) For any $\mathbf{a} \in \mathcal{A}$, it holds that all the eigenvalues of the symmetric matrix $\mathbf{V}$ are positive and finite.

For observations $\{Y(\mathbf{s}_j),\ j = 1, \cdots, N\}$, we may write down the Gaussian likelihood as a function of the parameters

$$L(\mathbf{a}) = \frac{1}{(\sqrt{2\pi})^N} |\mathbf{V}^{-1}|^{1/2} \exp\{-\frac{1}{2}[\sum_{j=1}^{N} a_{j,j}\ Y(\mathbf{s}_j)^2 - 2 \sum_{\substack{j,k=1, \\ j<k}}^{N} a_{j,k}\ Y(\mathbf{s}_j)Y(\mathbf{s}_k)]\},\ \mathbf{a} \in \mathcal{A}. \tag{5.2.6}$$

If the random variables $\{Y(\mathbf{s}_j),\ j = 1, \cdots, N\}$ are Gaussian, then we may find the sufficient statistic

$$\begin{aligned}\mathbf{S} &= [Y(\mathbf{s}_1)Y(\mathbf{s}_2), \cdots, Y(\mathbf{s}_1)Y(\mathbf{s}_N), Y(\mathbf{s}_2)Y(\mathbf{s}_3), \cdots, Y(\mathbf{s}_2)Y(\mathbf{s}_N), \cdots, Y(\mathbf{s}_{N-1})Y(\mathbf{s}_N), \\ &\quad Y(\mathbf{s}_1)^2, \cdots, Y(\mathbf{s}_N)^2]^\tau, \end{aligned} \tag{5.2.7}$$

which gives the mean-value parametrization

$$E\{\mathbf{S}\} = \boldsymbol{\gamma}. \tag{5.2.8}$$

This implies that

$$I(\boldsymbol{\gamma}) \equiv \mathrm{Var}\{\mathbf{S}\}^{-1} \tag{5.2.9}$$

is the Fisher information matrix for $\boldsymbol{\gamma}$. The Fisher information matrix for the natural parameter vector $\mathbf{a}$ is then equal to

$$I(\mathbf{a}) \equiv \mathbf{J} \cdot I(\boldsymbol{\gamma}) \cdot \mathbf{J}^\tau, \tag{5.2.10}$$

where

$$\mathbf{J} \equiv \partial \boldsymbol{\gamma}^\tau / \partial \mathbf{a}. \tag{5.2.11}$$

5.2.1 Interpretation of the inverse covariance matrix and best linear predictors

Besag (1975) tried to find a natural interpretation for the elements $a_{j,k},\ j,k = 1, \cdots, N$, of the inverse dispersion matrix $\mathbf{V}^{-1}$, rather than treat them as abstract quantities. The following proposition reveals how these elements can often be interpreted in terms of conditional expectations and variances. Of course, in this chapter we have considered the indexes $j = 1, \cdots, N$, to refer to specific locations of $\mathcal{R}^d$; once more we should not forget the case where the index j might refer to the points of a time series on $\mathcal{Z}$, and the next proposition could then be related to the classical smoothing for time series.

Proposition 5.1. Let $\mathbf{Y} = [Y(\mathbf{s}_1), \cdots, Y(\mathbf{s}_N)]^\tau$ be a $N \times 1$ real-valued random vector such that $E\{\mathbf{Y}\} = \mathbf{0}$ and $\mathbf{V} = \mathrm{Var}\{\mathbf{Y}\}$ is a positive definite matrix with elements $E\{Y(\mathbf{s}_j)Y(\mathbf{s}_k)\} = \gamma_{j,k},\ j,k = 1, \cdots, N$, such that $\gamma_{j,j} < \infty,\ j = 1, \cdots, N$. Then, if

$$E\{Y(\mathbf{s}_j)|Y(\mathbf{s}_k) = y(\mathbf{s}_k),\ k = 1, \cdots, N,\ k \neq j\} = \sum_{\substack{k=1, \\ k \neq j}}^{N} \beta_{j,k}\, y(\mathbf{s}_k),\ j = 1, \cdots, N, \tag{5.2.12}$$

and

$$\mathrm{Var}\{Y(\mathbf{s}_j)|Y(\mathbf{s}_k)=y(\mathbf{s}_k),\ k=1,\cdots,N,\ k\neq j\}=\nu_j\in(0,\infty),\ j=1,\cdots,N, \quad (5.2.13)$$

it holds that

$$\mathbf{V}^{-1}=\mathbf{\Lambda}^{-1}\mathbf{B}, \quad (5.2.14)$$

where $\mathbf{\Lambda}=\mathrm{diag}[\nu_j]_{j=1}^N$ and $\mathbf{B}$ is the $N\times N$ matrix with diagonal elements 1 and elements $-\beta_{j,k}$ at the j-th row and k-th column, $j,k=1,\cdots,N,\ j\neq k$.

Proof. For any $j,k=1,\cdots,N$, and $j\neq k$, it holds that

$$\begin{aligned}\gamma_{j,k} &= E\{Y(\mathbf{s}_j)Y(\mathbf{s}_k)\}=E\{E\{Y(\mathbf{s}_j)Y(\mathbf{s}_k)|Y(\mathbf{s}_m)=y(\mathbf{s}_m),\ m=1,\cdots,N,\ m\neq j\}\}\\ &= E\{Y(\mathbf{s}_k)\ E\{Y(\mathbf{s}_j)|Y(\mathbf{s}_m)=y(\mathbf{s}_m),\ m=1,\cdots,N,\ m\neq j\}\}\\ &= E\{Y(\mathbf{s}_k)\sum_{\substack{m=1,\\ m\neq j}}^{N}\beta_{j,m}\ Y(\mathbf{s}_m)\}=\sum_{\substack{m=1,\\ m\neq j}}^{N}\beta_{j,m}\ E\{Y(\mathbf{s}_k)Y(\mathbf{s}_m)\}=\sum_{\substack{m=1,\\ m\neq j}}^{N}\beta_{j,m}\cdot\gamma_{m,k}\end{aligned}$$

or

$$\gamma_{j,k}-\sum_{\substack{m=1,\\ m\neq j}}^{N}\beta_{j,m}\cdot\gamma_{m,k}=0 \quad (5.2.15)$$

On the other hand, we can write for $j=1,\cdots,N$,

$$\begin{aligned}\nu_j &= E\{Y(\mathbf{s}_j)^2|Y(\mathbf{s}_k)=y(\mathbf{s}_k),\ k=1,\cdots,N,\ k\neq j\}\\ &- E\{Y(\mathbf{s}_j)|Y(\mathbf{s}_k)=y(\mathbf{s}_k),\ k=1,\cdots,N,\ k\neq j\}^2\\ &= E\{Y(\mathbf{s}_j)^2|Y(\mathbf{s}_k)=y(\mathbf{s}_k),\ k=1,\cdots,N,\ k\neq j\}\\ &- [\sum_{\substack{m=1,\\ m\neq j}}^{N}\beta_{j,m}\ y(\mathbf{s}_m)][\sum_{\substack{k=1,\\ k\neq j}}^{N}\beta_{j,k}\ y(\mathbf{s}_k)]\end{aligned}$$

or

$$\begin{aligned}E\{\nu_j\} &= E\{Y(\mathbf{s}_j)^2\}-\sum_{\substack{m=1,\\ m\neq j}}^{N}\beta_{j,m}\sum_{\substack{k=1,\\ k\neq j}}^{N}\beta_{j,k}\ E\{Y(\mathbf{s}_m)Y(\mathbf{s}_k)\}\\ &= \gamma_{j,j}-\sum_{\substack{m=1,\\ m\neq j}}^{N}\beta_{j,m}\sum_{\substack{k=1,\\ k\neq j}}^{N}\beta_{j,k}\cdot\gamma_{k,m}\end{aligned}$$

Since $\nu_j\in(0,\infty)$ and, according to (5.2.15), for $m\neq j$, it holds that $\sum_{k=1,k\neq j}^{N}\beta_{j,k}\cdot\gamma_{k,m}=\gamma_{j,m}$, we can conclude

$$\nu_j=\gamma_{j,j}-\sum_{\substack{m=1,\\ m\neq j}}^{N}\beta_{j,m}\cdot\gamma_{m,j}$$

or

$$\gamma_{j,j}\{1/\nu_j\} - \sum_{\substack{m=1,\\ m\neq j}}^{N} \gamma_{m,j} \cdot \beta_{j,m}\{1/\nu_j\} = 1. \tag{5.2.16}$$

Re-writing (5.2.15) we get

$$\gamma_{j,k}\{1/\nu_j\} - \sum_{\substack{m=1,\\ m\neq j}}^{N} \gamma_{m,k} \cdot \beta_{j,m}\{1/\nu_j\} = 0,\ j,k = 1,\cdots,N,\ j \neq k. \tag{5.2.17}$$

Now, if we write the matrix

$$\mathbf{\Lambda}^{-1}\mathbf{B} = \begin{bmatrix} 1/\nu_1 & -\beta_{1,2}/\nu_1 & \cdots & -\beta_{1,N}/\nu_1 \\ -\beta_{2,1}/\nu_2 & 1/\nu_2 & \cdots & -\beta_{2,N}/\nu_2 \\ & & \ddots & \\ -\beta_{N,1}/\nu_N & -\beta_{N,2}/\nu_N & \cdots & 1/\nu_N \end{bmatrix} \tag{5.2.18}$$

and multiply with $\mathbf{V}$ from the right side, we find the unit matrix $\mathbf{I}_N$ due to equations (5.2.16) and (5.2.17). ■

Remark 5.1. (i) Some consistency conditions must be imposed to make the scheme valid. The symmetry condition implies that

$$\beta_{j,k}/\nu_j = \beta_{k,j}/\nu_k,\ j,k = 1,\cdots,N,\ j \neq k, \tag{5.2.19}$$

whilst positive-definiteness can, in general, only be checked once the coefficients are known numerically (Besag, 1975).

(ii) Every positive-definite dispersion matrix $\mathbf{V}$, with finite elements in the diagonal, determines unique values for $\mathbf{B}$ and $\mathbf{\Lambda}$. Indeed, those elements might be found immediately if we write the unique inverse matrix $\mathbf{V}^{-1}$ as a product of two matrices $\mathbf{\Lambda}^{-1}\mathbf{B}$, where $\mathbf{\Lambda}$ is a diagonal matrix and $\mathbf{B}$ has elements unity in the main diagonal. Then the matrices $\mathbf{\Lambda}$ and $\mathbf{B}$ are unique, too. Still, that does not imply that the elements of $\mathbf{B}$ and $\mathbf{\Lambda}$ are necessarily coefficients of linear conditional expectations and conditional variances, respectively. After Remark 5.1, we will see exactly what those elements represent, in general, and how Proposition 5.1 is a special case of that.

(iii) In the Gaussian case, the class of all valid conditional schemes is, in fact, equivalent to that of all multivariate normal schemes and one can always find the conditional expectations (5.2.12) and conditional variances (5.2.13) from the unique decomposition $\mathbf{\Lambda}^{-1}\mathbf{B}$ of the inverse variance matrix. Thus, the conditional formulation does not imply any inherent simplification of the parameter space (Besag, 1975).

(iv) According to Besag (1975, pp.182-183), the equations (5.2.12) and (5.2.13) allow us to use the term auto-normal scheme for $\{Y(\mathbf{s}_j),\ j = 1, \cdots, N\}$, as they can be obtained by Gaussian random variables. This should be, by no means, confused with the (weakly) stationary auto-normal formulation, as defined in the previous chapter, where the conditional expectations and variances are the same for all sites $\mathbf{s}^\tau \in \mathcal{Z}^d$, given the values of all other points on the lattice. In other words, for any N Gaussian random variables $\{Y(\mathbf{s}_j),\ j = 1, \cdots, N\}$ with second-order properties, which are reflected in the covariances $\gamma_{j,k},\ j,k = 1, \cdots, N$, and in the variance matrix $\mathbf{V}$, the inverse variance matrix comes from Proposition 5.1, while Chapter 4 refers to the second-order properties of some (weakly) stationary processes, which might also be Gaussian.

Proposition 5.1 shows how the representations (5.2.12) and (5.2.13) result in the decomposition (5.2.14). Of course, even when (5.2.12) and (5.2.13) do not hold, the unique inverse variance matrix $\mathbf{V}^{-1}$ can be uniquely decomposed as

$$\mathbf{V}^{-1} = \mathbf{\Lambda}^{-1}\mathbf{B}, \tag{5.2.20}$$

where $\mathbf{\Lambda}$ is diagonal and $\mathbf{B}$ has 1 on the main diagonal, as we explained in Remark 5.1(ii). Next, we answer the question what do the elements of the matrices $\mathbf{B}$ and $\mathbf{\Lambda}$ represent in general.

For any location $j = 1, \cdots, N$, we define

$$\widehat{Y}(\mathbf{s}_j) \equiv \sum_{\substack{k=1,\\ k\neq j}}^{N} \beta_{j,k}\, Y(\mathbf{s}_k), \tag{5.2.21}$$

to be the best linear predictor of $Y(\mathbf{s}_j)$, based on all other sites available $k = 1, \cdots, N,\ k \neq j$, in the sense that

$$E\{Y(\mathbf{s}_j) - \widehat{Y}(\mathbf{s}_j)\}^2 \equiv \min_{\{\psi_{j,k}\}} E\{Y(\mathbf{s}_j) - \sum_{\substack{k=1,\\ k\neq j}}^{N} \psi_{j,k}\, Y(\mathbf{s}_k)\}^2. \tag{5.2.22}$$

Then it holds that

$$\mathrm{Cov}\{Y(\mathbf{s}_j) - \widehat{Y}(\mathbf{s}_j), Y(\mathbf{s}_k)\} = 0,\ k = 1, \cdots, N,\ k \neq j, \tag{5.2.23}$$

or, similarly,

$$\gamma_{j,k} - \sum_{\substack{m=1,\\ m\neq j}}^{N} \beta_{j,m} \cdot \gamma_{m,k} = 0,\ k = 1, \cdots, N,\ k \neq j. \tag{5.2.24}$$

On the other hand, we can define the prediction variances as

$$\nu_j \equiv \text{Var}\{Y(\mathbf{s}_j) - \widehat{Y}(\mathbf{s}_j)\} = \text{Cov}\{Y(\mathbf{s}_j) - \widehat{Y}(\mathbf{s}_j), Y(\mathbf{s}_j)\} = \gamma_{j,j} - \sum_{\substack{m=1,\\ m\neq j}}^{N} \beta_{j,m} \cdot \gamma_{m,j}, \quad (5.2.25)$$

for $j = 1, \cdots, N$. Now, we may re-write (5.2.24) and (5.2.25) as

$$\gamma_{j,k}\{1/\nu_j\} - \sum_{\substack{m=1,\\ m\neq j}}^{N} \gamma_{m,k} \cdot \beta_{j,m}\{1/\nu_j\} = 0, \; j,k = 1, \cdots, N, \; k \neq j, \quad (5.2.26)$$

$$\gamma_{j,j}\{1/\nu_j\} - \sum_{\substack{m=1,\\ m\neq j}}^{N} \gamma_{m,j} \cdot \beta_{j,m}\{1/\nu_j\} = 1, \; j = 1, \cdots, N. \quad (5.2.27)$$

Equations (5.2.26) and (5.2.27) do imply the decomposition (5.2.20), as when we multiply $\mathbf{\Lambda}^{-1}\mathbf{B}$ by $\mathbf{V}$ from the right side, we will come up with the identity matrix $\mathbf{I}_N$. Indeed, we may verify the symmetry condition, as follows. For any $j, k = 1, \cdots, N$, and $j \neq k$, it holds that

$$\begin{aligned}
\text{Cov}\{Y(\mathbf{s}_j) - \widehat{Y}(\mathbf{s}_j), Y(\mathbf{s}_k) - \widehat{Y}(\mathbf{s}_k)\} &= \text{Cov}\{Y(\mathbf{s}_j), Y(\mathbf{s}_k) - \widehat{Y}(\mathbf{s}_k)\} \\
&- \text{Cov}\{\widehat{Y}(\mathbf{s}_j), Y(\mathbf{s}_k) - \widehat{Y}(\mathbf{s}_k)\} \\
&= -\text{Cov}\{\widehat{Y}(\mathbf{s}_j), Y(\mathbf{s}_k) - \widehat{Y}(\mathbf{s}_k)\} \\
&= -\beta_{j,k} \, \text{Cov}\{Y(\mathbf{s}_k), Y(\mathbf{s}_k) - \widehat{Y}(\mathbf{s}_k)\} \\
&= -\beta_{j,k} \, \text{Var}\{Y(\mathbf{s}_k) - \widehat{Y}(\mathbf{s}_k)\} = -\beta_{j,k} \cdot \nu_k,
\end{aligned}$$

due to the fact that $Y(\mathbf{s}_k) - \widehat{Y}(\mathbf{s}_k)$ is uncorrelated to $Y(\mathbf{s}_j)$ and that $\widehat{Y}(\mathbf{s}_k)$ is a linear function of $Y(\mathbf{s}_m)$, $m = 1, \cdots, N$, $m \neq k$, and so, uncorrelated to $Y(\mathbf{s}_k) - \widehat{Y}(\mathbf{s}_k)$ too. In the same way, we can write that

$$\text{Cov}\{Y(\mathbf{s}_j) - \widehat{Y}(\mathbf{s}_j), Y(\mathbf{s}_k) - \widehat{Y}(\mathbf{s}_k)\} = -\beta_{k,j} \cdot \nu_j,$$

and combine both to derive that

$$\beta_{j,k} \cdot \nu_k = \beta_{k,j} \cdot \nu_j. \quad (5.2.28)$$

Equations (5.2.26) and (5.2.27) together with (5.2.28) perform the decomposition of the inverse dispersion matrix ($\mathbf{V}^{-1}$) now in terms of coefficients of the best linear predictors of each random variable based on all other variables ($\mathbf{B}$) and the prediction variances ($\mathbf{\Lambda}$).

5.2.2 Computation of the inverse covariance matrix and the innovations algorithm

Suppose that for a set of observations $\{Y(\mathbf{s}_j),\ j = 1, \cdots, N\}$, we want to write down the exact Gaussian likelihood (5.2.6), when we know the covariances between the random variables of interest, expressed in the parameter vector $\boldsymbol{\gamma}$. This involves the computation of the inverse variance matrix $\mathbf{V}^{-1}$ and its determinant.

We have seen before how the unique decomposition (5.2.20) generates the equations (5.2.24) and (5.2.25). One has to take three steps. The first step is to find the coefficients $\beta_{j,m},\ j, m = 1, \cdots, N,\ j \neq m$, by solving the equations in (5.2.24). Next, from (5.2.25) to give the solutions for ν_j immediately. The third step gives the elements of the inverse dispersion matrix

$$a_{j,k} = \begin{cases} 1/\nu_j, & j = k \\ \beta_{j,k}/\nu_j = \beta_{k,j}/\nu_k, & j \neq k \end{cases} . \tag{5.2.29}$$

Looking back at the first step, one has to find the $N(N-1)$ different coefficients from the same number of equations. Actually, this can be simplified to N independent linear systems, $j = 1, \cdots, N$, with $N-1$ unknowns each. For example, for $j = 1$ the first system is

$$\begin{bmatrix} \gamma_{2,2} & \gamma_{3,2} & \cdots & \gamma_{N,2} \\ \gamma_{2,3} & \gamma_{3,3} & & \gamma_{N,3} \\ & & \ddots & \\ \gamma_{2,N} & \gamma_{3,N} & \cdots & \gamma_{N,N} \end{bmatrix} \cdot \begin{bmatrix} \beta_{1,2} \\ \beta_{1,3} \\ \vdots \\ \beta_{1,N} \end{bmatrix} = \begin{bmatrix} \gamma_{1,2} \\ \gamma_{1,3} \\ \vdots \\ \gamma_{1,N} \end{bmatrix} \tag{5.2.30}$$

and then

$$\begin{bmatrix} \beta_{1,2} \\ \beta_{1,3} \\ \vdots \\ \beta_{1,N} \end{bmatrix} = \begin{bmatrix} \gamma_{2,2} & \gamma_{3,2} & \cdots & \gamma_{N,2} \\ \gamma_{2,3} & \gamma_{3,3} & & \gamma_{N,3} \\ & & \ddots & \\ \gamma_{2,N} & \gamma_{3,N} & \cdots & \gamma_{N,N} \end{bmatrix}^{-1} \cdot \begin{bmatrix} \gamma_{1,2} \\ \gamma_{1,3} \\ \vdots \\ \gamma_{1,N} \end{bmatrix}, \tag{5.2.31}$$

which involves the inverse of a $(N-1) \times (N-1)$ covariance matrix this time. One may proceed the same way by eliminating one row and one column each time until the matrix is partitioned into 1×1 elements. This method for the computation of the inverse of a non-singular matrix is mainly known from mathematics and, here, it enjoys the privilege of a statistical interpretation too.

Instead, for the computation of the inverse $\mathbf{V}^{-1}$ and its determinant, we may resort to prewhitening the data, *i.e.* creating N uncorrelated random variables from the original set of correlated random variables $\{Y(\mathbf{s}_j),\ j = 1, \cdots, N\}$. A detailed account on prewhitening is available in Section 3.2 of Fan and Yao (2003).

We consider each location $j = 1, \cdots, N$, and we define

$$\widetilde{Y}(\mathbf{s}_1) \equiv 0 \tag{5.2.32}$$

and for any $j = 2, \cdots, N$,

$$\widetilde{Y}(\mathbf{s}_j) \equiv \sum_{k=1}^{j-1} \phi_{j,k}\, Y(\mathbf{s}_k), \tag{5.2.33}$$

to be the best linear predictor of $Y(\mathbf{s}_j)$ based on all the sites that have a smaller label $k < j,\ k = 1, \cdots, N-1$, in the sense that

$$E\{Y(\mathbf{s}_j) - \widetilde{Y}(\mathbf{s}_j)\}^2 \equiv \min_{\{\psi_{j,k}\}} E\{Y(\mathbf{s}_j) - \sum_{k=1}^{j-1} \psi_{j,k}\, Y(\mathbf{s}_k)\}^2,\ j = 2, \cdots, N. \tag{5.2.34}$$

Then it holds that

$$\mathrm{Cov}\{Y(\mathbf{s}_j) - \widetilde{Y}(\mathbf{s}_j), Y(\mathbf{s}_k)\} = 0,\ j = 2, \cdots, N,\ k = 1, \cdots, j-1, \tag{5.2.35}$$

or, similarly,

$$\gamma_{j,k} - \sum_{m=1}^{j-1} \phi_{j,m} \cdot \gamma_{m,k} = 0,\ j = 2, \cdots, N,\ k = 1, \cdots, j-1. \tag{5.2.36}$$

On the other hand, we can define the prediction variances as

$$r_j \equiv \mathrm{Var}\{Y(\mathbf{s}_j) - \widetilde{Y}(\mathbf{s}_j)\} = \begin{cases} \gamma_{1,1}, & j = 1 \\ \gamma_{j,j} - \sum_{m=1}^{j-1} \phi_{j,m} \cdot \gamma_{m,j}, & j = 2, \cdots, N \end{cases}. \tag{5.2.37}$$

If we stack together the predictors in a vector

$$\widetilde{\mathbf{Y}} \equiv [\widetilde{Y}(\mathbf{s}_1), \cdots, \widetilde{Y}(\mathbf{s}_N)]^\tau, \tag{5.2.38}$$

then it holds that

$$\mathrm{Var}\{\mathbf{Y} - \widetilde{\mathbf{Y}}\} = \mathbf{R}, \tag{5.2.39}$$

where

$$\mathbf{R} \equiv \mathrm{diag}[r_j]_{j=1}^N. \tag{5.2.40}$$

But we have defined

$$\widetilde{\mathbf{Y}} = \mathbf{\Phi} \cdot \mathbf{Y}, \tag{5.2.41}$$

where

$$\mathbf{\Phi} \equiv \begin{bmatrix} 0 & 0 & \cdots & 0 \\ \phi_{2,1} & 0 & & 0 \\ & & \ddots & \\ \phi_{N,1} & \phi_{N,2} & & 0 \end{bmatrix}. \tag{5.2.42}$$

We can re-write (5.2.41) as

$$\mathbf{Y} - \widetilde{\mathbf{Y}} = (\mathbf{I}_N - \mathbf{\Phi}) \cdot \mathbf{Y}, \tag{5.2.43}$$

which combined with (5.2.39) gives

$$(\mathbf{I}_N - \mathbf{\Phi}) \cdot \mathbf{V} \cdot (\mathbf{I}_N - \mathbf{\Phi})^\tau = \mathbf{R} \tag{5.2.44}$$

or

$$\mathbf{V} = (\mathbf{I}_N - \mathbf{\Phi})^{-1} \cdot \mathbf{R} \cdot (\mathbf{I}_N - \mathbf{\Phi})^{\tau - 1} \tag{5.2.45}$$

or, finally,

$$\mathbf{V}^{-1} = (\mathbf{I}_N - \mathbf{\Phi})^\tau \cdot \mathbf{R}^{-1} \cdot (\mathbf{I}_N - \mathbf{\Phi}). \tag{5.2.46}$$

When $\mathbf{V}$ is non-singular, the decomposition (5.2.46) of the inverse dispersion matrix is unique and the elements of the matrices $\mathbf{\Phi}$ and $\mathbf{R}$ can be computed via the innovations algorithm, or further the Durbin-Levinson recursions (Brockwell and Davis, pp.172-173). More specifically, the algorithm dictates that we first write

$$\begin{aligned} r_1 &= \gamma_{1,1} \\ \phi_{2,1} &= \gamma_{2,1}/r_1 \\ r_2 &= \gamma_{2,2} - \phi_{2,1}^2 \cdot r_1 \end{aligned}$$

and

$$\phi_{j,1} = \gamma_{j,1}/r_1, \ j = 3, \cdots, N.$$

Next, for fixed $j = 3, \cdots, N$, we may write

$$\phi_{j,k} = (\gamma_{j,k} - \sum_{l=1}^{k-1} \phi_{j,l} \cdot \phi_{k,l} \cdot r_l)/r_k, \ k = 2, \cdots, j-1,$$

and

$$r_j = \gamma_{j,j} - \sum_{l=1}^{j-1} \phi_{j,l}^2 \cdot r_l,$$

from which we compute $\phi_{3,2}$, r_3, $\phi_{4,2}$, $\phi_{4,3}$, r_4, $\phi_{N,2}, \cdots, \phi_{N,N-1}$, r_N.

Remark 5.2. A lot of speculation has involved the computation of the determinant $|\mathbf{B}|$ (Besag, 1975, p.193). Writing down the two decompositions of the inverse variance matrix

$$\mathbf{\Lambda}^{-1}\mathbf{B} = (\mathbf{I}_N - \mathbf{\Phi})^{\tau} \cdot \mathbf{R}^{-1} \cdot (\mathbf{I}_N - \mathbf{\Phi}), \tag{5.2.47}$$

implies that

$$\prod_{j=1}^{N} \frac{1}{\nu_j} |\mathbf{B}| = \prod_{j=1}^{N} \frac{1}{r_j} \tag{5.2.48}$$

or

$$|\mathbf{B}| = \prod_{j=1}^{N} \frac{\nu_j}{r_j}. \tag{5.2.49}$$

In general, if there is at least one pair $j, k = 1, \cdots, N,\ j \neq k$, such that it holds

$$\mathrm{Cov}\{Y(\mathbf{s}_j), Y(\mathbf{s}_k)\} \neq 0,$$

then we may see, using a conventional ordering of locations, that there is at least one $l = 1, \cdots, N-1$, such that

$$\nu_l < r_l. \tag{5.2.50}$$

It also holds that

$$\nu_N = r_N, \tag{5.2.51}$$

since

$$\widehat{Y}(\mathbf{s}_N) = \widetilde{Y}(\mathbf{s}_N). \tag{5.2.52}$$

We may, in general, write

$$|\mathbf{B}| = \prod_{j=1}^{N-1} \frac{\nu_j}{r_j} \leq 1, \tag{5.2.53}$$

where the equality holds if and only if

$$\mathrm{Cov}\{Y(\mathbf{s}_j), Y(\mathbf{s}_k)\} = 0,\ j, k = 1, \cdots, N,\ j \neq k,$$

since the eigenvalues of $\mathbf{V}$ are away from zero.

Equations (5.2.51) and (5.2.52) relate the best linear predictor coefficients and prediction variances coming from the two decompositions for the last site $j = N$. If we wish to move from one decomposition to the other for all sites $j = 1, \cdots, N$, then directly from (5.2.47) and after we have the coefficients $\varphi_{k,j},\ j = 1, \cdots, N-1,\ k = j+1, \cdots, N$, and the variances $r_j,\ j = 1, \cdots, N$ computed, we may proceed as follows.

