TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

DISSERTATION

VAR Systems: g-Identifiability and Asymptotic Properties of Parameter Estimates for the Mixed-Frequency Case

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines Doktors der technischen Wissenschaften unter der Leitung von

Em.o. Univ.-Prof. Dr.techn. Manfred Deistler
E105
Institut für Stochastik und Wirtschaftsmathematik

eingereicht an der Technischen Universität Wien
bei der Fakultät für Mathematik und Geoinformatik

von

Dipl.-Ing. Lukas Kölbl
e0725946

Wien, am 25. Juli 2015

Danksagung

Ich möchte mich recht herzlich bei meinem Betreuer Prof. Manfred Deistler bedanken, der mich während meiner Arbeit intensiv unterstützte und immer ein offenes Ohr für meine Anliegen hatte. Ebenso gilt mein aufrichtiger Dank den Herren (in alphabetischer Reihenfolge) Prof. Brian D.O. Anderson (ANU, Canberra), Prof. Marco Lippi (Einaudi Institute for Economics and Finance, Rom) und Prof. Wolfgang Scherrer (Forschungsgruppe Ökonometrie und Systemtheorie, TU Wien) für ihre wertvollen fachlichen Anregungen.

Des Weiteren danke ich meinen Forschungsgruppenkollegen Alexander Braumann, Elisabeth Felsenstein, Christoph Flamm, Diego Fresoli, Bernd Funovits und Andreas Graef für die Ratschläge und zahlreichen interessanten Diskussionen in den vergangenen drei Jahren.

Ein großer Dank gilt meiner Familie für die Unterstützung während meines gesamten Studiums. Zu guter Letzt möchte ich mich bei meiner Frau Sabrina bedanken, die mich davor bewahrt hat Versicherungsvertreter (!?!) zu werden und mich so unterstützte, dass ich mich voll und ganz auf mein Studium konzentrieren konnte.

Support by the FWF (Austrian Science Fund under contracts L630/N18 and P24198/N18) and the ARC (Australian Research Council under Discovery Project Grant DP1092571) is gratefully acknowledged.

Lukas Kölbl

Wien, am 25. Juli 2015

Deutsche Kurzfassung

Diese Dissertation beschäftigt sich mit mehrdimensionalen autoregressiven Systemen im Fall von Beobachtungen mit unterschiedlichen Abtastraten, sogenannten mixed-frequency (MF) Beobachtungen. In dieser Arbeit wird ausschließlich der Fall behandelt, bei der die Outputvariable in eine schnelle (high-frequency) und eine langsame (low-frequency) Komponente separiert werden kann. Ein Beispiel für so ein Beobachtungsschema wäre eine zweidimensionale Zeitreihe bei welcher die erste Komponente monatlich, z.B. Arbeitslosigkeit, und die zweite quartalsweise, z.B. Bruttoinlandsprodukt, beobachtet wird. Dem zugrundeliegenden System wird dabei unterstellt, dass der gesamte Output zu jedem schnellen Zeitpunkt generiert, jedoch der Output der langsamen Komponente nicht zu jedem Zeitpunkt beobachtet wird. Hierbei sei erwähnt, dass die mixed-frequency Annahme ein regelmäßiges Beobachtungsschema voraussetzt.

Dieses spezielle Beobachtungsschema spielt bei hochdimensionalen Zeitreihen eine immer wichtigere Rolle, wo die Verfügbarkeit von univariaten Zeitreihen zu unterschiedlichen Abtastraten gegeben ist. Als beliebter Modellierungsansatz haben sich die verallgemeinerten linearen dynamischen Faktormodelle etabliert, bei denen sich die statischen Faktoren typischerweise mit einem singulären autoregressiven System modellieren lassen.

Ein zentraler Teil der Arbeit befasst sich mit der Identifizierbarkeit und dem asymptotischen Verhalten von Parameterschätzern der hochfrequenten autoregressiven Systeme gegeben mixed-frequency Beobachtungen. Dabei werden bei der langsamen Komponente zwei Fälle berücksichtigt: die Fluss- und die Bestandsgröße. Wie sich herausstellt sind nicht alle Systeme identifizierbar, jedoch eine „große“ Teilmenge des Parameterraumes. Aufgrund von regelmäßig fehlenden Beobachtungen können gewisse Autokovarianzen nicht beobachtet und somit die klassischen Yule-Walker Gleichungen für die Schätzung nicht verwendet werden. Deswegen werden die erweiterten Yule-Walker (XYW) Gleichungen eingeführt. Diese XYW Gleichungen stellen in gewissem Sinn das mixed-frequency Analogon zu den klassischen Yule-Walker Gleichungen dar.

In weiterer Folge wird der XYW Schätzer und der verallgemeinerten Momentenschätzer (GMM) diskutiert und gezeigt, dass diese unter gewissen Annahmen asymptotisch normalverteilt sind. Hierfür wird eine Verallgemeinerung der Bartlett Formeln für den mixed-frequency Fall benötigt. Des Weiteren wird der Maximum Likelihood (ML) Schätzer eingegangen und die exakte asymptotische Varianz für den speziellen AR(1) Fall hergeleitet. Wie anhand von Beispielen gezeigt wird, ist im Allgemeinen der GMM Schätzer asymptotisch nicht effizient. Des Weiteren werden einige endliche Stichprobeneigenschaften anhand von Simulationen untersucht.

Abstract

This thesis deals with multivariate autoregressive systems in the case of observations with different sampling rates, the so-called mixed-frequency (MF) observations. In this work the case where the output variable can be separated into a fast (high-frequency) and a slow (low-frequency) component will be given attention. An example of such an observation scheme would be a two-dimensional time series in which the first component is observed monthly, for instance unemployment, and the second quarterly, for instance GDP. It is assumed that the underlying system generates the output at each time point, the so-called high-frequency, however, the output of the slow component is only observed at an integer multiple of the high-frequency. It is worth mentioning that the mixed-frequency case assumes a uniform observation pattern.

This special observation pattern plays an important role in high-dimensional time series, where the availability of the univariate time series is given at different sampling rates. A popular approach to model high-dimensional time series are the generalized linear dynamic factor models, where the static factors can be typically modeled by a singular autoregressive system.

A central part of this thesis is concerned with identifiability and with the asymptotic behavior of parameter estimators of the high-frequency autoregressive system given mixed-frequency observations. Here two cases for the slow component are considered: the stock and the flow case. It turns out that not all systems are identifiable, however, a "large" subset of the parameter space is still identifiable. Due to missing observations certain autocovariances cannot be observed and thus the standard Yule-Walker equations for the estimation cannot be used. Therefor the extended Yule-Walker (XYW) equations are introduced. These XYW equations represent, in a certain sense, the mixed-frequency analogue to the standard Yule-Walker equations.

Furthermore, the XYW estimator and the generalized method of moments estimator (GMM) are discussed and it is shown that they are asymptotically normal under certain assumptions. In order to achieve this, a generalization of Bartlett's formula for the mixed-frequency case is required. Also the maximum likelihood (ML) estimator is treated and the exact asymptotic variance for the special AR (1) case is derived. As shown by examples, the GMM estimator is, in general, not asymptotically efficient. In addition, some finite sample properties are investigated through simulations.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

In the last decades the availability of time series, especially of high-dimensional time series, has extremely increased. One of the most popular approaches to model such time series are the generalized linear dynamic factor models (GDFMs), which were introduced in Forni et al. (2000) and Stock and Watson (2002). The main reason why this high-dimensional time series cannot be modeled by a standard AR or ARMA system is that the number of parameters which have to be estimated quadratically increases with the number of output variables. As an example of such high-dimensional time series we mention the Stock and Watson (2008) data set, which contains 144 different time series from industrial production over prize indices to exchange rates. In contrast to the classical time series analysis, in which the cross-sectional dimension is small or only one-dimensional, we are interested in the structure of the relationship between the individual time series. The aim of this modeling approach is to analyze the driving forces of the high-dimensional time series, whose dimension is in general rather small compared to the dimension of observations, and to use that information to forecast in an optimal way. In GDFMs typically the dimension of the static factors is greater than the dimension of the dynamic factors and thus this static factors can be modeled by singular autoregressive models (see Deistler et al. (2010)). It is worth mentioning that most of the standard literature of estimating GDFMs is based on the assumption that all observations are observed at the same time points (single-frequency).

For high-dimensional time series it is very likely that the univariate components are observed at different sampling rates. We call this mixed-frequency observations. For instance, in the data set above there occur time series, which are sampled monthly, for example industrial production, as well as time series which are sampled quarterly, for example exports. One solution to overcome the problem of mixed-frequency observations, which was also used in Stock and Watson (2008), is to obtain "single"-frequency data by aggregating the monthly data to a quarterly. Of course, such aggregation will lead to an information loss.

In this thesis the emphasis is placed on autoregressive systems in the mixed-frequency case, which is in a certain sense a part of modeling GDFMs in the mixed-frequency case. It is assumed that the multivariate time series can be separated into two sub-series: On the one hand in the high-frequency series, which can be observed at any time point, and on the other hand in the low-frequency series, which are only observed at an integer multiple of the high-frequency time points. Furthermore, it is assumed that the underlying autoregressive system generates the outputs at the highest frequency, but the data are only available in the missing observation pattern described above. Throughout the whole

thesis we assume that the autoregressive system of dimension n is stable, i.e. all roots lie outside the unit circle, the lag order p of the autoregressive polynomial is known, the covariance matrix of the innovations has rank $q \leq n$ and that the dimension of the high-frequency series is at least one.

The main focus of this work is on identifiability of the autoregressive systems based on mixed-frequency observations and the associated autocovariances. Due to the mixed-frequency structure not all autocovariances are observable and therefore the standard identifiability results for autoregressive systems are not valid. For this reason, the extended Yule-Walker equations are introduced and with them the identifiability problem will be considered in more detail. If we replace the population second moments in the extended Yule-Walker equations by their sample counterparts, we directly obtain the extended Yule-Walker estimators and in a wider sense the generalized method of moments estimator. Furthermore, we investigate the statistical properties of these estimators and especially their asymptotic behavior. Moreover, we introduce the (Gaussian) maximum likelihood estimator and the EM algorithm for state space models, which is based on the Gaussian likelihood. All of these estimation methods have the disadvantage that they might lead to unstable estimates of the system parameters and do not fulfill the desired rank condition of the innovation covariance matrix. Thus, in a second step, we show how to project these estimated parameters back on the parameter space.

It is worth mentioning that there exists an enormous number of publications which deals with mixed-frequency observations, see e.g. Harvey and Pierse (1984); Kohn and Ansley (1986); Zadrozny (1988); Chen and Zadrozny (1998); Mariano and Murasawa (2003); Ghysels et al. (2004, 2007); Kuzin et al. (2011). A good overview of further mixed-frequency work is given in Wohlrabe (2008). In particular, we have to emphasize the work of Chen and Zadrozny (1998), which is used as a starting point. Another approach which has become very popular in the last years is the Mixed Data Sampling (MIDAS) regression, which is presented in Ghysels et al. (2004). The main idea behind this approach is that the slow component will be projected onto the space spanned by fast lagged components. Of course, this approach is fundamentally different to the approach used in this thesis and thus a comparison between these two can only be made by comparing the one-step-ahead prediction error covariance matrix.

The thesis is organized as follows: In Chapter 2 stationary processes where the emphasis lies on autoregressive systems in the single-frequency case are introduced. Two different estimation procedures, which are very important, are discussed in detail, namely the Yule-Walker and (Gaussian) maximum likelihood estimation. Furthermore, the asymptotic behavior of these estimators is investigated. Chapter 3 deals with identifiability of system and noise parameters of autoregressive systems in the mixed-frequency case. For the AR(1) case, which is investigated separately, the non-identifiable set can be described explicitly. In addition, for the general AR(p) case, the extended Yule-Walker equations are introduced as well as a so-called genericity property. Both cases where the slow components are observed as stock and as flow variables are discussed. In Chapter 4 four different estimation

procedures for the mixed-frequency case are presented. In Chapter 5 asymptotic properties of the autocovariance estimators, the extended Yule-Walker estimator and the generalized method of moments estimator are derived. Chapter 6 deals with the problem of projecting the estimated system and noise parameters back on the parameter space. In Chapter 7 some finite sample properties are investigated through simulations. In Appendix A the maximum likelihood estimator for the AR(1) case is derived. Appendix B contains the proofs of two lemmas from Chapter 5. Appendix C summarizes a few properties of the Kronecker product and derivation rules. Furthermore, tables of the simulations study are displayed.

CHAPTER 2

Vector Autoregressive Systems

Vector autoregressive (VAR) models are probably the most popular class for modeling time series since they are well understood and very easy to deal with. An autoregressive model depends on a finite number of system parameters as well as, for instance in the Gaussian case, on a finite number of noise parameters, which can be estimated via several methods, e.g. (Gaussian) maximum likelihood, Yule-Walker equations or Burg algorithm (see Hannan (1970); Burg (1975); Brockwell and Davis (1987); Reinsel (1993); Anderson (1994); Lütkepohl (2005); Hannan and Deistler (2012)).

In this chapter the focus is on single-frequency VAR models, i.e. VAR models in which all variables are observed at a single sampling rate. The case of mixed-frequency observations is treated in Chapter 3 to Chapter 5. Furthermore, we do not consider the case where missing data can occur.

We split this chapter into two parts. The first part introduces basic terminology of stationary time series and presents two popular estimators for autoregressive processes, namely the maximum likelihood estimator and the Yule-Walker estimator. The second part provides insights into convergence concepts and their corresponding properties as well as the asymptotic behavior of the mean, the autocovariance, the Yule-Walker and the maximum likelihood estimators.

2.1. Introduction to Autoregressive Systems

In this section well known concepts about multivariate (weak) stationary time series and autoregressive processes are introduced. Whereas most of the standard literature (see Rozanov (1967); Hannan (1970); Anderson (1994); Lütkepohl (2005); Hannan and Deistler (2012)) only considers the case where the autoregressive model is regular, i.e. the covariance of the innovation matrix is non-singular, we include in our analysis the case where the autoregressive model is singular. Thereby we introduce the density of a multivariate normal distributed variable where the covariance matrix is singular, which was, as far as we know, first considered in Khatri (1961). Furthermore, we represent the autoregressive model in state space form so that system theoretical approaches can be applied (see Kailath (1980)). Finally, we deal with the Yule-Walker equations for the regular and singular case. For further discussion, the reader is referred to Hannan (1970); Chen et al. (2011); Filler (2010).

Let $(\Omega, A, \mathbb{P})$ be a probability space and consider the stochastic process $(y_t)_{t\in\mathbb{Z}}$, which is a sequence of random vectors $y_t : \Omega \to \mathbb{R}^n$ defined on the same probability space $(\Omega, A, \mathbb{P})$.

DEFINITION 2.1.1. A process $(y_t)_{t\in\mathbb{Z}}$ is said to be (weakly) stationary if

(1) $\mathbb{E}\left(y_t^T y_t\right) < \infty$ for all $t \in \mathbb{Z}$
(2) $\mathbb{E}\left(y_t\right) = \mu$ for all $t \in \mathbb{Z}$

(3) $\gamma(t,s) = \gamma(t+r,s+r)$ for all $t,s,r \in \mathbb{Z}$ where $\gamma(t,s) = \mathbb{E}\left((y_t-\mu)(y_s-\mu)^T\right)$ is the autocovariance function.

For a stationary process the autocovariance function only depends on the difference $s-t$ rather than on (t,s) itself. This is the reason why we write the autocovariance function for a stationary process as $\gamma(t) = \gamma(t,0)$.

DEFINITION 2.1.2. A stochastic process is called strictly stationary if the distribution of

$$y_{t_1},\, y_{t_2},\ldots,y_{t_n}$$

and of

$$y_{t_1+t},\, y_{t_2+t},\ldots,y_{t_n+t}$$

is the same for every finite set of integers $(t_1,t_2,\ldots,t_n)$ and for every $t \in \mathbb{Z}$.

It immediately follows that a strictly stationary process with finite second moments is also (weakly) stationary. Note that the converse of this conclusion is not true. Indeed, a process which is independent identically Cauchy distributed is strictly stationary but not weakly stationary.

Let $L_2(\Omega,A,\mathbb{P})$ be the set of univariate square-integrable random variables defined over $(\Omega,A,\mathbb{P})$. We say that two elements $y_1,y_2 \in L_2(\Omega,A,\mathbb{P})$ are equivalent if $y_1 = y_2$ $a.s.$ Let $\mathcal{L}_2(\Omega,A,\mathbb{P})$ be the space of these equivalent classes and note that $\mathcal{L}_2(\Omega,A,\mathbb{P})$ is a Hilbert space. Further let $H_y(t)$ be the Hilbert space spanned by $\left\{y_{t-i}^j : i \geq 0,\, j = 1,...,n\right\}$ in $\mathcal{L}_2(\Omega,A,\mathbb{P})$. A stationary process is called linearly regular if $\mathbb{E}(y_t) = 0$ and

$$\lim_{s\to\infty} \mathbb{E}\left(y_{t+s|t}^T y_{t+s|t}\right) = 0 \tag{2.1.1}$$

holds, and linearly singular if

$$y_{t+s|t} = y_{t+s} \quad a.s. \tag{2.1.2}$$

holds for one (t,s), $s>0$ and therefore for all (t,s) and where $y_{t+s|t}$ is the best linear least squares predictor of y_{t+s} based on $y_i, i \leq t$. Denote z as the backward shift operator on $\mathbb{Z}$, i.e. $z(y_t)_{t\in\mathbb{Z}} = (y_{t-1})_{t\in\mathbb{Z}}$, as well as a complex variable. The next theorem is called Wold Decomposition (see Wold (1938); Rozanov (1967); Hannan (1970); Hannan and Deistler (2012)):

THEOREM 2.1.3 (Wold Decomposition Theorem). *Every stationary process $(x_t)_{t\in\mathbb{Z}}$ can be represented in a unique way as*

$$x_t = y_t + z_t \tag{2.1.3}$$

where $y_t, z_t \in H_x(t)$ and

$$\begin{aligned} y_t &\quad \textit{is linearly regular} \\ z_t &\quad \textit{is linearly singular} \\ \mathbb{E}\left(y_t z_s^T\right) &= 0 \quad \forall\, t,s \in \mathbb{Z}. \end{aligned}$$

Every linearly regular process can be represented as

$$y_t = \sum_{j=0}^{\infty} k_j \nu_{t-j} = \underbrace{\sum_{j=0}^{\infty} k_j z^j}_{k(z)} v_t \tag{2.1.4}$$

where $H_y(t) = H_\nu(t)$ *and where* $(\nu_t)_{t\in\mathbb{Z}}$ *is white noise with covariance* Σ_ν, $\sum_{j=0}^{\infty} \|k_j\|^2 < \infty$ *and* $k_0 = I_n$. *Thus the* ν_t *are the innovations of* y_t. *Here* $\|.\|$ *stands for an arbitrary matrix norm.*

For a detailed discussion about linearly singular processes see Hannan and Deistler (2012). Furthermore, the spectral density of a linearly regular process always exists and is given by

$$f_y(\omega) = \frac{1}{2\pi} k\left(e^{-i\omega}\right) \Sigma_\nu k\left(e^{i\omega}\right)^T. \tag{2.1.5}$$

DEFINITION 2.1.4. The process $(y_t)_{t\in\mathbb{Z}}$ of dimension n is said to be an autoregressive process if it is a stationary solution of an AR(p) system

$$y_t = A_1 y_{t-1} + \ldots + A_p y_{t-p} + \nu_t \quad t \in \mathbb{Z} \tag{2.1.6}$$

where p is an integer, $A_i \in \mathbb{R}^{n\times n}$ and $(\nu_t)_{t\in\mathbb{Z}}$ is white noise with zero mean and a covariance $\Sigma_\nu = \mathbb{E}\left(\nu_t \nu_t^T\right)$.