Step 1 : We set

$$\nu_N = r_N \tag{5.2.54}$$

and

$$\beta_{N,j} = \varphi_{N,j},\ j = 1, \cdots, N-1. \tag{5.2.55}$$

Step 2 : For $j = 1, \cdots, N-1$, we write

$$\nu_j = \left\{ \frac{1}{r_j} + \sum_{l=j+1}^{N} \frac{\varphi_{l,j}^2}{r_l} \right\}^{-1} \tag{5.2.56}$$

and

$$\beta_{j,N} = \frac{\nu_j}{\nu_N} \cdot \beta_{N,j}. \tag{5.2.57}$$

Step 3 : For $j = 1, \cdots, N-2$, and $k = j+1, \cdots, N-1$, we write

$$\beta_{k,j} = \nu_k \cdot \left[\frac{\varphi_{k,j}}{r_k} - \sum_{l=k+1}^{N} \frac{\varphi_{l,j} \cdot \varphi_{l,k}}{r_l} \right] \tag{5.2.58}$$

and

$$\beta_{j,k} = \frac{\nu_j}{\nu_k} \cdot \beta_{k,j}. \tag{5.2.59}$$

5.3 Spatio-temporal modeling

5.3.1 Models

We propose a spatio-temporal model for fitting data which is recorded regularly in time and irregularly over space. We aim for modeling both the spatial interdependence and the serial dependence structure.

Let $\{Y_t(\mathbf{s}_j),\ t \in \mathcal{Z},\ \mathbf{s}_j^\tau \in \mathcal{R}^d,\ j = 1, \cdots, N\}$ be a real-valued process observed over time and on N fixed locations. We consider for every location $j = 1, \cdots, N$, a causal auto-regression model:

$$Y_t(\mathbf{s}_j) \equiv b_{1,j}\ Y_{t-1}(\mathbf{s}_j) + \cdots + b_{p,j}\ Y_{t-p}(\mathbf{s}_j) + \varepsilon_t(\mathbf{s}_j), \tag{5.3.1}$$

where

$$\boldsymbol{\varepsilon}_t \equiv [\varepsilon_t(\mathbf{s}_1), \cdots, \varepsilon_t(\mathbf{s}_N)]^\tau \sim WN(\mathbf{0}, \mathbf{V}), \tag{5.3.2}$$

and $\mathbf{V}$ is a positive-definite dispersion matrix with finite eigenvalues.

For convenience, we stack the parameter vectors

$$\mathbf{b}_j \equiv [b_{1,j}, \cdots, b_{p,j}]^\tau \tag{5.3.3}$$

and

$$\mathbf{b} \equiv [\mathbf{b}_1^\tau, \cdots, \mathbf{b}_N^\tau]^\tau. \tag{5.3.4}$$

We also consider the parameter vector

$$\mathbf{a} \equiv [a_{1,2}, \cdots, a_{1,N}, a_{2,3}, \cdots, a_{2,N}, \cdots, a_{N-1,N}, a_{1,1}, \cdots, a_{N,N}]^\tau \tag{5.3.5}$$

with the elements of the inverse dispersion matrix

$$\mathbf{V}^{-1} \equiv \begin{bmatrix} a_{1,1} & -a_{1,2} & \cdots & -a_{1,N} \\ -a_{1,2} & a_{2,2} & & -a_{2,N} \\ \vdots & & \ddots & \\ -a_{1,N} & -a_{2,N} & & a_{N,N} \end{bmatrix}, \tag{5.3.6}$$

similar to that given in Section 5.2.

5.3.2 Multivariate Time Series context

For any $t \in \mathcal{Z}$, we let the vector

$$\mathbf{Y}_t \equiv [Y_t(\mathbf{s}_1), \cdots, Y_t(\mathbf{s}_N)]^\tau \tag{5.3.7}$$

and write

$$\mathbf{Y}_t - \mathbf{\Phi}_1 \, \mathbf{Y}_{t-1} - \cdots - \mathbf{\Phi}_p \, \mathbf{Y}_{t-p} = \boldsymbol{\varepsilon}_t, \; \{\boldsymbol{\varepsilon}_t\} \sim WN(\mathbf{0}, \mathbf{V}), \tag{5.3.8}$$

where

$$\mathbf{\Phi}_i \equiv \text{diag}[b_{i,j}]_{j=1}^N, \; i = 1, \cdots, p. \tag{5.3.9}$$

In (5.3.8), we have written (5.3.1) as a multivariate auto-regression. In fact, this is a seemingly unrelated auto-regression, according to Harvey (1989). This is because all the matrices (5.3.9) are diagonal and all their elements $b_{i,j}$, $i = 1, \cdots, p$, refer to one location $j = 1, \cdots, N$, only. In other words, any relation between two locations $j, k = 1, \cdots, N$, $j \neq k$, has only been expressed via the variance matrix $\mathbf{V}$. We could also see that from the univariate versions (5.3.1), where all the lagged values $Y_{t-1}(\mathbf{s}_j), \cdots, Y_{t-p}(\mathbf{s}_j)$ of $Y_t(\mathbf{s}_j)$, also referred to the same location j.

We consider the polynomial

$$\mathbf{\Phi}(z) \equiv \mathbf{I}_N - \mathbf{\Phi}_1 \, z - \cdots - \mathbf{\Phi}_p \, z^p \tag{5.3.10}$$

or

$$\mathbf{\Phi}(z) = \text{diag}[1 - \sum_{i=1}^{p} b_{i,j} \, z^i]_{j=1}^N. \tag{5.3.11}$$

It holds that $\{\mathbf{Y}_t,\ t \in \mathcal{Z}\}$ is a causal auto-regression if and only if

$$\det\{\mathbf{\Phi}(z)\} \neq 0 \Leftrightarrow \prod_{j=1}^{N}(1 - \sum_{i=1}^{p} b_{i,j}\ z^i) \neq 0 \tag{5.3.12}$$

for all complex numbers z such that $|z| \leq 1$. Now, it is clear that (5.3.12) implies that $\mathbf{b} \in \mathcal{B}^N$, which is exactly the same condition for causality as the one imposed before, *i.e.* that all the N auto-regressions are causal, and so $\mathbf{b}_j \in \mathcal{B}$ for all $j = 1, \cdots, N$.

For $j = 1, \cdots, N$, we define the polynomials

$$\psi_j(z) \equiv (1 - \sum_{i=1}^{p} b_{i,j}\ z^i)^{-1} \equiv 1 + \sum_{i=1}^{\infty} \psi_{i,j}\ z^i,\ \sum_{i=1}^{\infty} |\psi_{i,j}| < \infty. \tag{5.3.13}$$

Then for any $i = 1, 2, \cdots$, we define the matrices

$$\mathbf{\Psi}_i \equiv \text{diag}[\psi_{i,j}]_{j=1}^{N}, \tag{5.3.14}$$

such that we can write

$$\mathbf{Y}_t = \boldsymbol{\varepsilon}_t + \sum_{i=1}^{\infty} \mathbf{\Psi}_i\ \boldsymbol{\varepsilon}_{t-i} \tag{5.3.15}$$

and

$$\text{Var}\{\mathbf{Y}_t\} = \mathbf{V} + \sum_{i=1}^{p} \mathbf{\Psi}_i \mathbf{V} \mathbf{\Psi}_i. \tag{5.3.16}$$

On the one hand, it holds that for any $\boldsymbol{\lambda} \in \mathcal{R}^N$

$$\boldsymbol{\lambda}^\tau \mathbf{V}\ \boldsymbol{\lambda} > 0, \tag{5.3.17}$$

since $\mathbf{V}$ is positive-definite. Also, it holds that $\text{Var}\{\mathbf{\Psi}_i \boldsymbol{\varepsilon}_t\}$ is positive definite and

$$|\mathbf{\Psi}_i \mathbf{V} \mathbf{\Psi}_i| = \{\prod_{j=1}^{N} \psi_{i,j}^2\}\ |\mathbf{V}| \geq 0,\ i = 1, 2, \cdots. \tag{5.3.18}$$

Thus, for any $\boldsymbol{\lambda} \in \mathcal{R}^N$

$$\boldsymbol{\lambda}^\tau \text{Var}\{\mathbf{Y}_t\}\ \boldsymbol{\lambda} = \boldsymbol{\lambda}^\tau [\mathbf{V} + \sum_{i=1} \mathbf{\Psi}_i \mathbf{V} \mathbf{\Psi}_i]\ \boldsymbol{\lambda} > 0 \tag{5.3.19}$$

and $\text{Var}\{\mathbf{Y}_t\}$ is positive-definite with eigenvalues greater than 0.

On the other hand, if we write $\text{Var}\{\mathbf{Y}_t\} = [v_{j,k}]_{j,k=1}^{N}$, then it holds that

$$v_{j,k} = [1 + \sum_{i=1}^{\infty} \psi_{i,j} \psi_{i,k}]\ a_{j,k},\ j, k = 1, \cdots, N, \tag{5.3.20}$$

and the trace of $\text{Var}\{\mathbf{Y}_t\}$ and sum of its eigenvalues is equal to

$$\sum_{j=1}^{N} v_{j,j} = \sum_{j=1}^{N} [1 + \sum_{i=1}^{\infty} \psi_{i,j}^2]\ a_{j,j} < \infty,$$

where the inequality holds for fixed N. As a result, all the eigenvalues of the matrix are away from ∞.

Moreover since the multivariate auto-regression is causal, the variance matrix

$$\mathrm{Var}\{[\mathbf{Y}_1, \cdots, \mathbf{Y}_T]^\tau\}$$

has all its eigenvalues bounded away from 0 and ∞ for any positive integer T. This is a multivariate analogue of Remark 2.3.

5.3.3 Gaussian likelihood estimators

We observe $\{Y_t(\mathbf{s}_j),\ t = 1-p, \cdots, T,\ j = 1, \cdots, N\}$ and we want to estimate the true parameter vector that has generated these observations. We assume that the following condition holds.

(C9) The parameter space $\mathcal{A} \times \mathcal{B}^N \subseteq \mathcal{R}^{q+Np}$ is a compact set containing the true value $[\mathbf{a}_0^\tau, \mathbf{b}_0^\tau]^\tau$ as an inner point. Further, for any $\mathbf{a} \in \mathcal{A}$ all the eigenvalues of $\mathbf{V}^{-1}$ are positive and finite and for any $\mathbf{b}_j \in \mathcal{B},\ j = 1, \cdots, N$, a causal auto-regression (5.3.1) is defined.

For any $\mathbf{a} \in \mathcal{A}$ and $\mathbf{b} \in \mathcal{B}^N$, we write the Gaussian likelihood function

$$L(\mathbf{a}, \mathbf{b}) \equiv \frac{1}{(\sqrt{2\pi})^{NT}} |\mathbf{V}^{-1}|^{T/2} \exp\{ - \frac{1}{2} \sum_{t=1}^{T} [\sum_{j=1}^{N} a_{j,j}\ \varepsilon_t(\mathbf{s}_j, \mathbf{b}_j)^2 - 2 \sum_{\substack{j,k=1, \\ j<k}}^{N} a_{j,k}\ \varepsilon_t(\mathbf{s}_j, \mathbf{b}_j)\varepsilon_t(\mathbf{s}_k, \mathbf{b}_k)]\}, \quad (5.3.21)$$

where we define

$$\varepsilon_t(\mathbf{s}_j, \mathbf{b}_j) \equiv Y_t(\mathbf{s}_j) - \mathbf{X}_t^\tau(\mathbf{s}_j) \cdot \mathbf{b}_j = \varepsilon_t(\mathbf{s}_j) - \mathbf{X}_t^\tau(\mathbf{s}_j) \cdot [\mathbf{b}_j - \mathbf{b}_{j,0}],\ t \in \mathcal{Z},\ j = 1, \cdots, N, \quad (5.3.22)$$

with

$$\mathbf{X}_t^\tau(\mathbf{s}_j) \equiv [Y_{t-1}(\mathbf{s}_j), \cdots, Y_{t-p}(\mathbf{s}_j)],\ t \in \mathcal{Z},\ j = 1, \cdots, N \quad (5.3.23)$$

and

$$\varepsilon_t(\mathbf{s}_j, \mathbf{b}_{j,0}) \equiv \varepsilon_t(\mathbf{s}_j),\ t \in \mathcal{Z},\ j = 1, \cdots, N. \quad (5.3.24)$$

Like in the previous chapter, we have not assumed that our observations have been generated by Gaussian random variables, but we have chosen to write down, and later

to maximize, a Gaussian likelihood. Again we refer to White (1982) and to the quasi-maximum likelihood estimators defined then, in order to highlight the relevance with our method. To set our estimators that maximize the likelihood (5.3.21), we may first write down the natural logarithm of the likelihood

$$l(\mathbf{a},\mathbf{b}) = \frac{T}{2}\log\{|\mathbf{V}^{-1}|\} - \frac{1}{2}\sum_{t=1}^{T}[\sum_{j=1}^{N} a_{j,j}\ \varepsilon_t(\mathbf{s}_j,\mathbf{b}_j)^2 - 2\sum_{\substack{j,k=1,\\ j<k}}^{N} a_{j,k}\ \varepsilon_t(\mathbf{s}_j,\mathbf{b}_j)\varepsilon_t(\mathbf{s}_k,\mathbf{b}_k)], \tag{5.3.25}$$

for all $\mathbf{a} \in \mathcal{A},\ \mathbf{b} \in \mathcal{B}^N$, and maximize this instead. For the differentiation of the deterministic part of the likelihood, we will need the following proposition.

Proposition 5.2. For the symmetric, non-singular $N \times N$ matrices $\mathbf{V} = [\gamma_{j,k}]_{j,k=1}^{N}$ and $\mathbf{V}^{-1}$ as defined in (5.3.6), it holds that

- $\partial \log|\mathbf{V}^{-1}|/\partial a_{j,j} = \gamma_{j,j},\ j = 1,\cdots,N,$
- $\partial \log|\mathbf{V}^{-1}|/\partial a_{j,k} = -2\gamma_{j,k},\ j,k = 1,\cdots,N,\ j \neq k.$

Proof. We know that

$$\mathbf{V} = \frac{1}{|\mathbf{V}^{-1}|}\begin{bmatrix} A_{1,1} & -A_{1,2} & \cdots & (-1)^{N+1}A_{1,N} \\ -A_{1,2} & A_{2,2} & & (-1)^{N+2}A_{2,N} \\ \vdots & & \ddots & \\ (-1)^{N+1}A_{1,N} & (-1)^{N+2}A_{2,N} & & A_{N,N} \end{bmatrix}, \tag{5.3.26}$$

where $A_{j,k},\ j,k = 1,\cdots,N$, is the determinant of the $(N-1)\times(N-1)$ matrix that remains when we exclude the j-th row and the k-th column from the matrix $\mathbf{V}^{-1}$. We also know that

$$\mathbf{V} = \begin{bmatrix} \gamma_{1,1} & \gamma_{1,2} & \cdots & \gamma_{1,N} \\ \gamma_{1,2} & \gamma_{2,2} & & \gamma_{2,N} \\ \vdots & & \ddots & \\ \gamma_{1,N} & \gamma_{2,N} & & \gamma_{N,N} \end{bmatrix}. \tag{5.3.27}$$

Since

$$\partial \log|\mathbf{V}^{-1}|/\partial a_{j,k} = \frac{1}{|\mathbf{V}^{-1}|}\partial|\mathbf{V}^{-1}|/\partial a_{j,k},\ j,k = 1,\cdots,N, \tag{5.3.28}$$

all we need to show is that

$$\partial|\mathbf{V}^{-1}|/\partial a_{j,j} = A_{j,j},\ j = 1,\cdots,N, \tag{5.3.29}$$

and that

$$\partial|\mathbf{V}^{-1}|/\partial a_{j,k} = (-2)(-1)^{j+k}A_{j,k} = 2(-1)^{j+k-1}A_{j,k},\ j,k = 1,\cdots,N,\ j \neq k. \quad (5.3.30)$$

The proof of (5.3.29) is very simple. For any $j = 1,\cdots,N$, we may write

$$|\mathbf{V}^{-1}| = a_{j,j}A_{j,j} - \sum_{\substack{k=1,\\ k\neq j}}^{N} a_{j,k}(-1)^{j+k}A_{j,k}, \quad (5.3.31)$$

from which we can see immediately that (5.3.29) is true since all $A_{j,k},\ k = 1,\cdots,N$, do not depend on $a_{j,j}$. To show (5.3.30) and without loss of generality, we consider the derivatives with respect to $a_{1,2}$. Then all we need to show is that

$$\partial|\mathbf{V}^{-1}|/\partial a_{1,2} = 2A_{1,2}. \quad (5.3.32)$$

It holds that

$$|\mathbf{V}^{-1}| = a_{1,1}A_{1,1} - (-a_{1,2})A_{1,2} + (-a_{1,3})A_{1,3} - \cdots + (-1)^{N+1}(-a_{1,N})A_{1,N} \quad (5.3.33)$$

or

$$|\mathbf{V}^{-1}| = a_{1,1}A_{1,1} + a_{1,2}A_{1,2} - a_{1,3}A_{1,3} + \cdots - (-1)^{N+1}a_{1,N}A_{1,N}, \quad (5.3.34)$$

so that we can write

$$\partial|\mathbf{V}^{-1}|/\partial a_{1,2} = A_{1,2} + a_{1,2}\,\partial A_{1,2}/\partial a_{1,2} - a_{1,3}\,\partial A_{1,3}/\partial a_{1,2} + \cdots - (-1)^{N+1}a_{1,N}\,\partial A_{1,N}/\partial a_{1,2} \quad (5.3.35)$$

since both $a_{1,1},\ A_{1,1}$ do not depend on $a_{1,2}$. Looking at (5.3.33) and (5.3.35), the next step is to show that

$$A_{1,2} = a_{1,2}\,\partial A_{1,2}/\partial a_{1,2} - a_{1,3}\,\partial A_{1,3}/\partial a_{1,2} + \cdots - (-1)^{N+1}a_{1,N}\,\partial A_{1,N}/\partial a_{1,2}. \quad (5.3.36)$$

Since

$$A_{1,2} = \begin{vmatrix} -a_{1,2} & -a_{2,3} & \cdots & -a_{2,N} \\ -a_{1,3} & a_{3,3} & & -a_{3,N} \\ \vdots & & \ddots & \\ -a_{1,N} & -a_{3,N} & & a_{N,N} \end{vmatrix} \quad (5.3.37)$$

we may write

$$A_{1,2} = (-a_{1,2})M_{1,2} - (-a_{1,3})M_{1,3} + \cdots + (-1)^{N}(-a_{1,N})M_{1,N} \quad (5.3.38)$$

or

$$A_{1,2} = a_{1,2}(-M_{1,2}) - a_{1,3}(-M_{1,3}) + \cdots - (-1)^{N+1}a_{1,N}(-M_{1,N}), \quad (5.3.39)$$

where $M_{1,k}$, $k = 2, \cdots, N$, is the determinant of the $(N-2) \times (N-2)$ matrix that remains if we exclude the first column and the $(k-1)$ row from the matrix form of $A_{1,2}$. And, of course,

$$\partial A_{1,2}/\partial a_{1,2} = -M_{1,2}. \tag{5.3.40}$$

We may also see that

$$\partial A_{1,k}/\partial a_{1,2} = -M_{1,k}, \ k = 3, \cdots, N, \tag{5.3.41}$$

which proves (5.3.36) and the required result. ■

For any $i = 1, \cdots, p$, $j = 1, \cdots, N$, we write the derivatives

$$\partial l(\mathbf{a}, \mathbf{b})/\partial b_{i,j} = \sum_{t=1}^{T} \{[a_{j,j} \ \varepsilon_t(\mathbf{s}_j, \mathbf{b}_j) - \sum_{\substack{k=1, \\ k \neq j}}^{N} a_{j,k} \ \varepsilon_t(\mathbf{s}_k, \mathbf{b}_k)] Y_{t-i}(\mathbf{s}_j)\}. \tag{5.3.42}$$

Using Proposition 5.2, we write for $j = 1, \cdots, N$,

$$\partial l(\mathbf{a}, \mathbf{b})/\partial a_{j,j} = \frac{T}{2} \ \gamma_{j,j} - \frac{1}{2} \sum_{t=1}^{T} \varepsilon_t(\mathbf{s}_j, \mathbf{b}_j)^2 \tag{5.3.43}$$

and for $j, k = 1, \cdots, N$, $j < k$,

$$\partial l(\mathbf{a}, \mathbf{b})/\partial a_{j,k} = -T \ \gamma_{j,k} + \sum_{t=1}^{T} \varepsilon_t(\mathbf{s}_j, \mathbf{b}_j) \varepsilon_t(\mathbf{s}_k, \mathbf{b}_k). \tag{5.3.44}$$

We set the Gaussian likelihood estimators $\widehat{\mathbf{a}}$ and $\widehat{\mathbf{b}}$ and write

$$\widehat{\varepsilon}_t(\mathbf{s}_j) \equiv Y_t(\mathbf{s}_j) - \mathbf{X}_t^T(\mathbf{s}_j) \cdot \widehat{\mathbf{b}}_j, \tag{5.3.45}$$

such that

$$\sum_{t=1}^{T} \{[\widehat{a}_{j,j} \ \widehat{\varepsilon}_t(\mathbf{s}_j) - \sum_{\substack{k=1, \\ k \neq j}}^{N} \widehat{a}_{j,k} \ \widehat{\varepsilon}_t(\mathbf{s}_k)] Y_{t-i}(\mathbf{s}_j)\} \equiv 0, \ i = 1, \cdots, p, \ j = 1, \cdots, N, \tag{5.3.46}$$

and

$$\widehat{\gamma}_{j,k} \equiv \frac{1}{T} \sum_{t=1}^{T} \widehat{\varepsilon}_t(\mathbf{s}_j) \widehat{\varepsilon}_t(\mathbf{s}_k), \ j, k = 1, \cdots, N. \tag{5.3.47}$$

And, of course, if we write $\widehat{\mathbf{V}} = [\widehat{\gamma}_{j,k}]_{j,k=1}^{N}$ we define

$$\{\widehat{\mathbf{V}}\}^{-1} \equiv \begin{bmatrix} \widehat{a}_{1,1} & -\widehat{a}_{1,2} & \cdots & -\widehat{a}_{1,N} \\ -\widehat{a}_{1,2} & \widehat{a}_{2,2} & & -\widehat{a}_{2,N} \\ \vdots & & \ddots & \\ -\widehat{a}_{1,N} & -\widehat{a}_{2,N} & & \widehat{a}_{N,N} \end{bmatrix}. \tag{5.3.48}$$

Next we prove the properties of the estimators set in (5.3.46) and (5.3.47) provided that the next condition is true.

(C10) It holds that $\{\varepsilon_t,\ t \in \mathcal{Z}\}$ is a sequence of independent and identically distributed zero mean random vectors with variance matrix $\mathrm{Var}\{\varepsilon_t\} = \mathbf{V}$.

Consistency

Theorem 5.1. Under conditions (C8)-(C10) and as $T \to \infty$ (N fixed), it holds that

$$\begin{bmatrix} \widehat{\mathbf{b}} \\ \widehat{\mathbf{a}} \end{bmatrix} \xrightarrow{P} \begin{bmatrix} \mathbf{b}_0 \\ \mathbf{a}_0 \end{bmatrix}. \tag{5.3.49}$$

Proof. First we will show that $\widehat{\mathbf{b}} \xrightarrow{P} \mathbf{b}_0$. It holds that

$$\partial l(\mathbf{a},\mathbf{b})/\partial b_{i,j} = -\frac{1}{2}\frac{\partial}{\partial b_{i,j}} \sum_{t=1}^{T}\{\sum_{j=1}^{N}[a_{j,j}\ \varepsilon_t(\mathbf{s}_j,\mathbf{b}_j)^2 - 2 \sum_{\substack{j,k=1,\\ j<k}}^{N} a_{j,k}\ \varepsilon_t(\mathbf{s}_j,\mathbf{b}_j)\varepsilon_t(\mathbf{s}_k,\mathbf{b}_k)]\}, \tag{5.3.50}$$

for all $i = 1,\cdots,p$, and $j = 1,\cdots,N$. Then

$$\begin{aligned}
&\limsup_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\{\sum_{j=1}^{N}\widehat{a}_{j,j}\ \widehat{\varepsilon}_t(\mathbf{s}_j)^2 - 2\sum_{\substack{j,k=1,\\ j<k}}^{N}\widehat{a}_{j,k}\ \widehat{\varepsilon}_t(\mathbf{s}_j)\widehat{\varepsilon}_t(\mathbf{s}_k)\} \\
\leq\ &\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\{\sum_{j=1}^{N}\widehat{a}_{j,j}\ \varepsilon_t(\mathbf{s}_j)^2 - 2\sum_{\substack{j,k=1,\\ j<k}}^{N}\widehat{a}_{j,k}\ \varepsilon_t(\mathbf{s}_j)\varepsilon_t(\mathbf{s}_k)\} \\
=\ &\sum_{j=1}^{N}\widehat{a}_{j,j}\ \{\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(\mathbf{s}_j)^2\} - \sum_{\substack{j,k=1,\\ j<k}}^{N}\widehat{a}_{j,k}\ \{\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(\mathbf{s}_j)\varepsilon_t(\mathbf{s}_k)\} \\
=\ &\sum_{j=1}^{N}\widehat{a}_{j,j}\cdot\gamma_{j,j} - 2\sum_{\substack{j,k=1,\\ j<k}}\widehat{a}_{j,k}\cdot\gamma_{j,k},
\end{aligned} \tag{5.3.51}$$

where the last equality holds under (C10), since then we can write

$$\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(\mathbf{s}_j)\varepsilon_t(\mathbf{s}_k) \xrightarrow{P} \gamma_{j,k},\ j,k = 1,\cdots,N. \tag{5.3.52}$$

On the other hand, for any $\mathbf{b}_j \in \mathcal{B},\ j = 1,\cdots,N$, since it holds that

$$\varepsilon_t(\mathbf{s}_j,\mathbf{b}_j) = \varepsilon_t(\mathbf{s}_j) - \mathbf{X}_t^T(\mathbf{s}_j)[\mathbf{b}_j - \mathbf{b}_{j,0}], \tag{5.3.53}$$

we may write

$$\begin{aligned}
&\liminf_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\{\sum_{j=1}^{N} \widehat{a}_{j,j}\ \varepsilon_t(\mathbf{s}_j,\mathbf{b}_j)^2 - 2\sum_{\substack{j,k=1,\\ j<k}}^{N} \widehat{a}_{j,k}\ \varepsilon_t(\mathbf{s}_j,\mathbf{b}_j)\varepsilon_t(\mathbf{s}_k,\mathbf{b}_k)\} \\
= &\sum_{j=1}^{N} \widehat{a}_{j,j}\ \{\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(\mathbf{s}_j)^2\} - 2\sum_{j=1}^{N}\widehat{a}_{j,j}\ \{\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(\mathbf{s}_j)\mathbf{X}_t^{\tau}(\mathbf{s}_j)\}\ [\mathbf{b}_j-\mathbf{b}_{j,0}] \\
+ &\sum_{j=1}^{N} \widehat{a}_{j,j}\ [\mathbf{b}_j-\mathbf{b}_{j,0}]^{\tau}\ \{\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t(\mathbf{s}_j)\mathbf{X}_t^{\tau}(\mathbf{s}_j)\}\ [\mathbf{b}_j-\mathbf{b}_{j,0}] \\
- &2\sum_{\substack{j,k=1,\\ j<k}}^{N} \widehat{a}_{j,k}\ \{\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(\mathbf{s}_j)\varepsilon_t(\mathbf{s}_k)\} \\
+ &2\sum_{\substack{j,k=1,\\ j<k}}^{N} \widehat{a}_{j,k}\ \{\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(\mathbf{s}_j)\mathbf{X}_t^{\tau}(\mathbf{s}_k)\}\ [\mathbf{b}_k-\mathbf{b}_{k,0}] \\
+ &2\sum_{\substack{j,k=1,\\ j<k}}^{N} \widehat{a}_{j,k}\ \{\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(\mathbf{s}_k)\mathbf{X}_t^{\tau}(\mathbf{s}_j)\}\ [\mathbf{b}_j-\mathbf{b}_{j,0}] \\
- &2\sum_{\substack{j,k=1,\\ j<k}}^{N} \widehat{a}_{j,k}\ [\mathbf{b}_j-\mathbf{b}_{j,0}]^{\tau}\ \{\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t(\mathbf{s}_j)\mathbf{X}_t^{\tau}(\mathbf{s}_k)\}[\mathbf{b}_k-\mathbf{b}_{k,0}]. \qquad (5.3.54)
\end{aligned}$$

And, of course, like for (5.3.52) and under (C10), it holds that

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t(\mathbf{s}_j)\mathbf{X}_t^{\tau}(\mathbf{s}_k) \xrightarrow{P} E\{\mathbf{X}_t(\mathbf{s}_j)\mathbf{X}_t^{\tau}(\mathbf{s}_k)\},\ j,k=1,\cdots,N. \qquad (5.3.55)$$

It also holds for any $i>0$ and $j,k=1,\cdots,N$, that

$$\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(\mathbf{s}_j)Y_{t-i}(\mathbf{s}_k) \xrightarrow{P} E\{\varepsilon_t(\mathbf{s}_j)Y_{t-i}(\mathbf{s}_k)\}=0, \qquad (5.3.56)$$

thanks to the assumption of causality. As a result, we can re-write from (5.3.54) that for any $\mathbf{b}\in\mathcal{B}^N$

$$\begin{aligned}
&\liminf_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\{\sum_{j=1}^{N}\widehat{a}_{j,j}\ \varepsilon_t(\mathbf{s}_j,\mathbf{b}_j)^2 - 2\sum_{\substack{j,k=1,\\ j<k}}^{N}\widehat{a}_{j,k}\ \varepsilon_t(\mathbf{s}_j,\mathbf{b}_j)\varepsilon_t(\mathbf{s}_k,\mathbf{b}_k)\} \\
= &\sum_{j=1}^{N}\widehat{a}_{j,j}\cdot\gamma_{j,j} - 2\sum_{\substack{j,k=1,\\ j<k}}^{N}\widehat{a}_{j,k}\cdot\gamma_{j,k}+\delta(\mathbf{b}), \qquad (5.3.57)
\end{aligned}$$

where

$$\begin{aligned}\delta(\mathbf{b}) &= \sum_{j=1}^{N} \widehat{a}_{j,j}\ [\mathbf{b}_j - \mathbf{b}_{j,0}]^\tau E\{\mathbf{X}_t(\mathbf{s}_j)\mathbf{X}_t^\tau(\mathbf{s}_j)\}[\mathbf{b}_j - \mathbf{b}_{j,0}] \\ &- 2\sum_{\substack{j,k=1,\\ j<k}}^{N} \widehat{a}_{j,k}\ [\mathbf{b}_j - \mathbf{b}_{j,0}]^\tau E\{\mathbf{X}_t(\mathbf{s}_j)\mathbf{X}_t^\tau(\mathbf{s}_k)\}[\mathbf{b}_k - \mathbf{b}_{k,0}]. \\ &= [\mathbf{b} - \mathbf{b}_0]^\tau\ \mathbf{M}(\widehat{\mathbf{a}})\ [\mathbf{b} - \mathbf{b}_0] \geq \lambda_{\min}\{\mathbf{M}(\widehat{\mathbf{a}})\}||\mathbf{b} - \mathbf{b}_0||^2, \end{aligned} \tag{5.3.58}$$

with the random matrix $\mathbf{M}(\widehat{\mathbf{a}})$, set to be equal to

$$\begin{bmatrix} \widehat{a}_{1,1}E\{\mathbf{X}_t(\mathbf{s}_1)\mathbf{X}_t^\tau(\mathbf{s}_1)\} & -\widehat{a}_{1,2}E\{\mathbf{X}_t(\mathbf{s}_1)\mathbf{X}_t^\tau(\mathbf{s}_2)\} & \cdots & -\widehat{a}_{1,N}E\{\mathbf{X}_t(\mathbf{s}_1)\mathbf{X}_t^\tau(\mathbf{s}_N)\} \\ -\widehat{a}_{1,2}E\{\mathbf{X}_t(\mathbf{s}_2)\mathbf{X}_t^\tau(\mathbf{s}_1)\} & \widehat{a}_{2,2}E\{\mathbf{X}_t(\mathbf{s}_2)\mathbf{X}_t^\tau(\mathbf{s}_2)\} & & -\widehat{a}_{2,N}E\{\mathbf{X}_t(\mathbf{s}_2)\mathbf{X}_t^\tau(\mathbf{s}_N)\} \\ \vdots & & \ddots & \\ -\widehat{a}_{1,N}E\{\mathbf{X}_t(\mathbf{s}_N)\mathbf{X}_t^\tau(\mathbf{s}_1)\} & -\widehat{a}_{2,N}E\{\mathbf{X}_t(\mathbf{s}_N)\mathbf{X}_t^\tau(\mathbf{s}_2)\} & & \widehat{a}_{N,N}E\{\mathbf{X}_t(\mathbf{s}_N)\mathbf{X}_t^\tau(\mathbf{s}_N)\} \end{bmatrix} \tag{5.3.59}$$

and its minimum eigenvalue $\lambda_{\min}\{\mathbf{M}(\widehat{\mathbf{a}})\}$. Next, we will show that the eigenvalue is away from 0 and ∞.

Similarly to (5.3.59), we may consider for any $\mathbf{a} \in \mathcal{A}$ the matrix $\mathbf{M}(\mathbf{a})$. For convenience and without loss of generality, let us try to construct the variance matrix

$$\mathbf{M} \equiv \mathbf{M}(\mathbf{a}_0). \tag{5.3.60}$$

For $t \in \mathcal{Z}$, we define the random variables

$$Z_t(\mathbf{s}_j) \equiv a_{(j,j),0}\ \varepsilon_t(\mathbf{s}_j) - \sum_{\substack{k=1,\\ k\neq j}}^{N} a_{(j,k),0}\ \varepsilon_t(\mathbf{s}_k),\ j = 1, \cdots, N, \tag{5.3.61}$$

and the random vector

$$\mathbf{Z}_t \equiv [Z_t(\mathbf{s}_1), \cdots, Z_t(\mathbf{s}_N)]^\tau. \tag{5.3.62}$$

It holds that

$$\mathbf{Z}_t = \mathbf{V}^{-1}\boldsymbol{\varepsilon}_t \tag{5.3.63}$$

and

$$\mathrm{Var}\{\mathbf{Z}_t\} = \mathbf{V}^{-1}. \tag{5.3.64}$$

It also holds for $i, l > 0$

$$\begin{aligned}\text{Cov}\{Z_t(\mathbf{s}_j)Y_{t-i}(\mathbf{s}_j), Z_t(\mathbf{s}_k)Y_{t-l}(\mathbf{s}_k)\} &= E\{Z_t(\mathbf{s}_j)Y_{t-i}(\mathbf{s}_j)Z_t(\mathbf{s}_k)Y_{t-l}(\mathbf{s}_k)\} \\ &- E\{Z_t(\mathbf{s}_j)Y_{t-i}(\mathbf{s}_j)\}\, E\{Z_t(\mathbf{s}_k)Y_{t-l}(\mathbf{s}_k)\} \\ &= E\{Z_t(\mathbf{s}_j)Y_{t-i}(\mathbf{s}_j)Z_t(\mathbf{s}_k)Y_{t-l}(\mathbf{s}_k)\} \\ &= E\{Z_t(\mathbf{s}_j)Z_t(\mathbf{s}_k)\}\, E\{Y_{t-i}(\mathbf{s}_j)Y_{t-l}(\mathbf{s}_k)\}, \end{aligned} \tag{5.3.65}$$

thanks to the assumption of causality. As a result, we can write

$$\mathbf{M} = \text{Var}\{[Z_t(\mathbf{s}_1)\mathbf{X}_t^\tau(\mathbf{s}_1), \cdots, Z_t(\mathbf{s}_N)\mathbf{X}_t^\tau(\mathbf{s}_N)]^\tau\}. \tag{5.3.66}$$

Now suppose that we were interested in the variance matrix

$$\begin{aligned} &\text{Var}\{[Z_t(\mathbf{s}_1)\mathbf{X}_t^\tau(\mathbf{s}_1), \cdots, Z_t(\mathbf{s}_1)\mathbf{X}_t^\tau(\mathbf{s}_N), \cdots, Z_t(\mathbf{s}_N)\mathbf{X}_t^\tau(\mathbf{s}_1), \cdots, Z_t(\mathbf{s}_N)\mathbf{X}_t^\tau(\mathbf{s}_N)]^\tau\} \\ = \;&\text{Var}\{[Z_t(\mathbf{s}_1), \cdots, Z_t(\mathbf{s}_N)]^\tau\} \otimes \text{Var}\{[\mathbf{X}_t^\tau(\mathbf{s}_1), \cdots, \mathbf{X}_t^\tau(\mathbf{s}_N)]^\tau\}, \end{aligned} \tag{5.3.67}$$

whose eigenvalues are equal to all possible products of eigenvalues of the two variance matrices. Indeed, it holds that

$$\text{Var}\{[Z_t(\mathbf{s}_1), \cdots, Z_t(\mathbf{s}_N)]^\tau\} = \mathbf{V}$$

has all its eigenvalues positive. Similarly, if we put in a different order the Np random variables of the second matrix, we might write the variance matrix

$$\text{Var}\{[\mathbf{Y}_{t-1}^\tau, \cdots, \mathbf{Y}_{t-p}^\tau]^\tau$$

which also has all its eigenvalues positive under the assumption of causality and the fact that N is fixed. Thus, the Kronecker product variance matrix (5.3.67) has positive eigenvalues and positive determinant.

On the other hand, the determinant of a variance matrix can be seen as a product of the prediction variances, computed via the innovations algorithm. The positive determinant of (5.3.67) means that all the prediction variances are positive. Still, the innovations algorithm can be applied for any desired ordering of the variables. Thus, the Np variables

$$Z_t(\mathbf{s}_1)Y_{t-1}(\mathbf{s}_1), \cdots, Z_t(\mathbf{s}_1)Y_{t-p}(\mathbf{s}_1), \cdots, Z_t(\mathbf{s}_N)Y_{t-1}(\mathbf{s}_N), \cdots, Z_t(\mathbf{s}_N)Y_{t-p}(\mathbf{s}_N)$$

might be ordered first and generate the same prediction variances as for the variance matrix $\mathbf{M}$. But if the determinant of (5.3.67) is positive, so are all the prediction variances of $\mathbf{M}$. As a result,

$$|\mathbf{M}| > 0 \tag{5.3.68}$$

and the minimum eigenvalue of $\mathbf{M}$ must be a positive number. An identical argument might be applied for $\widehat{\mathbf{a}}$ since $\widehat{\mathbf{a}} \in \mathcal{A}$.

Looking back at (5.3.58), we may write

$$\delta(\mathbf{b}) \geq 0 \tag{5.3.69}$$

and the equality holds if and only if $\mathbf{b} = \mathbf{b}_0$. We define $\mathcal{C} = \{\liminf_{T\to\infty} ||\widehat{\mathbf{b}} - \mathbf{b}_0|| > 0\}$. For any $\omega \in \mathcal{C}$, there exists a subsequence of $\{T\}$, which we will still denote as $\{T\}$, for which $\widehat{\mathbf{b}}(\omega) \to \mathbf{b} \in \mathcal{B}^N$ and $\mathbf{b} \neq \mathbf{b}_0$. Combining (5.3.51), (5.3.57) and (5.3.69), we can write $P(\mathcal{C}) = 0$. As a result,

$$\widehat{\mathbf{b}} \xrightarrow{P} \mathbf{b}_0. \tag{5.3.70}$$

Then it holds for $j = 1, \cdots, N$, that

$$\widehat{\varepsilon}_t(\mathbf{s}_j) \xrightarrow{P} \varepsilon_t(\mathbf{s}_j) \tag{5.3.71}$$

and directly from (5.3.47) and (5.3.48)

$$\widehat{\gamma}_{j,k} \xrightarrow{P} \gamma_{j,k},\ j,k = 1, \cdots, N, \tag{5.3.72}$$

and

$$\widehat{a}_{j,k} \xrightarrow{P} a_{j,k},\ j,k = 1, \cdots, N, \tag{5.3.73}$$

respectively. Thus,

$$\widehat{\mathbf{a}} \xrightarrow{P} \mathbf{a}_0. \tag{5.3.74}$$

■

Asymptotic Normality

For any fixed $t \in \mathcal{Z}$ and any $\mathbf{b} \in \mathcal{B}^N$ and $\mathbf{a} \in \mathcal{A}$, we define the random variables

$$Z_t(\mathbf{s}_j, \mathbf{b}, \mathbf{a}) \equiv a_{j,j}\ \varepsilon_t(\mathbf{s}_j, \mathbf{b}_j) - \sum_{\substack{k=1,\\ k\neq j}}^{N} a_{j,k}\ \varepsilon_t(\mathbf{s}_k, \mathbf{b}_k),\ j = 1, \cdots, N, \tag{5.3.75}$$

and the $((Np) \times 1)$ random vector $\mathbf{R}_t(\mathbf{b}, \mathbf{a})$

$$\mathbf{R}_t(\mathbf{b}, \mathbf{a}) \equiv [Z_t(\mathbf{s}_1, \mathbf{b}, \mathbf{a})\mathbf{X}_t^\tau(\mathbf{s}_1), Z_t(\mathbf{s}_2, \mathbf{b}, \mathbf{a})\mathbf{X}_t^\tau(\mathbf{s}_2), \cdots, Z_t(\mathbf{s}_N, \mathbf{b}, \mathbf{a})\mathbf{X}_t^\tau(\mathbf{s}_N)]^\tau. \quad (5.3.76)$$

For the true parameter vector $\mathbf{a}_0$, we may replace $a_{j,k}$ by $a_{(j,k),0}$ and write $Z_t(\mathbf{s}_j, \mathbf{b}, \mathbf{a}_0) \equiv Z_t(\mathbf{s}_j, \mathbf{b})$ and $\mathbf{R}_t(\mathbf{b}, \mathbf{a}_0) \equiv \mathbf{R}_t(\mathbf{b})$. We then define the variance matrix

$$I(\mathbf{b}) \equiv \mathrm{Var}\{\mathbf{R}_t(\mathbf{b})\}. \quad (5.3.77)$$

We let

$$Z_t(\mathbf{s}_j, \mathbf{b}_0) \equiv Z_t(\mathbf{s}_j),\ t \in \mathcal{Z},\ j = 1, \cdots, N, \quad (5.3.78)$$

and

$$\mathbf{R}_t(\mathbf{b}_0) \equiv \mathbf{R}_t,\ t \in \mathcal{Z}. \quad (5.3.79)$$

We can see that $\{Z_t(\mathbf{s}_j),\ j = 1, \cdots, N\}$ and $\{\varepsilon_t(\mathbf{s}_k),\ k = 1, \cdots, N\}$ are two sets of uncorrelated random variables for any $j \neq k$. Moreover, according to (5.3.65), it holds for integers $i, l > 0$ that

$$\mathrm{Cov}\{Z_t(\mathbf{s}_j)Y_{t-i}(\mathbf{s}_j), Z_t(\mathbf{s}_k)Y_{t-l}(\mathbf{s}_k)\} = E\{Z_t(\mathbf{s}_j)Z_t(\mathbf{s}_k)\}E\{Y_{t-i}(\mathbf{s}_j)Y_{t-l}(\mathbf{s}_k)\}, \quad (5.3.80)$$

where

$$E\{Z_t(\mathbf{s}_j)Z_t(\mathbf{s}_k)\} = \mathrm{Cov}\{Z_t(\mathbf{s}_j), Z_t(\mathbf{s}_k)\} = \begin{cases} a_{(j,j),0}, & j = k \\ -a_{(j,k),0}, & j \neq k \end{cases}. \quad (5.3.81)$$

As a result, we may write $I(\mathbf{b}_0)$ to be equal to

$$\begin{bmatrix} a_{(1,1),0}E\{\mathbf{X}_t(\mathbf{s}_1)\mathbf{X}_t^\tau(\mathbf{s}_1)\} & -a_{(1,2),0}E\{\mathbf{X}_t(\mathbf{s}_1)\mathbf{X}_t^\tau(\mathbf{s}_2)\} & \cdots & -a_{(1,N),0}E\{\mathbf{X}_t(\mathbf{s}_1)\mathbf{X}_t^\tau(\mathbf{s}_N)\} \\ -a_{(1,2),0}E\{\mathbf{X}_t(\mathbf{s}_2)\mathbf{X}_t^\tau(\mathbf{s}_1)\} & a_{(2,2),0}E\{\mathbf{X}_t(\mathbf{s}_2)\mathbf{X}_t^\tau(\mathbf{s}_2)\} & & -a_{(2,N),0}E\{\mathbf{X}_t(\mathbf{s}_2)\mathbf{X}_t^\tau(\mathbf{s}_N)\} \\ \vdots & & \ddots & \\ -a_{(1,N),0}E\{\mathbf{X}_t(\mathbf{s}_N)\mathbf{X}_t^\tau(\mathbf{s}_1)\} & -a_{(2,N),0}E\{\mathbf{X}_t(\mathbf{s}_N)\mathbf{X}_t^\tau(\mathbf{s}_2)\} & & a_{(N,N),0}E\{\mathbf{X}_t(\mathbf{s}_N)\mathbf{X}_t^\tau(\mathbf{s}_N)\} \end{bmatrix}. \quad (5.3.82)$$

Finally, we define

$$\mathbf{W}(\mathbf{b}, \mathbf{a}) \equiv \begin{bmatrix} I(\mathbf{b}) & \mathbf{O}_{(Np)\times q} \\ \mathbf{O}_{q\times(Np)} & I(\mathbf{a}) \end{bmatrix}, \quad (5.3.83)$$

where $I(\mathbf{a})$ was defined back in Section 5.2.

Theorem 5.2. Under conditions (C8)-(C10) and as $T \to \infty$ (N fixed), it holds that

$$T^{1/2}\begin{bmatrix} \widehat{\mathbf{b}} - \mathbf{b}_0 \\ \widehat{\mathbf{a}} - \mathbf{a}_0 \end{bmatrix} \xrightarrow{D} N(\mathbf{0}, \mathbf{W}^{-1}(\mathbf{b}_0, \mathbf{a}_0)). \quad (5.3.84)$$

Proof. For any $j,k=1,\cdots,N$, we can write immediately from (5.3.47)

$$\begin{aligned}\widehat{\gamma}_{j,k}-\gamma_{(j,k),0} &= \frac{1}{T}\{\sum_{t=1}^{T}[\varepsilon_t(\mathbf{s}_j)\varepsilon_t(\mathbf{s}_k)-\gamma_{(j,k),0}]-\sum_{t=1}^{T}[\varepsilon_t(\mathbf{s}_k)\mathbf{X}_t^{\tau}(\mathbf{s}_j)]\ [\widehat{\mathbf{b}}_j-\mathbf{b}_{j,0}]\\ &- \sum_{t=1}^{T}[\varepsilon_t(\mathbf{s}_j)\mathbf{X}_t^{\tau}(\mathbf{s}_k)]\ [\widehat{\mathbf{b}}_k-\mathbf{b}_{k,0}]\\ &+ [\widehat{\mathbf{b}}_k-\mathbf{b}_{k,0}]^{\tau}\sum_{t=1}^{T}[\mathbf{X}_t(\mathbf{s}_k)\mathbf{X}_t^{\tau}(\mathbf{s}_j)]\ [\widehat{\mathbf{b}}_j-\mathbf{b}_{j,0}]\}\end{aligned} \tag{5.3.85}$$

and then we can stack all the q equations (5.3.85) in a vector as

$$\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0=\frac{1}{T}\sum_{t=1}^{T}\mathbf{S}_t^{*}-\frac{1}{T}\ \mathbf{H}_1(T)\ [\widehat{\mathbf{b}}-\mathbf{b}_0] \tag{5.3.86}$$

where

$$\mathbf{S}_t^{*}=\mathbf{S}_t-\boldsymbol{\gamma}_0 \tag{5.3.87}$$

and

$$\mathbf{S}_t=[\varepsilon_t(\mathbf{s}_1)\varepsilon_t(\mathbf{s}_2),\cdots,\varepsilon_t(\mathbf{s}_{N-1})\varepsilon_t(\mathbf{s}_N),\varepsilon_t(\mathbf{s}_1)^2,\cdots,\varepsilon_t(\mathbf{s}_N)^2]^{\tau}. \tag{5.3.88}$$

The matrix $\mathbf{H}_1(T)$ is such that its row vector corresponding to (j,k), $j,k=1,\cdots,N$, has as its element

$$\sum_{t=1}^{T}\varepsilon_t(\mathbf{s}_k)Y_{t-i}(\mathbf{s}_j)+O_P(T||\widehat{\mathbf{b}}-\mathbf{b}_0||)$$

if multiplied by $(\widehat{b}_{i,j}-b_{(i,j),0})$, it has element

$$\sum_{t=1}^{T}\varepsilon_t(\mathbf{s}_j)Y_{t-i}(\mathbf{s}_k)+O_P(T||\widehat{\mathbf{b}}-\mathbf{b}_0||)$$

if multiplied by $(b_{i,k}-b_{(i,k),0})$, and it has zero elements anywhere else. Since

$$\frac{1}{T}[\sum_{t=1}^{T}\varepsilon_t(\mathbf{s}_k)Y_{t-i}(\mathbf{s}_j)+O_P(T||\widehat{\mathbf{b}}-\mathbf{b}_0||)]\xrightarrow{P}0, \tag{5.3.89}$$

for any $j,k=1,\cdots,N$ and $i>0$, we may conclude that

$$\frac{1}{T}\ \mathbf{H}_1(T)\xrightarrow{P}\mathbf{O}_{q\times(Np)}. \tag{5.3.90}$$

On the other hand, we may write a Taylor's expansion

$$\widehat{\boldsymbol{\gamma}}=\boldsymbol{\gamma}_0+(\mathbf{J}^{\tau}|_{\mathbf{a}=\widehat{\mathbf{a}}}+\mathbf{H}_2(\widehat{\mathbf{a}}))\ [\widehat{\mathbf{a}}-\mathbf{a}_0], \tag{5.3.91}$$

where $\mathbf{J}$ was defined in (5.2.11) for any $\mathbf{a}\in\mathcal{A}$ and $\mathbf{H}_2(\widehat{\mathbf{a}})$ is a matrix, such that as $T\to\infty$

$$\mathbf{H}_2(\widehat{\mathbf{a}})\xrightarrow{P}\mathbf{H}_2(\mathbf{a}_0)=\mathbf{O}_{q\times q}, \tag{5.3.92}$$

since $\widehat{\mathbf{a}} \xrightarrow{P} \mathbf{a}_0$. It also holds that

$$\mathbf{J}|_{\mathbf{a}=\widehat{\mathbf{a}}} \xrightarrow{P} \mathbf{J}, \tag{5.3.93}$$

because of the consistency of $\widehat{\mathbf{a}}$ and the fact that the derivatives are smooth functions of $\mathbf{a}$, where we have considered for simplicity

$$\mathbf{J}|_{\mathbf{a}=\mathbf{a}_0} \equiv \mathbf{J}.$$

If we combine (5.3.86) and (5.3.91), we may write

$$[\ \mathbf{H}_1(T)/T \quad \mathbf{J}^\tau|_{\mathbf{a}=\widehat{\mathbf{a}}} + \mathbf{H}_2(\widehat{\mathbf{a}})\ ] \cdot \begin{bmatrix} \widehat{\mathbf{b}} - \mathbf{b}_0 \\ \widehat{\mathbf{a}} - \mathbf{a}_0 \end{bmatrix} = \frac{1}{T} \sum_{t=1}^{T} \mathbf{S}_t^*. \tag{5.3.94}$$

We have also defined our estimators according to (5.3.46), which we may re-write for any $i = 1, \cdots, p$ and $j = 1, \cdots, N$ as

$$\begin{aligned} &\sum_{t=1}^{T} \{[\widehat{a}_{(j,j),0}\ \widehat{\varepsilon}_t(\mathbf{s}_j) - \sum_{\substack{k=1,\\ k\neq j}}^{N} \widehat{a}_{(j,k),0}\ \widehat{\varepsilon}_t(\mathbf{s}_k)] Y_{t-i}(\mathbf{s}_j)\} \\ + &\sum_{t=1}^{T} \{\widehat{\varepsilon}_t(\mathbf{s}_j) Y_{t-i}(\mathbf{s}_j)\}\ (\widehat{a}_{j,j} - a_{(j,j),0}) - \sum_{\substack{k=1,\\ k\neq j}}^{N} \{\widehat{\varepsilon}_t(\mathbf{s}_k) Y_{t-i}(\mathbf{s}_j)\}\ (\widehat{a}_{j,k} - a_{(j,k),0}) = 0. \end{aligned} \tag{5.3.95}$$

It holds for any $i = 1, \cdots, p$, and $j, k = 1, \cdots, N$, that

$$\frac{1}{T} \sum_{t=1}^{T} \widehat{\varepsilon}_t(\mathbf{s}_k) Y_{t-i}(\mathbf{s}_j) \xrightarrow{P} \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t(\mathbf{s}_k) Y_{t-i}(\mathbf{s}_j) \xrightarrow{P} 0, \tag{5.3.96}$$

as $T \to \infty$. It also holds that

$$\begin{aligned} &\sum_{t=1}^{T} \{[\widehat{a}_{(j,j),0}\ \widehat{\varepsilon}_t(\mathbf{s}_j) - \sum_{\substack{k=1,\\ k\neq j}}^{N} \widehat{a}_{(j,k),0}\ \widehat{\varepsilon}_t(\mathbf{s}_k)] Y_{t-i}(\mathbf{s}_j)\} = \sum_{t=1}^{T} Z_t(\mathbf{s}_j) Y_{t-i}(\mathbf{s}_j) \\ - &\{a_{(j,j),0} \sum_{t=1}^{T} [Y_{t-i}(\mathbf{s}_j) \mathbf{X}_t^\tau(\mathbf{s}_j)]\ [\widehat{\mathbf{b}}_j - \mathbf{b}_{j,0}] \\ - &\sum_{\substack{k=1,\\ k\neq j}}^{N} a_{(j,k),0} \sum_{t=1}^{T} [Y_{t-i}(\mathbf{s}_j) \mathbf{X}_t^\tau(\mathbf{s}_k)]\ [\widehat{\mathbf{b}}_k - \mathbf{b}_{k,0}]\}. \end{aligned} \tag{5.3.97}$$

As a result, we can re-write (5.3.95) as

$$\begin{aligned}
& a_{(j,j),0} \sum_{t=1}^{T} [Y_{t-i}(\mathbf{s}_j)\mathbf{X}_t^\tau(\mathbf{s}_j)]\ [\widehat{\mathbf{b}}_j - \mathbf{b}_{j,0}] \\
- & \sum_{\substack{k=1,\\ k\neq j}}^{N} a_{(j,k),0} \sum_{t=1}^{T} [Y_{t-i}(\mathbf{s}_j)\mathbf{X}_t^\tau(\mathbf{s}_k)]\ [\widehat{\mathbf{b}}_k - \mathbf{b}_{k,0}] \\
- & \sum_{t=1}^{T} \{\widehat{\varepsilon}_t(\mathbf{s}_j) Y_{t-i}(\mathbf{s}_j)\}\ (\widehat{a}_{j,j} - a_{(j,j),0}) + \sum_{\substack{k=1,\\ k\neq j}}^{N} \sum_{t=1}^{T} \{\widehat{\varepsilon}_t(\mathbf{s}_k) Y_{t-i}(\mathbf{s}_j)\}\ (\widehat{a}_{j,k} - a_{(j,k),0}) \\
= & \sum_{t=1}^{T} Z_t(\mathbf{s}_j) Y_{t-i}(\mathbf{s}_j). \qquad (5.3.98)
\end{aligned}$$