Let $a(z) = I_n - A_1 z - \ldots - A_p z^p$ be the so-called AR polynomial. Throughout this thesis we impose the stability condition

$$\det(a(z)) \neq 0, \quad |z| \leq 1. \tag{2.1.7}$$

Further we restrict ourselves to the stationary and causal solution of (2.1.6). This solution can be expressed as

$$\begin{aligned} y_t &= a(z)^{-1} \nu_t \\ &= \sum_{j=0}^{\infty} k_j \nu_{t-j} \end{aligned} \tag{2.1.8}$$

where $k(z) = a(z)^{-1}$ is the transfer function from $(\nu_t)_{t\in\mathbb{Z}}$ to $(y_t)_{t\in\mathbb{Z}}$. It is easy to see that $(y_t)_{t\in\mathbb{Z}}$ is linearly regular. It is worth noting that (2.1.8) is also the Wold representation of the process (2.1.6). The spectral density of (2.1.8) is given by

$$f_y(\omega) = \frac{1}{2\pi} a^{-1}\left(e^{-i\omega}\right) \Sigma_\nu a^{-1}\left(e^{i\omega}\right)^T. \tag{2.1.9}$$

The parameter space for the high-frequency models considered is:

$$\Theta = \underbrace{\{(A_1, \ldots, A_p) \mid \det(a(z)) \neq 0,\ |z| \leq 1\}}_{S} \times \underbrace{\left\{\Sigma_\nu \mid \Sigma_\nu = \Sigma_\nu^T, \Sigma_\nu \geq 0, \mathrm{rk}(\Sigma_\nu) = q \leq n\right\}}_{D}.$$

We do not allow cross restrictions between the system and noise parameters. Throughout this thesis we will assume that the integers p and q are given. Furthermore, we assume that the rank of Σ_ν is greater than or equal to one.

It is possible to rewrite $(\nu_t)_{t\in\mathbb{Z}}$ as $\nu_t = b\varepsilon_t$ where b is a $n \times q$ matrix with full column rank, $\Sigma_\nu = bb^T$ and $(\varepsilon_t)_{t\in\mathbb{Z}}$ is white noise with covariance $\mathbb{E}\left(\varepsilon_t \varepsilon_t^T\right) = I_q$. Notice that b can only be determined up

to post multiplication by a constant orthogonal matrix from the second moments of the observations. For a particular unique choice of b see Proposition 3.1.2 in Filler (2010). Then equation (2.1.6) can be written as

$$a(z)\, y_t = b\varepsilon_t. \tag{2.1.10}$$

By writing $A > B$ ($A \geq B$), where A and B are two matrices with the same dimension, we mean that $A - B$ is positive (semi)-definite. If $\Sigma_\nu > 0$, then clearly $q = n$ and the AR(p) system (2.1.6) is called regular. On the other hand, when $\operatorname{rk}(\Sigma_\nu) = q < n$ holds, we call the system singular. Singular AR(p) systems are very important for modeling the latent variable in generalized dynamic factor models (GDFMs) (see Forni et al. (2000), Stock and Watson (2002), Doz et al. (2011), Forni et al. (2015)). They occur if the dimension of the minimal static factors is greater than the dimension of the minimal dynamic factors. For further discussion see Deistler et al. (2010).

The system (2.1.6) can be represented in state space form as

$$\underbrace{\begin{pmatrix} y_t \\ \vdots \\ y_{t-p+1} \end{pmatrix}}_{=x_{t+1}} = \underbrace{\begin{pmatrix} A_1 & \cdots & A_{p-1} & A_p \\ I_n & & & 0 \\ & \ddots & & \vdots \\ & & I_n & 0 \end{pmatrix}}_{=\mathcal{A}} \underbrace{\begin{pmatrix} y_{t-1} \\ \vdots \\ y_{t-p} \end{pmatrix}}_{=x_t} + \underbrace{\begin{pmatrix} b \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{=\mathcal{B}} \varepsilon_t \tag{2.1.11}$$

$$y_t = \underbrace{(A_1 \cdots A_p)}_{\mathcal{C}} x_t + b\varepsilon_t, \tag{2.1.12}$$

where (2.1.11) is the state equation and (2.1.12) the observation equation. In this context the matrix $\mathcal{A}$ is called the companion form of the system parameters and x_t is called the state. We say that $\mathcal{A}$ is stable if all its eigenvalues are inside the unit circle. Note also that if $a(z)$ is stable, then $\mathcal{A}$ is stable as well, and vice versa, since the eigenvalues of $\mathcal{A}$ are the roots of $\det\left(z^p a\left(z^{-1}\right)\right)$ (see Hannan and Deistler (2012), p. 19). The system (2.1.11) and (2.1.12) is called minimal if the dimension of the state x_t is minimal among all realizations of the transfer function (see Hannan and Deistler (2012)).

The solution of (2.1.11), (2.1.12) is of the form

$$y_t = a(z)^{-1}\nu_t = a(z)^{-1} b\varepsilon_t = \left(\mathcal{C}\left(I - \mathcal{A}z\right)^{-1}\mathcal{B}z + b\right)\varepsilon_t, \tag{2.1.13}$$

where $k(z)b = \mathcal{C}\left(I - \mathcal{A}z\right)^{-1}\mathcal{B}z + b$ is the transfer function from $(\varepsilon_t)_{t\in\mathbb{Z}}$ to $(y_t)_{t\in\mathbb{Z}}$. The spectral density can analogously be represented as

$$f_y(\omega) = \frac{1}{2\pi}\left(\mathcal{C}\left(I - \mathcal{A}e^{-i\omega}\right)^{-1}\mathcal{B}e^{-i\omega} + b\right)\left(\mathcal{C}\left(I - \mathcal{A}e^{i\omega}\right)^{-1}\mathcal{B}e^{i\omega} + b\right)^T. \tag{2.1.14}$$

The Lyapunov equation for the system (2.1.11) is

$$\Gamma_p - \mathcal{A}\Gamma_p\mathcal{A}^T = \mathcal{B}\mathcal{B}^T, \tag{2.1.15}$$

where $\Gamma_p = \mathbb{E}\left(x_t x_t^T\right)$. The system (2.1.15) has a unique solution for given $(\mathcal{A}, \mathcal{B})$ under the stability assumption (2.1.7) and the covariance of (x_t) can be expressed as

$$\Gamma_p = \sum_{j=0}^{\infty} \mathcal{A}^j \mathcal{B}\mathcal{B}^T \left(\mathcal{A}^T\right)^j. \tag{2.1.16}$$

DEFINITION 2.1.5. The pair $(\mathcal{C}, \mathcal{A})$ where $\mathcal{A} \in \mathbb{R}^{m\times m}$ and $\mathcal{C} \in \mathbb{R}^{n\times m}$ is called observable if

$$\operatorname{rk} \begin{pmatrix} \mathcal{C} \\ \mathcal{C}\mathcal{A} \\ \vdots \\ \mathcal{C}\mathcal{A}^{m-1} \end{pmatrix} = m. \tag{2.1.17}$$

The pair $(\mathcal{A}, \mathcal{B})$ where $\mathcal{A} \in \mathbb{R}^{m\times m}$ and $\mathcal{B} \in \mathbb{R}^{m\times q}$ is called controllable if

$$\operatorname{rk}\left(\mathcal{B}, \mathcal{A}\mathcal{B}, \ldots, \mathcal{A}^{m-1}\mathcal{B}\right) = m.$$

It can be shown that a system is minimal if and only if it is observable and controllable (see Hannan and Deistler (2012), p. 48).

The system (2.1.11) and (2.1.12) is controllable if and only if $\Gamma_p > 0$ since

$$\Gamma_p = \left(\mathcal{B}, \mathcal{A}\mathcal{B}, \ldots, \mathcal{A}^{np-1}\mathcal{B}, \ldots\right) \begin{pmatrix} \mathcal{B}^T \\ \mathcal{B}^T\mathcal{A}^T \\ \vdots \\ \mathcal{B}^T\left(\mathcal{A}^{np-1}\right)^T \\ \vdots \end{pmatrix}. \tag{2.1.18}$$

The next theorem, called Popov-Belevitch-Hautus (PBH) test, gives an easy condition to determine if a system is observable or controllable (see Kailath (1980), p. 135).

THEOREM 2.1.6 (PBH Test). *Let $\mathcal{A} \in \mathbb{R}^{m\times m}$, $\mathcal{B} \in \mathbb{R}^{m\times q}$ and $\mathcal{C} \in \mathbb{R}^{n\times m}$. Then the pair $(\mathcal{C}, \mathcal{A})$ is observable if and only if the matrix*

$$\begin{pmatrix} \mathcal{A} - I_m\lambda \\ \mathcal{C} \end{pmatrix} \tag{2.1.19}$$

has full column rank for all $\lambda \in \mathbb{C}$ and the pair $(\mathcal{A}, \mathcal{B})$ is controllable if and only if the matrix

$$\begin{pmatrix} \mathcal{A} - I_m\lambda, & \mathcal{B} \end{pmatrix} \tag{2.1.20}$$

has full row rank for all $\lambda \in \mathbb{C}$.

The following lemma gives us insights into the structure of the eigenvectors of the companion form:

LEMMA 2.1.7. *Let $\mathcal{A} \in \mathbb{R}^{np\times np}$ be the companion form of the polynomial $a(z)$. Let*

$$p_i = \left(\left(p_i^1\right)^T, \ldots, \left(p_i^p\right)^T\right)^T$$

where $p_i^j \in \mathbb{C}^n$ and p_i be a right eigenvector of the matrix $\mathcal{A}$ with the corresponding eigenvalue $\lambda_i \neq 0$. Then $p_i^1 \neq 0$, $p_i^j = p_i^1 \lambda_i^{-j+1}$ for $j = 2, \ldots, p$ and p_i^1 lies in the right kernel of $a\left(\lambda_i^{-1}\right)$.

PROOF. see Anderson et al. (2012), Lemma 2. □

Considering the system (2.1.11) and (2.1.12), it is easy to conclude that this system is observable if and only if A_p is nonsingular. Indeed if A_p is singular the system is not observable since every vector which lies in the kernel of $\mathcal{A}$ is also orthogonal to the rows of $\mathcal{C}$ and thus the PBH test fails. On the other hand, if A_p is nonsingular, Lemma 2.1.7 guarantees that the first n entries of an eigenvector of $\mathcal{A}$ are not all equal to zero. Thus there exists no nontrivial vector which lies in the kernel of the matrix (2.1.19). Therefore we have proved the following lemma:

LEMMA 2.1.8. *The system (2.1.11) and (2.1.12) is minimal if and only if* $\Gamma_p > 0$ *and* A_p *is nonsingular.*

The next lemma gives us a relation between the positive definiteness property of Γ_p and the parameters of the autoregressive system.

LEMMA 2.1.9. *The covariance matrix* Γ_p *is positive definite if and only if* (A_p, b) *is of full row rank and* $(a(z), b)$ *is left co-prime, i.e. if the only left divisors of* $(a(z), b)$ *are unimodular ones.*

PROOF. see Felsenstein (2014), Lemma 1.1.6. □

Thus the autocovariance matrix Γ_p is always nonsingular for a regular autoregressive system.

2.1.1. Forecasting and Interpolation. If we project y_t onto the space $H_y(t-1)$, it is easy to see that the best linear predictor is given by

$$y_{t|t-1} = P_{H_y(t-1)}(y_t) = A_1 y_{t-1} + \cdots + A_p y_{t-p}, \tag{2.1.21}$$

since the components of $\left(y_t - y_{t|t-1}\right)$ are uncorrelated with the components of y_{t-i}, $i > 0$, i.e.

$$\mathbb{E}\left(\left(y_t - y_{t|t-1}\right) y_{t-i}^T\right) = 0,\ i > 0.$$

The next theorem, which considers the problem of interpolation, is a generalization of Pourahmadi (1988) to the multivariate case. For further discussion about interpolation see Grenander and Rosenblatt (1957); Friedman (1962); Rozanov (1967); Brubacher and Tunnicliffe Wilson (1976); Gomez et al. (1999). For the purpose of simplicity we assume that Σ_ν is nonsingular, which implies that $\sum_{k=0}^{p} A_k^T \Sigma_\nu^{-1} A_k$, where $A_0 = -I_n$, is also nonsingular.

THEOREM 2.1.10. *Assume that* Σ_ν *is nonsingular and let* $\rho_j = \sum_{k=0}^{p-j} A_k^T \Sigma_\nu^{-1} A_{k+j}$ *for* $j = 0, \ldots, p$ *where* $A_0 = -I_n$ *and* $H_y^o(t)$ *be the Hilbert space spanned by*

$$\left\{y_{t-i}^j : i > 0,\ j = 1, \ldots, n\right\} \text{ and } \left\{y_{t+i}^j : i > 0,\ j = 1, \ldots, n\right\}.$$

Then the projection of y_t *on the space* $H_y^o(t)$ *is given by*

$$P_{H_y^o(t)}(y_t) = \rho_0^{-1}\left(-\sum_{i=1}^{p} \rho_i y_{t-i} - \sum_{i=1}^{p} \rho_i^T y_{t+i}\right). \tag{2.1.22}$$

Furthermore, the interpolation error covariance matrix is given by ρ_0^{-1}.

PROOF. In a first step we want to split the Hilbert space $H_y^o(t)$ into the sum of three orthogonal subspaces. Let $H_y^+(t+1)$ be the Hilbert space spanned by $\left\{y_{t+1+i}^j : i \geq 0,\, j = 1,\dots,n\right\}$ and observe that $y_{t+p+1|t+p} = A_1 y_{t+p} + \cdots + A_p y_{t+1}$. Thus y_{t+p+1} can be represented as a linear combination of $\{y_{t+p},\dots,y_{t+1},\nu_{t+p+1}\}$. Also observe that

$$\begin{aligned} y_{t+p+2|t+p+1} &= A_1 y_{t+p+1} + \cdots + A_p y_{t+2} \\ &= A_1 y_{t+p+1|t+p} + \cdots + A_p y_{t+2} + k_1 \nu_{t+p+1} \end{aligned} \tag{2.1.23}$$

and thus y_{t+p+2} can be represented as a linear combination of $\{y_{t+p},\dots,y_{t+1},\nu_{t+p+1},\nu_{t+p+2}\}$ and so on. Therefore we can represent $H_y^+(t+1)$ as $\overline{\text{span}}\{y_{t+1},\dots,y_{t+p},\nu_{t+p+1},\nu_{t+p+2},...\}$ where $\bar{A}$ is the closure of A. Let $H_y^p(t+1)$ be the Hilbert space spanned by $\{y_{t+1},\dots,y_{t+p}\}$. By construction we see that $H_y^+(t+1) = H_y^p(t+1) \oplus H_\nu^+(t+p+1)$ where $H_\nu^+(t+p+1)$ is the Hilbert space spanned by $\{\nu_{t+p+1},\nu_{t+p+2},\dots\}$.