Stacking all the (Np) equations together, we may write

$$\boldsymbol{\Gamma}(T)\ [\widehat{\mathbf{b}} - \mathbf{b}_0] + \mathbf{H}_3(T)\ [\widehat{\mathbf{a}} - \mathbf{a}_0] = \sum_{t=1}^{T} \mathbf{R}_t, \qquad (5.3.99)$$

where $\boldsymbol{\Gamma}(T)$ is set to be equal to

$$\sum_{t=1}^{T} \begin{bmatrix}
a_{(1,1),0}\ \mathbf{X}_t(\mathbf{s}_1)\mathbf{X}_t^\tau(\mathbf{s}_1) & -a_{(1,2),0}\ \mathbf{X}_t(\mathbf{s}_1)\mathbf{X}_t^\tau(\mathbf{s}_2) & \cdots & -a_{(1,N),0}\ \mathbf{X}_t(\mathbf{s}_1)\mathbf{X}_t^\tau(\mathbf{s}_N) \\
-a_{(1,2),0}\ \mathbf{X}_t(\mathbf{s}_2)\mathbf{X}_t^\tau(\mathbf{s}_1) & a_{(2,2),0}\ \mathbf{X}_t(\mathbf{s}_2)\mathbf{X}_t^\tau(\mathbf{s}_2) & & -a_{(2,N),0}\ \mathbf{X}_t(\mathbf{s}_2)\mathbf{X}_t^\tau(\mathbf{s}_N) \\
\vdots & & \ddots & \\
-a_{(1,N),0}\ \mathbf{X}_t(\mathbf{s}_N)\mathbf{X}_t^\tau(\mathbf{s}_1) & -a_{(2,N),0}\ \mathbf{X}_t(\mathbf{s}_N)\mathbf{X}_t^\tau(\mathbf{s}_2) & & a_{(N,N),0}\ \mathbf{X}_t(\mathbf{s}_N)\mathbf{X}_t^\tau(\mathbf{s}_N)
\end{bmatrix} \qquad (5.3.100)$$

and the matrix $\mathbf{H}_3(T)$ is such that

$$\frac{1}{T}\ \mathbf{H}_3(T) \xrightarrow{P} \mathbf{O}_{(Np)\times q} \qquad (5.3.101)$$

as $T \to \infty$. We can see immediately that

$$\frac{1}{T}\ \boldsymbol{\Gamma}(T) \xrightarrow{P} I(\mathbf{b}_0). \qquad (5.3.102)$$

When we put (5.3.94) and (5.3.99) together, we may write

$$T^{1/2} \begin{bmatrix} \widehat{\mathbf{b}} - \mathbf{b}_0 \\ \widehat{\mathbf{a}} - \mathbf{a}_0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Gamma}(T)/T & \mathbf{H}_3(T)/T \\ \mathbf{H}_1(T)/T & \mathbf{J}^\tau|_{\mathbf{a}=\widehat{\mathbf{a}}} + \mathbf{H}_2(\widehat{\mathbf{a}}) \end{bmatrix}^{-1} T^{-1/2} \sum_{t=1}^{T} \mathbf{U}_t, \qquad (5.3.103)$$

where

$$\mathbf{U}_t \equiv [\mathbf{R}_t^\tau, \mathbf{S}_t^{*\tau}]^\tau. \qquad (5.3.104)$$

Recalling (5.3.90), (5.3.92), (5.3.93), (5.3.101) and (5.3.102), we can write that as $T \to \infty$

$$\begin{aligned}\begin{bmatrix} \boldsymbol{\Gamma}(T)/T & \mathbf{H}_3(T)/T \\ \mathbf{H}_1(T)/T & \mathbf{J}^\tau|_{\mathbf{a}=\widehat{\mathbf{a}}} + \mathbf{H}_2(\widehat{\mathbf{a}}) \end{bmatrix}^{-1} &\xrightarrow{P} \begin{bmatrix} I(\mathbf{b}_0) & \mathbf{O}_{(Np)\times q} \\ \mathbf{O}_{q\times(Np)} & \mathbf{J}^\tau \end{bmatrix}^{-1} \\ &= \begin{bmatrix} I(\mathbf{b}_0)^{-1} & \mathbf{O}_{(Np)\times q} \\ \mathbf{O}_{q\times(Np)} & \mathbf{J}^{\tau-1} \end{bmatrix}. \end{aligned} \tag{5.3.105}$$

Next, we will show that for any $\boldsymbol{\lambda} \in \mathcal{R}^{Np+q}$, it holds that

$$T^{-1/2}\sum_{t=1}^{T} \boldsymbol{\lambda}^\tau \mathbf{U}_t \xrightarrow{D} N\left(0, \boldsymbol{\lambda}^\tau \begin{bmatrix} I(\mathbf{b}_0) & \mathbf{O}_{(Np)\times q} \\ \mathbf{O}_{q\times(Np)} & I(\boldsymbol{\gamma}_0)^{-1} \end{bmatrix} \boldsymbol{\lambda}\right), \tag{5.3.106}$$

where $I(\boldsymbol{\gamma})$ was defined back in Section 5.2. We can see that $E\{\mathbf{S}_t^*\} = \mathbf{0}$ and that $\mathrm{Var}\{\mathbf{S}_t^*\} = \mathrm{Var}\{\mathbf{S}_t\} = I(\boldsymbol{\gamma}_0)^{-1}$. Similarly, $E\{\mathbf{R}_t\} = \mathbf{0}$ and $\mathrm{Var}\{\mathbf{R}_t\} = I(\mathbf{b}_0)$. We write for any $i = 1, \cdots, p$, and any $j, k, m = 1, \cdots, N$,

$$\begin{aligned}\mathrm{Cov}\{\varepsilon_t(\mathbf{s}_j)\varepsilon_t(\mathbf{s}_k), Z_t(\mathbf{s}_m)Y_{t-i}(\mathbf{s}_m)\} &= E\{\varepsilon_t(\mathbf{s}_j)\varepsilon_t(\mathbf{s}_k)Z_t(\mathbf{s}_m)Y_{t-i}(\mathbf{s}_m)\} - \gamma_{i,j}\cdot 0 \\ &= E\{\varepsilon_t(\mathbf{s}_j)\varepsilon_t(\mathbf{s}_k)Z_t(\mathbf{s}_m)\}E\{Y_{t-i}(\mathbf{s}_m)\} = 0, \end{aligned} \tag{5.3.107}$$

due to independence and causality. We may write

$$E\{\mathbf{U}_t\} = \mathbf{0} \tag{5.3.108}$$

and

$$\mathrm{Var}\{\mathbf{U}_t\} = \begin{bmatrix} I(\mathbf{b}_0) & \mathbf{O}_{(Np)\times q} \\ \mathbf{O}_{q\times(Np)} & I(\boldsymbol{\gamma}_0)^{-1} \end{bmatrix}. \tag{5.3.109}$$

We recall the MA(∞) representations of (5.3.15) and define for fixed $K = 1, 2, \cdots$, the random variables

$$Y_t^{(K)}(\mathbf{s}_j) = \varepsilon_t(\mathbf{s}_j) + \sum_{i=1}^{K} \psi_{i,j}\ \varepsilon_{t-i}(\mathbf{s}_j),\ j = 1, \cdots, N, \tag{5.3.110}$$

and

$$\mathbf{X}_t^{(K)}(\mathbf{s}_j) = [Y_{t-1}^{(K)}(\mathbf{s}_j), \cdots, Y_{t-p}^{(K)}(\mathbf{s}_j)]^\tau,\ j = 1, \cdots, N. \tag{5.3.111}$$

As a result, we let

$$\mathbf{R}_t^{(K)} = [Z_t(\mathbf{s}_1)\mathbf{X}_t^{(K)\tau}(\mathbf{s}_1), \cdots, Z_t(\mathbf{s}_N)\mathbf{X}_t^{(K)\tau}(\mathbf{s}_N)]^\tau \tag{5.3.112}$$

and

$$\mathbf{U}_t^{(K)} = [\mathbf{R}_t^{(K)^\tau}, \mathbf{S}_t^{*\tau}]^\tau. \tag{5.3.113}$$

We can write that $\{\boldsymbol{\lambda}^\tau \mathbf{U}_t^{(K)},\ t \in \mathcal{Z}\}$ is a K-dependent white noise sequence and it holds as $T \to \infty$

$$T^{-1/2}\sum_{t=1}^{T}\boldsymbol{\lambda}^\tau \mathbf{U}_t^{(K)} \xrightarrow{D} \boldsymbol{\lambda}^\tau \mathbf{V}^{(K)} \sim N\left(0, \boldsymbol{\lambda}^\tau \begin{bmatrix} \mathrm{Var}\{\mathbf{R}_t^{(K)}\} & \mathbf{O}_{(Np)\times q} \\ \mathbf{O}_{q\times(Np)} & I(\boldsymbol{\gamma}_0)^{-1} \end{bmatrix} \boldsymbol{\lambda}\right), \quad (5.3.114)$$

since again

$$\mathrm{Cov}\{\varepsilon_t(\mathbf{s}_j)\varepsilon_t(\mathbf{s}_k), Z_t(\mathbf{s}_m)Y_{t-i}^{(K)}(\mathbf{s}_m)\} = 0, \quad (5.3.115)$$

for any $i = 1, \cdots, p$, and any $j, k, m = 1, \cdots, N$. Since

$$\mathrm{Var}\{\mathbf{R}_t^{(K)}\} \to \mathrm{Var}\{\mathbf{R}_t\} \quad (5.3.116)$$

as $K \to \infty$, we have that

$$\boldsymbol{\lambda}^\tau \mathbf{V}^{(K)} \xrightarrow{D} \boldsymbol{\lambda}^\tau \mathbf{V} \sim N\left(0, \boldsymbol{\lambda}^\tau \begin{bmatrix} \mathrm{Var}\{\mathbf{R}_t\} & \mathbf{O}_{(Np)\times q} \\ \mathbf{O}_{q\times(Np)} & I(\boldsymbol{\gamma}_0)^{-1} \end{bmatrix} \boldsymbol{\lambda}\right). \quad (5.3.117)$$

Also, it is easy to check that

$$T^{-1}\mathrm{Var}\{\boldsymbol{\lambda}^\tau \sum_{t=1}^{T}(\mathbf{U}_t^{(K)} - \mathbf{U}_t)\} = \boldsymbol{\lambda}^\tau E\{(\mathbf{U}_t^{(K)} - \mathbf{U}_t)(\mathbf{U}_t^{(K)} - \mathbf{U}_t)^\tau\}\boldsymbol{\lambda} \to 0 \quad (5.3.118)$$

as $K \to \infty$. As a result, we may conclude from the Cramer-Wold device that

$$T^{-1/2}\sum_{t=1}^{T}\mathbf{U}_t \xrightarrow{D} N\left(0, \begin{bmatrix} I(\mathbf{b}_0) & \mathbf{O}_{(Np)\times q} \\ \mathbf{O}_{q\times(Np)} & I(\boldsymbol{\gamma}_0)^{-1} \end{bmatrix}\right). \quad (5.3.119)$$

From (5.3.103), (5.3.105) and (5.3.119), we may conclude that

$$T^{1/2}\begin{bmatrix} \widehat{\mathbf{b}} - \mathbf{b}_0 \\ \widehat{\mathbf{a}} - \mathbf{a}_0 \end{bmatrix} \xrightarrow{D} N(\mathbf{0}, \mathbf{W}^{-1}(\mathbf{b}_0, \mathbf{a}_0)), \quad (5.3.120)$$

where

$$\mathbf{W}(\mathbf{b}_0, \mathbf{a}_0) = \begin{bmatrix} I(\mathbf{b}_0) & \mathbf{O}_{(Np)\times q} \\ \mathbf{O}_{q\times(Np)} & I(\mathbf{a}_0) \end{bmatrix}$$

and $I(\mathbf{a}_0) = \mathbf{J}I(\boldsymbol{\gamma}_0)\mathbf{J}^\tau$. ■

Remark 5.3. If we consider the $(q \times 1)$ vector

$$\boldsymbol{\beta} \equiv [\beta_{1,2}, \cdots, \beta_{1,N}, \beta_{2,3}, \cdots, \beta_{2,N}, \cdots, \beta_{N-1,N}, \nu_1, \cdots, \nu_N]^\tau \quad (5.3.121)$$

then we may recall that

$$\beta_{j,k} = \frac{a_{j,k}}{a_{j,j}},\ j \neq k,\ j, k = 1, \cdots, N,$$

and

$$\nu_j = \frac{1}{a_{j,j}},\ j = 1, \cdots, N,$$

in order to write $\boldsymbol{\beta}$ as a function of the original parameter vector $\mathbf{a}$. According to Proposition 6.4.3 of Brockwell and Davis (1991, p.211), we may then write as $T \to \infty$ that

$$T^{1/2} \begin{bmatrix} \widehat{\mathbf{b}} - \mathbf{b}_0 \\ \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \end{bmatrix} \xrightarrow{D} N\left(\mathbf{0}, \begin{bmatrix} I(\mathbf{b}_0)^{-1} & \mathbf{O}_{(Np)\times q} \\ \mathbf{O}_{q\times(Np)} & \mathbf{D}^\tau I(\mathbf{a}_0)^{-1}\mathbf{D} \end{bmatrix}\right), \tag{5.3.122}$$

where $\mathbf{D}^\tau = \partial\boldsymbol{\beta}/\partial\mathbf{a}^\tau|_{\mathbf{a}=\mathbf{a}_0}$. It holds that

$$\partial\beta_{j,k}/\partial a_{j,k} = \frac{1}{a_{j,j}} = \nu_j \tag{5.3.123}$$

and

$$\partial\beta_{j,k}/\partial a_{j,j} = -\frac{a_{j,k}}{a_{j,j}^2} = -\nu_j \cdot \beta_{j,k}. \tag{5.3.124}$$

It also holds that

$$\partial\nu_j/\partial a_{j,j} = -\frac{1}{a_{j,j}^2} = -\nu_j^2. \tag{5.3.125}$$

5.4 Hypothesis testing

5.4.1 Tests for the serial dependence

Test for the significance of the time factor

We want to test

$$H_0 : \{\mathbf{Y}_t\} \sim IID(\mathbf{0}, \mathbf{V})$$

$$H_1 : \text{otherwise.}$$

In other words, the null hypothesis assumes that $\boldsymbol{\Phi}_1 = \cdots = \boldsymbol{\Phi}_p = \mathbf{O}_{N\times N}$.

We observe $\{\mathbf{Y}_t,\ t = 1-p, \cdots, T\}$ and we estimate the auto-covariance matrices as

$$\widehat{\boldsymbol{\Sigma}}(i) = \frac{1}{T}\sum_{t=1-p+i}^{T} \mathbf{Y}_t\mathbf{Y}_{t-i}^\tau,\ i = 0, \cdots, T-1. \tag{5.4.1}$$

According to Reinsel (1997, p.151), under H_0 and as $T \to \infty$

$$\frac{T}{\sqrt{T-i}}\ \mathrm{vec}\{\widehat{\boldsymbol{\Sigma}}(\mathrm{i})\} \xrightarrow{\mathrm{D}} \mathrm{N}(\mathbf{0}, \mathbf{V}\otimes\mathbf{V}), \tag{5.4.2}$$

for $i = 1, \cdots, T-1$. Suggestion (5.4.2) can be slightly changed if we ignore the factor $T/(T-i)$ which is asymptotically equal to 1 for fixed i. Moreover, under H_0 and as

$T \to \infty$, the random vectors $\text{vec}\{\widehat{\boldsymbol{\Sigma}}(\text{i})\}$ and $\text{vec}\{\widehat{\boldsymbol{\Sigma}}(\text{l})\}$ are asymptotically independent for $i \neq l$ and $i, l = 1, \cdots, T-1$. Thus,

$$\begin{aligned} & T^2 \sum_{i=1}^{d} \frac{1}{T-i} \left[\text{vec}\{\widehat{\boldsymbol{\Sigma}}(\text{i})\}^\tau (\mathbf{V}^{-1} \otimes \mathbf{V}^{-1}) \text{vec}\{\widehat{\boldsymbol{\Sigma}}(\text{i})\}\right] \\ = \; & T^2 \sum_{i=1}^{d} \frac{1}{T-i} \text{tr}\{\mathbf{V}^{-1} \widehat{\boldsymbol{\Sigma}}(i)^\tau \mathbf{V}^{-1} \widehat{\boldsymbol{\Sigma}}(i)\} \xrightarrow{D} \mathcal{X}^2_{dN^2}, \end{aligned} \tag{5.4.3}$$

where d can be any integer between 1 and $T-1$. This asymptotic result, known as a multivariate portmanteau test, is not affected by the replacement of the unknown matrix $\mathbf{V}$ by its consistent estimator $\widehat{\boldsymbol{\Sigma}}(0) = \widehat{\mathbf{V}}$:

$$T^2 \sum_{i=1}^{d} \frac{1}{T-i} \text{tr}\{\widehat{\mathbf{V}}^{-1} \widehat{\boldsymbol{\Sigma}}(i)^\tau \widehat{\mathbf{V}}^{-1} \widehat{\boldsymbol{\Sigma}}(i)\} \xrightarrow{D} \mathcal{X}^2_{dN^2}, \; d = 1, \cdots, T-1. \tag{5.4.4}$$

Test for equal coefficients on different locations

We want to test

$$H_0 : \boldsymbol{\Phi}_i = b_i \cdot \mathbf{I}_N, \; i = 1, \cdots, p$$

$$H_1 : \text{otherwise.}$$

In other words, the null hypothesis assumes that we can write

$$Y_t(\mathbf{s}_j) = \varepsilon_t(\mathbf{s}_j) + \sum_{i=1}^{\infty} \psi_i \; \varepsilon_{t-i}(\mathbf{s}_j) \tag{5.4.5}$$

for every $j = 1, \cdots, N$. If we put them all together, that would be

$$\mathbf{Y}_t = b_1 \; \mathbf{Y}_{t-1} + \cdots + b_p \; \mathbf{Y}_{t-p} + \boldsymbol{\varepsilon}_t \tag{5.4.6}$$

and

$$\mathbf{Y}_t = \boldsymbol{\varepsilon}_t + \sum_{i=1}^{\infty} \psi_i \; \boldsymbol{\varepsilon}_{t-i}. \tag{5.4.7}$$

We may write under the null hypothesis

$$\mathbf{b}_j = \mathbf{b}, \; j = 1, \cdots, N. \tag{5.4.8}$$

As a result,

$$\text{Cov}\{Y_t(\mathbf{s}_j), Y_{t-i}(\mathbf{s}_j)\} = \text{Cov}\{Y_t(\mathbf{s}_k), Y_{t-i}(\mathbf{s}_k)\}, \; i \in \mathbb{Z}, \; j, k = 1, \cdots, N, \tag{5.4.9}$$

which implies that we meet some form of stationarity over space.

Thanks to Theorem 5.2, for observations $\{Y_t(\mathbf{s}_j),\ t = 1-p, \cdots, T\ j = 1, \cdots, N\}$ and when $\{\boldsymbol{\varepsilon}_t\} \sim IID(\mathbf{0}, \mathbf{V})$, we might make a decision using the ratio of conditional Gaussian likelihoods, which is based on the statistic

$$\lambda_{LR} = 2(l(\widehat{\mathbf{b}}, \widehat{\mathbf{a}}) - l_0(\widehat{\mathbf{b}}_0, \widehat{\mathbf{a}}_0)) \xrightarrow{D} \mathcal{X}^2_{(N-1)p} \tag{5.4.10}$$

as $T \to \infty$. In (5.4.10), the logarithm of the Gaussian likelihood $l(\mathbf{b}, \mathbf{a})$ was defined back in (5.3.25) and the logarithm of the Gaussian likelihood $l_0(\mathbf{b}, \mathbf{a})$ under the null hypothesis is

$$l_0(\mathbf{b}, \mathbf{a}) \equiv T/2 \log |\mathbf{V}^{-1}| - 1/2 \sum_{t=1}^{T} [\sum_{j=1}^{N} a_{j,j}\ \varepsilon_t(\mathbf{s}_j, \mathbf{b})^2 - 2 \sum_{\substack{j,k=1,\\ j<k}}^{N} a_{j,k}\ \varepsilon_t(\mathbf{s}_j, \mathbf{b}) \varepsilon_t(\mathbf{s}_k, \mathbf{b})], \tag{5.4.11}$$

where $\mathbf{b} \in \mathcal{B},\ \mathbf{a} \in \mathcal{A}$. The arguments $\widehat{\mathbf{b}} \in \mathcal{B}^N,\ \widehat{\mathbf{a}} \in \mathcal{A}$ and $\widehat{\mathbf{b}}_0 \in \mathcal{B},\ \widehat{\mathbf{a}}_0 \in \mathcal{A}$ maximize $l(\mathbf{b}, \mathbf{a})$ and $l_0(\mathbf{b}, \mathbf{a})$, respectively. We may then write under the restriction $\widehat{a}_{j,k} = \widehat{a}_{k,j},\ j, k = 1, \cdots, N,\ j \neq k$, that

$$\begin{aligned}
l(\widehat{\mathbf{b}}, \widehat{\mathbf{a}}) &= T/2\{\log |\widehat{\mathbf{V}}|^{-1}\} - \sum_{j=1}^{N} [\widehat{a}_{j,j}\ 1/T \sum_{t=1}^{T} \widehat{\varepsilon}_t(\mathbf{s}_j, \widehat{\mathbf{b}}_j)^2 \\
&- \sum_{\substack{k=1,\\ k\neq j}}^{N} \widehat{a}_{j,k}\ 1/T \sum_{t=1}^{T} \widehat{\varepsilon}_t(\mathbf{s}_j, \widehat{\mathbf{b}}_j) \widehat{\varepsilon}_t(\mathbf{s}_k, \mathbf{b}_k)]\} \\
&= T/2\{-\log\{|\widehat{\mathbf{V}}|\} - \sum_{j=1}^{N} [\widehat{a}_{j,j} \widehat{\gamma}_{j,j} - \sum_{\substack{k=1,\\ k\neq j}}^{N} \widehat{a}_{j,k} \widehat{\gamma}_{j,k}]\} \\
&= T/2\{-\log\{|\widehat{\mathbf{V}}|\} - \mathrm{tr}\{\widehat{\mathbf{V}} \widehat{\mathbf{V}}^{-1}\}\} \\
&= -T/2 \log |\widehat{\mathbf{V}}| - NT/2
\end{aligned} \tag{5.4.12}$$

We recall that $\widehat{\mathbf{V}} = [\widehat{\gamma}_{j,k}]_{j,k=1}^N$ with $\widehat{\gamma}_{j,k} = \frac{1}{T} \sum_{t=1}^{T} \widehat{\varepsilon}_t(\mathbf{s}_j, \widehat{\mathbf{b}}_j) \widehat{\varepsilon}_t(\mathbf{s}_k, \widehat{\mathbf{b}}_k)$ and $\{\widehat{\mathbf{b}}_j \in \mathcal{B},\ j = 1, \cdots, N\}$ are such that

$$\sum_{t=1}^{T} \{[\widehat{a}_{j,j}\ \widehat{\varepsilon}_t(\mathbf{s}_j, \widehat{\mathbf{b}}_j) - \sum_{\substack{k=1,\\ k\neq j}}^{N} \widehat{a}_{j,k}\ \widehat{\varepsilon}_t(\mathbf{s}_k, \widehat{\mathbf{b}}_k)] Y_{t-i}(\mathbf{s}_j)\} = 0$$

for all $i = 1, \cdots, p,\ j = 1, \cdots, N$. In a similar way, we can show that

$$l_0(\widehat{\mathbf{b}}_0, \widehat{\mathbf{a}}_0) = -T/2 \log |\widehat{\mathbf{V}}_0| - NT/2, \tag{5.4.13}$$

where

$$\widehat{\mathbf{V}}_0 = \frac{1}{T} [\sum_{t=1}^{T} (Y_t(\mathbf{s}_j) - \mathbf{X}_t^T(\mathbf{s}_j) \widehat{\mathbf{b}}_0)(Y_t(\mathbf{s}_k) - \mathbf{X}_t^T(\mathbf{s}_k) \widehat{\mathbf{b}}_0)]_{j,k=1}^N \tag{5.4.14}$$

and $\widehat{\mathbf{b}}_0$ is such that

$$\sum_{t=1}^{T}\sum_{j=1}^{N}\{2\,\widehat{a}_{(j,j),0}\,\varepsilon_t(\mathbf{s}_j,\widehat{\mathbf{b}}_0)Y_{t-i}(\mathbf{s}_j) - \sum_{\substack{k=1,\\ k\neq j}}^{N}\widehat{a}_{(j,k),0}\,[\varepsilon_t(\mathbf{s}_j,\widehat{\mathbf{b}}_0)Y_{t-i}(\mathbf{s}_k)+\varepsilon_t(\mathbf{s}_k,\widehat{\mathbf{b}}_0)Y_{t-i}(\mathbf{s}_j)]\} = 0, \tag{5.4.15}$$

for all $i = 1, \cdots, p$. As a result,

$$\lambda_{LR} = T\log\{|\widehat{\mathbf{V}}_0|/|\widehat{\mathbf{V}}|\} \xrightarrow{D} \mathcal{X}^2_{(N-1)p} \tag{5.4.16}$$

as $T \to \infty$ and H_0 holds.

5.4.2 Tests for the spatial interdependence

Test for smaller cliques

We take this opportunity to refer to the terms 'clique' and 'neighbour'. For Gaussian random variables, the sites $\{\mathbf{s}_j,\ j = 1, \cdots, N\}$ form a clique if and only if $a_{j,k} \neq 0$ for all $j \neq k,\ j,k = 1, \cdots, N$. Moreover the site $\mathbf{s}_j$ is a neighbour of $\mathbf{s}_k$, where $k \neq j$, if and only if $a_{j,k} \neq 0$, according to Besag (1975, pp. 180-181). In general when we do not deal with Gaussian random variables, two sites $\mathbf{s}_j$ and $\mathbf{s}_k$ with $j,k = 1, \cdots, N,\ j \neq k$, are called neighbours if the conditional distribution of site $\mathbf{s}_j$ given the values of all other sites, does depend on site $\mathbf{s}_k$, or vice versa. This is the essence of local or Markov independence. A clique then is consisted of sites that are all neighbours with each other.

We would like to test

$$H_0 : a_{j,k} = 0,\ k = 1, \cdots, N,\ k \neq j$$
$$H_1 : \text{otherwise,}$$

i.e. whether the site $\mathbf{s}_j$ belongs in the clique of all other sites. If we denote with $\mathbf{V}^{-1}_{(N-1)}$ the $((N-1)\times(N-1))$ inverse of the variance matrix that remains if we exclude the covariances on the j-th row and column, then we may write the logarithm of the likelihood under H_0, as

$$\begin{aligned} l_0(\mathbf{b}_0,\mathbf{a}_0) &= T/2\ \log|\mathbf{V}^{-1}_{(N-1)}| - 1/2\sum_{t=1}^{T}[\sum_{\substack{m=1,\\ m\neq j}} a_{m,m}\ \varepsilon_t(\mathbf{s}_m,\mathbf{b}_m)^2 \\ &- 2\sum_{\substack{m,k=1,\\ m<k,\ k\neq j}}^{N} a_{m,k}\ \varepsilon_t(\mathbf{s}_m,\mathbf{b}_m)\varepsilon_t(\mathbf{s}_k,\mathbf{b}_k)], \end{aligned} \tag{5.4.17}$$

with parameter vectors

$$\mathbf{a}_0 \equiv [\quad a_{1,2},\cdots,a_{1,j-1},a_{1,j+1},\cdots,a_{1,N},a_{2,3},\cdots,a_{2,j-1},a_{2,j+1},\cdots,a_{2,N},\cdots$$
$$a_{1,1},\cdots,a_{j-1,j-1},a_{j+1,j+1},\cdots,a_{N,N}\quad]^\tau \tag{5.4.18}$$

and

$$\mathbf{b}_0 \equiv [\mathbf{b}_1^\tau,\cdots,\mathbf{b}_{j-1}^\tau,\mathbf{b}_{j+1}^\tau,\cdots,\mathbf{b}_N^\tau]^\tau. \tag{5.4.19}$$

If we then consider

$$\widehat{\mathbf{b}}_0 \equiv [\widehat{\mathbf{b}}_{1,0}^\tau,\cdots,\widehat{\mathbf{b}}_{j-1,0}^\tau,\widehat{\mathbf{b}}_{j+1,0}^\tau,\cdots,\widehat{\mathbf{b}}_{N,0}^\tau]^\tau$$

and

$$\widehat{\mathbf{a}}_0 \equiv [\quad \widehat{a}_{(1,2),0},\cdots,\widehat{a}_{(1,j-1),0},\widehat{a}_{(1,j+1),0},\cdots,\widehat{a}_{(1,N),0},\widehat{a}_{(2,3),0},\cdots,\widehat{a}_{(2,j-1),0},\widehat{a}_{(2,j+1),0},$$
$$\cdots,\widehat{a}_{(2,N),0},\cdots,\widehat{a}_{(1,1),0},\cdots,\widehat{a}_{(j-1,j-1),0},\widehat{a}_{(j+1,j+1),0},\cdots,\widehat{a}_{(N,N),0}\quad]^\tau$$

the arguments that maximize it, it holds in a similar way to (5.4.12), that

$$l_0(\widehat{\mathbf{b}}_0,\widehat{\mathbf{a}}_0) = -T/2\log|\widehat{\mathbf{V}}_{(N-1)}| - (N-1)T/2, \tag{5.4.20}$$

where $\widehat{\mathbf{V}}_{(N-1)} = [\widehat{\gamma}_{(m,k),0}]_{\substack{m,k=1,\\ m,k\neq j}}^{N}$ with $\widehat{\gamma}_{(m,k),0} = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(\mathbf{s}_m,\widehat{\mathbf{b}}_{m,0})\varepsilon_t(\mathbf{s}_k,\widehat{\mathbf{b}}_{k,0})$ and it holds that

$$\sum_{t=1}^{T}\{[\widehat{a}_{(m,m),0}\ \varepsilon_t(\mathbf{s}_m,\widehat{\mathbf{b}}_{m,0}) - \sum_{\substack{k=1,\\ k\neq j,m}}\widehat{a}_{(m,k),0}\ \varepsilon_t(\mathbf{s}_k,\widehat{\mathbf{b}}_{k,0})]Y_{t-i}(\mathbf{s}_m)\} = 0, \tag{5.4.21}$$

for all $i = 1,\cdots,p$ and $m = 1,\cdots,j-1,j+1,\cdots,N$. If we combine (5.4.20) and (5.4.12), we may write the likelihood ratio statistic as

$$\lambda_{LR} = T\log\{|\widehat{\mathbf{V}}_{(N-1)}|/|\widehat{\mathbf{V}}|\} - T \xrightarrow{D} \mathcal{X}^2_{df_1} \tag{5.4.22}$$

as $T\to\infty$ and H_0 holds. The degrees of freedom are

$$df_1 = [Np + N(N+1)/2] - [(N-1)p + (N-1)N/2] = p + N.$$

Test for equality of conditional variances

We want to test

$$H_0 : \nu_j = \nu,\ j = 1,\cdots,N$$
$$H_1 : \text{otherwise}$$

and under the null hypothesis, it holds that

$$\mathbf{V}^{-1} = \frac{1}{\nu}\,\mathbf{B}. \tag{5.4.23}$$

We then write the log-likelihood as

$$\begin{aligned} l_0(\mathbf{b},\boldsymbol{\beta},\nu) &= -NT/2\ \log\nu + T/2\ \log|\mathbf{B}| - \frac{1}{2\nu}\sum_{t=1}^{T}[\sum_{j=1}^{N}\varepsilon_t(\mathbf{s}_j,\mathbf{b}_j)^2 \\ &- 2\sum_{\substack{j,k=1,\\ j<k}}^{N}\beta_{j,k}\ \varepsilon_t(\mathbf{s}_j,\mathbf{b}_j)\varepsilon_t(\mathbf{s}_k,\mathbf{b}_k)] \end{aligned} \tag{5.4.24}$$

and consider the arguments that maximize it $\widehat{\mathbf{b}}_0, \widehat{\boldsymbol{\beta}}_0$ and $\widehat{\nu}$. We can see immediately that

$$\widehat{\nu} = \frac{1}{NT}\sum_{t=1}^{T}[\sum_{j=1}^{N}\varepsilon_t(\mathbf{s}_j,\widehat{\mathbf{b}}_{j,0})^2 - 2\sum_{\substack{j,k=1,\\ j<k}}^{N}\widehat{\beta}_{(j,k),0}\ \varepsilon_t(\mathbf{s}_j,\widehat{\mathbf{b}}_{j,0})\varepsilon_t(\mathbf{s}_k,\widehat{\mathbf{b}}_{k,0})]. \tag{5.4.25}$$

Then it holds that

$$l_0(\widehat{\mathbf{b}}_0,\widehat{\boldsymbol{\beta}}_0,\widehat{\nu}) = -NT/2\ \log\widehat{\nu} + T/2\log|\widehat{\mathbf{B}}_0| - NT/2 \tag{5.4.26}$$

where

$$\widehat{\mathbf{B}}_0 \equiv \begin{bmatrix} 1 & -\widehat{\beta}_{(1,2),0} & \cdots & -\widehat{\beta}_{(1,N),0} \\ -\widehat{\beta}_{(1,2),0} & 1 & & -\widehat{\beta}_{(2,N),0} \\ & & \ddots & \\ -\widehat{\beta}_{(1,N),0} & -\widehat{\beta}_{(2,N),0} & & 1 \end{bmatrix} \tag{5.4.27}$$

Again, if we combine (5.4.26) with (5.4.12), we may write the statistic as

$$\lambda_{LR} = -T\log|\widehat{\mathbf{V}}| + NT\log\widehat{\nu} - T\log|\widehat{\mathbf{B}}_0| = T\ \log\{|\widehat{\nu}\{\widehat{\mathbf{B}}_0\}^{-1}|/|\widehat{\mathbf{V}}|\} \xrightarrow{D} \mathcal{X}^2_{df_2} \tag{5.4.28}$$

as $T \to \infty$ and H_0 holds. For the degrees of freedom it holds that

$$df_2 = Np + (N(N+1))/2 - [Np + N(N-1)/2 + 1] = N - 1.$$

5.5 Forms of prediction and kriging

If the causal multivariate auto-regression defined back in (5.3.8) is the model of interest, then we know that the best linear predictor of $\mathbf{Y}_t$ based on at least p lags from its recent past $\mathbf{Y}_{t-1}, \cdots, \mathbf{Y}_{t-p}$, is equal to

$$\widetilde{\mathbf{Y}}_t \equiv \sum_{i=1}^{p}\boldsymbol{\Phi}_i\ \mathbf{Y}_{t-i}. \tag{5.5.1}$$

Moreover, if it holds that $\{\boldsymbol{\varepsilon}_t\} \sim IID(\mathbf{0}, \mathbf{V})$, then (5.5.1) takes the form of a conditional expectation and becomes the optimal predictor. We always use the words 'best' or 'optimal' for the case of minimum squared error loss. In other words, for any other predictor

$$\mathbf{Y}_t^* = f(\mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \cdots),$$

and any $\boldsymbol{\lambda} \in \mathcal{R}^N$, it holds that

$$\boldsymbol{\lambda}^\tau [E\{(\mathbf{Y}_t - \mathbf{Y}_t^*)(\mathbf{Y}_t - \mathbf{Y}_t^*)^\tau\} - \mathbf{V}]\boldsymbol{\lambda} \geq 0, \tag{5.5.2}$$

where we may write

$$\mathbf{V} = E\{(\mathbf{Y}_t - \sum_{i=1}^{p} \boldsymbol{\Phi}_i \, \mathbf{Y}_{t-i})(\mathbf{Y}_t - \sum_{i=1}^{p} \boldsymbol{\Phi}_i \, \mathbf{Y}_{t-i})^\tau\} = E\{(\mathbf{Y}_t - \widetilde{\mathbf{Y}}_t)(\mathbf{Y}_t - \widetilde{\mathbf{Y}}_t)^\tau\}.$$

Although the form of interdependence expressed in the covariance matrix $\mathbf{V}$ is of a spatial nature, we have not derived our predictors in any different way than for any other time series. The word 'kriging' is synonymous with 'optimal prediction' (Cressie, 1993, p.119) but is mostly used when we are interested in the value of a random variable $Y(\mathbf{s})$ at a fixed location $\mathbf{s}^\tau \in \mathcal{R}^d$, based on available observations $y(\mathbf{s}_k)$, $k = 1, \cdots, N$. When there are linear relationships between the random variables and we use linear predictors, we call this ordinary kriging (Cressie, 1993, p.120).

We assume that we can write

$$\mathbf{Y}_t = \sum_{i=1}^{p} b_i \, \mathbf{Y}_{t-i} + \boldsymbol{\varepsilon}_t, \; \{\boldsymbol{\varepsilon}_t\} \sim WN(\mathbf{0}, \mathbf{V}), \tag{5.5.3}$$

which means that all the N auto-regressions taking place on the different locations use the same p parameters. Then, according to (5.3.16) we can write

$$\text{Var}\{\mathbf{Y}_t\} = [1 + \sum_{i=1}^{\infty} \psi_i^2] \cdot \mathbf{V} \tag{5.5.4}$$

where

$$\psi(z) \equiv 1 + \sum_{i=1}^{\infty} \psi_i \cdot z^i \equiv [1 - \sum_{i=1}^{p} b_i \cdot z^i]^{-1}.$$

For the decomposition

$$\mathbf{V}^{-1} = \boldsymbol{\Lambda}^{-1}\mathbf{B},$$

as it was expressed before, we can also write after (5.5.4)

$$\text{Var}\{\mathbf{Y}_t\}^{-1} = \boldsymbol{\Lambda}^{*-1}\mathbf{B}, \tag{5.5.5}$$

where

$$\mathbf{\Lambda}^* = [1 + \sum_{i=1}^{\infty} \psi_i^2] \cdot \mathbf{\Lambda}. \tag{5.5.6}$$

According to Section 5.2.1, the representation (5.5.6) implies that

$$\widehat{Y}_t(\mathbf{s}_j) \equiv \sum_{\substack{k=1,\\ k\neq j}}^{N} \beta_{j,k}\ Y_t(\mathbf{s}_k) \tag{5.5.7}$$

is the best linear predictor of $Y_t(\mathbf{s}_j)$ at time t and location j based on the observations $Y_t(\mathbf{s}_k),\ k = 1, \cdots, N,\ j \neq k$, at the same time and all other locations. The prediction variance is due to (5.5.6), equal to

$$E\{Y_t(\mathbf{s}_j) - \widehat{Y}_t(\mathbf{s}_j)\}^2 = [1 + \sum_{i=1}^{\infty} \psi_i^2] \cdot \nu_j. \tag{5.5.8}$$

Finally, we may be interested in predicting the value of $Y_t(\mathbf{s}_j)$ from both the observations from its recent past $\mathbf{Y}_{t-1}, \cdots, \mathbf{Y}_{t-p}$, and the observations that have occurred at the same time but on the other locations $\mathbf{s}_k,\ k = 1, \cdots, N,\ k \neq j$. For this, we may take advantage of the fact that

$$\mathrm{Var}\{\boldsymbol{\varepsilon}_t\} = \mathbf{V}. \tag{5.5.9}$$

If we were observing $\{\varepsilon_t(\mathbf{s}_k),\ k = 1, \cdots, N,\ k \neq j\}$, and were interested in predicting the value of $\varepsilon_t(\mathbf{s}_j)$, then straight from (5.2.21) and (5.5.9), we would write

$$\widehat{\varepsilon}_t(\mathbf{s}_j) \equiv \sum_{\substack{k=1,\\ k\neq j}}^{N} \beta_{j,k}\ \varepsilon_t(\mathbf{s}_k) \tag{5.5.10}$$

the best linear predictor and the prediction variance

$$E\{\varepsilon_t(\mathbf{s}_j) - \widehat{\varepsilon}_t(\mathbf{s}_j)\}^2 = \nu_j. \tag{5.5.11}$$

We may transform (5.5.10) into

$$Y_t^*(\mathbf{s}_j) - \sum_{i=1}^{p} b_{i,j}\ Y_{t-i}(\mathbf{s}_j) \equiv \sum_{\substack{k=1,\\ k\neq j}}^{N} \beta_{j,k}\ [Y_t(\mathbf{s}_k) - \sum_{i=1}^{p} b_{i,k}\ Y_{t-i}(\mathbf{s}_k)] \tag{5.5.12}$$

or

$$Y_t^*(\mathbf{s}_j) \equiv \sum_{i=1}^{p} b_{i,j}\ Y_{t-i}(\mathbf{s}_j) + \sum_{\substack{k=1,\\ k\neq j}}^{N} \beta_{j,k}\ [Y_t(\mathbf{s}_k) - \sum_{i=1}^{p} b_{i,k}\ Y_{t-i}(\mathbf{s}_k)], \tag{5.5.13}$$

which is the best linear predictor. Indeed, it holds that

$$Y_t(\mathbf{s}_j) - Y_t^*(\mathbf{s}_j) = \varepsilon_t(\mathbf{s}_j) - \sum_{\substack{k=1,\\ k\neq j}}^{N} \beta_{j,k}\ \varepsilon_t(\mathbf{s}_k), \tag{5.5.14}$$

which is uncorrelated with $Y_{t-i}(\mathbf{s}_k), i > 0$, because of the causality of the auto-regressions, and it is also uncorrelated with $\varepsilon_t(\mathbf{s}_k),\ k = 1, \cdots, N,\ k \neq j$, because of (5.5.10). As a result, it is uncorrelated with any $Y_t(\mathbf{s}_k),\ k = 1, \cdots, N,\ k \neq j$, since

$$Y_t(\mathbf{s}_k) = \varepsilon_t(\mathbf{s}_k) + \sum_{i=1}^{\infty} \psi_{i,k}\ \varepsilon_{t-i}(\mathbf{s}_k).$$

As for the prediction variance, it holds due to (5.5.14), (5.5.10) and (5.5.11) that

$$E\{Y_t(\mathbf{s}_j) - Y_t^*(\mathbf{s}_j)\}^2 = \nu_j. \tag{5.5.15}$$

5.6 Mink and Muskrat spatio-temporal data

Following the example of Zhang, Yao, Tong and Stenseth (2003), in this section we try to model the food-chain interaction between mink and muskrat in Canada. We have available the annual numbers of mink and muskrat fur sales on 82 different locations and for 25 consecutive years, *i.e.* from 1925 to 1949. With our model and using the methods described in this chapter, we are interested in showing that there is indeed a food-chain interaction between the mink and muskrat as predator and prey, respectively. For a more detailed statistical analysis, which takes into account the special nature of the data, we refer again to the work of Zhang, Yao, Tong and Stenseth (2003).

We write $N = 82$ for the number of posts available in the sample and $T = 25 - 1 = 24$ years. We also write $Y_t(\mathbf{s}_j)$ and $X_t(\mathbf{s}_j)$ for the mink observation and the muskrat observation, respectively, on a natural logarithmic scale for any location $j = 1, \cdots, N$, and any year $t = 0, 1, \cdots, T$, in the dataset. Then, for each fixed post $j = 1, \cdots, N$, we find the sample means

$$\bar{Y}_j \equiv \frac{1}{T+1} \sum_{t=0}^{T} Y_t(\mathbf{s}_j), \tag{5.6.1}$$

$$\bar{X}_j \equiv \frac{1}{T+1} \sum_{t=0}^{T} X_t(\mathbf{s}_j), \tag{5.6.2}$$

and the sample variances

$$\widehat{\sigma}^2_{Y,j} \equiv \frac{1}{T} [\sum_{t=0}^{T} Y_t^2(\mathbf{s}_j) - (T+1) \cdot \bar{Y}_j^2], \tag{5.6.3}$$

$$\widehat{\sigma}^2_{X,j} \equiv \frac{1}{T} [\sum_{t=0}^{T} X_t^2(\mathbf{s}_j) - (T+1) \cdot \bar{X}_j^2]. \tag{5.6.4}$$

We define the standardized values

$$Y_t^{(c)}(\mathbf{s}_j) \equiv \frac{Y_t(\mathbf{s}_j) - \bar{Y}_j}{\widehat{\sigma}_{Y,j}}, \tag{5.6.5}$$

$$X_t^{(c)}(\mathbf{s}_j) \equiv \frac{X_t(\mathbf{s}_j) - \bar{X}_j}{\widehat{\sigma}_{X,j}}, \tag{5.6.6}$$

for all $j = 1, \cdots, N$, and $t = 0, \cdots, T$.

Next, we will need to assume that a time series is taking place over $t \in \mathcal{Z}$. Since we are dealing with two time series $\{Y_t,\ t \in \mathcal{Z}\}$ and $\{X_t,\ t \in \mathcal{Z}\}$, we need to decide which one should play the role of dependent and which of independent set of variables. The interaction between the two sets, though, does not have an obvious direction; the mink counts on the presence of the muskrat to survive and, so does the muskrat count on the absence of the mink to survive. In order to avoid the inevitable cause-and-effect formulation implied by a univariate time series, we preferred to assume that the following bivariate time series is taking place instead. We write for every $j = 1, \cdots, N$, the causal first-order auto-regression

$$\begin{bmatrix} Y_t^{(c)}(\mathbf{s}_j) \\ X_t^{(c)}(\mathbf{s}_j) \end{bmatrix} = \begin{bmatrix} b_j & c_j \\ -c_j & b_j \end{bmatrix} \cdot \begin{bmatrix} Y_{t-1}^{(c)}(\mathbf{s}_j) \\ X_{t-1}^{(c)}(\mathbf{s}_j) \end{bmatrix} + \begin{bmatrix} \varepsilon_t^{(Y)}(\mathbf{s}_j) \\ \varepsilon_t^{(X)}(\mathbf{s}_j) \end{bmatrix},\ t \in \mathcal{Z}, \tag{5.6.7}$$

where

$$\boldsymbol{\varepsilon}_t^{(Y)} \equiv [\varepsilon_t^{(Y)}(\mathbf{s}_1), \cdots, \varepsilon_t^{(Y)}(\mathbf{s}_N)]^\tau \sim WN(\mathbf{0}, \mathbf{V}_Y), \tag{5.6.8}$$

$$\boldsymbol{\varepsilon}_t^{(X)} \equiv [\varepsilon_t^{(X)}(\mathbf{s}_1), \cdots, \varepsilon_t^{(X)}(\mathbf{s}_N)]^\tau \sim WN(\mathbf{0}, \mathbf{V}_X), \tag{5.6.9}$$

and

$$\boldsymbol{\varepsilon}_t = \begin{bmatrix} \boldsymbol{\varepsilon}_t^{(Y)} \\ \boldsymbol{\varepsilon}_t^{(X)} \end{bmatrix} \sim WN(\mathbf{0}, \mathbf{V}), \tag{5.6.10}$$

for any $t \in \mathcal{Z}$.

In (5.6.7) we have used the matrix

$$\begin{bmatrix} b_j & c_j \\ -c_j & b_j \end{bmatrix},\ j = 1, \cdots, N,$$

instead of a matrix that uses four different elements. This is in order to reduce the number of parameters of the model. We take advantage of the fact that both series have been standardized. For the two elements $\pm c_j$ that have different signs, we expect that the more the muskrats of the region the more the minks, while the fewer the minks of the area the more the muskrats there.

We keep aiming at reducing the number of parameters in (5.6.7). Thus, we assume that

$$\mathbf{V} \equiv \begin{bmatrix} \mathbf{V}_Y & \mathbf{O}_{N\times N} \\ \mathbf{O}_{N\times N} & \mathbf{V}_X \end{bmatrix} \tag{5.6.11}$$

and any interaction between mink and muskrat is now clearly expressed only via the parameters $c_j,\ j = 1, \cdots, N$.

Finally, we put the N posts in the same three categories, namely 'west' (N_{west} = 29 posts), 'east' (N_{east} = 9 posts) and 'centre' (N_{centre} = 44 posts), as it has been explained in the (2003) paper, and we proceed further with a parameter reduction. The N parameters $b_j,\ j = 1, \cdots, N$, are reduced to only three parameters $b_{west},\ b_{east},\ b_{centre}$ and so are $c_j,\ j = 1, \cdots, N$, replaced by $c_{west},\ c_{east}$ and c_{centre}. We do the same for the inverse variance matrices $\mathbf{V}_Y^{-1}$ and $\mathbf{V}_X^{-1}$. We set the matrix $\mathbf{V}_Y^{-1}$ equal to

$$\begin{bmatrix} a^{(Y)}_{west} \cdot \mathbf{1}_{N_{west}\times N_{west}} & -a^{(Y)}_{west,east} \cdot \mathbf{1}_{N_{west}\times N_{east}} & -a^{(Y)}_{west,centre} \cdot \mathbf{1}_{N_{west}\times N_{centre}} \\ -a^{(Y)}_{west,east} \cdot \mathbf{1}_{N_{east}\times N_{west}} & a^{(Y)}_{east} \cdot \mathbf{1}_{N_{east}\times N_{east}} & -a^{(Y)}_{east,centre} \cdot \mathbf{1}_{N_{east}\times N_{centre}} \\ -a^{(Y)}_{west,centre} \cdot \mathbf{1}_{N_{centre}\times N_{west}} & -a^{(Y)}_{centre,east} \cdot \mathbf{1}_{N_{centre}\times N_{east}} & a^{(Y)}_{centre} \cdot \mathbf{1}_{N_{centre}\times N_{centre}} \end{bmatrix}, \tag{5.6.12}$$

where $\mathbf{1}_{n\times m}$ is the matrix with all $(n \cdot m)$ elements equal to 1. Similarly, we define $\mathbf{V}_X^{-1}$ with variance-covariance parameters $a^{(X)}_{west}$, $a^{(X)}_{east}$, $a^{(X)}_{centre}$, and covariance parameters $a^{(X)}_{west,east}$, $a^{(X)}_{west,centre}$ and $a^{(X)}_{east,centre}$. Overall, our model is using 18 parameters. On the other hand, the likelihood is using $2 \cdot (NT) = 3,936$ observations. We write

$$\begin{aligned} -2\,l \;\equiv\; & \sum_{t=1}^{T} [a^{(Y)}_{west} \sum_{j,k} \varepsilon_t^{(Y)}(\mathbf{s}_j)\varepsilon_t^{(Y)}(\mathbf{s}_k) + a^{(Y)}_{east} \sum_{j,k} \varepsilon_t^{(Y)}(\mathbf{s}_j)\varepsilon_t^{(Y)}(\mathbf{s}_k) \\ & + a^{(Y)}_{centre} \sum_{j,k} \varepsilon_t^{(Y)}(\mathbf{s}_j)\varepsilon_t^{(Y)}(\mathbf{s}_k) \\ & - a^{(Y)}_{west,east} \sum_{\substack{j,k \\ j\neq k}} \varepsilon_t^{(Y)}(\mathbf{s}_j)\varepsilon_t^{(Y)}(\mathbf{s}_k) - a^{(Y)}_{west,centre} \sum_{\substack{j,k \\ j\neq k}} \varepsilon_t^{(Y)}(\mathbf{s}_j)\varepsilon_t^{(Y)}(\mathbf{s}_k) \\ & - a^{(Y)}_{east,centre} \sum_{\substack{j,k \\ j\neq k}} \varepsilon_t^{(Y)}(\mathbf{s}_j)\varepsilon_t^{(Y)}(\mathbf{s}_k) + a^{(X)}_{west} \sum_{j,k} \varepsilon_t^{(X)}(\mathbf{s}_j)\varepsilon_t^{(X)}(\mathbf{s}_k) \\ & + a^{(X)}_{east} \sum_{j,k} \varepsilon_t^{(X)}(\mathbf{s}_j)\varepsilon_t^{(X)}(\mathbf{s}_k) \\ & + a^{(X)}_{centre} \sum_{j,k} \varepsilon_t^{(X)}(\mathbf{s}_j)\varepsilon_t^{(X)}(\mathbf{s}_k) - a^{(X)}_{west,east} \sum_{\substack{j,k \\ j\neq k}} \varepsilon_t^{(X)}(\mathbf{s}_j)\varepsilon_t^{(X)}(\mathbf{s}_k) \\ & - a^{(X)}_{west,centre} \sum_{\substack{j,k \\ j\neq k}} \varepsilon_t^{(X)}(\mathbf{s}_j)\varepsilon_t^{(X)}(\mathbf{s}_k) - a^{(X)}_{east,centre} \sum_{\substack{j,k \\ j\neq k}} \varepsilon_t^{(X)}(\mathbf{s}_j)\varepsilon_t^{(X)}(\mathbf{s}_k)] \\ & - T\,[\log|\mathbf{V}_Y^{-1}| + \log|\mathbf{V}_X^{-1}|], \end{aligned} \tag{5.6.13}$$

for the quantity we will minimize, where we consider $\varepsilon_t^{(Y)}(\mathbf{s}_j), \varepsilon_t^{(X)}(\mathbf{s}_j)$, $j = 1, \cdots, N$, to be functions of the six parameters $b_{west}, \cdots, c_{centre}$, and the data.

For the computation of $|\mathbf{V}_Y^{-1}|$ (or $|\mathbf{V}_X^{-1}|$), we keep in mind that

$$\text{Var}\{\mathbf{V}_Y^{-1} \cdot \boldsymbol{\varepsilon}_t^{(Y)}\} = \mathbf{V}_Y^{-1}, \tag{5.6.14}$$

which is also a variance matrix. Thus, we can place the six elements $a_{west}^{(Y)}, \cdots, a_{east,centre}^{(Y)}$, in the inverse, and perform the innovations algorithm to find the prediction variances, for their product is equal to the determinant of interest. The next subsection examines which are these values of the parameters that we are allowed to try, in order to compute (5.6.13).

5.6.1 Restrictions on the parameters

Restrictions on the temporal parameters

First, we find the restrictions for the parameters $b_{west}, \cdots, c_{centre}$. For any $j = 1, \cdots, N$, the auto-regression defined by (5.6.7) is causal if

$$\begin{aligned} &\det\left\{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} b_j & c_j \\ -c_j & b_j \end{bmatrix} z\right\} = \det\left\{\begin{bmatrix} 1 - b_j\, z & -c_j\, z \\ c_j\, z & 1 - b_j\, z \end{bmatrix}\right\} \\ = \;& (1 - b_j\, z)^2 + c_j^2\, z^2 = 1 - 2\, b_j\, z + (b_j^2 + c_j^2)\, z^2 \neq 0, \end{aligned} \tag{5.6.15}$$

for all $|z| \leq 1$, where z is a complex number.

Indeed, we may define

$$\Delta \equiv 4\, b_j^2 - 4\, (b_j^2 + c_j^2) = -4\, c_j^2, \tag{5.6.16}$$

and find the roots

$$z = \frac{2\, b_j \pm 2\, c_j \cdot i}{2\, (b_j^2 + c_j^2)} = \frac{b_j}{b_j^2 + c_j^2} \pm i \cdot \frac{c_j}{b_j^2 + c_j^2}, \tag{5.6.17}$$

where $i = \sqrt{-1}$. Thus, we want that

$$|z| = \sqrt{\frac{b_j^2 + c_j^2}{(b_j^2 + c_j^2)^2}} = \frac{1}{\sqrt{b_j^2 + c_j^2}} > 1,$$

or that

$$b_j^2 + c_j^2 < 1, \tag{5.6.18}$$

for the three areas 'west', 'east' and 'centre'. As a result, we computed the quadratic form of the likelihood for all the values

$$b_j,\ c_j = 0, \pm 0.1, \cdots, \pm 0.9,$$

under the restriction (5.6.18).

Restrictions on the spatial parameters

We discuss the case of $\mathbf{V}_Y^{-1}$, as we may work out the case of $\mathbf{V}_X^{-1}$ identically. According to the decomposition

$$\mathbf{V}_Y^{-1} = \mathbf{\Lambda}_Y^{-1} \cdot \mathbf{B}_Y, \tag{5.6.19}$$

as this was described before, we may count that the elements

$$\nu_j^{(Y)} = 1/a_{j,j}^{(Y)},\ j = 1, \cdots, N,$$

are such that

$$\nu_j^{(Y)} \leq \mathrm{Var}\{\varepsilon_t^{(Y)}(\mathbf{s}_j)\}. \tag{5.6.20}$$

In the case of Gaussian random variables, ν are conditional variances; otherwise they are just prediction variances. We follow (5.6.20) and write

$$\begin{aligned}
\nu_j^{(Y)} &\leq \mathrm{Var}\{Y_t^{(c)}(\mathbf{s}_j) - b_j\ Y_{t-1}^{(c)}(\mathbf{s}_j) - c_j\ X_{t-1}^{(c)}(\mathbf{s}_j)\} \\
&= (1 + b_j^2)\ \mathrm{Var}\{Y_t^{(c)}(\mathbf{s}_j)\} + c_j^2\ \mathrm{Var}\{X_t^{(c)}(\mathbf{s}_j)\} - 2\ b_j\ \mathrm{Corr}\{Y_t^{(c)}(\mathbf{s}_j), Y_{t-1}^{(c)}(\mathbf{s}_j)\} \\
&- 2\ c_j\ \mathrm{Corr}\{Y_t^{(c)}(\mathbf{s}_j), X_{t-1}^{(c)}(\mathbf{s}_j)\} + 2\ b_j \cdot c_j\ \mathrm{Corr}\{Y_{t-1}^{(c)}(\mathbf{s}_j), X_{t-1}^{(c)}(\mathbf{s}_j)\} \\
&\leq 1 + (b_j^2 + c_j^2) + 2\ |b_j| + 2\ |c_j| + 2\ |b_j \cdot c_j|,
\end{aligned}$$

since

$$\mathrm{Var}\{Y_t^{(c)}(\mathbf{s}_j)\} = \mathrm{Var}\{X_t^{(c)}(\mathbf{s}_j)\} = 1.$$

Finally, since (5.6.18) holds and $|b_j|,\ |c_j| < 1$, we may write that

$$\nu_j^{(Y)} \leq 1 + 1 + 2 + 2 + 2 = 8. \tag{5.6.21}$$

We tried the values

$$\nu_j^{(Y)} = 1, 2, \cdots, 8,$$

for the three areas 'west', 'east' and 'centre'. Since it holds that

$$|a_{j,k}^{(Y)}| \leq \frac{1}{\{\nu_j^{(Y)} \cdot \nu_k^{(Y)}\}^{1/2}}, \tag{5.6.22}$$

we tried the values

$$a_{j,k}^{(Y)} = 0, \pm 0.1, \cdots, \pm 1,$$

for the three covariance parameters 'west-east', 'west-centre' and 'east-centre'.