Since $y_{t+i} - y_{t+i|t-1} = \sum_{j=0}^{i} k_j \nu_{t+i-j}$ for $i \geq 0$ and ν_{t+k}, $k \geq 0$ is orthogonal to $H_y(t-1)$ we can conclude that $H_y^o(t) = H_y(t-1) \oplus H_{t-1}^p(t+1) \oplus H_\nu^+(t+p+1)$ where $H_{t-1}^p(t+1)$ is spanned by

$$\left\{ \underbrace{\nu_{t+1} + k_1\nu_t}_{\text{one-step-ahead pred. error}}, \underbrace{\nu_{t+2} + k_1\nu_{t+1} + k_2\nu_t}_{\text{two-step-ahead pred. error}}, \dots \right\} = \left\{ y_{t+1} - y_{t+1|t-1}, y_{t+2} - y_{t+2|t-1}, \dots \right\}.$$

Now it follows that

$$\begin{aligned} P_{H_y^o(t)}(y_t) &= P_{H_y(t-1)}(y_t) + P_{H_{t-1}^p(t+1)}(y_t) + \underbrace{P_{H_\nu^+(t+p+1)}(y_t)}_{=0} \\ &= A_1 y_{t-1} + \cdots + A_p y_{t-p} + P_{H_{t-1}^p(t+1)}(y_t). \end{aligned} \tag{2.1.24}$$

We can conclude that projecting on the space spanned by $\{y_{t-i}, |i| = 1,\dots,p\}$ is the same as projecting on $H_y^o(t)$. Finally, we have to show that (2.1.22) is indeed the best linear interpolation, i.e. $\mathbb{E}\left(\left(y_t - P_{H_y^o(t)}(y_t)\right) y_{t-i}^T\right) = 0$ for $|i| = 1,\dots,p$. For this purpose we consider the linear system of equations which has to be fulfilled:

$$(\gamma(1),\dots,\gamma(p),\gamma(-1),\dots,\gamma(-p)) = (H_1,\dots,H_{2p}) \begin{pmatrix} \gamma(0) & \cdots & \gamma(p-1) & \gamma(-2) & \cdots & \gamma(-p-1) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \gamma(-p+1) & \cdots & \gamma(0) & \gamma(-p-1) & \cdots & \gamma(-2p) \\ \gamma(2) & \cdots & \gamma(p+1) & \gamma(0) & \cdots & \gamma(-p+1) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \gamma(p+1) & \cdots & \gamma(2p) & \gamma(p-1) & \cdots & \gamma(0) \end{pmatrix}.$$

Inserting the coefficient of (2.1.22) into the above system of equations we obtain for $i = 1, \ldots, p$,

$$\begin{aligned}
\rho_0\gamma(i) &\stackrel{!}{=} -\sum_{j=1}^{p}\rho_j\gamma(i-j) - \sum_{j=1}^{p}\rho_j^T\gamma(i+j) \\
&= -\sum_{j=1}^{p}\sum_{k=0}^{p-j}A_k^T\Sigma_\nu^{-1}A_{k+j}\gamma(i-j) - \sum_{j=1}^{p}\sum_{k=0}^{p-j}A_{k+j}^T\Sigma_\nu^{-1}A_k\gamma(i+j) \\
&= -\sum_{j=1}^{p}\sum_{k=0}^{p-j}A_k^T\Sigma_\nu^{-1}A_{k+j}\gamma(i-j) - \sum_{j=1}^{p}\sum_{k=j}^{p}A_k^T\Sigma_\nu^{-1}A_{k-j}\gamma(i+j).
\end{aligned}$$

Using the fact that $\sum_{j=1}^{p-k}A_{k+j}\gamma(i-j) + \sum_{j=1}^{k}A_{k-j}\gamma(i+j) = -A_k\gamma(i)$ for $i > 0$, we can rewrite the above term as follows

$$\begin{aligned}
\rho_0\gamma(i) &\stackrel{!}{=} -\sum_{k=0}^{p-1}\sum_{j=1}^{p-k}A_k^T\Sigma_\nu^{-1}A_{k+j}\gamma(i-j) - \sum_{k=1}^{p}\sum_{j=1}^{k}A_k^T\Sigma_\nu^{-1}A_{k-j}\gamma(i+j) \\
&= -\sum_{k=1}^{p-1}A_k^T\Sigma_\nu^{-1}\left(\sum_{j=1}^{p-k}A_{k+j}\gamma(i-j) + \sum_{j=1}^{k}A_{k-j}\gamma(i+j)\right) \\
&\quad -\sum_{j=1}^{p}A_0^T\Sigma_\nu^{-1}A_j\gamma(i-j) - \sum_{j=1}^{p}A_p^T\Sigma_\nu^{-1}A_{p-j}\gamma(i+j) \\
&= \sum_{k=1}^{p-1}A_k^T\Sigma_\nu^{-1}A_k\gamma(i) + \Sigma_\nu^{-1}\gamma(i) + A_p^T\Sigma_\nu^{-1}A_p\gamma(i) \\
&= \rho_0\gamma(i).
\end{aligned}$$

In an analogous way one can prove the case $i = -1, \ldots, -p$. The interpolation error covariance matrix can be rewritten into $\mathbb{V}\left(y_t - P_{H_y^o(t)}(y_t)\right) = \rho_0^{-1}\mathbb{V}\left(\rho_0 y_t + \sum_{i=1}^{p}\rho_i y_{t-i} + \sum_{i=1}^{p}\rho_i^T y_{t+i}\right)\rho_0^{-1}$. Furthermore, it is straightforward to show that $\rho_0 y_t + \sum_{i=1}^{p}\rho_i y_{t-i} + \sum_{i=1}^{p}\rho_i^T y_{t+i} = -\sum_{k=0}^{p}A_k^T\Sigma_\nu^{-1}\nu_{t+k}$. Thus $\mathbb{V}\left(\rho_0 y_t + \sum_{i=1}^{p}\rho_i y_{t-i} + \sum_{i=1}^{p}\rho_i^T y_{t+i}\right) = \rho_0$ and therefore $\mathbb{V}\left(y_t - P_{H_y^o(t)}(y_t)\right) = \rho_0^{-1}$ which completes our proof. □

Note that if we define the process $y_t^D = a^T(z)\eta_t$ where $(\eta_t)_{t\in\mathbb{Z}}$ is white noise with covariance Σ_ν^{-1} and let $\gamma^D(i)$ be the corresponding autocovariance function of lag i, we see that $\gamma^D(i) = \rho_i^T$.

Example 2.1.11. Consider the AR(1) case where Σ_ν is nonsingular. Then the best linear interpolation is given by

$$P_{H_y^o(t)}(y_t) = \left(\Sigma_\nu^{-1} + A_1^T\Sigma_\nu^{-1}A_1\right)^{-1}\left(\Sigma_\nu^{-1}A_1 y_{t-1} + A_1^T\Sigma_\nu^{-1}y_{t+1}\right)$$

and the interpolation error is

$$y_t - P_{H_y^o(t)}(y_t) = \left(\Sigma_\nu^{-1} + A_1^T\Sigma_\nu^{-1}A_1\right)^{-1}\left(\Sigma_\nu^{-1}\nu_t - A_1^T\Sigma_\nu^{-1}\nu_{t+1}\right).$$

Thus the covariance matrix of the interpolation error is

$$\mathbb{V}\left(y_t - P_{H_y^o(t)}(y_t)\right) = \left(\Sigma_\nu^{-1} + A_1^T\Sigma_\nu^{-1}A_1\right)^{-1} = \rho_0^{-1}.$$

PROPOSITION 2.1.12. *Let* $y_t = \begin{pmatrix} y_t^1 \\ y_t^2 \end{pmatrix}$, $y_t^1 = (I_{n_1}, 0)\, y_t$, $y_t^2 = (0, I_{n_2})\, y_t$, $n_1 + n_2 = n$ *and assume that* Σ_ν *is nonsingular. In addition, let* $\rho_j = \sum_{k=0}^{p-j} A_k^T \Sigma_\nu^{-1} A_{k+j}$ *for* $j = 0, \ldots, p$ *where* $A_0 = -I_n$ *and* $H_y^{o,1}(t)$ *be the Hilbert space spanned by* $\left\{ y_{t-i}^j : i > 0,\, j = 1, \ldots, n \right\}$, $\left\{ y_{t+i}^j : i > 0,\, j = 1, \ldots, n \right\}$ *and* $\left\{ y_t^1 \right\}$. *Then the projection of* y_t^2 *on the space* $H_y^{o,1}(t)$ *is given by*

$$P_{H_y^{o,1}(t)}\left(y_t^2\right) = \left(\rho_0^{22}\right)^{-1} \left(-(0, I_{n_2}) \sum_{i=1}^{p} \rho_i y_{t-i} - (0, I_{n_2}) \sum_{i=1}^{p} \rho_i^T y_{t+i} - (0, I_{n_2})\, \rho_0 \begin{pmatrix} y_t^1 \\ 0 \end{pmatrix} \right) \tag{2.1.25}$$

where $\rho_0^{22} = (0, I_{n_2})\, \rho_0\, (0, I_{n_2})^T$ *is nonsingular. Furthermore, the interpolation error covariance matrix is given by* $\left(\rho_0^{22}\right)^{-1}$.

PROOF. The proof follows the same steps as in Theorem 2.1.10. □

2.1.2. Maximum Likelihood Estimation. Before we derive the (Gaussian) maximum likelihood estimator for system (2.1.6) we introduce the density of a multivariate normally distributed variable for the nonsingular and singular case. Whereas in the nonsingular case the density is uniquely defined except for a Lebesgue null set, the density in the singular case does not exist in the usual way. The reason for this is that the covariance is singular and thus its inverse does not exist.

Let x_0 be univariate normally distributed with mean μ and covariance $0 < \sigma^2 < \infty$. In this thesis we use the notation $x_0 \sim \mathcal{N}_1\left(\mu, \sigma^2\right)$ where the subscript denotes the dimension of the random variable. The following definition for the multivariate case is due to Rao (1972), p. 522. Note that this definition also includes the case where the covariance matrix, say Σ, might be singular.

DEFINITION 2.1.13. A n-dimensional vector x_0 is called multivariate normally distributed with $\mathbb{E}(x_0) = \mu \in \mathbb{R}^n$, $\mathbb{V}(x_0) = \Sigma \in \mathbb{R}^{n \times n}$ and $\operatorname{rk}(\Sigma) = q \leq n$ if it can be expressed as

$$x_0 = \mu + b x_1 \tag{2.1.26}$$

where $b \in \mathbb{R}^{n \times q}$ has rank q, $bb^T = \Sigma$ and x_1 is a q-dimensional vector of univariate independently normally distributed variables with zero mean and variance one.

Note that this definition is equivalent to say that x_0 is normally distributed if for any vector $c \in \mathbb{R}^n$ with $c^T \Sigma c > 0$ the linear combination $c^T x_0$ is normally distributed with mean $c^T \mu$ and variance $c^T \Sigma c$.

It is obvious that the factorization of Σ is not unique so that we can find another $b_1 \neq b$ for which $b_1 b_1^T = \Sigma$ holds. In the case where the covariance matrix is singular x_0 is called singular normally distributed, otherwise it is called regular normally distributed. For a detailed discussion see Rao (1972), p. 516. or Anderson (1994), p. 29.

If $\det(\Sigma) \neq 0$, x_0 has a density

$$f_{x_0}(x) = \left(\frac{1}{2\pi}\right)^{n/2} \det(\Sigma)^{-1/2} \exp\left(-(x - \mu)^T \Sigma^{-1} (x - \mu) \right), \quad x = (x_1, \ldots, x_n)^T. \tag{2.1.27}$$

In this case one can use the linear transformation $y_0 = b^{-1}(x_0 - \mu)$, where $\Sigma = bb^T$, to obtain a standardized normal distribution, $y_0 \sim \mathcal{N}_n(0, I_n)$.

If $\mathrm{rk}(\Sigma) = q < n$, then of course the density in the usual sense does not exist. Nevertheless, the density of x_0 can be constructed as follows (see Khatri (1961), p. 275 or Rao (1972), p. 527): Let $b_{\mathrm{sp}} \in \mathbb{R}^{n\times q}$ be a matrix of orthonormal column vectors belonging to the span of Σ. For instance, this matrix can be found by choosing the q eigenvectors of Σ associated with the q largest (and positive) eigenvalues. Furthermore, let $b_{\mathrm{ke}} \in \mathbb{R}^{n\times(n-q)}$ be a matrix so that $b_{\mathrm{ke}}^T\Sigma = 0$. Now consider the linear transformation

$$\begin{pmatrix} y_{\mathrm{sp}} \\ y_{\mathrm{ke}} \end{pmatrix} = \begin{pmatrix} b_{\mathrm{sp}}^T \\ b_{\mathrm{ke}}^T \end{pmatrix} x_0.$$

It follows that $\mathbb{E}(y_{\mathrm{ke}}) = b_{\mathrm{ke}}^T\mu$ and $\mathbb{V}(y_{\mathrm{ke}}) = 0$ so that $y_{\mathrm{ke}} = b_{\mathrm{ke}}^T\mu$ with probability one. For y_{sp} one can conclude that $\mathbb{E}(y_{\mathrm{sp}}) = b_{\mathrm{sp}}^T\mu$ and $\mathbb{V}(y_{\mathrm{sp}}) = b_{\mathrm{sp}}^T\Sigma b_{\mathrm{sp}}$, so that $y_{\mathrm{sp}} \sim \mathcal{N}_q\left(b_{\mathrm{sp}}^T\mu, b_{\mathrm{sp}}^T\Sigma b_{\mathrm{sp}}\right)$. Thus, the density of y_{sp} can be written as

$$f_{y_{\mathrm{sp}}}(y) = \left(\frac{1}{2\pi}\right)^{q/2} \frac{1}{\sqrt{\lambda_1\cdots\lambda_q}} \exp\left(-\left(y - b_{\mathrm{sp}}^T\mu\right)^T \left(b_{\mathrm{sp}}^T\Sigma b_{\mathrm{sp}}\right)^{-1}\left(y - b_{\mathrm{sp}}^T\mu\right)\right) \tag{2.1.28}$$

where λ_i, $i = 1,\ldots,q$ are the positive eigenvalues of Σ. Note that this choice of the density is not unique.

Now one choice of the density of x_0 would be

$$f_{x_0}(x) = \left(\frac{1}{2\pi}\right)^{q/2} \frac{1}{\sqrt{\lambda_1\cdots\lambda_q}} \exp\left(-(x-\mu)^T b_{\mathrm{sp}} \left(b_{\mathrm{sp}}^T\Sigma b_{\mathrm{sp}}\right)^{-1} b_{\mathrm{sp}}^T(x-\mu)\right)$$

provided $b_{\mathrm{ke}}^T x_0 = b_{\mathrm{ke}}^T\mu$ with probability one.

Let vec(.) denote columnwise vectorization and $\otimes$ the Kronecker product, see Appendix C for more details. First, we derive the likelihood function for a regular autoregressive system: We assume that the ν_t are independently normally distributed where $\Sigma_\nu > 0$ (see Lütkepohl (2005)):

$$v \quad = \quad \mathrm{vec}(\nu_1, \nu_2, \ldots, \nu_T) \sim \mathcal{N}_{Tn}(0, I_T \otimes \Sigma_\nu).$$

Therefore, this variable has the density

$$f_v(v) = \frac{1}{(2\pi)^{nT/2}} |I_T \otimes \Sigma_\nu|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} v^T \left(I_T \otimes \Sigma_\nu^{-1}\right) v\right).$$

Let $Y = (y_1, y_2, \ldots, y_T)$, $X = (x_1, x_2, \ldots, x_T)$ and $A = (A_1, \ldots, A_p)$. Defining $y = \mathrm{vec}(Y)$ and using $v = y - (X^T \otimes I_n)\mathrm{vec}(A)$, we get the density of y as

$$\begin{aligned} f_y(y) \quad &= \quad \underbrace{\left|\frac{\partial v}{\partial y^T}\right|}_{1} f_v(v) = \frac{1}{(2\pi)^{nT/2}} |I_T \otimes \Sigma_\nu|^{-\frac{1}{2}} \\ &\exp\left(-\frac{1}{2}\left(y - (X^T \otimes I_n)\mathrm{vec}(A)\right)^T \left(I_T \otimes \Sigma_\nu^{-1}\right)\left(y - (X^T \otimes I_n)\mathrm{vec}(A)\right)\right). \end{aligned} \tag{2.1.29}$$

To simplify matters we have assumed that the initial values, i.e. $x_1 = \left(y_0^T, \ldots, y_{-p+1}^T\right)^T$, are given fixed numbers. Taking the logarithm of the above formula and taking into account that the last term in (2.1.29) can be rewritten into $\exp\left(\mathrm{tr}\left((Y - AX)^T \Sigma_\nu^{-1}(Y - AX)\right)\right)$ we obtain the likelihood for

the AR(p) case:

$$\ln\left(l\left(A,\Sigma_\nu\right)\right) = c-\frac{T}{2}\ln\left|\Sigma_\nu\right|-\frac{1}{2}\mathrm{tr}\left(\left(Y-AX\right)^T\Sigma_\nu^{-1}\left(Y-AX\right)\right). \tag{2.1.30}$$

If we assume that $\mathrm{rk}\left(\Sigma_\nu\right)=q<n$, we choose the unique factorization of $\Sigma_\nu = bb^T$ in the free parameters, which was proposed in Filler (2010): Let $\Sigma_\nu = O\Lambda O^T = O_1\Lambda_1 O_1^T + O_2\Lambda_2 O_2^T$ where $O=(O_1,O_2)$ are the orthonormal eigenvectors and Λ is a diagonal matrix containing the eigenvalues of Σ_ν and Λ_1 is a diagonal matrix containing the q positive eigenvalues. Now define $b=O_1\Lambda_1^{1/2}Q^T$ where the orthonormal matrix Q comes from the LQ decomposition of the first nonsingular submatrix of $O_1\Lambda_1^{1/2}$ (we assume that the diagonal elements of the L matrix are all positive, which guarantees uniqueness). For notational convenience we assume that the first $q\times q$ submatrix is nonsingular. Now b has the structure

$$b=\begin{pmatrix}b_1\\ b_2\end{pmatrix}\in\mathbb{R}^{n\times q}$$

where $b_1\in\mathbb{R}^{q\times q}$ is a lower triangular matrix and $b_2\in\mathbb{R}^{(n-q)\times q}$. It is easy to see that $\det\left(b^Tb\right)=\det\left(Q\Lambda_1Q^T\right)=\det\left(\Lambda_1\right)$ and $\Sigma^\dagger=\left(bb^T\right)^\dagger=b\left(b^Tb\right)^{-2}b^T$. Thus, the likelihood function for the singular case can be represented as

$$\begin{aligned}\ln\left(l\left(A,\Sigma_\nu\right)\right) &= c-\frac{T}{2}\ln\left|\Lambda_1\right|-\frac{1}{2}\mathrm{tr}\left(\left(Y-AX\right)^T\Sigma_\nu^\dagger\left(Y-AX\right)\right)\\ &= c-\frac{T}{2}\ln\left|b^Tb\right|-\frac{1}{2}\mathrm{tr}\left(\left(Y-AX\right)^Tb\left(b^Tb\right)^{-2}b^T\left(Y-AX\right)\right)\end{aligned} \tag{2.1.31}$$

THEOREM 2.1.14. *Let $\Gamma_p>0$. Then the maximum likelihood estimators for system (2.1.6) in the nonsingular and singular case are*

$$\begin{aligned}\hat{A}_{ML} &= YX^T\left(XX^T\right)^{-1} \qquad (2.1.32)\\ \hat{\Sigma}_{ML} &= \frac{1}{T}\left(Y-\hat{A}_{ML}X\right)\left(Y-\hat{A}_{ML}X\right)^T.\end{aligned}$$

PROOF. The proof for the nonsingular case, $n=q$, is given in Lütkepohl (2005) p. 90 and for the singular case see Srivastava and von Rosen (2002). □

REMARK 2.1.15. In the singular case $\hat{\Sigma}_{\mathrm{ML}}$ has the property that $b_{\mathrm{ke}}^T\hat{\Sigma}_{\mathrm{ML}}=0$ since

$$\begin{aligned}b_{\mathrm{ke}}^T\hat{\Sigma}_{\mathrm{ML}} &= \frac{1}{T}b_{\mathrm{ke}}^TY\left(I-X^T\left(XX^T\right)^{-1}X\right)Y^T\\ &= \frac{1}{T}b_{\mathrm{ke}}^TV\left(I-X^T\left(XX^T\right)^{-1}X\right)V^T=0\end{aligned} \tag{2.1.33}$$

where $V=(\nu_1,\nu_2,\ldots,\nu_T)$. Note that we assumed that the initial values are given fixed numbers, which is a simplifying assumption. If, in particular, we replace these initial values by zeros, then it follows that $b_{\mathrm{ke}}^T\hat{\Sigma}_{\mathrm{ML}}\neq 0$ and thus we have to project $\hat{\Sigma}_{\mathrm{ML}}$ back on D, see Chapter 6.