5.6.2 Testing the interaction of mink and muskrat

If we consider the indexes $j = 1, \cdots, 29$, corresponding to the locations of the western sites, and also the indexes $j = 30, \cdots, 38$, and $j = 39, \cdots, 82$, corresponding to the locations of eastern and central locations, respectively, we may then summarize our full model, for any $t \in \mathbb{Z}$, by the equations

$$\begin{bmatrix} Y_t^{(c)}(\mathbf{s}_j) \\ X_t^{(c)}(\mathbf{s}_j) \end{bmatrix} = \begin{bmatrix} b_{west} & c_{west} \\ -c_{west} & b_{west} \end{bmatrix} \cdot \begin{bmatrix} Y_{t-1}^{(c)}(\mathbf{s}_j) \\ X_{t-1}^{(c)}(\mathbf{s}_j) \end{bmatrix} + \begin{bmatrix} \varepsilon_t^{(Y)}(\mathbf{s}_j) \\ \varepsilon_t^{(X)}(\mathbf{s}_j) \end{bmatrix}, \ j = 1, \cdots, 29,$$

$$\begin{bmatrix} Y_t^{(c)}(\mathbf{s}_j) \\ X_t^{(c)}(\mathbf{s}_j) \end{bmatrix} = \begin{bmatrix} b_{east} & c_{east} \\ -c_{east} & b_{east} \end{bmatrix} \cdot \begin{bmatrix} Y_{t-1}^{(c)}(\mathbf{s}_j) \\ X_{t-1}^{(c)}(\mathbf{s}_j) \end{bmatrix} + \begin{bmatrix} \varepsilon_t^{(Y)}(\mathbf{s}_j) \\ \varepsilon_t^{(X)}(\mathbf{s}_j) \end{bmatrix}, \ j = 30, \cdots, 38,$$

$$\begin{bmatrix} Y_t^{(c)}(\mathbf{s}_j) \\ X_t^{(c)}(\mathbf{s}_j) \end{bmatrix} = \begin{bmatrix} b_{centre} & c_{centre} \\ -c_{centre} & b_{centre} \end{bmatrix} \cdot \begin{bmatrix} Y_{t-1}^{(c)}(\mathbf{s}_j) \\ X_{t-1}^{(c)}(\mathbf{s}_j) \end{bmatrix} + \begin{bmatrix} \varepsilon_t^{(Y)}(\mathbf{s}_j) \\ \varepsilon_t^{(X)}(\mathbf{s}_j) \end{bmatrix}, \ j = 39, \cdots, 82,$$

under the restrictions

$$b_{west}^2 + c_{west}^2, \ b_{east}^2 + c_{east}^2, \ b_{centre}^2 + c_{centre}^2 < 1.$$

Moreover, we write

$$\begin{aligned} \boldsymbol{\varepsilon}_t^{(Y)} &\equiv [\varepsilon_t^{(Y)}(\mathbf{s}_1), \cdots, \varepsilon_t^{(Y)}(\mathbf{s}_{82})]^\tau \sim IID(\mathbf{0}, \mathbf{V}_Y), \\ \boldsymbol{\varepsilon}_t^{(X)} &\equiv [\varepsilon_t^{(X)}(\mathbf{s}_1), \cdots, \varepsilon_t^{(X)}(\mathbf{s}_{82})]^\tau \sim IID(\mathbf{0}, \mathbf{V}_X), \end{aligned}$$

and

$$\boldsymbol{\varepsilon}_t = \begin{bmatrix} \boldsymbol{\varepsilon}_t^{(Y)} \\ \boldsymbol{\varepsilon}_t^{(X)} \end{bmatrix} \sim IID(\mathbf{0}, \mathbf{V}),$$

for any $t \in \mathbb{Z}$, with

$$\mathbf{V} \equiv \begin{bmatrix} \mathbf{V}_Y & \mathbf{O}_{82\times 82} \\ \mathbf{O}_{82\times 82} & \mathbf{V}_X \end{bmatrix}.$$

Both the matrices $\mathbf{V}_X^{-1}$ and $\mathbf{V}_Y^{-1}$ involve six parameters, as it was described in (5.6.12), and they are restricted to be covariance matrices.

We have tested

$$H_0 : c_{west} = c_{east} = c_{centre} = 0$$

$$H_1 : \text{otherwise}$$

and have found the minimum under the null hypothesis equal to

$$-2\ l_0 = (24) * (321.037), \tag{5.6.23}$$

and the minimum under the null or alternative hypothesis equal to

$$-2\ l = (24) * (302.096). \tag{5.6.24}$$

Thus, we may verify that their difference is a very extreme point of the chi-square distribution with 3 degrees of freedom and that there is a food-chain interaction between the mink and muskrat in Canada.

Chapter 6

Exact Gaussian likelihoods for observations from spatial quarter ARMA models

6.1 Introduction

Generalizing the theory of the one-dimensional time series to include the two-dimensional spatial processes is often a very difficult task. We saw back in Chapter 3 the obstacles one has to overcome when attempting to prove the properties of the maximum Gaussian likelihood estimators for the parameters of an ARMA model because of the edge-effect. While the estimators for the parameters of the one-dimensional ARMA obtain all desirable properties for large sample sizes, one cannot say the same for the case of the spatial parameters. Up to these days, it looks like a large number of locations available in the sample cannot guarantee that the absolute bias of estimators moves fast enough to zero. In order to prove the asymptotic unbiasedness and normality of the estimators, corrected versions of the likelihood have been maximized rather than the likelihood in its genuine form.

In this chapter, we first find and write down the theoretical form of the exact Gaussian likelihood function for a specific class of two-dimensional auto-regressions, which are both simple and flexible enough to provide a reasonable representation of planar auto-correlation, *i.e.* auto-correlation on $\mathcal{Z}^2$. The main advantage of these processes is that they can easily factorize their deterministic parts, such as the auto-correlation function

and the spectral density, into two distinct parts arising from one-dimensional processes. Special attention to these processes has been paid by Martin (1979). The doubly geometric process (Martin, 1979, p.211) may be seen as an example. As it has been very well described by Martin (1979, p.212), under certain conditions, such processes can be considered as auto-linear formulations and they can be very useful for spatial prediction. The way they are defined originally, does not make this very clear and it can be misleading.

Next, we simulate observations from a two-dimensional moving-average process, which uses the first-order filter with two spatial parameters, in order to verify that the exact Gaussian likelihood estimators have a bias of order greater than the one desired. The observations lie on a rectangle and we maximize the exact likelihood using the innovations algorithm. We compare our results with simulated results from some corrected estimators, for which we may rest that they have the desired properties; we showed back in Chapter 3 that for a moving-average, one can cut a finite number of observations from each dimension; in Section 6.3.2 we also refer to the modification of Yao and Brockwell (2006), who suggested a more general correction to include all the two-dimensional processes that possess an AR(∞) representation via the innovations algorithm. Questions to be answered are which estimators dominate in terms of the mean square error, whether the selection of observations to reduce the absolute bias compensates for the inflation of the variance and, whether the bias moves more slowly to zero indeed when no selection takes place.

6.2 Linear-by-linear auto-regressions

We give the definition of a linear-by-linear process as given by Martin (1979, p.210).

Definition 6.1. A linear-by-linear process $\{X(u,v),\ u,v \in \mathcal{Z}\}$ is defined as a (weakly) stationary planar lattice process for which the spectral density function exists and is proportional to the product of two spectral density functions corresponding to two one-dimensional processes.

We can now define a linear-by-linear and causal auto-regression $\{X(u,v),\ u,v \in \mathcal{Z}\}$ satisfying the following equation

$$\theta_1(B_1) \cdot \theta_2(B_2)\ X(u,v) \equiv e^{(1)}(u,v),\ \{e^{(1)}(u,v)\} \sim WN(0,1). \qquad (6.2.1)$$

We summarize the auto-regressive polynomial as

$$\theta(z_1, z_2) \equiv \theta_1(z_1) \cdot \theta_2(z_2) \equiv \sum_{[k,l] \in \mathcal{J}_Q} \theta_{k,l}\ z_1^k \cdot z_2^l, \tag{6.2.2}$$

where

$$\theta_1(z) \equiv 1 + \sum_{k=1}^{q_1} \theta_k^{(1)}\ z^k, \tag{6.2.3}$$

$$\theta_2(z) \equiv 1 + \sum_{l=1}^{q_2} \theta_l^{(2)}\ z^l. \tag{6.2.4}$$

Thus, the auto-regressions defined by (6.2.1) are a special case of the causal auto-regressions defined on $\mathcal{Z}^2$. For the index set

$$\mathcal{J}_Q \equiv \{[k,l] : k = 0, \cdots, q_1,\ l = 0, \cdots, q_2\} \tag{6.2.5}$$

it holds that

$$\theta_{k,l} = \theta_k^{(1)} \cdot \theta_l^{(2)},\ [k,l] \in \mathcal{J}_Q, \tag{6.2.6}$$

where

$$\theta_0^{(1)} = \theta_0^{(2)} \equiv 1. \tag{6.2.7}$$

The spectral density of $\{X(u,v),\ u,v \in \mathcal{Z}\}$ as defined in (6.2.1) can be summarized by

$$g_X(\omega_1, \omega_2) = \frac{1}{(2\pi)^2} \cdot \frac{1}{\theta_1(e^{i\omega_1}) \cdot \theta_2(e^{i\omega_2}) \cdot \theta_1(e^{-i\omega_1}) \cdot \theta_2(e^{-i\omega_2})},\ -\pi \le \omega_1,\ \omega_2 \le \pi, \tag{6.2.8}$$

which verifies that we have defined a linear-by-linear process. Also, it holds that

$$\theta_1(z)^{-1} \equiv 1 + \sum_{k=1}^{\infty} \Theta_k^{(1)}\ z^k,\ \sum_{k=1}^{\infty} |\Theta_k^{(1)}| < \infty, \tag{6.2.9}$$

$$\theta_2(z)^{-1} \equiv 1 + \sum_{l=1}^{\infty} \Theta_l^{(2)}\ z^l,\ \sum_{l=1}^{\infty} |\Theta_l^{(2)}| < \infty, \tag{6.2.10}$$

since we want the auto-regression to be causal. We may write

$$\theta(z_1, z_2)^{-1} \equiv \sum_{k,l=0}^{\infty} \Theta_{k,l}\ z_1^k \cdot z_2^l = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \Theta_k^{(1)}\ \Theta_l^{(2)}\ z_1^k \cdot z_2^l, \tag{6.2.11}$$

where

$$\Theta_{0,0} = \Theta_0^{(1)} = \Theta_0^{(2)} \equiv 1. \tag{6.2.12}$$

Thanks to (2.4.24), it holds that

$$\Theta_k^{(1)} + \theta_1^{(1)}\Theta_{k-1}^{(1)} + \cdots + \theta_{q_1}^{(1)}\Theta_{k-q_1}^{(1)} = 0,\ k > 0, \tag{6.2.13}$$

$$\Theta_l^{(2)} + \theta_1^{(2)}\Theta_{l-1}^{(2)} + \cdots + \theta_{q_2}^{(2)}\Theta_{l-q_2}^{(2)} = 0,\ l > 0. \tag{6.2.14}$$

There are also three more sequences of uncorrelated random variables, such that we can write the auto-regression as their unilateral representation. It holds that

$$\theta_1(B_1)\cdot\theta_2(B_2^{-1})\ X(u,v)\equiv e^{(2)}(u,v),\ \{e^{(2)}(u,v)\}\sim WN(0,1), \tag{6.2.15}$$

is a representation that expresses the auto-regression as a function of $e^{(2)}(u-k,v+l)$, $k,l\geq 0$ only, and the unilateral counterparts of (6.2.1) and (6.2.15) are

$$\theta_1(B_1^{-1})\cdot\theta_2(B_2^{-1})\ X(u,v) \equiv u^{(1)}(u,v),\ \{u^{(1)}(u,v)\}\sim WN(0,1), \tag{6.2.16}$$

$$\theta_1(B_1^{-1})\cdot\theta_2(B_2)\ X(u,v) \equiv u^{(2)}(u,v),\ \{u^{(2)}(u,v)\}\sim WN(0,1), \tag{6.2.17}$$

respectively.

Looking back at the set $\mathcal{J}_Q$, all the representations (6.2.1), (6.2.15), (6.2.16) and (6.2.17) express $\{X(u,v),\ u,v\in\mathcal{Z}\}$ as a quarter process, *i.e.* possessing an MA(∞) representation from one of the four quarters only. In general, a unilateral representation on two dimensions involves two quarters.

Theorem 6.1. From the process defined by (6.2.1) and for positive integers N_1, N_2, we let the random vector

$$\mathbf{X}\equiv\begin{bmatrix} X(1,1), & X(1,2), & \cdots & X(1,N_2),\\ X(2,1), & X(2,2), & \cdots & X(2,N_2),\\ \vdots & & & \vdots\\ X(N_1,1), & X(N_1,2), & \cdots & X(N_1,N_2)\end{bmatrix}^{\tau}. \tag{6.2.18}$$

If $N_k > 2q_k,\ k=1,2$, then it holds that

$$\mathrm{Var}\{\mathbf{X}\}^{-1}=\boldsymbol{\Sigma}_1^{-1}\otimes\boldsymbol{\Sigma}_2^{-1}, \tag{6.2.19}$$

where we set the matrix $\boldsymbol{\Sigma}_k^{-1}$, to be equal to

$$\begin{bmatrix}
1 & \theta_1^{(k)} & \cdots & \theta_{q_k}^{(k)} & 0 & & \cdots & & 0\\
\theta_1^{(k)} & 1+\theta_1^{(k)2} & \cdots & (\theta_{q_k-1}^{(k)}+\theta_1^{(k)}\theta_{q_k}^{(k)}) & \theta_{q_k}^{(k)} & & & & 0\\
\vdots & & & \vdots & & & & & \\
\theta_{q_k}^{(k)} & & & 1+\cdots+\theta_{q_k}^{(k)2} & & & & & \\
 & & & & & \ddots & & & \\
 & & & & & & 1+\cdots+\theta_{q_k}^{(k)2} & & \theta_{q_k}^{(k)}\\
 & & & & & & \vdots & & \vdots\\
0 & & & & \theta_{q_k}^{(k)} & (\theta_{q_k-1}^{(k)}+\theta_1^{(k)}\theta_{q_k}^{(k)}) & \cdots & 1+\theta_1^{(k)2} & \theta_1^{(k)}\\
0 & & \cdots & & 0 & \theta_{q_k}^{(k)} & \cdots & \theta_1^{(k)} & 1
\end{bmatrix}$$

for $k = 1, 2$.

Proof. We let the set

$$\mathcal{S} \equiv \{[u, v] : u = 1, \cdots, N_1,\ v = 1, \cdots, N_2\}. \tag{6.2.20}$$

For the first location $[1, 1]$, it holds that

$$\theta_1(B_1^{-1}) \cdot \theta_2(B_2^{-1})\ X(1,1) = u^{(1)}(1,1) \tag{6.2.21}$$

or

$$X(1,1)+\sum_{k=1}^{q_1} \theta_k^{(1)}\ X(1+k,1)+\sum_{l=1}^{q_2} \theta_l^{(2)}\ X(1,1+l)+\sum_{k=1}^{q_1}\sum_{l=1}^{q_2} \theta_k^{(1)}\cdot\theta_l^{(2)}\ X(1+k,1+l) = u^{(1)}(1,1). \tag{6.2.22}$$

It holds because of the causality assumption that

$$E\{X(u,v) \cdot u^{(1)}(1,1)\} = \begin{cases} 1, & [u,v] = [1,1] \\ 0, & [u,v] \in \mathcal{S} - \{[1,1]\} \end{cases} \tag{6.2.23}$$

and according to (6.2.22), it holds that $u^{(1)}(1,1)$ can be written as a linear combination of $X(u,v)$, $[u,v] \in \mathcal{S}$ since $N_1 > 2q_1$ and $N_2 > 2q_2$.

For convenience, we write for any $[u,v]$, $[u+k, v+l] \in \mathcal{S}$

$$\sigma_{[u,v],[k,l]} = \sigma_{[u+k,v+l],[-k,-l]} \tag{6.2.24}$$

for the element of the matrix $\mathrm{Var}\{\mathbf{X}\}^{-1}$ that is in the same row and column as the covariance of the random variables $X(u,v)$ and $X(u+k, v+l)$, in $\mathrm{Var}\{\mathbf{X}\}$. It holds that

$$\sigma_{[1,1],[k,l]} = \begin{cases} 1, & [k,l] = [0,0] \\ \theta_k^{(1)}, & [k,l] = [k,0],\ k = 1, \cdots, q_1 \\ \theta_l^{(2)}, & [k,l] = [0,l],\ l = 1, \cdots, q_2 \\ \theta_k^{(1)}\theta_l^{(2)}, & k = 1, \cdots, q_1,\ l = 1, \cdots, q_2 \\ 0, & k = q_1+1, \cdots, N_1, \text{ or } l = q_2+1, \cdots, N_2 \end{cases} \tag{6.2.25}$$

Similarly, for the second location $[1, 2]$, it holds that

$$\theta_1(B_1^{-1}) \cdot \theta_2(B_2^{-1})\ [X(1,2) + \theta_1^{(2)} X(1,1)] = u^{(1)}(1,2) + \theta_1^{(2)} u^{(1)}(1,1) \tag{6.2.26}$$

and

$$E\{X(u,v) \cdot [u^{(1)}(1,2) + \theta_1^{(2)} u^{(1)}(1,1)]\} = \begin{cases} 0, & [u,v] = [1,1] \\ 1, & [u,v] = [1,2] \\ 0, & [u,v] \in \mathcal{S} - \{[1,1],[1,2]\} \end{cases}, \tag{6.2.27}$$

where the first part of (6.2.27) holds since

$$X(1,1) = u^{(1)}(1,1) + \Theta_1^{(2)} u^{(1)}(1,2) + \sum_{l=2}^{\infty} \Theta_l^{(2)} u^{(1)}(1,1+l) + \sum_{k=1}^{\infty} \Theta_k^{(1)} u^{(1)}(1+k,1)$$
$$+ \sum_{k=1}^{\infty}\sum_{l=1}^{\infty} \Theta_k^{(1)}\Theta_l^{(2)} u^{(1)}(1+k,1+l) \qquad (6.2.28)$$

and

$$E\{X(1,1)\cdot[u^{(1)}(1,2) + \theta_1^{(2)} u^{(1)}(1,1)]\} = \Theta_1^{(2)} + \theta_1^{(2)} = 0, \qquad (6.2.29)$$

according to (6.2.14). Again, in the left hand side of (6.2.26), we have expressed the random quantity $u^{(1)}(1,2) + \theta_1^{(2)} u^{(1)}(1,1)$ as a linear function of $X(u,v)$, $[u,v] \in \mathcal{S}$. As a result, thanks to (6.2.27), we may consider as $\sigma_{[1,2],[k,l]} = \sigma_{[1+k,2+l],[-k,-l]}$, with $[1+k,2+l] \in \mathcal{S}$, the coefficient of $X(1+k,2+l)$ in this linear representation (6.2.26).

We may keep going the same way by writing

$$\theta_1(B_1^{-1})\cdot\theta_2(B_2^{-1})\,[X(1,3) + \theta_1^{(2)} X(1,2) + \theta_2^{(2)} X(1,1)]$$
$$= u^{(1)}(1,3) + \theta_1^{(2)} u^{(1)}(1,2) + \theta_2^{(2)} u^{(1)}(1,1)$$
$$\equiv \sum_{[1+k,3+l]\in\mathcal{S}} \sigma_{[1,3],[k,l]}\, X(1+k,3+l) \qquad (6.2.30)$$

for the location $[1,3]$, and for the locations $[1,v]$, $v = 1,\cdots,q_2$, in general, we can write

$$\theta_1(B_1^{-1})\cdot\theta_2(B_2^{-1})\,[X(1,v) + \sum_{l=1}^{v-1} \theta_l^{(2)} X(1,v-l)]$$
$$= u^{(1)}(1,v) + \sum_{l=1}^{v-1} \theta_l^{(2)} u^{(1)}(1,v-l)$$
$$\equiv \sum_{[1+k,v+l]\in\mathcal{S}} \sigma_{[1,v],[k,l]}\, X(1+k,v+l). \qquad (6.2.31)$$

For all locations $[1,v]$, $v = q_2+1,\cdots,N_2-q_2$, we may write

$$\theta_1(B_1^{-1})\cdot\theta_2(B_2^{-1})\cdot\theta_2(B_2)\, X(1,v) = \theta_2(B_2)\, u^{(1)}(1,v) = \theta_2(B_2^{-1})\, u^{(2)}(1,v)$$
$$\equiv \sum_{[1+k,v+l]\in\mathcal{S}} \sigma_{[1,v],[k,l]}\, X(1+k,v+l). \qquad (6.2.32)$$

We may start going backwards for the locations $[1,v]$, $v = N_2-q_2+1,\cdots,N_2-1$, by writing

$$\theta_1(B_1^{-1})\cdot\theta_2(B_2)\,[X(1,v) + \sum_{l=1}^{N_2-v} \theta_l^{(2)} X(1,v+l)] = u^{(2)}(1,v) + \sum_{l=1}^{N_2-v} \theta_l^{(2)} u^{(2)}(1,v+l)$$
$$\equiv \sum_{[1+k,v+l]\in\mathcal{S}} \sigma_{[1,v],[k,l]}\, X(1+k,v+l). \qquad (6.2.33)$$

Finally, for the location $[1, N_2]$, we can write

$$\theta_1(B_1^{-1}) \cdot \theta_2(B_2)\ X(1, N_2) \equiv \sum_{[1+k, N_2+l] \in \mathcal{S}} \sigma_{[1,N_2],[k,l]}\ X(1+k, N_2+l) = u^{(2)}(1, N_2). \tag{6.2.34}$$

Following the same sequel, we may define for all $[u, v] \in \mathcal{S}$

$$\begin{aligned} &\sum_{[u+k, v+l] \in \mathcal{S}} \sigma_{[u,v],[k,l]}\ X(u+k, v+l) \\ &\equiv \sum_{k=0}^{u-1} \theta_k^{(1)} \sum_{k^*=0}^{N_1-u} \theta_{k^*}^{(1)} \sum_{l=0}^{v-1} \theta_l^{(2)} \sum_{l^*=0}^{N_2-v} \theta_{l^*}^{(2)}\ X(u-k+k^*, v-l+l^*) \end{aligned} \tag{6.2.35}$$

where, of course,

$$\theta_k^{(1)} \equiv 0,\ k = q_1+1, q_1+2, \cdots \tag{6.2.36}$$

and

$$\theta_l^{(2)} \equiv 0,\ l = q_2+1, q_2+2, \cdots. \tag{6.2.37}$$

Equation (6.2.35) might be expressed as

- $$\sum_{k=0}^{u-1} \theta_k^{(1)} \sum_{l=0}^{v-1} \theta_l^{(2)}\ u^{(1)}(u-k, v-l),$$

  for $u = 1, \cdots, q_1,\ v = 1, \cdots, q_2,$

- $$\sum_{k=0}^{u-1} \theta_k^{(1)} \sum_{l^*=0}^{N_2-v} \theta_{l^*}^{(2)}\ u^{(2)}(u-k, v+l^*),$$

  for $u = 1, \cdots, q_1,\ v = N_2 - q_2 + 1, \cdots, N_2,$

- $$\sum_{k^*=0}^{N_1-u} \theta_{k^*}^{(1)} \sum_{l=0}^{v-1} \theta_l^{(2)}\ e^{(2)}(u+k^*, v-l),$$

  for $u = N_1 - q_1 + 1, \cdots, N_1,\ v = 1, \cdots, q_2,$

- $$\sum_{k^*=0}^{N_1-u} \theta_{k^*}^{(1)} \sum_{l^*=0}^{N_2-v} \theta_{l^*}^{(2)}\ e^{(1)}(u+k^*, v+l^*),$$

  for $u = N_1 - q_1 + 1, \cdots, N_1,\ v = N_2 - q_2 + 1, \cdots, N_2,$

- $$\sum_{k=0}^{u-1} \theta_k^{(1)}\ \theta_2(B_2)\ u^{(1)}(u-k, v) = \sum_{k=0}^{u-1} \theta_k^{(1)}\ \theta_2(B_2^{-1})\ u^{(2)}(u-k, v),$$

  for $u = 1, \cdots, q_1,\ v = q_2+1, \cdots, N_2 - q_2,$

- $$\sum_{l^*=0}^{N_2-v} \theta_{l^*}^{(2)}\ \theta_1(B_1)\ u^{(2)}(u, v+l^*) = \sum_{l^*=0}^{N_2-v} \theta_{l^*}^{(2)}\ \theta_1(B_1^{-1})\ e^{(1)}(u, v+l^*),$$

  for $u = q_1+1, \cdots, N_1-q_1,\ v = N_2-q_2+1, \cdots, N_2,$

- $$\sum_{k^*=0}^{N_1-u} \theta_{k^*}^{(1)} \theta_2(B_2^{-1})\ e^{(1)}(u+k^*, v) = \sum_{k^*=0}^{N_1-u} \theta_{k^*}^{(1)} \theta_2(B_2)\ e^{(2)}(u+k^*, v),$$

  for $u = N_1-q_1+1, \cdots, N_1,\ v = q_2+1, \cdots, N_2-q_2,$

- $$\sum_{l=0}^{v-1} \theta_l^{(2)}\ \theta_1(B_1^{-1})\ e^{(2)}(u, v-l) = \sum_{l=0}^{v-1} \theta_l^{(2)}\ \theta_1(B_1)\ u^{(1)}(u, v-l),$$

  for $u = q_1+1, \cdots, N_1-q_1,\ v = 1, \cdots, q_2.$

- $$\begin{aligned} & \theta_1(B_1)\cdot\theta_2(B_2)\cdot\theta_1(B_1^{-1})\cdot\theta_2(B_2^{-1})\ X(u,v) \\ = \ & \theta_1(B_1^{-1})\cdot\theta_2(B_2^{-1})\ e^{(1)}(u,v) = \theta_1(B_1)\cdot\theta_2(B_2)\ u^{(1)}(u,v) \\ = \ & \theta_1(B_1^{-1})\cdot\theta_2(B_2)\ e^{(2)}(u,v) = \theta_1(B_1)\cdot\theta_2(B_2^{-1})\ u^{(2)}(u,v), \end{aligned}$$

  for $u = q_1+1, \cdots, N_1-q_1,\ v = q_2+1, \cdots, N_2-q_2.$

The proof of the theorem follows after comparing (6.2.35) to (6.2.19). ■

Remark 6.1. (i) The form (6.2.19) verifies Martin (1979, p.211) who suggests that both the variance matrix for the observations $\{X(u,v),\ [u,v] \in \mathcal{S}\}$ and its inverse should be written as Kronecker products of two variance matrices and their inverses, respectively; two processes taking place on the line transect might have produced such variance matrices. Similarly, it holds for the determinant

$$|\mathrm{Var}\{\mathbf{X}\}^{-1}| = |\boldsymbol{\Sigma}_1^{-1}|^{N_2} \cdot |\boldsymbol{\Sigma}_2^{-1}|^{N_1}. \tag{6.2.38}$$

(ii) According to Remark 2.3, since we have assumed that (6.2.1) is a causal autoregression, the variance matrix $\mathrm{Var}\{\mathbf{X}\}$ is non-singular and so are the variance matrices $\boldsymbol{\Sigma}_1,\ \boldsymbol{\Sigma}_2$. The elements of $\mathrm{Var}\{\mathbf{X}\}^{-1}$ are the unique coefficients of the best linear predictors for each observation based on all other observations in $\mathcal{S}$.

(iii) Definition 6.1 and Theorem 6.1 can also be applied for the general case of any dimensionality d. Also, for the special case where $d = 1$, the theorem gives the form of

the inverse variance matrix of N consecutive observations from a causal auto-regression, when N is greater than twice the order of the auto-regression. Examples of similar work for the general case of an ARMA(p,q) on $\mathcal{Z}$, may be found in Ansley (1979) or Penzer and Shea (1997).