2.1.3. Yule-Walker Estimation. One of the most important estimation procedures for the parameters of an AR(p) process is based on the so-called Yule-Walker equations. If we postmultiply

equation (2.1.6) by y_{t-i}, $i = 1, \ldots, p$ and take the expectation, we obtain

$$\underbrace{(\gamma(1), \ldots, \gamma(p))}_{\gamma_1} = (A_1, \ldots, A_p)\,\Gamma_p. \tag{2.1.34}$$

Analogously, if we postmultiply equation (2.1.6) by y_t and take the expectation we obtain

$$\Sigma_\nu = \gamma(0) - (A_1, \ldots, A_p)\,(\gamma(1), \cdots, \gamma(p))^T \tag{2.1.35}$$

which are the Yule-Walker equations in the single-frequency case.

DEFINITION 2.1.16. The autocovariance estimator for the lag $k \geq 0$ where we include the mean correction is defined as

$$\begin{aligned} \hat{\gamma}(k) &= \frac{1}{T}\sum_{t=k+1}^{T}(y_t - \bar{y}_T)(y_{t-k} - \bar{y}_T)^T \\ \hat{\gamma}(-k) &= \hat{\gamma}(k)^T \end{aligned} \tag{2.1.36}$$

where $\bar{y}_T$ is the mean estimator. Let $\hat{\gamma}_1$ be the estimator for γ_1 and $\hat{\Gamma}_p$ be the estimator for Γ_p. If we assume that Γ_p is nonsingular, which is always the case for nonsingular Σ_ν and is fulfilled in an open and dense subset of the parameter space Θ for a singular Σ_ν (see Lemma 3.1.8), the estimator $\hat{\Gamma}_p$ is nonsingular from a certain T_0 onwards, too. Thus, we can define the Yule-Walker estimators for the system and noise parameters for the nonsingular and singular case as

$$\hat{A}_{\mathrm{YW}} = \hat{\gamma}_1\hat{\Gamma}_p^{-1} \tag{2.1.37}$$

$$\hat{\Sigma}_{\mathrm{YW}} = \hat{\gamma}(0) - \hat{\gamma}_1\hat{\Gamma}_p^{-1}\hat{\gamma}_1^T. \tag{2.1.38}$$

For the case where Γ_p is singular, see Deistler et al. (2011); Chen et al. (2011).

Whereas the maximum likelihood estimator for the noise covariance matrix fulfills the desired rank condition, provided that the initial values are given fixed numbers, the Yule-Walker estimator does not have this property. Nevertheless, the Yule-Walker estimator for the system parameters always leads to a stable system, provided that $\hat{\Gamma}_p > 0$ holds (see Deistler et al. (2010)), which is not the case for the maximum likelihood estimator. If the condition number of Γ_p is poor, i.e. $\kappa(A) = \|A\| \left\|A^{-1}\right\|$ is large, one should use the Burg estimator instead of the Yule-Walker estimator since a small bias on the autocovariances may lead to a complete different model, see De Hoon et al. (1996).

2.2. Asymptotic Behavior in the Single-Frequency Case

2.2.1. Convergence Concepts and their Properties. In this subsection we will introduce some definitions and repeat well known concepts about the convergence of random variables. Throughout this section let $(x_t)_{t\in\mathbb{N}}$ be a sequence of multivariate random variables of dimension n which are all defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

DEFINITION 2.2.1. A random sequence of vectors $(x_t)_{t\in\mathbb{N}}$ with distribution function F_t is said to converge in distribution to a vector x_0 with distribution function F_0 if

$$\lim_{t\to\infty} F_t(x) = \lim_{t\to\infty} \mathbb{P}(x_t \leq x) = \mathbb{P}(x_0 \leq x) = F_0(x) \tag{2.2.1}$$

for every continuity point x of F_0.

To simplify matters we will denote the convergence in distribution with $\xrightarrow{d}$, e.g. $x_t \xrightarrow{d} x_0$. The limit x_0 is in general non-unique. Let $\|x\| : \mathbb{R}^n \to \mathbb{R}$ be a norm, e.g. the Euclidean norm.

DEFINITION 2.2.2. A random sequence of vectors $(x_t)_{t\in\mathbb{N}}$ is said to converge in probability to x_0 if for every $\epsilon > 0$

$$\lim_{t\to\infty} \mathbb{P}\left(\|x_t - x_0\| > \epsilon\right) = 0. \tag{2.2.2}$$

This convergence is denoted by $x_t \xrightarrow{p} x_0$.

DEFINITION 2.2.3. A random sequence of vectors $(x_t)_{t\in\mathbb{N}}$ is said to converge almost surely to x_0 if

$$\mathbb{P}\left(\lim_{t\to\infty} \|x_t - x_0\| = 0\right) = 1. \tag{2.2.3}$$

This convergence is denoted by $x_t \xrightarrow{a.s.} x_0$.

DEFINITION 2.2.4. A random sequence of vectors $(x_t)_{t\in\mathbb{N}}$ is said to converge in the mean square sense to x_0 if $\mathbb{E}\left(x_0^T x_0\right) < \infty$ and

$$\lim_{t\to\infty} \mathbb{E}\left(\left(x_t - x_0\right)^T \left(x_t - x_0\right)\right) = 0. \tag{2.2.4}$$

This convergence is denoted by $\operatorname*{l.i.m.}_{t\to\infty} x_t = x_0$.

The next two lemmas give relations between the different concepts of convergence and their properties with respect to continuous functions. Further, we will introduce Slutsky's Lemma, which is a consequence of these lemmas.

LEMMA 2.2.5 (Continuous Mapping Theorem). *Let $(x_t)_{t\in\mathbb{N}}$ and x_0 be random vectors and let $f : \mathbb{R}^n \to \mathbb{R}^k$ be a function which is continuous at every point of a set C with the property that $\mathbb{P}(x_0 \in C) = 1$.*

(1) *If $x_t \xrightarrow{d} x_0$, then $f(x_t) \xrightarrow{d} f(x_0)$.*
(2) *If $x_t \xrightarrow{p} x_0$, then $f(x_t) \xrightarrow{p} f(x_0)$.*
(3) *If $x_t \xrightarrow{a.s.} x_0$, then $f(x_t) \xrightarrow{a.s.} f(x_0)$.*

PROOF. see van der Vaart (2000), Theorem 2.3.

□

LEMMA 2.2.6. *Let $(x_t)_{t\in\mathbb{N}}$, x_0, $(y_t)_{t\in\mathbb{N}}$ and y_0 be random vectors and c a constant vector of the same dimension.*

(1) *If $\underset{t\to\infty}{l.i.m.}\, x_t = x_0$ then $x_t \xrightarrow{p} x_0$.*
(2) *If $x_t \xrightarrow{a.s.} x_0$ then $x_t \xrightarrow{p} x_0$.*
(3) *If $x_t \xrightarrow{p} x_0$ then $x_t \xrightarrow{d} x_0$.*
(4) *$x_t \xrightarrow{p} c$ if and only if $x_t \xrightarrow{d} c$.*
(5) *If $x_t \xrightarrow{d} x_0$ and $x_t - y_t \xrightarrow{p} 0$, then $y_t \xrightarrow{d} x_0$.*
(6) *If $x_t \xrightarrow{d} x_0$ and $y_t \xrightarrow{p} c$, then $\left(x_t^T, y_t^T\right)^T \xrightarrow{d} \left(x_0^T, c^T\right)^T$.*
(7) *If $x_t \xrightarrow{p} x_0$ and $y_t \xrightarrow{p} y_0$, then $\left(x_t^T, y_t^T\right)^T \xrightarrow{p} \left(x_0^T, y_0^T\right)^T$.*

PROOF. 1. follows immediately from Chebyshev's inequality and for 2.-7. see van der Vaart (2000), Theorem 2.7.

□

The following lemma, which is called Slutsky's Lemma, is a special case of Lemma 2.2.6 and will provide us a useful tool for calculating the asymptotic covariance of specific estimators in Section 5.

LEMMA 2.2.7 (Slutsky's Lemma). *Let $(x_t)_{t\in\mathbb{N}}$, x_0 be random vectors and $(y_t)_{t\in\mathbb{N}}$ be random vectors or random matrices. If $x_t \xrightarrow{d} x_0$ and $y_t \xrightarrow{d} c$ for a constant vector or matrix c, then*

(1) *$x_t + y_t \xrightarrow{d} x_0 + c$.*
(2) *$y_t x_t \xrightarrow{d} c x_0$.*
(3) *$y_t^{-1} x_t \xrightarrow{d} c^{-1} x_0$ if c is nonsingular.*

PROOF. All three statements are special cases of Lemma 2.2.5 together with Lemma 2.2.6 (3), (5). For more details, see van der Vaart (2000), Lemma 2.8.

□

The next theorem, which is called Cramér-Wold Device, reduces the problem of multivariate convergence to a univariate one.

THEOREM 2.2.8 (Cramér-Wold Device). *Let $(x_t)_{t\in\mathbb{N}}$ and x_0 be random variables. Then $x_t \xrightarrow{d} x_0$ if and only if $\lambda^T x_t \xrightarrow{d} \lambda^T x_0$ for all $\lambda \in \mathbb{R}^n$.*

PROOF. see discussion after Proposition 2.17 in van der Vaart (2000), p. 16.

□

With the help of the last theorem and the discussion after Definition 2.1.13 it is intuitively clear how to define the asymptotic normality of multivariate random variables in an alternative way:

DEFINITION 2.2.9. Let $(x_t)_{t\in\mathbb{N}}$ be a sequence of multivariate random vectors. This sequence of vectors is said to be asymptotic normally distributed with mean μ and Σ if

$$\frac{c^T(x_t-\mu)}{\sqrt{c^T\Sigma c}} \xrightarrow{d} x_0 \sim \mathcal{N}_1(0,1) \tag{2.2.5}$$

for every $c\in\mathbb{R}^n$ so that $c^T\Sigma c>0$. This definition includes the case where Σ is singular. The next theorem is called Delta Method and can be used to calculate the asymptotic distribution of a random statistic under a (in general non linear) function ϕ which is differentiable at a certain point.

THEOREM 2.2.10 (Delta Method). *Let $S\subset\mathbb{R}^n$ and $\phi: S\mapsto\mathbb{R}^m$ be a function which is differentiable at θ_0. Assume that $(x_T)_{T\in\mathbb{N}}$ be random vectors taking their values in the domain of ϕ. Further assume that*

$$\sqrt{T}(x_T-\theta_0) \xrightarrow{d} \mathcal{N}_n(0,\Sigma)$$

for $T\to\infty$. Then it follows that

$$\sqrt{T}(\phi(x_T)-\phi(\theta_0)) \xrightarrow{d} \mathcal{N}_m\left(0,(\nabla\phi(\theta_0))^T\Sigma(\nabla\phi(\theta_0))\right) \tag{2.2.6}$$

where $\nabla\phi(\theta_0)$ is the derivation of ϕ at the point θ_0.

PROOF. see Theorem 3.1 in van der Vaart (2000), p. 26.

□

2.2.2. Asymptotic Properties of the Mean and Autocovariance Estimators. In this subsection we want to give insights into the asymptotic properties of the estimates of the mean and the autocovariance estimators in the single- or high-frequency case. Note that there exists a wide range of literature about these estimates and their properties, see e.g. Rozanov (1967); Hannan (1970); Brockwell and Davis (1987); Anderson (1994); Su and Lund (2011).

We commence with the definition of the mean estimator

$$\bar{y}_T = \frac{1}{T}\sum_{t=1}^{T} y_t \tag{2.2.7}$$

and observe that this estimator is an unbiased one. It converges in the mean square sense to a stochastic variable:

THEOREM 2.2.11. *Let $(y_t)_{t\in\mathbb{Z}}$ be stationary and*

$$y_t = \int_{-\pi}^{\pi} \exp(-it\omega)\, z(d\omega)$$

where $(z(\omega)\,|\omega\in[-\pi,\pi])$ is the process of orthogonal increments (for a detailed discussion see Rozanov (1967)). Then it follows that

$$\underset{T\to\infty}{l.i.m.}\,\bar{y}_T = z(0)-z\left(0^-\right) \tag{2.2.8}$$

where $z\left(\omega_0^-\right)$ denotes the left limit at frequency ω_0.

PROOF. compare Theorem 4 in Hannan (1970), p. 204. □

It immediately follows that the mean estimator converges to $\mu = \mathbb{E}(y_t)$ (and is therefore consistent) if and only if the covariance of $z(0) - z(0^-)$ is zero. Now we will introduce the definition of a linear process:

DEFINITION 2.2.12. A process $(y_t)_{t\in\mathbb{Z}}$ is called a linear process if

$$y_t = \sum_{j=-\infty}^{\infty} k_j \nu_{t-j}, \tag{2.2.9}$$

$k_j \in \mathbb{R}^{n\times n}$, $\sum_{j=-\infty}^{\infty} \|k_j\| < \infty$ and $(\nu_t)_{t\in\mathbb{Z}}$ are independent identically distributed with zero mean and finite covariance Σ_ν or $(\nu_t)_{t\in\mathbb{Z}} \sim IID_n(0, \Sigma_\nu)$.

One property of a linear process is that $\sum_{j=-\infty}^{\infty} \|k_j\| < \infty$ implies $\sum_{j=-\infty}^{\infty} \|\gamma(j)\| < \infty$. This follows from the following inequalities:

$$\begin{aligned}\sum_{j=-\infty}^{\infty} \|\gamma(j)\| &\leq \|\Sigma_\nu\| \sum_{j=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} \|k_i\| \|k_{i+j}\| \\ &\leq \|\Sigma_\nu\| \left(\sum_{j=-\infty}^{\infty} \|k_j\| \right)^2 < \infty.\end{aligned}$$

Furthermore, it can be easily seen that

$$\sum_{j=-\infty}^{\infty} \|\gamma(j) \otimes \gamma(j)\| = \sum_{j=-\infty}^{\infty} \|\gamma(j)\|^2 < \infty.$$

One can conclude that a linear process always has a spectral density

$$\begin{aligned}f_y(\omega) &= \frac{1}{2\pi} k\left(e^{-i\omega}\right) \Sigma_\nu k\left(e^{i\omega}\right)^T \\ &= \sum_{j=-\infty}^{\infty} \gamma(j) e^{-ij\omega}\end{aligned}$$

where $k(z) = \sum_{j=-\infty}^{\infty} k_j z^j$.

EXAMPLE 2.2.13. Under the further assumption that $\nu_t \sim IID_n(0, \Sigma_\nu)$, the process $(y_t)_{t\in\mathbb{Z}}$, which is the output of system (2.1.6), is a linear process. Indeed the process has the representation

$$\begin{aligned}y_t &= a(z)^{-1} \nu_t \\ &= \sum_{j=-\infty}^{\infty} k_j \nu_{t-j}\end{aligned}$$

where $k_j = 0$ for $j < 0$.

THEOREM 2.2.14. *Let $(y_t)_{t\in\mathbb{Z}}$ be a linear process with spectral density $f_y(.)$. Then*

$$\lim_{T\to\infty} T\mathbb{E}\left(\bar{y}_T \left(\bar{y}_T\right)^T\right) = 2\pi f_y(0). \tag{2.2.10}$$

PROOF. see Corollary 4 in Hannan (1970), p. 208, where a more general theorem is stated. □

THEOREM 2.2.15. *Let $(y_t)_{t\in\mathbb{Z}}$ be a linear process with spectral density $f_y(.)$. Then*

$$\sqrt{T}\bar{y}_T \xrightarrow{d} \mathcal{N}_n\left(0, 2\pi f_y(0)\right) \tag{2.2.11}$$

PROOF. compare Theorem 11 in Hannan (1970), p. 221, where a more general theorem is stated. □

Reconsidering the estimator for $\gamma(k)$ in (2.1.36) it is worth mentioning that for the asymptotic analysis we will use the slightly different estimator

$$\begin{aligned} \hat{\hat{\gamma}}(k) &= \frac{1}{T}\sum_{t=1}^{T} y_t y_{t-k}^T \\ \hat{\hat{\gamma}}(-k) &= \hat{\hat{\gamma}}(k)^T, \end{aligned} \tag{2.2.12}$$

which is a non feasible estimator for the sample $y_1, \ldots, y_T$. The estimators in (2.1.36) and in (2.2.12) differ only by the mean correction and a finite number or to be more precisely by k summands, i.e. $y_t y_{t-k}^T$ for $t = 1, \ldots, k$. Whereas $\hat{\gamma}(k)$ is a biased estimator for $\gamma(k)$, $\hat{\hat{\gamma}}(k)$ is not. If k is fixed, the next lemma states that for a linear process the two different estimators for the autocovariance, i.e. $\hat{\gamma}(k)$ and $\hat{\hat{\gamma}}(k)$, do not differ asymptotically to the rate $\sqrt{T}$.

LEMMA 2.2.16. *Let $(y_t)_{t\in\mathbb{Z}}$ be a linear process and $k \in \mathbb{Z}$. Then*

$$\sqrt{T}\left(\hat{\gamma}(k) - \hat{\hat{\gamma}}(k)\right) \xrightarrow{p} 0. \tag{2.2.13}$$

PROOF. We are following the lines of Hannan (1970), p. 329. Rewriting $\sqrt{T}\left(\hat{\gamma}(k) - \hat{\hat{\gamma}}(k)\right)$ leads to

$$\sqrt{T}\left(\left(1 - \frac{k}{T}\right)\bar{y}_T\bar{y}_T^T - \frac{1}{T}\sum_{t=k+1}^{T} y_t\bar{y}_T^T - \frac{1}{T}\sum_{t=k+1}^{T}\bar{y}_T y_{t-k}^T - \frac{1}{T}\sum_{t=1}^{k} y_t y_{t-k}^T\right) \tag{2.2.14}$$

With the help of Theorem 2.2.15 it follows that $\sqrt{T}\bar{y}_T \xrightarrow{d} \mathcal{N}_n\left(0, 2\pi f_y(0)\right)$ and Theorem 2.2.11 implies $\left(1 - \frac{k}{T}\right)\bar{y}_T \xrightarrow{p} 0$ and $\frac{1}{T}\sum_{t=k+1}^{T} y_t \xrightarrow{p} 0$. Thus, with Lemmas 2.2.6 and 2.2.7 the first three terms in (2.2.14) converge to zero. The fourth term converges to zero, too, which can be seen from the Markov's inequality

$$\begin{aligned} \mathbb{P}\left(\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{k} y_t^i y_{t-k}^j\right| \geq \epsilon\right) &\leq \frac{\mathbb{E}\left|\sum_{t=1}^{k} y_t^i y_{t-k}^j\right|}{\sqrt{T}\epsilon} \\ &\leq \frac{c}{\sqrt{T}} \end{aligned}$$

for $i, j = 1, \ldots, n$. This proves the desired result. □

Analogously to the mean estimator for a linear process the autocovariance estimator is consistent, too.

THEOREM 2.2.17. *Let $(y_t)_{t\in\mathbb{Z}}$ be a linear process and $k \in \mathbb{Z}$. Then it follows that*

$$\hat{\gamma}(k) \xrightarrow{p} \gamma(k). \tag{2.2.15}$$

PROOF. see Hannan (1970), p. 231. □

Note that the assumption of a linear process is a bit restrictive and can be relaxed, see e.g. Hannan and Heyde (1972); Hall and Heyde (1980), where the i.i.d. assumption of the inputs is replaced by

$$\begin{aligned} \mathbb{E}\left(\nu_t|\mathcal{F}_{t-1}\right) &= 0 \\ \mathbb{E}\left(\nu_t\nu_t^T|\mathcal{F}_{t-1}\right) &= \Sigma_\nu, \end{aligned} \tag{2.2.16}$$

where $\mathcal{F}_t$ is the σ-algebra generated by ν_s for $s \leq t$.

The next lemma is important for obtaining the asymptotic covariance of the covariance estimators $\hat{\gamma}(k)$. Let $\eta = \mathbb{E}\left(\nu_t\nu_t^T \otimes \nu_t\nu_t^T\right)$ and $\kappa = \eta - \operatorname{vec}(\Sigma_\nu)\operatorname{vec}(\Sigma_\nu)^T - (\Sigma_\nu \otimes \Sigma_\nu) - K_{n,n}(\Sigma_\nu \otimes \Sigma_\nu)$ where $K_{n,n}$ is a commutation matrix (see Lemma C.0.2).

LEMMA 2.2.18. *Let $(y_t)_{t\in\mathbb{Z}}$ be a linear process and assume that $\eta = \mathbb{E}\left(\nu_t\nu_t^T \otimes \nu_t\nu_t^T\right)$ exists. Then we obtain*

$$\lim_{T\to\infty} TCov\left(vec\left(\hat{\gamma}(p)\right), vec\left(\hat{\gamma}(q)\right)\right) = R_{p,q} + S_{p,q} \tag{2.2.17}$$

for $p, q \in \mathbb{Z}$ where

$$\begin{aligned} R_{p,q} &= \sum_{k=-\infty}^{\infty} \left(\gamma(k+q-p)\otimes\gamma(k)\right) + K_{n,n}\left(\gamma(k+q)\otimes\gamma(k-p)\right) \\ S_{p,q} &= \sum_{k=-\infty}^{\infty}\sum_{r=-\infty}^{\infty} \left(k_{k-p}\otimes k_k\right)\kappa\left(k_{r+k-q}\otimes k_{r+k}\right)^T. \end{aligned}$$

Note that $S_{p,q}$ can be obtained through

$$vec(S_{p,q}) = \left(\tilde{\gamma}(p)\otimes\tilde{\gamma}(q)\right) vec(\kappa) \tag{2.2.18}$$

where $\tilde{\gamma}(p) = \sum_{r=-\infty}^{\infty}\left(k_{r-p}\otimes k_r\right)$.

PROOF. This lemma has been proved in Su and Lund (2011), where the line-by-line vectorization is used. □

In a next step we want to derive the asymptotic distribution of the autocovariance estimators. In order to prove this, we first need to introduce the concept of an M-dependent sequence.

DEFINITION 2.2.19. A sequence $(y_t)_{t\in\mathbb{Z}}$ is said to be M-dependent if for each t the sets of variables $\{y_j, j \leq t\}$ and $\{y_j, j \geq t+M+1\}$ are independent.

The next theorem is from Anderson (1994), p. 429, which shows the asymptotic normality of the mean estimator in the case of M-dependent sequences. Another version of this theorem can be found in Brockwell and Davis (1987), p. 213.

THEOREM 2.2.20. *Let $(y_t)_{t\in\mathbb{Z}}$ be a strictly stationary M-dependent process with zero mean. Then $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} y_t$ converges to a normal distribution with zero mean and a covariance matrix given by $\Sigma_M = \sum_{i=-M}^{M}\gamma(i)$, i.e.*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} y_t \xrightarrow{d} \mathcal{N}_n(0, \Sigma_M). \tag{2.2.19}$$

The following lemma leads to a useful tool for calculating asymptotic distributions. In Hannan (1970); Hall and Heyde (1980) it is called Bernstein's Lemma (see also Anderson (1958)):

LEMMA 2.2.21. *Consider the sequence $(a_T)_{T\in\mathbb{N}}$ of random vectors with zero mean. Let $a_T = b_T^K + c_T^K$ for $T, K \in \mathbb{N}$. If for every $\epsilon > 0$ and $\eta > 0$ there exist a K_0 so that for $K > K_0$*

$$b_T^K \xrightarrow{d} \mathcal{N}_n\left(0, \Sigma_K\right),$$

$\lim_{K\to\infty} \Sigma_K = \Sigma$ *and* $\mathbb{P}\left(\left(c_T^K\right)^T c_T^K > \eta\right) < \epsilon$ *for all T, then*

$$a_T \xrightarrow{d} \mathcal{N}_n\left(0, \Sigma\right). \tag{2.2.20}$$

PROOF. see Hannan (1970) p. 242 or Anderson (1958) p.425. □

Assuming that we want to calculate the asymptotic distribution of a random sequence $(a_T)_{T\in\mathbb{N}}$, but, for instance, we can only show that the "nearby" sequence $\left(b_T^K\right)_{T\in\mathbb{N}}$ is asymptotically normal, then, under the assumptions of the last lemma, we can conclude that also $(a_T)_{T\in\mathbb{N}}$ is asymptotically normally distributed. The next theorem is a multivariate version of Bartlett's formula.

THEOREM 2.2.22 (Bartlett's Formula). *Let $(y_t)_{t\in\mathbb{Z}}$ be a linear process, $s \in \mathbb{N}$ and assume that $\eta = \mathbb{E}\left(\nu_t\nu_t^T \otimes \nu_t\nu_t^T\right)$ exists. Then we obtain*

$$\sqrt{T}\left(\begin{pmatrix} vec\left(\hat{\gamma}\left(0\right)\right) \\ \vdots \\ vec\left(\hat{\gamma}\left(s\right)\right) \end{pmatrix} - \begin{pmatrix} vec\left(\gamma\left(0\right)\right) \\ \vdots \\ vec\left(\gamma\left(s\right)\right) \end{pmatrix}\right) \xrightarrow{d} \mathcal{N}_{n^2(s+1)}\left(0, \Sigma_\gamma\right) \tag{2.2.21}$$

where

$$\Sigma_\gamma \quad = \quad \left(R_{p,q} + S_{p,q}\right)_{p,q=0,\ldots,s}$$

can be obtained from Lemma 2.2.18.

PROOF. We are following the lines of the proof of Theorem 14 of Hannan (1970), p. 228: First of all note that using Lemma 2.2.16 we only have to consider $\hat{\hat{\gamma}}\left(p\right)$ instead of $\hat{\gamma}\left(p\right)$, since $\sqrt{T}\left(\hat{\gamma}\left(p\right) - \hat{\hat{\gamma}}\left(p\right)\right) \xrightarrow{p} 0$. We are considering a truncation of the linear process y_t, say y_t^K, depending on K

$$y_t^K = \sum_{j=-K}^{K} k_j \nu_{t-j}.$$

Now the vectorized unbiased autocovariance estimator of this truncated process is

$$\begin{aligned} \operatorname{vec}\left(\hat{\hat{\gamma}}_K\left(p\right)\right) &= \frac{1}{T}\sum_{t=1}^{T} \operatorname{vec}\left(y_t^K\left(y_{t-p}^K\right)^T\right) \\ &= \frac{1}{T}\sum_{t=1}^{T}\left(y_{t-p}^K \otimes y_t^K\right) \end{aligned}$$

or $\operatorname{vec}\left(\hat{\hat{\gamma}}_K(p)\right) = \frac{1}{T}\sum_{t=1}^{T} z_{t,p}$ where $z_{t,p} = \left(y_{t-p}^K \otimes y_t^K\right)$. Note that $z_{t,p}$ and $z_{t+h,p}$ are independent if $h > 2K+p$ and so $z_{t,p}$ is a $(2K+p)$-dependent process. Let $M_K = 2K+s$ and

$$\Gamma = \begin{pmatrix} \operatorname{vec}(\gamma(0)) \\ \vdots \\ \operatorname{vec}(\gamma(s)) \end{pmatrix}.$$

In an analogous way let us define $\hat{\hat{\Gamma}}$ as the corresponding sample estimator. Further let

$$\Gamma_K = \begin{pmatrix} \operatorname{vec}(\gamma_K(0)) \\ \vdots \\ \operatorname{vec}(\gamma_K(s)) \end{pmatrix}$$

where $\gamma_K(h)$ is the autocovariance of the truncated process y_t^K and let $\hat{\hat{\Gamma}}_K$ be the sample estimator of Γ_K. Thus, with the help of Theorem 2.2.20, we obtain that

$$\sqrt{T}\left(\hat{\hat{\Gamma}}_K - \Gamma_K\right) \xrightarrow{d} \mathcal{N}_{n^2(s+1)}\left(0, \Sigma_\gamma^{M_K}\right)$$

where $\Sigma_\gamma^{M_K} = \sum_{i=-M_K}^{M_K} \tilde{\Gamma}_{M_K}(i) = 2\pi f_{M_K}(0)$, $\tilde{\Gamma}_{M_K}(i)$ is the autocovariance and $f_{M_K}(\lambda)$ is the spectral density of

$$\begin{pmatrix} \operatorname{vec}\left(y_t^K \left(y_t^K\right)^T\right) \\ \vdots \\ \operatorname{vec}\left(y_t^K \left(y_{t-s}^K\right)^T\right) \end{pmatrix}.$$

The last blocked process is, of course, strictly stationary. Now we want to apply Lemma 2.2.21 to prove the normality for the non-truncated process $(y_t)_{t\in\mathbb{Z}}$. We define: $a_T = \sqrt{T}\left(\hat{\hat{\Gamma}} - \Gamma\right)$, $b_T^K = \sqrt{T}\left(\hat{\hat{\Gamma}}_K - \Gamma_K\right)$ and $c_T^K = b_T^K - a_T$. By taking the limes of $\Sigma_\gamma^{M_K}$ we obtain that $\lim_{K\to\infty} \Sigma_\gamma^{M_K} = \Sigma_\gamma = 2\pi f(0)$ where $f(\lambda)$ is the spectral density of

$$\begin{pmatrix} \operatorname{vec}\left(y_t y_t^T\right) \\ \vdots \\ \operatorname{vec}\left(y_t y_{t-s}^T\right) \end{pmatrix}.$$

Note that, under our assumptions, this spectral density exists. Thus, by the Chebyshev's inequality, it remains to prove that

$$\lim_{T\to\infty} T\mathbb{V}\left(\left(\hat{\hat{\Gamma}} - \hat{\hat{\Gamma}}_K\right)_m\right) = 0, \quad m = 1, \ldots, n^2(s+1).$$

This step is tedious and will be omitted here (see p. 244 in Hannan (1970)). This proves the asymptotic normality for the autocovariance estimators.

In a last step we can conclude that the covariance of the asymptotic distribution is the same as the right hand side of equation (2.2.10) in Theorem 2.2.14. So we can use Lemma 2.2.18 to calculate the desired asymptotic covariance. □

REMARK 2.2.23. The last theorem can be extended to any set of autocovariance lags including negative ones. As an example, one can use the lag set $(-p+1,\ldots,p)$, which includes all autocovariances needed for the Yule-Walker estimator (2.1.34).

Note, too, that in the Gaussian case, i.e. when the ν_t are assumed to be normally distributed, $\eta = \operatorname{vec}(\Sigma_\nu)\operatorname{vec}(\Sigma_\nu)^T + (\Sigma_\nu \otimes \Sigma_\nu) + K_{n,n}(\Sigma_\nu \otimes \Sigma_\nu)$ holds and clearly in this case $S_{p,q}$ vanishes.

Of course, Σ_γ can be a singular covariance matrix. For instance, if we want to derive the asymptotic distribution of the estimator $\hat{\gamma}(0) = \hat{\gamma}(0)^T$ for the case $n > 1$, the asymptotic covariance has rank $\frac{n(n+1)}{2} < n^2$ and thus is singular. This is one of the reasons why we introduced singular normal distributions.

2.2.3. Asymptotic Properties of the Yule-Walker and Maximum Likelihood Estimator. In a next step we want to derive the asymptotic behavior of the maximum likelihood estimator and the Yule-Walker estimator. It will turn out that, under our assumptions, the Yule-Walker estimator has the same asymptotic covariance as the maximum likelihood estimator and thus is asymptotically efficient, i.e. the asymptotic covariance of the Yule-Walker estimator equals the Cramer Rao bound.

THEOREM 2.2.24. *Let $(y_t)_{t\in\mathbb{Z}}$ be the output of system (2.1.6) with inputs $(\nu_t)_{t\in\mathbb{Z}} \sim IID_n(0,\Sigma_\nu)$. Assume that Σ_ν is nonsingular, $\eta = \mathbb{E}\left(\nu_t\nu_t^T \otimes \nu_t\nu_t^T\right)$ exists and let $\theta = \left(vec(A)^T, vech(\Sigma_\nu)^T\right)^T$. Then the maximum likelihood estimator for the AR(p) case is asymptotically normally distributed, i.e.*

$$\sqrt{T}\left(\hat{\theta}_{ML} - \theta\right) \xrightarrow{d} \mathcal{N}_{pn^2+n(n+1)/2}(0,\Sigma_{ML}) \tag{2.2.22}$$

where $\hat{\theta}_{ML} = \left(vec\left(\hat{A}_{ML}\right)^T, vech\left(\hat{\Sigma}_{ML}\right)^T\right)^T$ and has a covariance matrix given by

$$\Sigma_{ML} = \begin{pmatrix} \left(\Gamma_p^{-1} \otimes \Sigma_\nu\right) & 0 \\ 0 & D_n^\dagger\left(\eta - vec(\Sigma_\nu)\,vec(\Sigma_\nu)^T\right)\left(D_n^\dagger\right)^T \end{pmatrix}$$

where $D_n^\dagger$ is the generalized inverse of the duplication matrix D_n (see Appendix C). In the Gaussian case this variance changes to

$$\Sigma_{ML} = \begin{pmatrix} \left(\Gamma_p^{-1} \otimes \Sigma_\nu\right) & 0 \\ 0 & 2D_n^\dagger(\Sigma_\nu \otimes \Sigma_\nu)\left(D_n^\dagger\right)^T \end{pmatrix}.$$

PROOF. see Mann and Wald (1943), Anderson (1994) p. 183 or Lütkepohl (2005), p. 93. □

The next theorem states that the Yule-Walker estimators are asymptotically normally distributed.

THEOREM 2.2.25. *Let $(y_t)_{t\in\mathbb{Z}}$ be the output of system (2.1.6) with inputs $(\nu_t)_{t\in\mathbb{Z}} \sim IID_n(0,\Sigma_\nu)$ and assume $\Sigma_\nu > 0$. Then $\sqrt{T}\left(vec\left(\hat{A}_{YW}\right) - vec(A)\right)$ is asymptotically normally distributed with zero mean and the covariance matrix given by $\left(\Gamma_p^{-1} \otimes \Sigma_\nu\right)$, i.e.*

$$\sqrt{T}\left(vec\left(\hat{A}_{YW}\right) - vec(A)\right) \xrightarrow{d} \mathcal{N}_{n^2p}\left(0,\Gamma_p^{-1} \otimes \Sigma_\nu\right). \tag{2.2.23}$$

In addition, assume that $\eta = \mathbb{E}\left(\nu_t \nu_t^T \otimes \nu_t \nu_t^T\right)$ *exists. It follows that* $\sqrt{T}\left(vech\left(\hat{\Sigma}_{YW}\right) - vech\left(\Sigma_\nu\right)\right)$ *is asymptotically normally distributed, i.e.*

$$\sqrt{T}\left(vech\left(\hat{\Sigma}_{YW}\right) - vech\left(\Sigma_\nu\right)\right) \quad \xrightarrow{d} \quad \mathcal{N}_{n(n+1)/2}\left(0, D_n^\dagger\left(\eta - vec\left(\Sigma_\nu\right) vec\left(\Sigma_\nu\right)^T\right)\left(D_n^\dagger\right)^T\right). \tag{2.2.24}$$

PROOF. The proof for the system parameters is from Hannan (1970), p. 328 and can be separated into two parts:

(1) Show that $\sqrt{T}\left(\hat{\Gamma}_p \otimes I_n\right)\left(\text{vec}\left(\hat{A}_{\text{YW}}\right) - \text{vec}\left(A\right)\right) = \sqrt{T}\text{vec}\left(\hat{e}\right)$, where $\hat{e}$ is defined below.
(2) Show that $\sqrt{T}\text{vec}\left(\hat{e}\right)$ is asymptotically normally distributed with zero mean and the covariance matrix given by $\left(\Gamma_p \otimes \Sigma_\nu\right)$.