(iv) For

$$\{e^{(1)}(u,v)\} \sim WN(0,1)$$

the process

$$X(u,v) \equiv \lambda X(u-1,v) + \nu X(u,v-1) - \lambda\nu X(u-1,v-1) + e^{(1)}(u,v), \qquad (6.2.39)$$

or an equivalent having the same second-order properties is called the doubly-geometric process (Martin, 1979, p. 211). This process is linear-by-linear and the two corresponding one-dimensional processes are AR(1) on the line transect.

(v) Theorem 6.1 is an example of Section 2.4.1 and, more specifically, of the subsection on the conditional variance matrix of a random vector from a moving-average. Indeed, we saw there that some conditional variance matrices of this type are equal to variance matrices of random variables from auto-regressions. The elements of our inverse variance matrix $\mathrm{Var}\{\mathbf{X}\}^{-1}$ in the theorem, closely resemble the auto-covariances of the invertible moving-average process defined by the equation

$$Y(u,v) \equiv \theta_1(B_1^{-1}) \cdot \theta(B_2^{-1})\, e^{(1)}(u,v),\ \{e^{(1)}(u,v)\} \sim WN(0,1). \qquad (6.2.40)$$

Indeed, we saw in the end of the proof of the theorem that if $u = q_1 + 1, \cdots, N_1 - q_1$ and $v = q_2 + 1, \cdots, N_2 - q_2$, then the random variable

$$\begin{aligned} &\theta_1(B_1) \cdot \theta_2(B_2) \cdot \theta_1(B_1^{-1}) \cdot \theta_2(B_2^{-1})\ X(u,v) \\ = &\sum_{k,k^*=0}^{q_1} \sum_{l,l^*=0}^{q_2} \theta_k^{(1)} \cdot \theta_{k^*}^{(1)} \cdot \theta_l^{(2)} \cdot \theta_{l^*}^{(2)}\ X(u-k+k^*, v-l+l^*) \end{aligned}$$

is uncorrelated with all random variables on the locations of the set $\mathcal{S}$, apart from the random variable $X(u,v)$ itself. As a result, elements of the inverse $\mathrm{Var}\{\mathbf{X}\}^{-1}$ there would be of the form

$$\sum_{k-k^*} \sum_{l-l^*} \theta_k^{(1)} \cdot \theta_{k^*}^{(1)} \cdot \theta_l^{(2)} \cdot \theta_{l^*}^{(2)},$$

which are also auto-covariances from model (6.2.40). Why we cannot say the same for the locations on the edges of the rectangle, when $u = 1, \cdots, q_1$, or $u = N_1 - q_1 + 1, \cdots, N_1$, or $v = 1, \cdots, q_2$, or $v = N_2 - q_2 + 1, \cdots, N_2$, can be related to the fact that we only

expect a conditional variance matrix of random variables from (6.2.40) on the locations of $\mathcal{S}$ and not the exact variance matrix.

6.2.1 Gaussian likelihoods

For the parameter vectors

$$\boldsymbol{\theta}_1 \equiv [\theta_1^{(1)}, \cdots, \theta_{q_1}^{(1)}]^\tau \tag{6.2.41}$$

and

$$\boldsymbol{\theta}_2 \equiv [\theta_1^{(2)}, \cdots, \theta_{q_2}^{(2)}]^\tau \tag{6.2.42}$$

and for the observations $\{X(u,v),\ [u,v] \in \mathcal{S}\}$ from (6.2.1), where $\mathcal{S}$ is given in (6.2.20), we may write the Gaussian likelihood

$$L(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \propto \left\{ \frac{1}{[\prod_{u=1}^{q_1} r^{(1)}(u)]^{N_2}} \cdot \frac{1}{[\prod_{v=1}^{q_2} r^{(2)}(v)]^{N_1}} \right\}^{1/2} \exp\{-1/2\ \mathbf{X}^\tau [\boldsymbol{\Sigma}_1^{-1} \otimes \boldsymbol{\Sigma}_2^{-1}] \mathbf{X}\}, \tag{6.2.43}$$

where $\boldsymbol{\theta}_1 \in \Theta_1$, $\boldsymbol{\theta}_2 \in \Theta_2$ and $\Theta_1 \subseteq \mathcal{R}^{q_1}$ and $\Theta_2 \subseteq \mathcal{R}^{q_2}$ are parameter spaces, such that the process $\{X^{(1)}(u),\ u \in \mathcal{Z}\}$

$$\theta_1(B)\ X^{(1)}(u) \equiv e^{(1)}(u),\ \{e^{(1)}(u)\} \sim WN(0,1), \tag{6.2.44}$$

and the process $\{X^{(2)}(v),\ v \in \mathcal{Z}\}$

$$\theta_2(B)\ X^{(2)}(v) \equiv e^{(2)}(v),\ \{e^{(2)}(v)\} \sim WN(0,1) \tag{6.2.45}$$

are defined as causal auto-regressions taking place on line transects.

If we consider q_1 consecutive observations from (6.2.44), say $X^{(1)}(1), \cdots, X^{(1)}(q_1)$, then we may define

$$\widetilde{X}^{(1)}(1) \equiv 0 \tag{6.2.46}$$

and

$$\widetilde{X}^{(1)}(u) \equiv \sum_{k=1}^{u-1} \phi_{u,k}^{(1)}\ X^{(1)}(u-k),\ u = 2, \cdots, q_1, \tag{6.2.47}$$

to be the best linear predictor of $X^{(1)}(u)$ based on previous observations $X^{(1)}(1), \cdots, X^{(1)}(u-1)$. We may also define the prediction variances

$$r^{(1)}(u) \equiv E\{X^{(1)}(u) - \widetilde{X}^{(1)}(u)\}^2,\ u = 1, \cdots, q_1, \tag{6.2.48}$$

and compute them from the innovations algorithm.

Similarly, for the observations $X^{(2)}(1), \cdots, X^{(2)}(q_2)$, from (6.2.45), we define

$$\widetilde{X}^{(2)}(1) \equiv 0 \tag{6.2.49}$$

and

$$\widetilde{X}^{(2)}(v) \equiv \sum_{l=1}^{v-1} \phi_{v,l}^{(2)} \; X^{(2)}(v-l), \; v = 2, \cdots, q_2, \tag{6.2.50}$$

to be the best linear predictor of $X^{(2)}(v)$ based on $X^{(2)}(1), \cdots, X^{(2)}(v-1)$. The prediction variances can be defined as

$$r^{(2)}(v) \equiv E\{X^{(2)}(v) - \widetilde{X}^{(2)}(v)\}^2, \; v = 1, \cdots, q_2, \tag{6.2.51}$$

and can be computed via the innovations algorithm.

6.3 Quarter moving-averages

6.3.1 First-order filters

We consider the process $\{Y(u,v), \; u,v \in \mathcal{Z}\}$ to satisfy

$$Y(u,v) \equiv e(u,v) + a \; e(u-1,v) + b \; e(u,v-1), \; \{e(u,v)\} \sim IID(0,1), \tag{6.3.1}$$

where

$$|a| + |b| < 1. \tag{6.3.2}$$

Due to the last condition the process is invertible. Condition (6.3.2) is not just sufficient for the stability of the filter $1 + a \; z_1 + b \; z_2, \; |z_1|, |z_2| \leq 1$ (Guyon, 1982, p.102) but also necessary according to Huang (1972, p.162). The auto-covariance function of the process is

$$\gamma(k,l) \equiv E\{X(u,v)X(u-k,v-l)\} = \begin{cases} 1 + a^2 + b^2, & [k,l] = [0,0] \\ a, & [k,l] = [k,0], \; k = \pm 1 \\ b, & [k,l] = [0,l], \; l = \pm 1 \\ ab, & [k,l] = [1,-1] \text{ or } [k,l] = [-1,1] \\ 0, & \text{otherwise} \end{cases} . \tag{6.3.3}$$

For $N = N_1 N_2$ observations on the locations

$$\mathcal{S} \equiv \{[u,v] : u = 1, \cdots, N_1, \; v = 1, \cdots, N_2\},$$

from the process (6.3.1) with unknown values for the parameters a and b, the maximum likelihood estimators, say $\widehat{a}$ and $\widehat{b}$, may be searched over all the possible values $|a|+|b| < 1$. The absolute value of the bias of both the maximum likelihood estimators $\widehat{a}$ and $\widehat{b}$ is of order $N^{-1/2}$ (Guyon, 1982). This means that, if multiplied by $N^{1/2}$, it can be bounded away from ∞ but cannot be guaranteed to reach 0 for large sample sizes. Since the standard error of the estimators is also of order $N^{-1/2}$, we cannot derive both the asymptotic unbiasedness and asymptotic normality of the estimators. This, of course, is a special case of a general problem that affects the processes of any dimensionality d and is a direct consequence of the edge-effect. When $d > 2$, the bias multiplied by $N^{1/2}$ cannot be guaranteed to be bounded away from ∞ either, as it is, in general, of order $N^{-1/d}$. No wonder why the problem does not become apparent when $d = 1$.

For the special case when $d = 2$, Yao and Brockwell (2006) gave a solution to the problem of estimating the parameters of any process that can possess an AR(∞) representation. Suppose then that we can write

$$Y(u,v) \equiv \sum_{[k,l]>\mathbf{0}} \varphi_{[k,l]}\, Y(u-k, v-l) + e(u,v),\ \{e(u,v)\} \sim IID(0,1),\ \sum_{[k,l]>\mathbf{0}} |\varphi_{[k,l]}| < \infty. \tag{6.3.4}$$

In (6.3.4) we may find the unique coefficients that give the best linear predictor of an observation based on all other observations from its 'past'. In a sample set like $\mathcal{S}$ though, one can never have all the observations needed from the 'past', unless the auto-regression (6.3.4) is of finite order in its causal representation. The Cholesky decomposition for the $(N \times N)$ variance matrix uses for every location $[u,v] \in \mathcal{S}$, the best linear predictors

$$\widetilde{Y}(u,v) \equiv \sum_{\substack{[k,l]>\mathbf{0},\\ [u-k,v-l]\in\mathcal{S}}} \varphi_{[k,l]}^{[u,v]} Y(u-k, v-l) \tag{6.3.5}$$

and the prediction variances

$$r(u,v) \equiv E\{Y(u,v) - \widetilde{Y}(u,v)\}^2 \tag{6.3.6}$$

instead. It has been proven that $\varphi_{[k,l]}^{[u,v]}$ is closer to $\varphi_{[k,l]}$ as the location $[u,v]$ moves away from the edges of the rectangle and there is more and more information about its 'past'. The primary index u must be as close to N_1 as possible while the index v must be away from both sides of length N_1 of the rectangle (Yao and Brockwell, 2006). In other words

$$\varphi_{[k,l]}^{[u,v]} \to \varphi_{[k,l]} \tag{6.3.7}$$

as $\min\{u, v, N_2 - v\} \to \infty$ for the sampling set $\mathcal{S}$.

For one-dimensional processes we may rest that, as we obtain more and more observations, we also have more and more information available from the 'past' of every new observation generated. The same cannot be said for $d \geq 2$ as it was described in Section 3.2.2. Like for the coefficients $\varphi_{[k,l]}^{[u,v]}$ (6.3.7) holds, an inequality with similar meaning can be obtained, *i.e.*

$$E|[Y(u,v) - \widetilde{Y}(u,v)] - e(u,v)| \leq C\ [\alpha^u + \alpha^v + \alpha^{N_2 - v}], \tag{6.3.8}$$

for constants $C > 0$ and $\alpha \in (0,1)$.

To remedy (6.3.7) and (6.3.8), Yao and Brockwell (2006) proceeded with a selection of N^* observations

$$\mathcal{S}^* \equiv \{[u,v] : u = n_1 + 1, \cdots, N_1,\ v = n_2 + 1, \cdots, N_2 - n_2\}, \tag{6.3.9}$$

where $n_1, n_2 \to \infty$ but also $n_1/N_1 \to 0$ and $n_2/N_2 \to 0$, as the selection takes place over an increasing number of locations available in the sample $\mathcal{S}$. Then, they proved that maximizing the modified Gaussian likelihood

$$L^* \propto \frac{1}{\prod_{[u,v]\in\mathcal{S}^*} r(u,v)^{1/2}}\ \exp\{-1/2[\sum_{[u,v]\in\mathcal{S}^*} \frac{[Y(u,v) - \widetilde{Y}(u,v)]^2}{r(u,v)}]\} \tag{6.3.10}$$

generates consistent and asymptotically unbiased and normal estimators for the unknown parameters of (6.3.1).

6.3.2 Simulations

We have produced $N = N_1 N_2$ observations $\{Y(u,v),\ u = 1, 2, \cdots, N_1,\ v = 1, \cdots, N_2\}$ from the process satisfying

$$Y(u,v) \equiv e(u,v) + 0.1\ e(u-1,v) + 0.2\ e(u,v-1),\ \{e(u,v)\} \sim NID(0,1), \tag{6.3.11}$$

for $N_1 = 5, 10, \cdots, 30$, and $N_2 = 10$. As a result, we have used the true parameters

$$a \equiv 0.1, \quad b \equiv 0.2, \tag{6.3.12}$$

and a normally distributed error sequence in the process defined by (6.3.1).

Next we consider that the observations available have been generated by a process satisfying (6.3.1), but with unknown parameters a and b. Using the innovations algorithm, we have computed the exact Gaussian likelihood as a function of the values

$a, b = 0, \pm 0.01, \cdots, \pm 0.99$, such that $|a| + |b| < 1$. First, we have considered the pair $\widehat{a}, \widehat{b}$, such that the likelihood there, is greater than all the other computed values. The estimators $\widehat{a}$, $\widehat{b}$ will be called estimators without selection.

On the other hand, we have considered the estimators $\widetilde{a}, \widetilde{b}$, after selecting locations according to (6.3.9). We have then found the maximum of the modified Gaussian likelihood (6.3.10) for the pair of values, say $\widetilde{a}$ and $\widetilde{b}$. We have considered two cases of estimators $\widetilde{a}, \widetilde{b}$, with selection. The first one is a fixed selection, for which we always use

$$n_1 = n_2 = 1 \tag{6.3.13}$$

for any N_1, N_2. We call the second type of selection a square root selection, as

$$n_1 = [\sqrt{N_1}],\ n_2 = [\sqrt{N_2/2}] \tag{6.3.14}$$

then.

While the square root selection does obey to the rule proposed by Yao and Brockwell (2006), *i.e.* that $n_1, n_2 \to \infty$ and $n_1/N_1, n_2/N_2 \to 0$ as $\min\{N_1, N_2\} \to \infty$, the fixed selection does not. We should remember here that it holds

$$E\{Y(u,v)Y(u-k, v-l)\} = 0,\ k, l = \pm 2, \pm 3, \cdots, \tag{6.3.15}$$

according to (6.3.3). This argument could be used as in Chapter 3, in order to guarantee that maximizing certain quantities gives unbiased estimators. Unfortunately, (6.3.10) is not one of these quantities; the selection (6.3.13) would work if the original process was using an auto-regressive rather than a moving-average polynomial

$$\theta(z_1, z_2) \equiv 1 + a\ z_1 + b\ z_2.$$

Still, if we use (6.3.7), (6.3.8) and mathematical arguments like the ones provided by Yao and Brockwell (2006), we should expect the absolute bias of the fixed selection estimators to be smaller than that of the original estimators. The observations used for the computation of the fixed selection estimators all miss information from more than one step away from their 'past'. For large sample sizes, such as $N = (30 * 10) = 300$, when we may rely on computing the bias of the fixed estimators from $(29 * 8)$ locations, it would be interesting to see whether there seem to be a big difference between those estimators and the original ones. If not, that would mean that excluding one step only from the past has a minimal effect, as we increase the number of observations, and more observations need to be excluded.

We are interested in comparing the performance of the original estimators without selection and their converted versions, in order to see whether the modification proposed for the Gaussian likelihood is essential to the analysis and beneficial regarding the edge-effect. Thus, we have repeated the same steps 100 times. We write $\widehat{a}[i]$ for the estimator without selection of the parameter $a = 0.1$ from the i-th replication, where $i = 1, \cdots, 100$. Similarly, we write $\widetilde{a}[i]$ and $\widehat{b}[i]$, $\widetilde{b}[i]$ for the parameter $b = 0.2$. We have computed the mean

$$\widehat{E}(\widehat{a}) \equiv \frac{1}{100} \sum_{i=1}^{100} \widehat{a}[i] \tag{6.3.16}$$

and the quantity

$$\widehat{Bias}(\widehat{a}) \equiv \widehat{E}(\widehat{a}) - a = \widehat{E}(\widehat{a}) - 0.1, \tag{6.3.17}$$

which we expect to be close enough to the bias of the estimator $\widehat{a}$. Moreover, we have computed

$$\widehat{\text{Var}}\{\widehat{a}\} \equiv \frac{1}{99}[\sum_{i=1}^{100} \widehat{a}[i]^2 - 100 \ \{\widehat{E}(\widehat{a})\}^2] \tag{6.3.18}$$

to be close enough to the true variance of $\widehat{a}$. Thus, we write

$$\widehat{MSE}(\widehat{a}) \equiv \widehat{\text{Var}}\{\widehat{a}\} + \{\widehat{Bias}(\widehat{a})\}^2 \tag{6.3.19}$$

to approximate the mean square error of the estimator $\widehat{a}$. We also approximate the expected value, the bias, the variance and the mean square error (MSE) of the other estimators of interest in exactly the same way from the 100 replications.

We have chosen to fix $N_2 = 10$ and let N_1 only take several values. Later on, we will also discuss some cases when $N_1 = N_2 = n$ and $n = 10, 15, 20$. When $N_1 \to \infty$ only, the number of observations $N = 50, 100, \cdots, 300$ increases to infinity, though at the speed of N_1 only. Since the model (6.3.1) treats both its parameters a and b equally, it would be interesting to see how the estimation of its parameters is affected when the number of recordings on one dimension only is changing. Especially for the estimators computed with fixed selection, it is a question of interest whether they can compete with the standard estimators then. In the contrary, the square root selection estimators have been proven by Yao and Brockwell (2006) to become asymptotically unbiased and normal as both $N_1, N_2 \to \infty$. Here, we will see whether they manage to equalize the performance of the standard estimators, in terms of the Mean Square Error, when $N_1 \to \infty$ only.

In Figure 6.1, we look at the absolute value of the bias of the three different estimators of the parameter $a = 0.1$. We can see immediately that the estimator without

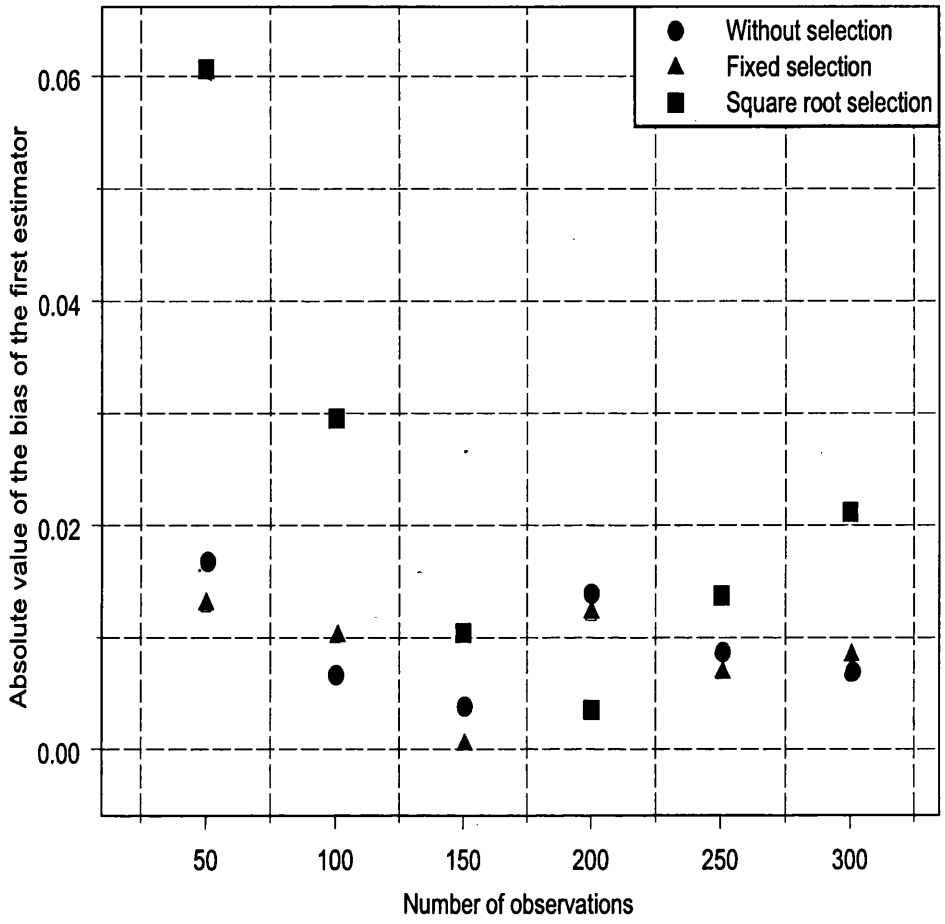


Figure 6.1: Absolute value of the bias of the estimators $\widehat{a}$, $\widetilde{a}$ versus the number of observations from 100 replications.

selection and the fixed selection estimator exhibit a similar performance regarding their bias. Except from the case when we have generated $N = 200$ observations, where the square root selection estimator of a has the smallest absolute bias, in all other cases the absolute bias of the two estimators are smaller and close enough; for the case of $N = 300$ observations, the genuine estimator has an absolute bias 0.0067, while the fixed selection one has an absolute bias equal to 0.0082. Especially the square root selection estimator shows a dramatic trend downwards from $N_1 = 5$ to $N_1 = 200$, but this seems

to stop when it reaches the performance of the other two estimators. The two other estimators do not exhibit the same trend, but they are not outperformed by the square root selection estimator either, especially when $N = 300$.

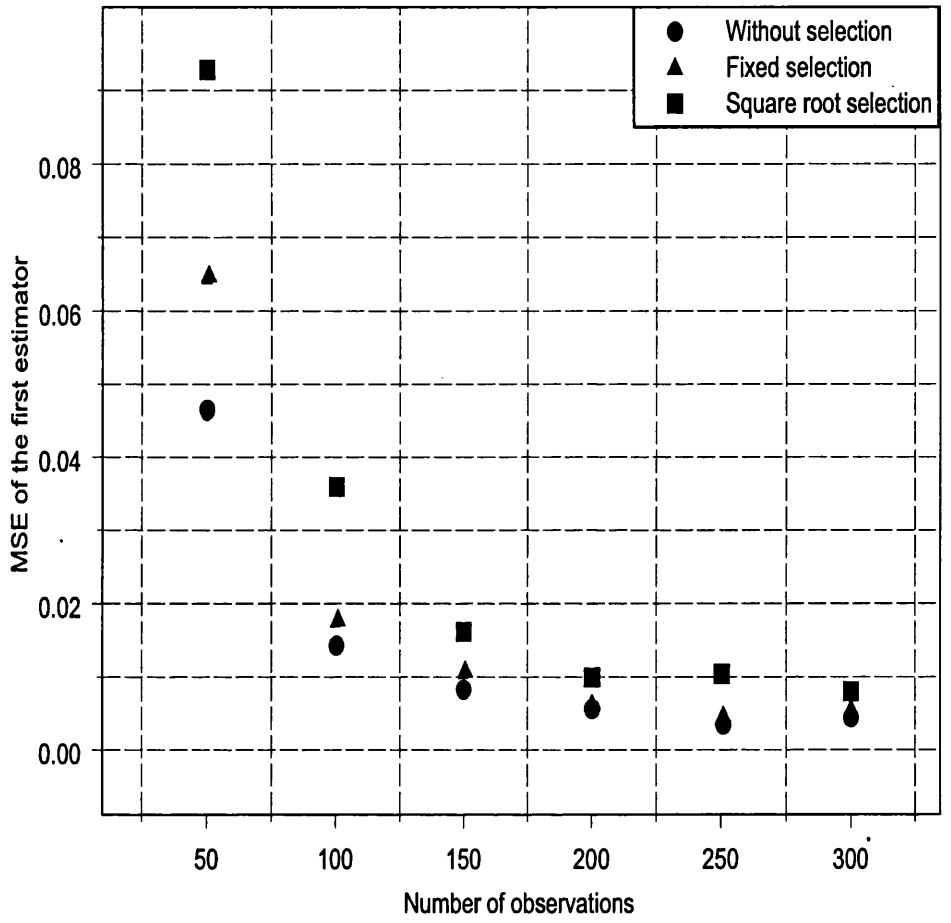


Figure 6.2: Mean Square Error of the estimators $\widehat{a}$, $\widetilde{a}$ versus the number of observations from 100 replications.

Looking at the mean square error of the estimators in Figure 6.2 is expected to be a combination of the picture we saw in Figure 6.1 and the variance of the estimators, according to (6.3.19). The standard estimator enjoys the privilege of using all the N observations available, while when a selection takes place, some observations are only

used to reduce the absolute bias but not the variance. Thus, we know in advance that the standard estimator has a smaller variance than the fixed selection estimator, and the fixed selection estimator has a smaller variance than any estimator that chooses to leave out more and more locations from the edges. The question is whether the standard estimator will manage to dominate in terms of the mean square error, or whether the estimators that have omitted some locations will manage to compete with it successfully.

The answer to that comes straight from Figure 6.2, which reveals that all the three estimators have asymptotically almost the same mean square error. Especially the fixed selection estimator seems to have an equal performance as the estimator without selection; this might be attributed first to the fact that the estimator with selection had often a smaller absolute bias according to Figure 6.1, as well as to the fact that only one location per dimension, which is left out of use for the reduction of the variance, is asymptotically negligible as $N \to \infty$. Of course, since

$$\frac{n_1}{N_1} = \frac{\sqrt{N_1}}{N_1} = \frac{1}{\sqrt{N_1}} \to 0$$

as $N_1 \to \infty$, it holds that omitting $\sqrt{N_1}$ locations is also asymptotically negligible, which is the reason why the square root selection estimator is so close to the other two estimators when $N_1 = 30$.

As we have explained before, the model (6.3.1) treats its two parameters a and b in exactly the same way. As a result, any differences detected in the performance of the two original estimators $\widehat{a}$ and $\widehat{b}$, should clearly be attributed to the fact that $N_1 \to \infty$ but $N_2 = 10$ is fixed. On the other hand, the way Yao and Brockwell (2006) proceeded with a selection of observations favors the dimension $u \in \mathcal{Z}$, as the primary dimension, and leaves $v \in \mathcal{Z}$ to be the secondary dimension. Consequently, both the fixed and square root selection estimators will also reflect this hierarchy between the two axes as a difference in the performances of $\widetilde{a}$ and $\widetilde{b}$.

Indeed, the picture we get from Figure 6.3 is that the points are scattered without any specific structure compared to Figure 6.1. That must be the effect of the fact that $N_2 = 10$ is fixed. Nevertheless, when $N = 300$ observations are generated, all the three estimators manage to reduce their absolute bias effectively. The absolute value of the bias of the estimator without selection $\widehat{b}$ and the fixed and square root selection estimators $\widetilde{b}$ are $0.0016, 0.0010$ and 0.0054, respectively. These are smaller numbers than the ones written down for the absolute value of $\widehat{a}$ and $\widetilde{a}$. The value 0.0054 of

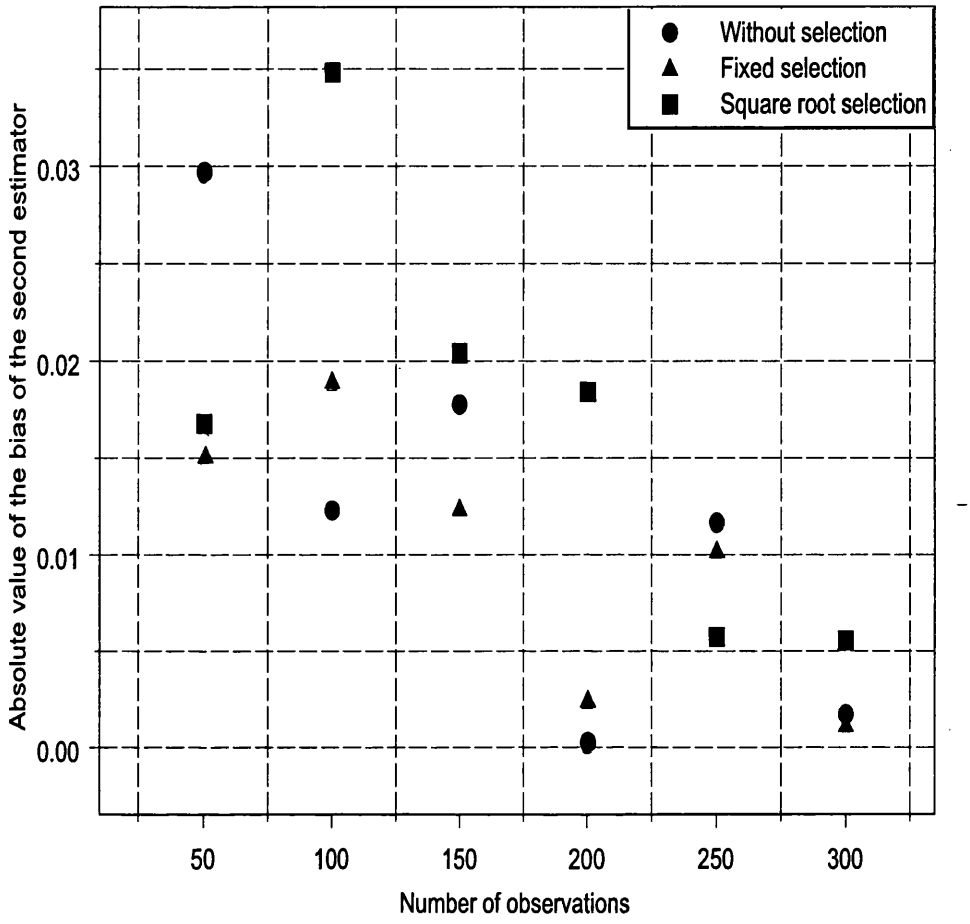


Figure 6.3: Absolute value of the bias of the estimators $\widehat{b}$, $\widetilde{b}$ versus the number of observations from 100 replications.

the square root selection estimator is bigger than the values for the absolute bias for the remaining two estimators, which make us wonder whether it is worth proceeding with such a selection of observations. Let us not forget that, given that making a selection of locations automatically implies that smaller variances of the estimators will be derived, we should at least demand to balance that with a smaller absolute bias. Further comments on this will be made when we will allow for $N = n^2$ and $n \to \infty$ later.