Consider

$$\sqrt{T}\hat{\gamma}_1 = \sqrt{T}\left(A_1, ..., A_p\right)\hat{\Gamma}_p + \sqrt{T}\underbrace{\left(\hat{e}(1), ..., \hat{e}(p)\right)}_{\hat{e}} + \sqrt{T}c_T \tag{2.2.25}$$

where

$$\hat{e}\left(i\right) = \frac{1}{T}\sum_{t=1}^{T} \nu_t y_{t-i}^T \quad i = 1, \ldots, p$$

and $\sqrt{T}c_T \xrightarrow{p} 0$, see Lemma 2.2.16, and thus will be neglected. Columnwise vectorization of equation (2.2.25) leads to

$$\sqrt{T}\text{vec}\left(\hat{\gamma}_1\right) = \sqrt{T}\left(\hat{\Gamma}_p \otimes I_n\right)\text{vec}\left(A\right) + \sqrt{T}\text{vec}\left(\hat{e}\right),$$

compare rule 3 in Lemma C.0.1. Replacing $\hat{\gamma}_1$ with equation (2.1.37) we directly get

$$\begin{aligned}\sqrt{T}\left(\hat{\Gamma}_p \otimes I_n\right)\text{vec}\left(A\right) + \sqrt{T}\text{vec}\left(\hat{e}\right) &= \sqrt{T}\text{vec}\left(\hat{A}_{\text{YW}}\hat{\Gamma}_p\right) \\ &= \sqrt{T}\left(\hat{\Gamma}_p \otimes I_n\right)\text{vec}\left(\hat{A}_{\text{YW}}\right)\end{aligned}$$

or in another form:

$$\sqrt{T}\left(\hat{\Gamma}_p \otimes I_n\right)\left(\text{vec}\left(\hat{A}_{\text{YW}}\right) - \text{vec}\left(A\right)\right) = \sqrt{T}\text{vec}\left(\hat{e}\right). \tag{2.2.26}$$

Thus, it remains to prove that $\sqrt{T}\text{vec}\left(\hat{e}\right)$ is asymptotically normally distributed with zero mean and covariance $\left(\Gamma_p \otimes \Sigma_\nu\right)$. As it can be easily seen $\sqrt{T}\text{vec}\left(\hat{e}\right)$ has a zero mean and covariance $\left(\Gamma_p \otimes \Sigma_\nu\right)$, since for a particular element of $\mathbb{V}\left(\sqrt{T}\text{vec}\left(\hat{e}\right)\right)$ it follows that

$$\begin{aligned}\mathbb{E}\left(\frac{1}{T}\left(\sum_{t=1}^{T} y_{t-p} \otimes \nu_t\right)\left(\sum_{n=1}^{T} y_{n-q} \otimes \nu_n\right)^T\right) &= \frac{1}{T}\mathbb{E}\left(\sum_{t=1}^{T} y_{t-p} y_{t-q}^T \otimes \nu_t \nu_t^T\right) \\ &\quad + \underbrace{\frac{1}{T}\mathbb{E}\left(\sum_{t=1}^{T}\sum_{n=1, t\neq n}^{T} y_{t-p} y_{n-q}^T \otimes \nu_t \nu_n^T\right)}_{=0} \\ &= \gamma\left(p-q\right) \otimes \Sigma_\nu.\end{aligned}$$

Thus it is left to show that $\sqrt{T}\mathrm{vec}\left(\hat{e}\right)$ is asymptotically normal. As in the proof of Theorem 2.2.22 we define

$$\begin{aligned} y_t^K &= \sum_{j=0}^{K} k_j \nu_{t-j} \\ \mathrm{vec}\left(\hat{e}_K\left(i\right)\right) &= \frac{1}{T}\sum_{t=1}^{T}\left(y_{t-i}^K \otimes \nu_t\right), \quad i = 1, \ldots, p. \end{aligned}$$

or $\mathrm{vec}\left(\hat{e}_K\left(i\right)\right) = \frac{1}{T}\sum_{t=1}^{T} z_{t,i}$, where $z_{t,i} = \left(y_{t-i}^K \otimes \nu_t\right)$. Note that $z_{t,i}$ and $z_{t+h,i}$ are independent if $h > 2K + i$ and so $z_{t,i}$ is a $2K+i$-dependent process. We define $M_K = 2K + p$,

$$\hat{e}_K = \begin{pmatrix} \mathrm{vec}\left(\hat{e}_K\left(1\right)\right) \\ \vdots \\ \mathrm{vec}\left(\hat{e}_K\left(p\right)\right) \end{pmatrix}$$

and with Theorem 2.2.20 we obtain

$$\sqrt{T}\left(\hat{e}_K\right) \xrightarrow{d} \mathcal{N}_{n^2 p}\left(0, \Sigma_e^{M_K}\right)$$

where $\Sigma_e^{M_K} = \sum_{i=-M_K}^{M_K} \tilde{\Gamma}_{M_K}\left(i\right) = 2\pi f_{M_K}\left(0\right)$, $\tilde{\Gamma}_{M_K}\left(i\right)$ is the autocovariance and $f_{M_K}\left(\lambda\right)$ is the spectral density of

$$\begin{pmatrix} \mathrm{vec}\left(\nu_t\left(y_{t-1}^K\right)^T\right) \\ \vdots \\ \mathrm{vec}\left(\nu_t\left(y_{t-p}^K\right)^T\right) \end{pmatrix}.$$

Note that $\tilde{\Gamma}_{M_K}\left(i\right)$ does not need further assumptions on higher moments of $\left(\nu_t\right)_{t\in\mathbb{Z}}$ because

$$\begin{pmatrix} \mathrm{vec}\left(\nu_t\left(y_{t-1}^K\right)^T\right) \\ \vdots \\ \mathrm{vec}\left(\nu_t\left(y_{t-p}^K\right)^T\right) \end{pmatrix} = \sum_{j=0}^{K}\left(I_p \otimes \left(k_j \otimes I_n\right)\right) \begin{pmatrix} \left(\nu_{t-1-j} \otimes \nu_t\right) \\ \vdots \\ \left(\nu_{t-p-j} \otimes \nu_t\right) \end{pmatrix}$$

and thus ν_{t-i-j} and ν_t do not occur at the same time point. In order to apply Lemma 2.2.21 we set $a_T = \sqrt{T}\mathrm{vec}\left(\hat{e}\right)$, $b_T^K = \sqrt{T}\hat{e}_K$ and $c_T^K = b_T^K - a_T$. If we take the limes of $\Sigma_e^{M_K}$, we get $\lim_{K\to\infty}\Sigma_e^{M_K} = \Sigma_e = \left(\Gamma_p \otimes \Sigma_\nu\right)$. Then due to Chebyshev's inequality we only need to prove that

$$\lim_{T\to\infty} T\mathbb{V}\left(\left(\mathrm{vec}\left(\hat{e}\right) - \hat{e}_K\right)_m\right) = 0 \quad m = 1, \ldots, n^2 p$$

to fulfill the assumptions of Lemma 2.2.21. For a certain j, i and s this follows directly from

$$\begin{aligned} T\mathbb{E}\left(\left(\operatorname{vec}\left(\hat{e}\right)-\hat{e}_K\right)_m\right)^2 &= \frac{1}{T}\mathbb{E}\left(\sum_{t=1}^{T}\sum_{u=1}^{T}\nu_t^i e_j^T\left(y_{t-s}-y_{t-s}^K\right)\nu_u^i e_j^T\left(y_{u-s}-y_{u-s}^K\right)\right) \\ &= \frac{1}{T}\sum_{t=1}^{T}\sum_{u=1}^{T}\sum_{v=K+1}^{\infty}\sum_{w=K+1}^{\infty}\mathbb{E}\left(e_i^T\nu_t\nu_{t-s-v}^T\left(e_j^T k_v\right)^T e_i^T\nu_u\nu_{u-s-w}^T\left(e_j^T k_w\right)^T\right) \\ &= \frac{1}{T}\sum_{t=1}^{T}\sum_{w=K+1}^{\infty}\mathbb{E}\left(\left(e_i^T\nu_t\right)^2\nu_{t-s-w}^T\left(e_j^T k_w\right)^T\nu_{t-s-w}^T\left(e_j^T k_w\right)^T\right) \\ &= \sigma_{ii}^2\sum_{w=K+1}^{\infty}k_w^j\Sigma_\nu\left(k_w^j\right)^T \overset{K\to\infty}{\longrightarrow} 0 \end{aligned}$$

where e_i denotes the i-th unit vector. Now we apply Bernstein's Lemma to obtain that $\sqrt{T}\operatorname{vec}\left(\hat{e}\right)$ is asymptotically normal. Thereby $\sqrt{T}\left(\operatorname{vec}\left(\hat{A}_{\mathrm{YW}}\right)-\operatorname{vec}\left(A\right)\right)$ is asymptotically normal with the covariance matrix $\left(\Gamma_p^{-1}\otimes\Sigma_\nu\right)$. This follows immediately from Slutsky's Lemma 2.2.7 since $\left(\hat{\Gamma}_p\otimes I_n\right)\xrightarrow{p}\left(\Gamma_p\otimes I_n\right)$ and $\sqrt{T}\operatorname{vec}\left(\hat{e}\right)\xrightarrow{d}\mathcal{N}_{n^2p}\left(0,\Gamma_p\otimes\Sigma_\nu\right)$, so that

$$\sqrt{T}\left(\operatorname{vec}\left(\hat{A}_{\mathrm{YW}}\right)-\operatorname{vec}\left(A\right)\right)=\sqrt{T}\left(\hat{\Gamma}_p^{-1}\otimes I_n\right)\operatorname{vec}\left(\hat{e}\right)\xrightarrow{d}\mathcal{N}_{n^2p}\left(0,\Gamma_p^{-1}\otimes\Sigma_\nu\right).$$

For the asymptotic behavior of the estimator of the noise parameters, i.e. $\hat{\Sigma}_{\mathrm{YW}}=\hat{\gamma}\left(0\right)-\hat{\gamma}_1\hat{\Gamma}_p^{-1}\hat{\gamma}_1^T$, we can observe that $\hat{\gamma}\left(0\right)=\frac{1}{T}YY^T$ where again $Y=\left(y_1,\ldots,y_T\right)$ and that $\hat{\gamma}_1\hat{\Gamma}_p^{-1}\hat{\gamma}_1^T$ is approximately $YX^T\left(XX^T\right)^{-1}XY^T$ where again $X=\left(x_1,\ldots,x_T\right)$. Indeed, it is easy to show that

$$\sqrt{T}\left(\frac{1}{T}YX^T\left(\frac{1}{T}XX^T\right)^{-1}-\hat{\gamma}_1\hat{\Gamma}_p^{-1}\right)\xrightarrow{p}0,$$

which states that the ML and the Yule-Walker estimator for the system parameters are asymptotically equivalent. Furthermore, from Lemma 2.2.16 we can conclude that $\sqrt{T}\left(\frac{1}{T}YX^T-\hat{\gamma}_1\right)\xrightarrow{p}0$. Now with the aid of Lemma 2.2.7 we directly obtain that

$$\begin{aligned} \sqrt{T}\left(YX^T\left(XX^T\right)^{-1}XY^T-\hat{\gamma}_1\hat{\Gamma}_p^{-1}\hat{\gamma}_1^T\right) &= \sqrt{T}\left(\frac{1}{T}YX^T\left(\frac{1}{T}XX^T\right)^{-1}-\hat{\gamma}_1\hat{\Gamma}_p^{-1}\right)XY^T \\ &\quad+\sqrt{T}\hat{\gamma}_1\hat{\Gamma}_p^{-1}\left(\frac{1}{T}XY^T-\hat{\gamma}_1^T\right)\xrightarrow{p}0, \end{aligned}$$

which states that the ML and the Yule-Walker estimator for the noise parameters are asymptotically equivalent.

□

REMARK 2.2.26. Note that, as mentioned above, under our assumptions the Yule-Walker estimator has the same asymptotic covariance as the maximum likelihood estimator and thus is asymptotically efficient. Whereas these two estimators do not differ asymptotically, they have different finite sample properties: The Yule-Walker estimator always leads to a stable system, provided $\hat{\Gamma}_p>0$ holds, while the maximum likelihood estimator, in general, does not lead to a stable system. The two estimators for the noise covariance matrix are always positive semi-definite.

CHAPTER 3

Identifiability of AR Systems in the Mixed-Frequency Case

In this chapter we deal with the problem of identifiability of the high-frequency system (2.1.6) in the mixed-frequency case, i.e. uniquely determine the system and noise parameters from the population second moments which can be directly observed. We assume that at least one component of the process $(y_t)_{t\in\mathbb{Z}}$ is observed at every time point $t \in \mathbb{Z}$. It turns out that identifiability cannot be achieved on the whole set Θ but on a "large" subset. Therefore, we introduce a property for a subset called generic, see Section 3.1. Section 3.2 shows the generic identifiability first for the AR(1) and then for the AR(p) where $p \geq 1$ case. In the AR(1) case the non-identifiable set can be described explicitly. Identifiability of the fourth moment of the innovations (if it exists) is also discussed since it is needed for the analysis in Chapter 5. While from Section 3.2.1 to Section 3.2.3 the stock case is considered, in Section 3.2.4 the results are generalized to the flow case, i.e. where the slow component is aggregated by a known scheme. This chapter is based on the results obtained in Anderson et al. (2012, 2015a).

3.1. What is generic?

In Section 3.2 we want to discuss which subset of the parameter space Θ is identifiable from the second moments, which can be observed directly. It will turn out that not all parameters of the whole parameter space Θ are identifiable but at least on a generic subset:

DEFINITION 3.1.1. We say that a property holds generically on a space Θ if there exists an open and dense subset of Θ on which the property holds.

First, we start with a lemma, which will be essential for the following proofs.

LEMMA 3.1.2. *Let $f : \Theta \to \mathbb{R}$ be a polynomial function. If there exists a $\theta^* \in \Theta$ so that $f(\theta^*) \neq 0$, then the set of zeros of f is a proper algebraic set, i.e. an algebraic set of dimension smaller than the dimension of Θ, and in particular its complement in Θ is generic.*

PROOF. For the proof and a detailed discussion see e.g. Lee and Markus (1967); Wonham (1985); Bochnak et al. (1998). □

The following lemmas state some desirable properties which hold generically on the parameter space Θ.

LEMMA 3.1.3. *Let $A \in \mathbb{R}^{n\times n}$. Then A is generically nonsingular in $\mathbb{R}^{n\times n}$.*

PROOF. In order to apply Lemma 3.1.2 we define $f(A) = \det(A)$. Thus, we only have to find a point in the parameter space for which $f(A) \neq 0$ holds. By choosing $A^* = I_n$ it follows $f(A^*) \neq 0$

and thus the set of singular matrices is a proper algebraic set. The complement of this set is generic in $\mathbb{R}^{n\times n}$. □

LEMMA 3.1.4. *Let $z_i \in \mathbb{C}$ for $i = 1, \dots, m$, $m \in \mathbb{N}$ and $\mathcal{A}$ be the companion form of $(A_1, \dots, A_p)$. Then the subset of Θ on which z_i $i = 1, \dots, m$ are no eigenvalues of $\mathcal{A}$ is generic.*

PROOF. The finite intersection of open and dense subsets is still open and dense. Thus w.l.o.g. we only have to show that generically z_1 is not an eigenvalue of $\mathcal{A}$. Again we want to apply Lemma 3.1.2. Thus, we have to find a point in the parameter space so that

$$f(A) = \det(\mathcal{A} - z_1 I_{np}) \neq 0$$

holds. We choose the point, $\mathcal{A}^*$, where

$$A_i = \begin{cases} 0 & i = 1, \dots, p-1 \\ \rho^p C & i = p \end{cases} \tag{3.1.1}$$

and $\rho \in (0,1)$ with the further assumption that $\rho \neq |z_1|$ and where

$$C = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix} \tag{3.1.2}$$

is a so-called circulant matrix with eigenvalues $\omega_j = \exp\left(\frac{2\pi i j}{n}\right)$. It is easy to show that the eigenvalues λ_j of $\mathcal{A}^*$ have the property that $|\lambda_j| = \rho$. Therefore, $\det(\mathcal{A}^* - z_1 I_{np}) \neq 0$ and the same property holds generically on the parameter space Θ. □

LEMMA 3.1.5. *Let $\mathcal{A}$ be the companion form of $(A_1, \dots, A_p)$. Then the subset of Θ where the eigenvalues of $\mathcal{A}$ are distinct is generic.*

PROOF. A proof has been given in Felsenstein (2014), Lemma 3.4.2. □

Note that it immediately follows that the matrix $\mathcal{A}$ has generically the representation $\mathcal{A} = P\Lambda P^{-1}$ where $\Lambda \in \mathbb{R}^{np\times np}$ is a diagonal matrix which contains the eigenvalues λ_i of $\mathcal{A}$ and $P = (p_1, \dots, p_{np})$ where p_i is the eigenvector corresponding to λ_i. It is worth mentioning that this representation is generically also possible for unstable polynomials.

LEMMA 3.1.6. *Let $1 < N \in \mathbb{N}$ and let $\Theta_N \subset \Theta$ so that every $\mathcal{A}$ fulfills the property that for two eigenvalues $\lambda_i \neq \lambda_j$ of $\mathcal{A}$ it follows that $\lambda_i^N \neq \lambda_j^N$. Then Θ_N is generic in Θ.*

PROOF. The proof follows the same idea as Lemma 3.4.2 in Felsenstein (2014): We define $p_j(\lambda) = \det\left(\mathcal{A}^j - \lambda I\right)$ and $q_j(\lambda) = p_j'(\lambda) = \frac{\partial p_j(\lambda)}{\partial \lambda}$. It is well known that λ_i is a zero of $p_j(\lambda)$, $j \in \mathbb{N}$ with multiplicity larger than one if and only if it is a zero of $q_j(\lambda)$, too. Thus, using the result of Sylvester, which says that two polynomials, say $a(z)$ and $b(z)$, are co-prime if and only if $\det(S(a,b)) \neq 0$ where

$S(a,b)$ is the Sylvester resultant (compare Kailath (1980) p. 142), we only have to find a point in the parameter space Θ so that $f(A) = \det(S(p_1,q_1))\det(S(p_N,q_N)) \neq 0$. Note that $\det(S(p_1,q_1))$ guarantees that $\mathcal{A}$ has different eigenvalues and $\det(S(p_N,q_N))$ guarantees that $\lambda_i^N \neq \lambda_j^N$. An easy way to construct such a point, $f(A^*) \neq 0$, is: Let $z_i \in \mathbb{R}$, $i = 1,\ldots,np$ with $0 < |z_i| < 1$ and $|z_i| < |z_{i+1}|$. Furthermore, let $\tilde{a}(z) = \prod_{i=1}^{p}(a_i - I_n z)$ where a_i is a diagonal matrix where the j-th diagonal entry is $z_{(i-1)n+j}$. Then $\tilde{a}(z) = A_0 + A_1 z + \ldots + A_p z^p$ and by normalizing A_0 we obtain our point $a^*(z) = A_0^{-1}\tilde{a}(z)$. Of course, this point fulfills $f(A^*) \neq 0$ since $\mathcal{A}^*$ has the eigenvalues $1/z_i$, $i = 1,\ldots,np$ and thus $(\mathcal{A}^*)^N$ has the eigenvalues $1/z_i^N$ for $i = 1,\ldots,np$ which are of course distinct. □

We can conclude that if p_i is an eigenvector of $\mathcal{A}$ corresponding to the eigenvalue λ_i, then it is generically an eigenvector of $\mathcal{A}^N$ corresponding to the eigenvalue λ_i^N and vice versa.

LEMMA 3.1.7. *Let $\mathcal{A}$ be the companion form of $(A_1,\ldots,A_p)$. For $i = 1,\ldots,n$, the pair $\left(e_i^T,\mathcal{A}\right)$ where $e_i \in \mathbb{R}^{np}$ is the i-th unit vector is generically observable.*

PROOF. A proof has been given in Anderson et al. (2012), Lemma 3. Here we will present a different one: We have to find a point so that $f(A) = \det\left(\left(e_i, \mathcal{A}^T e_i, \ldots, \left(\mathcal{A}^T\right)^{np-1} e_i\right)\right) \neq 0$. We choose the point, say $\mathcal{A}^*$, where

$$A_i = \begin{cases} 0 & i = 1,\ldots,p-1 \\ \rho^p C & i = p \end{cases} \tag{3.1.3}$$

and where $\rho \in (0,1)$. It is a well known fact that C has the normalized right eigenvectors $q_j = \frac{1}{\sqrt{n}}\left(1,\omega_j,\omega_j^2,\ldots,\omega_j^{n-1}\right)^T$ where $\omega_j = \exp\left(\frac{2\pi i j}{n}\right)$ for $j = 1,\ldots,n$. Note that ω_j is the j-th eigenvalue of C.

Using the PBH test (see Theorem 2.1.6) we only have to show that $e_i^T p_j \neq 0$ for all eigenvectors p_j of $\mathcal{A}^*$. Note that because of Lemma 2.1.7 $p_j = \left(\left(p_j^1\right)^T, \left(p_j^1\right)^T \lambda_j^{-1}, \ldots, \left(p_j^1\right)^T \lambda_j^{-p+1}\right)^T$ and p_j^1 fulfills $a^*\left(\lambda_j^{-1}\right) p_j^1 = \left(I_n - \lambda_j^{-p}\rho^p C\right) p_j^1 = 0$. Thus, p_j^1 is an eigenvector of C and has no zero entry. To summarize, this pair is observable and thus this property holds on a generic subset of the parameter space Θ, too. □

LEMMA 3.1.8. *Let $\mathcal{A}$ be the companion form of $(A_1,\ldots,A_p)$ and $\mathcal{B}$ as in equation (2.1.11). Then Γ_p is generically nonsingular in Θ.*

PROOF. As mentioned before Theorem 2.1.6, Γ_p is nonsingular if and only if the pair $(\mathcal{A},\mathcal{B})$ is controllable, i.e.

$$\operatorname{rk} \underbrace{\left(\mathcal{B},\mathcal{A}\mathcal{B},\ldots,\mathcal{A}^{np-1}\mathcal{B}\right)}_{C} = np.$$

Thus it remains to prove that generically C has full row rank or $\det\left(CC^T\right)$, which is a polynomial function in the system and noise parameters, is generically unequal to zero.

Again we want to use Lemma 3.1.2 to establish our result: We choose the point, say θ^*, where $\mathcal{A}^*$ is the companion form where the system parameters are given in (3.1.3) and $\mathcal{B}^* = \left(e_1^T,0,\ldots,0\right)^T \in \mathbb{R}^{np}$

where $e_1 \in \mathbb{R}^n$ is the first unit vector. Then for an AR(p) process $(w_t)_{t\in\mathbb{Z}}$ of a system with parameters θ^* the covariance $\gamma^*(0)$ is diagonal and the covariances $\gamma^*(j)$, $j = 1, \ldots, p-1$ are zero, which can be easily seen by looking at the Wold decomposition $w_t = \sum_{j=0}^{\infty} \rho^{pj} C^j e_1 \varepsilon_{t-jp}$. Obviously, $\gamma^*(0) = \sum_{j=0}^{\infty} \rho^{2pj} C^j e_1 e_1^T \left(C^j\right)^T$ is nonsingular. Thus, $\Gamma_p^* > 0$ holds and therefore $\det\left(C^* \left(C^*\right)^T\right) > 0$, which leads to our desired result. □

It is a fallacy to claim that all roots of $\det(a(z))$ are generically complex with non-zero complex part since the real numbers are a one-dimensional subspace of the complex numbers. This can be easily demonstrated by assuming that np is odd. Then, of course, at least one root has to be real since complex roots have to occur in complex conjugate pairs.

3.2. g-Identifiability of AR Systems

In this section we consider the problem of identifiability, i.e. whether for given Θ the parameters A_i and Σ_ν of the high-frequency system are uniquely determined by the population second moments of the observations.

Let

$$y_t = \begin{pmatrix} y_t^f \\ y_t^s \end{pmatrix} \tag{3.2.1}$$

where the n_f-dimensional, say fast (or to be more precise high-frequency) component y_t^f is observed at the highest (sampling) frequency $t \in \mathbb{Z}$ and the n_s-dimensional slow (or to be more precise low-frequency) component y_t^s is observed only for $t \in N\mathbb{Z}$ (N being an integer $N > 1$), i.e. for every N-th time point. Throughout we assume $n_f \geq 1$ and that we deal with stock variables. The matrices

$$A_i = \begin{pmatrix} a_{ff}(i) & a_{fs}(i) \\ a_{sf}(i) & a_{ss}(i) \end{pmatrix}, \; i = 1, \ldots, p, \; \Sigma_\nu = \begin{pmatrix} \sigma_{ff} & \sigma_{fs} \\ \sigma_{sf} & \sigma_{ss} \end{pmatrix} \tag{3.2.2}$$

are partitioned accordingly.

The population second moments which can be directly observed are

$$\begin{aligned} \gamma^{ff}(h) &= \mathbb{E}\left(y_{t+h}^f \left(y_t^f\right)^T\right), \; h \in \mathbb{Z}, \\ \gamma^{sf}(h) &= \mathbb{E}\left(y_{t+h}^s \left(y_t^f\right)^T\right), \; h \in \mathbb{Z}, \\ \gamma^{ss}(h) &= \mathbb{E}\left(y_{t+h}^s \left(y_t^s\right)^T\right), \; h \in N\mathbb{Z}. \end{aligned} \tag{3.2.3}$$

EXAMPLE 3.2.1. Let $N = 2$, $n = 2$ and thus $n_f = n_s = 1$. Then we have the following observation scheme:

t	$\cdots$	1	2	3	4	5	$\cdots$
y_t^f	$\cdots$	y_1^f	y_2^f	y_3^f	y_4^f	y_5^f	$\cdots$
y_t^s	$\cdots$	$\times$	y_2^s	$\times$	y_4^s	$\times$	$\cdots$

(3.2.4)

where $\times$ means that this observation is not available. The second moments which can be observed are

$$\cdots \begin{pmatrix} \gamma^{ff}(0) & \gamma^{fs}(0) \\ \gamma^{sf}(0) & \gamma^{ss}(0) \end{pmatrix}, \begin{pmatrix} \gamma^{ff}(1) & \gamma^{fs}(1) \\ \gamma^{sf}(1) & \times \end{pmatrix}, \begin{pmatrix} \gamma^{ff}(2) & \gamma^{fs}(2) \\ \gamma^{sf}(2) & \gamma^{ss}(2) \end{pmatrix}, \begin{pmatrix} \gamma^{ff}(3) & \gamma^{fs}(3) \\ \gamma^{sf}(3) & \times \end{pmatrix}, \cdots$$

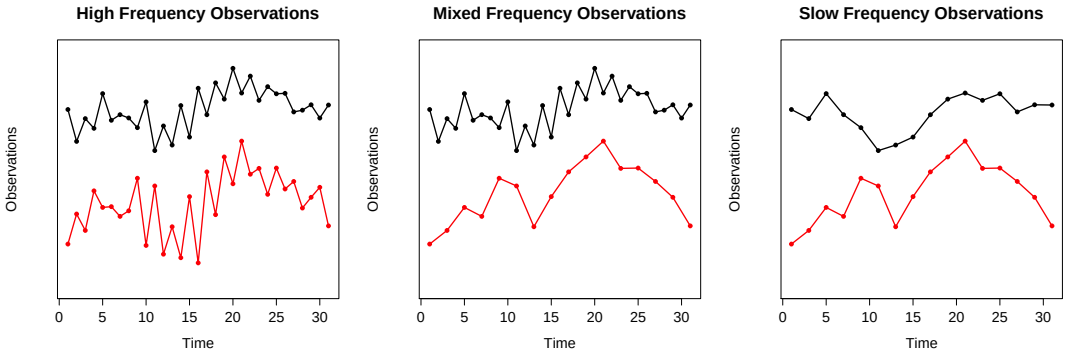

FIGURE 3.2.1. Different sampling rates of a two-dimensional process for $N = 2$

As mentioned in Chapter 1 there are many alternative ways to overcome the problem of mixed-frequency data. Two of these ways are motivated as follows: Since most of the standard literature in time series analysis is based on single-frequency observations one could, for instance, interpolate the slow component so that single/high-frequency "data" are available (see Friedman (1962)) and then apply procedures as described in Chapter 2. One drawback of using interpolated data will be discussed in Section 4.3. Another way would be to transform the observed data to the lowest frequency. One possible transformation would be to use the observations which are only available at the same time point. Figure 3.2.1 shows a two-dimensional time series with the fast component in black and the slow component in red. In the left picture the time series is observed at every time point, in the middle we have our mixed-frequency setting, where we assume that we only observe the slow component every second time point, and in the right picture we observe the whole time series only at every second time point. Of course, the behavior of the time series in the right picture is smoother than in the left one. This figure should indicate that omitting observed fast components leads to an information loss.

The next definition was first introduced in Anderson et al. (2015a):

DEFINITION 3.2.2. A parameter space is called g-identifiable if there exists a generic subset of the parameter space on which the parameters are identifiable.

Note that if identifiability holds and if, in addition, there is an algorithm for obtaining the parameters of the high-frequency system from the population second moments of the observations, we can reconstruct the missing second moments (see Section 3.2.3).

3.2.1. Identifiability of AR(1) Systems. The content of this section has been published in Anderson et al. (2015a) (except for Example 3.2.3 which is contained in Koelbl et al. (2015)). For further discussion see also Anderson et al. (2012).

We consider the special case of AR(1) systems. In addition, for simplicity, we restrict ourselves to the case $N = 2$, $n_f = n_s = 1$. Furthermore, we assume throughout this section that the AR(1) system is regular. This is done for two reasons. Firstly, it gives an example illustrating the problem. Secondly, as will be shown below, this analysis yields special results; in particular the subset Θ_I of Θ,

where identifiability is obtained (without imposing additional restrictions), can be described explicitly. We show that the complement of this set is a so-called semi-algebraic set, see Bochnak et al. (1998), Definition 2.1.4, i.e. a set of (multivariate) polynomial zeros where algebraic inequalities are imposed additionally, and so we conclude that for generic parameter values identifiability is obtained.

We first consider the case where in addition Σ_ν is diagonal. We can write

$$\begin{pmatrix} y_t^f \\ y_t^s \end{pmatrix} = \underbrace{\begin{pmatrix} a_{ff} & a_{fs} \\ a_{sf} & a_{ss} \end{pmatrix}}_{A_1} \begin{pmatrix} y_{t-1}^f \\ y_{t-1}^s \end{pmatrix} + \begin{pmatrix} \nu_t^f \\ \nu_t^s \end{pmatrix} \tag{3.2.5}$$

where we assume that every even time point is fully observed. Now the one-step-ahead predictor for y_{t-1}^f, $t-1$ odd, based on observed outputs is obtained from the following equation:

$$y_{t-1}^f = a_{ff} y_{t-2}^f + a_{fs} y_{t-2}^s + \nu_{t-1}^f$$

and the two-step-ahead predictor of y_t, t even, is obtained from

$$y_t = A_1^2 y_{t-2} + A_1 \nu_{t-1} + \nu_t.$$

Combining both equations gives a three-dimensional system on $2\mathbb{Z}$:

$$\underbrace{\begin{pmatrix} y_t^f \\ y_t^s \\ y_{t-1}^f \end{pmatrix}}_{\tilde{y}_t} = \underbrace{\begin{pmatrix} A_1^2 & 0 \\ a_{ff} & a_{fs} & 0 \end{pmatrix}}_{\tilde{\mathcal{A}}} \begin{pmatrix} y_{t-2}^f \\ y_{t-2}^s \\ y_{t-3}^f \end{pmatrix} + \underbrace{\begin{pmatrix} A_1 \nu_{t-1} + \nu_t \\ \nu_{t-1}^f \end{pmatrix}}_{\tilde{\nu}_t}. \tag{3.2.6}$$

Note that (3.2.6) is an AR(1) system on $2\mathbb{Z}$ whose outputs $\tilde{y}_t$ are the observed variables and thus may serve as a model for the mixed-frequency data.

As the components of $\tilde{y}_{t-2}$ are linearly independent by the regularity assumption, $\tilde{\mathcal{A}}$ and $\Sigma_{\tilde{\nu}} = \mathbb{E}\left(\tilde{\nu}_t \tilde{\nu}_t^T\right)$ are uniquely determined from the second moments of $(\tilde{y}_t)_{t \in 2\mathbb{Z}}$. However, not all entries in $\tilde{\mathcal{A}}$, $\Sigma_{\tilde{\nu}}$ are free, since

$$\tilde{\mathcal{A}} = \begin{pmatrix} a_{ff}^2 + a_{fs} a_{sf} & a_{ff} a_{fs} + a_{fs} a_{ss} & 0 \\ a_{sf} a_{ff} + a_{ss} a_{sf} & a_{sf} a_{fs} + a_{ss}^2 & 0 \\ a_{ff} & a_{fs} & 0 \end{pmatrix}, \tag{3.2.7}$$

$$\Sigma_{\tilde{\nu}} = \begin{pmatrix} \sigma_{ff} & 0 & 0 \\ 0 & \sigma_{ss} & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} a_{ff} & a_{fs} \\ a_{sf} & a_{ss} \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \sigma_{ff} & 0 \\ 0 & \sigma_{ss} \end{pmatrix} \begin{pmatrix} a_{ff} & a_{sf} & 1 \\ a_{fs} & a_{ss} & 0 \end{pmatrix} \tag{3.2.8}$$

must hold.

Here the high-frequency system has 6 free parameters, whereas a general AR(1) system for $n = 3$ has 15 free parameters. In order to analyze identifiability we solve (3.2.7), (3.2.8) for given $\tilde{\mathcal{A}}$, $\Sigma_{\tilde{\nu}}$ for the high-frequency parameters A_1 and Σ_ν. We see that if a_{fs} and a_{sf} are both zero, then only a_{ss}^2 is unique, otherwise A_1 and Σ_ν are unique and thus we have non-identifiability if and only if $a_{fs} = a_{sf} = 0$ and $a_{ss} \neq 0$ hold.

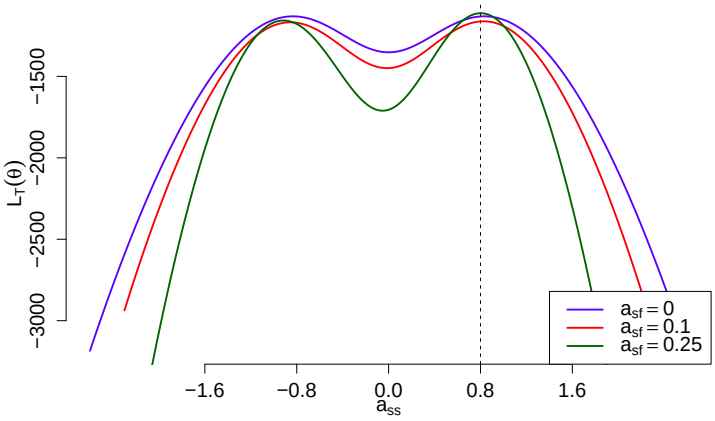

FIGURE 3.2.2. Sections of the likelihood functions for three different values of a_{sf}

It is interesting to note that we have identifiability whenever the two component processes $\left(y_t^f\right)_{t\in\mathbb{Z}}$ and $(y_t^s)_{t\in\mathbb{Z}}$ are not orthogonal.

In order to demonstrate the effects of being close to the non-identifiable subset, we consider a simple example (which is also contained in Koelbl et al. (2015)):

EXAMPLE 3.2.3. Assume that $p = 1$, $n_f = n_s = 1$, $N = 2$ and $\Sigma_\nu = I_2$. If we fix the parameters a_{ff}, a_{fs} and a_{ss}, e.g. as

$$y_t = \begin{pmatrix} 0.9 & 0 \\ a_{sf} & 0.8 \end{pmatrix} y_{t-1} + \nu_t \tag{3.2.9}$$

and if we vary $a_{sf} \in \{0, 0.1, 0.25\}$, we obtain the following sections of the likelihood shown in Figure 3.2.2, where only a_{ss} varies. We see that if $a_{sf} = 0$ the likelihood has two maximums.

If we drop the assumption $\sigma_{sf} = 0$, we obtain:

THEOREM 3.2.4. *Assume that $p = 1$, $n_f = n_s = 1$, $\Sigma_\nu > 0$ and $N = 2$. The system and noise parameters $\begin{pmatrix} a_{ff} & a_{fs} \\ a_{sf} & a_{ss} \end{pmatrix}$, σ_{ff}, σ_{sf} and σ_{ss} are not identifiable if and only if they satisfy the equations*

$$\begin{aligned} a_{fs} &= 0, \\ a_{sf} + \frac{\sigma_{sf}}{\sigma_{ff}}(a_{ss} - a_{ff}) &= 0, \\ a_{ss} &\neq 0. \end{aligned} \tag{3.2.10}$$

The complement of the set of solutions of (3.2.10) contains a generic subset of Θ.

PROOF. If σ_{sf} is not necessarily equal to zero, then (3.2.7) does not depend on whether or not σ_{sf} is zero and remains unaffected. Only (3.2.8) is changed to

$$\Sigma_{\tilde{\nu}} = \begin{pmatrix} \sigma_{ff} & \sigma_{sf} & 0 \\ \sigma_{sf} & \sigma_{ss} & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} a_{ff} & a_{fs} \\ a_{sf} & a_{ss} \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \sigma_{ff} & \sigma_{sf} \\ \sigma_{sf} & \sigma_{ss} \end{pmatrix} \begin{pmatrix} a_{ff} & a_{sf} & 1 \\ a_{fs} & a_{ss} & 0 \end{pmatrix}. \tag{3.2.11}$$

Thus a_{ff}, a_{fs} and σ_{ff} are unique for given $\tilde{\mathcal{A}}$, $\Sigma_{\tilde{\nu}}$.