Finally, Figure 6.4 gives a similar impression to Figure 6.2, as they both express

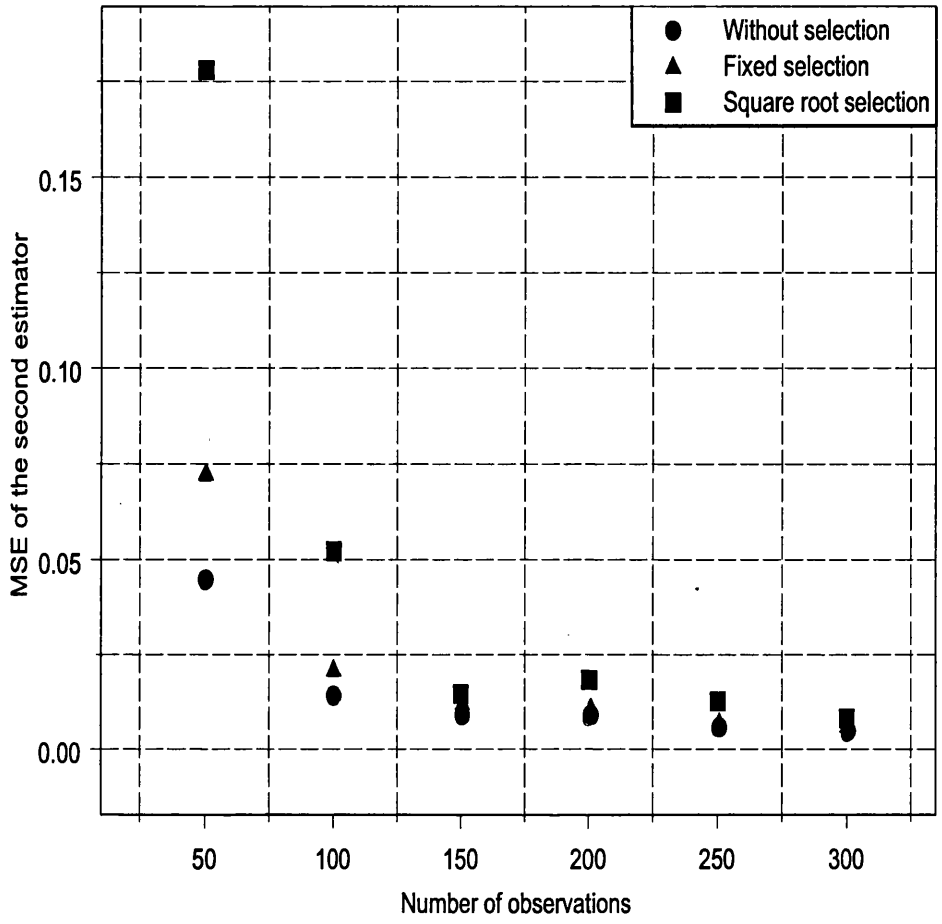


Figure 6.4: Mean Square Error of the estimators $\widehat{b}$, $\widetilde{b}$ versus the number of observations from 100 replications.

how the mean square error of the estimators is reducing as $N = 50, 100, \cdots, 300$. Moreover and when $N = 300$, the MSE of the three estimators $\widehat{a}$ and $\widetilde{a}$ are equal to $0.00407530, 0.00490841$ and 0.00772686, while for the estimators $\widehat{b}$ and $\widetilde{b}$, we have the values $0.00377371, 0.00490302$ and 0.00699162, which are all slightly smaller. This must be due to the small differences in the absolute bias. Similar conclusions might be made from Figure 6.4 and for $N = 300$, as all the three estimators exhibit a similar performance in terms of the mean square error.

The estimators proposed by Yao and Brockwell (2006) have been proven to be asymptotically normal as both $N_1, N_2 \to \infty$. We would like to see next, whether our decision to set $N_2 = 10$ has really deprived the square root selection estimators of a much faster reduction of the absolute bias than that of the exact likelihood estimators. Thus, we have generated $N = n^2$ observations from (6.3.11) and have found the exact likelihood and square root selection estimators $\widehat{a}, \widehat{b}$ and $\widetilde{a}, \widetilde{b}$, respectively. We have chosen $n = 10, 15, 20$, and we have repeated the same procedure 30 times.

Table 6.1: Bias of estimators $\widehat{a}, \widetilde{a}, \widehat{b}, \widetilde{b}$, from n^2 observations and 30 replications.

	Without selection $\widehat{a}$	With selection $\widetilde{a}$	Without selection $\widehat{b}$	With selection $\widetilde{b}$
$n = 10$	-0.0065	-0.0293	0.0122	0.0347
$n = 15$	-0.0084	-0.0009	-0.0124	-0.0077
$n = 20$	0.000666667	-0.00966667	-0.015	-0.018

Considering the estimators of the parameter a first, it seems that when $n = 15$ and $[\sqrt{15}] = 3$ locations have been excluded from the primary dimension, the selection estimator gives a smaller absolute bias than the genuine estimator. When $n = 20$ though, not only does $\widehat{a}$ have a smaller bias in absolute value, but also its bias and the bias of $\widetilde{a}$ have a different sign; since $[\sqrt{20}] = 4$, this might mean that we have excluded now too many locations, which gives a completely different impression than excluding 3 only. As a result, there is the question which one is the best selection we should trust. Maybe the square root selection excludes locations too fast and there is no need for that, as the exact likelihood estimator performs well too. Especially when there are $n^2 = 400$ observations available, the expected value of the estimator $\widehat{a}$ seems very close to $a = 0.1$. Thus, the mean of $\widehat{a}$ does not look to have reached 0.1 slower than the mean of $\widetilde{a}$ there.

The conclusions can be similar for the two estimators of b, though not that exaggerated. When $n = 15$ only, the absolute bias of $\widetilde{b}$ is smaller than that of $\widehat{b}$. In general, the biases of the two estimators are really close when $n = 10, 15, 20$, and they always have the same sign. Jumping from $n = 10$ to $n = 20$ does not really allow to $\widehat{b}$ to reduce its absolute bias, while $\widetilde{b}$ has reduced its absolute bias from 0.0347 to 0.018, *i.e.* there is a 48.127% decrease. That could be seen as a sign of a faster recovery of the absolute bias of $\widetilde{b}$ to zero. Still, the absolute bias of $\widehat{b}$ and that of $\widetilde{b}$ are really close even when $n = 20$.

We do not give the mean square error of the estimators and how it was approximated for the four estimators, as it was always smaller for the exact likelihood estimators $\widehat{a}$ and $\widehat{b}$.

As a result, we should wonder if a selection of locations is of any practical use. If a modified estimator cannot compete with the standard estimator by giving a much smaller absolute bias, how could we possibly single it out as the best estimator, in terms of the mean square error? Further investigation is required for even larger sample sizes and $n \to \infty$. Nevertheless, if the number of observations recorded is not that big, our investigation so far reveals that the exact Gaussian likelihood estimators have a good behavior, that is not likely to be improved if some locations are excluded. Furthermore, the type of selection followed to exclude locations is a question of interest. The solution proposed by Yao and Brockwell (2006) provides the mathematical convenience required for the proofs, but it is also very general and abstract. It is only preferred, given that we believe our exact likelihood estimators are worse, in terms of the order of the bias; we have not always seen that here.

In conclusion, we would like to urge the reader to remember all the estimators we proposed back in Chapter 3. If the process of interest is an ARMA(p, q) on $\mathcal{Z}^d$, then we could use the modified Gaussian likelihood estimators proposed there. They are computed in a simple way, guaranteed to be asymptotically unbiased and normal and they use a very specific and finite exclusion of locations, based on the order of the process of interest. We do hope that these estimators could perform at least equally to the standard Gaussian likelihood estimators in terms of the Mean Square Error, and they could be a satisfying remedy to the problem of estimation for the parameters of any ARMA process on $\mathcal{Z}^d$ with any positive integer d.

Chapter 7

Conclusion

A large part of this thesis referred to (weakly) stationary processes, which take place on $\mathcal{Z}^d$. Chapter 2 introduced ideas, which were essential later for the two chapters of the estimation and statistical testing, *i.e.* Chapters 3 and 4. The general Yule-Walker equations were introduced in Chapter 2 and they could be considered as a fundamental idea used repeatedly in the next two chapters. The general Yule-Walker equations relate the auto-covariance functions of two different processes, or they relate the numerator and denominator of the spectral density of the same process, or alternatively, they relate the auto-covariance function and the best linear predictor coefficients of the same process. For the cases of auto-regressions and moving-averages in particular, it is thanks to the general Yule-Walker equations that we have managed to discover the form of inverse variance matrices of random vectors from those processes. As a result, in Chapters 3 and 4 we have written down conditional Gaussian likelihoods explicitly.

Chapter 3 was dealing with the estimation of the parameters of an ARMA process. For the case of invertible moving-averages, the general Yule-Walker equations have also motivated a method of moments estimation that replaces the theoretical auto-covariances of the moving-average by their sample estimates. This is another contribution of the Yule-Walker equations. In general in Chapter 3, it was attempted to estimate the parameters of interest by maximizing modified versions of Gaussian likelihoods. The order of the bias of the estimators defined, was of great importance when $d \geq 2$, because of the edge-effect. All our estimators, which were computed from N observations and defined for any positive integer number d, achieved a bias of order N^{-1} and defeated the edge-effect. In that sense, we have managed similar results to Guyon (1982). Moreover,

all our steps followed the time domain, thanks to the explicit results of the previous chapter, regarding inverse conditional variance matrices. Guyon (1982) had served the spectral domain and he had used the Gaussian likelihood proposed by Whittle (1954). Furthermore in Chapter 3, we suggested a modification of the Gaussian likelihood, which is appropriate for ARMA processes only. This was because it was based on the finite cardinality of the set of lags, for which the auto-covariance function of a moving-average was not equal to zero. The moving-average would be created from the original ARMA with a finite transformation. Guyon's (1982) suggestion was using the estimates of auto-covariance function of the process of interest; as a result for an ARMA process, that would imply that the more the observations recorded, the more the lags that would need to be taken into account for the computation of the likelihood.

Our suggested modification on the Gaussian likelihood of observations from an ARMA process can apply to both the cases when the ARMA is unilateral or bilateral. Whittle (1954) had dealt with bilateral auto-regressions, but never before did we have a solution for the bilateral ARMA in its original form, rather than in its AR(∞) representation. The parametrization of the bilateral ARMA should be done with care, as the theoretical model should include positive or negative lags only; this is often unnatural for a bilateral process. The modified Gaussian likelihood estimators, which we defined, are asymptotically normal under the assumption of a finite fourth moment of the sequence of independent and identically distributed random variables of interest. This could be seen as a similar condition to that of Guyon (1982, p.100), *i.e.* that the fourth cumulant of the error sequence exists. Nevertheless, while we established the asymptotic normality, we did not manage to discover the elements of the variance matrix of the estimators, unless the process satisfies an extra condition. In the case of Gaussian processes, the conditions are satisfied and the estimators are asymptotically efficient. In the future, we do hope to improve the conditions of Proposition 2.6, which was the main tool that we used to find the elements of all variance matrices in Chapters 3 and 4.

The strong conditions of the theorems in Chapters 3 and 4 were not the only difficulty we confronted. In general, the nature of the problems of estimation when $d \geq 2$ are reflected in the edge-effect. Thus, in order to make sure that the edge-effect would not interfere with our methods, we often left out some of the observations, in the sense that we did not make full use of them. We saw this very clearly with our application in Section 3.7.3, where we had to reduce our sample size from $N = 8,500$ to $N^* = 4,050$.

When $d = 1$, the number of observations we may leave out is the same, no matter how large the sample size might be. As a result in time series, maximizing the exact Gaussian likelihood of all the N observations guarantees that the same results can be achieved as when a conditional version has been preferred. Unfortunately for $d \geq 2$, as N increases to infinity, it is only the number of indexes that should be left out per dimension that remains fixed; the number of rejected observations tends to infinity. Exact Gaussian likelihoods should not be used and we should be very careful if we are to involve the observations of the edges. We have not managed to prove in theory that there is any way to include these observations and, consequently, we have not managed to derive the asymptotic unbiasedness and normality of the estimators then. In Chapter 6 and for the special case when $d = 2$ only, we did use simulations to verify that maximizing the exact Gaussian likelihood might have a minimal difference from maximizing the modified version of Yao and Brockwell (2006), regarding the bias and, consequently, the mean square error of the estimators. When and how all the observations could be used in practice, what happens when our final sets of cardinality N^* are empty and how can we make the most of our observations without reviving the edge-effect, are all questions that have not yet been answered.

Special reference was made for the case of finite auto-regressions. The conditions of the theorems were relaxed then, as only a finite second moment was required from the sequence of independent and identically distributed random variables. That is the same condition also used by Yao and Brockwell (2006), since they had followed the infinite auto-regressive path of the process of interest, in order to approach the problem of estimation of the parameters. A finite auto-regression is a linear model and we could easily generalize or use all the established results for linear models. The conditional Gaussian likelihood estimators are also least squares estimators and, consequently, they are very easy to compute. In Section 3.7, we saw how a finite auto-regression may be used to model observations recorded regularly over the space $\mathcal{Z}^2$ and the time axis $\mathcal{Z}$. The statistical tests we proposed, were tests for nested linear models; in order to discover how powerful those tests can be for our special case, we may use the result for the asymptotic normality of our estimators. Even when the inverse variance matrix of the estimators is hard to find or when the original distribution of the errors is unlikely to have produced asymptotically normal estimators, combined together with a small sample size, we may resort to bootstrap techniques. Of course to apply bootstrap, the original N dependent

observations will be transformed to N^* realizations from independent random variables, in order to take random samples with replacement repeatedly; that is possible when a linear model has been assumed.

In Chapter 4, we dealt with a more general version of the auto-normal formulations, *i.e.* the auto-linear formulations on $\mathcal{Z}^d$. The auto-normal schemes on $\mathcal{Z}^2$ had been introduced by Besag (1974), who presented the second-order properties of stationary processes in terms of a new parametrization. Those parameters could not be related to the auto-covariance function of the process directly. As a result, it was difficult to find their sample analogues. The two methods of estimation that were proposed by Besag (1974), *i.e.* the pseudo-likelihood and coding techniques, had failed to capture the complexity of the problem and to provide estimators with known properties. For the first time, we have proven some of the properties of the Gaussian pseudo-likelihood estimators, but he have not stopped there. Thanks to the general Yule-Walker equations, we managed to relate the parameters of interest with the auto-covariance function of another (weakly) stationary process, which can be created from the original one with a finite linear transformation. As a result, we invented a new method of moments estimation for our unknown parameters, for which we managed to establish the consistency and asymptotic normality of the estimators, under certain conditions. Moreover, we expressed the elements of the variance matrix of the estimators defined, in terms of the unknown parameters. Statistical tests could be performed.

As in Chapter 3, the strong conditions required for the proof of the theorems, remain a weakness. The edge-effect also remains a menace, since we have dealt with processes on $\mathcal{Z}^d$; all the selections of observations we have made are mandatory for our results to hold, and this often implies that we have left a lot of information unused. Moreover, the method of moments we proposed, generates estimators that are solutions of the same number of equations. As we saw with our application in Section 4.7, it might be difficult to solve these equations in practice. In that same section, we found our estimates by minimizing a random quantity, rather than by solving the equations. Whether the estimators used are close enough to the estimators we studied before, is a question of interest. Furthermore when the variance of the process is also unknown, the quantity we should choose to minimize should not be dependent on the variance of the process. We proceeded with a very subjective selection of the quantity that we would minimize. Did that give the same solutions as we had studied theoretically? Did that allow us to

perform the statistical tests we established in Section 4.6? Our investigation so far has not answered these questions.

Next in Chapter 5 we changed the setting, in order to allow for our locations to be scattered anywhere on $\mathcal{R}^d$. We studied multivariate auto-regressions that could take place on these locations and on the time axis $\mathcal{Z}$. The price we paid for generalizing our methods from the spatial regular space $\mathcal{Z}^d$ to $\mathcal{R}^d$, would be to have to fix our locations; only the number of timings observed could increase to infinity. In order to model our observations over time, we considered a finite multivariate auto-regression, which, of course, also involved a multivariate sequence of uncorrelated random vectors. We used the elements of the inverse variance matrix of these vectors as our spatial parameters. Using a conditional Gaussian likelihood to define our estimators, we proved that those are consistent and asymptotically normal, under very weak conditions. We used the likelihood ratio to test statistical hypotheses and we applied our results on the mink and muskrat spatio-temporal data. It was then that we could see clearly that when the number of locations, say N, is large, we have an even larger number of spatial parameters, *i.e.* a number bounded by N^2. A possible grouping of the spatial sites could then be fruitful, as it could save a lot of computational time. Nevertheless, the grouping of the sites should not cause a simplistic but a meaningful reduction of parameters.

In the last chapter, we attempted to study a special case of two-dimensional auto-regressions, which are called linear-by-linear. The results of this section could project to higher dimensionalities, but mainly they reflect the properties of one-dimensional auto-regressions. We managed to write explicitly the inverse variance matrix of observations on a rectangle from such processes. This result came straight from an identical derivation for one-dimensional processes, *i.e.* writing explicitly the inverse variance matrix of consecutive observations from a finite auto-regression on $\mathcal{Z}$. Future work will include writing down an algorithm for the case of the inverse variance matrix of observations from an ARMA process; one-dimensional and then linear-by-linear.

We concluded the thesis with simulations, which aimed at comparing the performance of the exact Gaussian likelihood estimators versus the modified Gaussian likelihood estimators, where the modification was according to Yao and Brockwell (2006) and the model of interest was a first-order, two-dimensional moving-average. The performance of the estimators was evaluated in terms of their absolute bias, variance and mean square error. The same pattern should be followed for all the new estimators we proposed back

in Chapter 3, in order to confirm that they do perform well, compared to the genuine or other modified likelihood estimators. In the future, we hope to have not only theoretical but also practical evidence, in order to convince the reader that our modifications on Gaussian likelihoods are both necessary and reliable, especially when the problem becomes complex with more than two dimensions.

Acknowledgements

I would like to thank the Leverhulme Trust, which funded my studies for three years.
Professor Q. Yao has honored me with his trust and has been a very generous supervisor. The lessons he has taught me I will not forget. I admire him both as a person and as a statistician.
Doctor J. Penzer has contributed regularly with written notes and helpful comments. Doctor M. Knott has favored me with many discussions. Doctor J. Lu has provided all the datasets and has guided me in my first steps as a C++ programmer. I was also very lucky to have some interesting conversations with Professor A. Atkinson.
In general, I would like to thank everyone in the Department of Statistics of the London School of Economics for the nice years I spent with them.
I would like to send my particular thanks to Professor N. Cressie from the Ohio State University, who made some very interesting comments and suggestions for the improvement of Chapter 5, during a conference.
Professors E. Bura and I. Moustaki from the George Washington University and the Athens University of Economics and Business, respectively, have advised me repeatedly on matters of career and statistics.
This is in the memory of my good teachers Doctor P. Panopoulos and I. Vassilatos.
I would like to dedicate this thesis to those who inspire me with their talents and to those who encourage my efforts and dreams.
Finally, I thank my wonderful family for their love and support.

Bibliography

Anderson, B. D. O. and Jury, E. I. (1974). Stability of Multidimensional Digital Filters. *IEEE Transactions on Circuits and Systems*, **CAS-21**, 300-304.

Ansley, C. F. (1979). An algorithm for the exact likelihood of a mixed autoregressive-moving average process. *Biometrika*, **66**, 59-65.

Bartlett, M. S. (1967). Inference and Stochastic Processes. *J. Royal Statist. Soc.* **A**, **130**, 457-478.

Bartlett, M. S. (1968). A Further Note on Nearest Neighbour Models. *J. Royal Statist. Soc.* **A**, **131**, 579-580.

Bartlett, M. S. (1971). Physical Nearest-Neighbour Models and non-Linear Time-Series. *J. Appl. Prob.* **8**, 222-232.

Bartlett, M. S. (1972). Physical Nearest-Neighbour Models and non-Linear Time-Series. II Further Discussion of Approximate Solutions and Exact Equations. *J. Appl. Prob.*, **9**,76-86.

Besag, J. (1972). On the Correlation Structure of Some Two-Dimensional Stationary Processes. *Biometrika*, **59**, 43-48.

Besag, J. (1974). Spatial Interaction and the Statistical Analysis of Lattice Systems (with discussion). *J. Royal Statist. Soc.* **B**, **36**, 192-236.

Besag, J. (1975). Statistical Analysis of Non-Lattice Data. *The Statistician*, **24**, 179-195.

Besag, J. E. and Moran, P. A. P. (1975). On the Estimation and Testing of Spatial Interaction in Gaussian Lattice Processes. *Biometrika*, **62**, 555-562.

Besag, J. (1977). Errors-In-Variables-Estimation for Gaussian Lattice Schemes. *J. Royal Statist. Soc.* **B**, **39**, 73-78.

Besag, J. (1977). Efficiency of Pseudolikelihood Estimation for Simple Gaussian Fields. *Biometrika*, **64**, 616-618.

Box, G. E. P. and Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*, Holden-Day, San-Francisco.

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

Clifford, P. and Sudbury, A. (1973). A Model for Spatial Conflict. *Biometrika*, **60**, 581-588.

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

Dahlhaus, R. and Künsch, H. (1987). Edge Effects and Efficient Parameter Estimation for Stationary Random Fields. *Biometrika*, **74**, 877-882.

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series. Nonparametric and Parametric Methods.* Springer, New-York.

Goodman, D. (1977). Some Stability Properties of Two-Dimensional Linear Shift-Invariant Digital Filters. *IEEE Transactions on Circuits and Systems*, **CAS-24**, 201-208.

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.* Springer-Verlag, New York.

Hannan, E. J. (1973). The Asymptotic Theory of Linear Time-Series Models. *J. Appl. Prob.*, **10**, 130-145.

Harvey, A. C. (1989). *Forecasting. Structural Time Series Models and the Kalman Filter.* Cambridge University Press.

Helson, H. and Lowdenslager, D. (1958). Prediction Theory and Fourier Series in Several Variables. *Acta Math.*, **99**, 165-201.

Hjellvik, V. and Tjøstheim, D. (1999). Modelling Panels of Intercorrelated Autoregressive Time Series. *Biometrika*, **86**, 573-590.

Huang, D. and Anh, V.V. (1992). Estimation of Spatial ARMA Models. *Austral. J. Statist.*, **34**, 513-530.

Huang, T. S. (1972). Stability of Two-Dimensional Recursive Filters. *IEEE Transactions on Audio and Electroacoustics*, **AU-20**, 158-163.

Justice, J. H. and Shanks, J. L. (1973). Stability Criterion for N-Dimensional Digital Filters. *IEEE Transactions on Automatic Control*, **AC-18**, 284-286.

Kashyap, R. L. (1984). Characterization and Estimation of Two-Dimensional ARMA Models. *IEEE Transactions on Information Theory*, **IT-30**, 736-745.

Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*, 2nd ed. Springer-Verlag, New York.

Lehmann, E. L. (1999). *Elements of Large-Sample Theory.* Springer-Verlag, New York.

Lutkepøhl, H. (1991), *Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin · Heidelberg.

Ma, C. (2003). Families of Spatio-Temporal Stationary Covariance Models. *Journal of Statistical Planning and Inference*, **116**, 489-501.

Ma, C. (2003). Spatio-Temporal Stationary Covariance Models. *Journal of Multivariate Analysis*, **86**, 97-107.

Martin, R. J. (1979). A Subclass of Lattice Processes Applied to a Problem in Planar Sampling. *Biometrika*, **66**, 209-217.

Moran, P. A. P. (1973). A Gaussian Markovian Process on a Square Lattice. *J. Appl. Prob.*, **10**, 54-62.

Penzer, J. and Shea B. (1997). The exact likelihood of an autoregressive-moving average model with incomplete data. *Biometrika*, **84**, 919-928.

Perera, G. (2001). Random Fields on Z^d. Limit Theorems and Irregular Sets. *Spatial Statistics: Methodological Aspects and Applications*, M. Moore (edit.), p.57-82. Springer-Verlag, New York.

Reinsel, G. C. (1997). *Elements of Multivariate Time Series Analysis*, 2nd ed. Springer-Verlag, New York.

Rosanov, Y. A. (1967). *Stationary Random Process.* San Francisco, Holden Day.

Rosenblatt, M. (2000). *Gaussian and Non-Gaussian Linear Time Series and Random Fields.* Springer-Verlag, New York.

Shanks, J. L., Treitel, S. and Justice, J. H. (1972). Stability and Synthesis of Two-Dimensional Recursive Filters. *IEEE Transactions on Audio and Electroacoustics*, **AU-20**, 115-128.

Shiryayev, A. N. (1984). *Probability.* Springer-Verlag, New York.

Stein, M. L., Chi, Z. and Welty, L. J. (2004). Approximating Likelihoods for Large Spatial Data Sets. *J. Royal Statist. Soc.* **B**, **66**, 275-296.

Strintzis, M. G. (1977). Tests of Stability of Multidimensional Filters. *IEEE Transactions on Circuits and Systems*, **CAS-24**, 432-437.

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

Tjøstheim, D. (1983). Statistical Spatial Series Modelling II: Some Further Results in Unilateral Processes. *Advances in Applied Probability*, **15**, 562-584.

Wald, A. (1949). Note on the Consistency of the Maximum Likelihood Estimate. *The Annals of Mathematical Statistics*, **20**, 595-601.

Warnes, J. J. and Ripley B. D. (1987). Problems with Likelihood Estimation of Covariance Functions of Spatial Gaussian Processes. *Biometrika*, **74**, 640-642.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, **50**, 1-26.

Whittle, P. (1952). Tests of Fit in Time Series. *Biometrika*, **39**, 309-318.

Whittle, P. (1953). The Analysis of Multiple Stationary Time Series. *J. Royal Statist. Soc.* **B**, **15**, 125-139.

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

Whittle, P. (1962). Gaussian Estimation in Stationary Time Series. *Bull. Inst. Internat. Statist.* **39**, 105-129.

Wood, A. T. A. and Chan, G. (1994). Simulation of Stationary Gaussian Processes in $[0,1]^d$. *Journal of Computational and Graphical Statistics*, **3**, 409-432.

Yao, Q. and Brockwell, P. J. (2006). Gaussian Maximum Likelihood Estimation for ARMA models II: Spatial Processes. *Bernoulli*, to appear.

Zhang, W., Yao, Q., Tong, H. and Stenseth N. C. (2003). Smoothing for Spatio-Temporal Models and its Application in Modelling Muskrat-Mink Interaction. *Biometrics*, **59**, 813-821.

Zhu, J., Lahiri S. N. and Cressie N. (2002). Asymptotic Inference for Spatial CDFS over Time. *Statistica Sinica*, **12**, 843-861.