We are left with the problem to uniquely solve equation systems (3.2.7) and (3.2.11) in the variables a_{sf}, a_{ss}, σ_{sf}, and σ_{ss}. Thereto we distinguish two cases, namely the case $a_{fs} = 0$ and the case $a_{fs} \neq 0$ considering that we already know a_{fs}.

We start with the case $a_{fs} \neq 0$. It is easy to see that the missing parameters a_{sf}, a_{ss} and σ_{sf} can be recovered using (3.2.7) and (3.2.11). Subsequently, σ_{ss} can be recovered using equation (3.2.11).

In the event of $a_{fs} = 0$, then A_1^2 is lower triangular with $(2,2)$ entry a_{ss}^2. First, observe that the $(2,3)$ and $(2,1)$ entries of $\Sigma_{\tilde{\nu}}$ are respectively $\sigma_{ff}a_{sf} + \sigma_{sf}a_{ss}$ and $a_{ff}\left(\sigma_{ff}a_{sf} + \sigma_{sf}a_{ss}\right) + \sigma_{sf}$. It is obvious that σ_{sf} is available.

Next, if $a_{ss} = 0$, something which is known from the $(2,2)$ entry of A_1^2, then the $(2,3)$ entry of $\Sigma_{\tilde{\nu}}$ is simply $\sigma_{ff}a_{sf}$ and since σ_{ff} is obviously nonzero, the value of a_{sf} can be obtained. Also, the $(2,2)$ entry of $\Sigma_{\tilde{\nu}}$ is $a_{sf}\left(\sigma_{ff}a_{sf} + \sigma_{sf}a_{ss}\right) + \sigma_{ss}$ and one immediately has σ_{ss}.

It therefore remains to consider the situation where $a_{ss} \neq 0$. The following quantities α and β, corresponding to the $(2,1)$ entry of A_1^2 and the $(2,3)$ entry of $\Sigma_{\tilde{\nu}}$, are known:

$$a_{sf}a_{ff} + a_{ss}a_{sf} = \alpha$$
$$\sigma_{ff}a_{sf} + \sigma_{sf}a_{ss} = \beta$$

By eliminating a_{sf}, we obtain

$$-a_{ss}^2\sigma_{sf} + \left(\beta - a_{ff}\sigma_{sf}\right)a_{ss} + a_{ff}\beta = \alpha\sigma_{ff}.$$

Using this equation and the value for a_{ss}^2 available from A_1^2, it follows that a_{ss} is uniquely determined if and only if

$$\beta - a_{ff}\sigma_{sf} \neq 0.$$

Introducing the expression above for β, this yields:

$$a_{sf}\sigma_{ff} + a_{ss}\sigma_{sf} - a_{ff}\sigma_{sf} \neq 0.$$

To sum up, identification is possible except for parameters satisfying

$$a_{fs} = 0$$
$$a_{ss} \neq 0$$
$$a_{sf}\sigma_{ff} + a_{ss}\sigma_{sf} - a_{ff}\sigma_{sf} = 0.$$

The set of non-identifiable points as described by equations (3.2.10) is a so-called semi-algebraic set. Here, in particular, the set of all identifiable parameters, which is a complement of the semi-algebraic set above, contains a generic subset of the parameter space, viz. the complement of the set defined by the zeros of the polynomial equalities alone. □

An interesting interpretation of Theorem 3.2.4 is the following: The parameters of the underlying high-frequency model cannot be obtained if and only if there is a static linear transformation so that the transformed model has a diagonal innovation covariance and the transformed system matrix is diagonal with nonzero $(2,2)$ entry. Note that such a transformation must be of the form

$$T = \begin{pmatrix} 1 & 0 \\ -\sigma_{sf}\sigma_{ff}^{-1} & 1 \end{pmatrix}$$

and that for given T the conditions of non-identifiability arising are exactly the same as in (3.2.10). Thus, identifiability for systems with non-diagonal innovation covariance can be traced back to identifiability for systems with diagonal innovation covariance.

Note that equations (3.2.7), (3.2.8) may also be used for identifiability analysis for $n = q > 2$ though dealing with the various special cases is more intricate since scalars are replaced by matrices. We repeat that the advantage of the analysis given above is that the subset of identifiable parameters is explicitly given and that the genericity property stands out clearly.

On the other hand, the analysis above cannot be extended to the case $p > 1$ since the blocked process $(\tilde{y}_t)$ is in general no longer AR but ARMA.

3.2.2. g-Identifiability of System Parameters by Extended Yule-Walker Equations. In this section we establish the extended Yule-Walker equations, which are an analogue of the Yule-Walker equations for the mixed-frequency setting. This section with the exception of Example 3.2.5 has been published in Anderson et al. (2015a). For further discussion see also Anderson et al. (2012).

By postmultiplying equation (2.1.6) by $y_{t-j}^T,\ j = 1, \dots, p$ and forming expectations we obtain the Yule-Walker equations. The problem with these equations is that matrices on both the left and right hand side contain unobserved second moments. In order to overcome this problem we postmultiply equation (2.1.6) by $\left(y_{t-j}^f\right)^T,\ j > 0$ and form expectations. Thereby we obtain extended Yule-Walker (XYW) equations proposed in Chen and Zadrozny (1998) as

$$\mathbb{E}\left(y_t\left((y_{t-1}^f)^T, (y_{t-2}^f)^T, \dots\right)\right) = (A_1, \dots, A_p)\mathbb{E}\left(\begin{pmatrix} y_{t-1} \\ \vdots \\ y_{t-p} \end{pmatrix}\left((y_{t-1}^f)^T, (y_{t-2}^f)^T, \dots\right)\right). \tag{3.2.12}$$

Let

$$K := \mathbb{E}\left(x_t\left(y_{t-1}^f\right)^T\right) = \mathbb{E}\left(\begin{pmatrix} y_{t-1} \\ \vdots \\ y_{t-p} \end{pmatrix}\left(y_{t-1}^f\right)^T\right) = \Gamma_p \begin{pmatrix} I_{n_f} \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{3.2.13}$$

From equation (2.1.11), i.e. $x_{t+1} = \mathcal{A}x_t + \mathcal{B}\varepsilon_t$, we see that $x_t = \sum_{i=0}^{\infty} \mathcal{A}^i \mathcal{B}\varepsilon_{t-i-1}$ and $x_{t+s} = \mathcal{A}^s x_t + \sum_{i=0}^{s-1} \mathcal{A}^i \mathcal{B}\varepsilon_{t+s-i-1}$. The block columns of the second matrix on the right hand side of (3.2.12) are of the form

$$\begin{aligned}\mathbb{E}\left(x_t \left(y_{t-j-1}^f\right)^T\right) &= \mathbb{E}\left(x_{t+j}\left(y_{t-1}^f\right)^T\right) = \mathbb{E}\left(\left(\mathcal{A}^j x_t + \sum_{i=0}^{j-1} \mathcal{A}^i \mathcal{B}\varepsilon_{t+j-i-1}\right)\left(y_{t-1}^f\right)^T\right) \\ &= \mathcal{A}^j \mathbb{E}\left(x_t\left(y_{t-1}^f\right)^T\right) = \mathcal{A}^j K,\ j \geq 0.\end{aligned}$$

Thus the rightmost matrix in the extended Yule-Walker equations (3.2.12) can be written as

$$(K, \mathcal{A}K, \mathcal{A}^2 K, \dots). \tag{3.2.14}$$

From the Cayley-Hamilton Theorem and since $\mathcal{A} \in \mathbb{R}^{np \times np}$ we see that the second matrix on the right hand side of (3.2.12) has full row rank if and only if the matrix consisting of the first np blocks has full row rank. In this way we have obtained our XYW equations which are of the form

$$\underbrace{\mathbb{E}\left(y_t\left((y_{t-1}^f)^T, \dots, (y_{t-np}^f)^T\right)\right)}_{=Z_1} = (A_1, \dots, A_p)\, \underbrace{\mathbb{E}\left(\begin{pmatrix} y_{t-1} \\ \vdots \\ y_{t-p}\end{pmatrix}\left((y_{t-1}^f)^T, \dots, (y_{t-np}^f)^T\right)\right)}_{=Z_0}. \tag{3.2.15}$$

The crucial point is that the matrix Z_0 can be written as

$$Z_0 = (K, \mathcal{A}K, \mathcal{A}^2 K, \dots, \mathcal{A}^{np-1} K) \tag{3.2.16}$$

and therefore has the structure of a controllability matrix.

Clearly, the system parameters $(A_1, \dots, A_p)$ of (2.1.6) are identifiable if Z_0 has full row rank np or equivalently the pair $(\mathcal{A}, K)$ is controllable. Note, however, that contrary to usual controllability matrices here K depends on $\mathcal{A}$, which makes the task of verifying generic controllability more demanding.

EXAMPLE 3.2.5. We consider the two-dimensional AR(1) system

$$\begin{pmatrix} y_t^f \\ y_t^s \end{pmatrix} = \begin{pmatrix} 0.5 & 0.5 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} y_{t-1}^f \\ y_{t-1}^s \end{pmatrix} + \begin{pmatrix} \nu_t^f \\ \nu_t^s \end{pmatrix} \tag{3.2.17}$$

where $\Sigma_\nu = I_2$. It immediately follows that this system is identifiable since $a_{ss} = \sigma_{sf} = 0$. Nevertheless, the matrix Z_0 in equation (3.2.15) does not have full row rank:

$$\begin{pmatrix} 0.833 & 0.417 \\ 0 & 0 \end{pmatrix} = Z_1 = A_1 Z_0 = A_1 \begin{pmatrix} 1.667 & 0.833 \\ 0 & 0 \end{pmatrix}$$

This shows that the condition that $\operatorname{rk}(Z_0) = np$ holds is not necessary for identifiability. Indeed, the solutions of the above equation are given as

$$A_1 = Z_1 Z_0^\dagger + H\left(I_2 - Z_0 Z_0^\dagger\right) = \begin{pmatrix} 0.5 & 0 \\ 0 & 0 \end{pmatrix} + H\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \tag{3.2.18}$$

(see Rao and Mitra (1971)) where $H \in \mathbb{R}^{2\times 2}$ is an arbitrary matrix and $Z_0^\dagger$ is the Moore–Penrose pseudoinverse (see Golub and Van Loan (1996)). As we know, for the case $\sigma_{sf} = a_{fs} = a_{sf} = 0$ the classes of observationally equivalent parameters consist of two points (corresponding to the two choices for the square root of a_{ss}^2), whereas the solution set of the XYW equations is a nontrivial affine subset. This shows that the XYW equations do not use the full information contained in the second moments which are directly observed.

The parameter space in this section is the set Θ of all $((A_1, \ldots, A_p), \Sigma_\nu)$ where $(A_1, \ldots, A_p) \in S$ and $\Sigma_\nu \in D$. We analyze identifiability of system parameters first.

The next theorem shows that the matrix Z_0 in equation (3.2.15) is generically of full row rank and thus we have generic identifiability for $(A_1, \ldots, A_p)$. Note that this holds both for regular and singular AR systems for all sampling frequency ratios N and all $n^f \geq 1$.

THEOREM 3.2.6. *The matrix Z_0 in the extended Yule-Walker equations (3.2.15) has full row rank np on a generic subset of the parameter space Θ, and thus the system parameters are generically identifiable.*

PROOF. In a first step, we have to show that Z_0 is a rational function of $\theta \in \Theta$. It follows immediately that Z_0 is rational if we can show that K is a rational function of $\theta \in \Theta$. Vectorizing the Lyapunov equation (2.1.15) we obtain

$$\operatorname{vec}(\Gamma_p) = (\mathcal{A} \otimes \mathcal{A}) \operatorname{vec}(\Gamma_p) + \operatorname{vec}\left(\mathcal{B}\mathcal{B}^T\right)$$

and thus

$$\operatorname{vec}(\Gamma_p) = (I_{(np)^2} - (\mathcal{A} \otimes \mathcal{A}))^{-1} \operatorname{vec}\left(\mathcal{B}\mathcal{B}^T\right). \tag{3.2.19}$$

Note that the absolute value of all eigenvalues λ_j of $\mathcal{A}$ is smaller than one by the stability assumption (2.1.7). Therefore, the same holds for the eigenvalues of $(\mathcal{A} \otimes \mathcal{A})$ since the eigenvalues of $(\mathcal{A} \otimes \mathcal{A})$ are $\lambda_i \lambda_j$ $i, j = 1, \ldots, np$ and thus $\left(I_{(np)^2} - (\mathcal{A} \otimes \mathcal{A})\right)$ is nonsingular. This implies that $\operatorname{vec}(\Gamma_p)$ is a rational function in $((A_1, \ldots, A_p), \Sigma_\nu)$ having no poles in Θ. Thus, K and $\mathcal{A}^j K$ and subsequently Z_0 are rational in $((A_1, \ldots, A_p), \Sigma_\nu)$ on Θ.

Without loss of generality we may restrict ourselves to the case where K is a vector and thus Z_0 is square. Multiplying Z_0 by $\det\left(I_{(np)^2} - (\mathcal{A} \otimes \mathcal{A})\right)$ we obtain a polynomial in the entries of $((A_1, \ldots, A_p), \Sigma_\nu)$ since $\det\left(I_{(np)^2} - (\mathcal{A} \otimes \mathcal{A})\right)$ has no zeros on Θ. Thus, the set of zeros of the determinant of the polynomial matrix $\det\left(I_{(np)^2} - (\mathcal{A} \otimes \mathcal{A})\right) Z_0$ is the same as the set of zeros of the determinant of Z_0 and thus is an algebraic set in Θ (compare Bochnak et al. (1998) page 23).

Now consider a point θ^* in Θ given by

$$\mathcal{A}^* = \begin{pmatrix} 0 & \cdots & 0 & \rho^p C \\ I_n & & & \\ & \ddots & & \\ & & I_n & 0 \end{pmatrix}, \; \mathcal{B}^* = \mathcal{E}_1 = \begin{pmatrix} e_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{3.2.20}$$

where $\rho \in (0, 1)$, C is again the circulant matrix (see Lemma 3.1.4) and $e_1 \in \mathbb{R}^n$ is the first unit vector. We will show that for this point in the parameter space $\det(Z_0) \neq 0$ holds:

As can be seen in the proof of Lemma 3.1.8 for the point θ^*, Γ_p is a diagonal and nonsingular matrix which implies that $\left(\mathcal{B}, \mathcal{AB}, \ldots, \mathcal{A}^{np-1}\mathcal{B}\right)$ is of full row rank. Now it is immediate that Z_0 is of full row rank since, as Γ_p is diagonal, $Z_0 = \left(\Gamma_p\mathcal{E}_1, \mathcal{A}\Gamma_p\mathcal{E}_1, \ldots, \mathcal{A}^{np-1}\Gamma_p\mathcal{E}_1\right)$ is a multiple of $\left(\mathcal{B}, \mathcal{AB}, \ldots, \mathcal{A}^{np-1}\mathcal{B}\right)$ and therefore $\det(Z_0) \neq 0$ holds.

Thus, the set of zeros of $\det(Z_0)$ is a proper algebraic set, i.e. an algebraic set of dimension smaller than the dimension of Θ. Therefore, its complement in the parameter space, which corresponds to all controllable pairs, is the complement of a proper algebraic set and hence is generic in the parameter space. □

REMARK 3.2.7. Instead of postmultiplying $\left(y_t^f\right)^T$ in equation (3.2.15) only the first or at least one component of y_t^f is needed to guarantee that the system parameters are g-identifiable. This issue can be easily seen in the proof of Theorem 3.2.6. Of course, this causes a further information loss.

3.2.3. g-Identifiability of the Noise Parameters and of the 4th Moments. In this section we consider g-identifiability of the noise covariance matrix and of the fourth moment (if it exists). The first part of this section until Remark 3.2.9 has been published in Anderson et al. (2015a), see also Anderson et al. (2012). Let us define $\mathcal{G} = (I_n, 0, \ldots, 0)$.

THEOREM 3.2.8. *The noise parameters Σ_ν are g-identifiable in Θ from*

$$vec(\Sigma_\nu) = \left((\mathcal{G}\otimes\mathcal{G})\left(I_{(np)^2} - (\mathcal{A}\otimes\mathcal{A})\right)^{-1}\left(\mathcal{G}^T\otimes\mathcal{G}^T\right)\right)^{-1} vec(\gamma(0)). \tag{3.2.21}$$

PROOF. We commence from identifiable system parameters $(A_1, \ldots, A_p)$. Through columnwise vectorization of

$$\gamma(0) = \mathbb{E}\left(y_t y_t^T\right) = \mathcal{G}\Gamma_p\mathcal{G}^T$$

we obtain

$$\mathrm{vec}(\gamma(0)) = (\mathcal{G}\otimes\mathcal{G})\mathrm{vec}(\Gamma_p).$$

This together with (3.2.19) leads to

$$\mathrm{vec}(\gamma(0)) = (\mathcal{G}\otimes\mathcal{G})\left(I_{(np)^2} - (\mathcal{A}\otimes\mathcal{A})\right)^{-1}(\mathcal{G}^T\otimes\mathcal{G}^T)\mathrm{vec}(\Sigma_\nu) \tag{3.2.22}$$

where we used $\mathcal{B}\mathcal{B}^T = \mathcal{G}^T\Sigma_\nu\mathcal{G}$.

Note that $\left(I_{(np)^2} - (\mathcal{A}\otimes\mathcal{A})\right)$ is nonsingular. $(\mathcal{G}\otimes\mathcal{G})\left(I_{(np)^2} - (\mathcal{A}\otimes\mathcal{A})\right)^{-1}\left(\mathcal{G}^T\otimes\mathcal{G}^T\right)$ is a function rational in $(A_1, \ldots, A_p)$ having no poles. For $A_1 = \cdots = A_p = 0$, the matrix $\left(I_{(np)^2} - (\mathcal{A}\otimes\mathcal{A})\right)^{-1}$ is triangular with ones on its diagonal. Thus, in view of the particular form of $\mathcal{G}$

$$(\mathcal{G}\otimes\mathcal{G})\left(I_{(np)^2} - (\mathcal{A}\otimes\mathcal{A})\right)^{-1}\left(\mathcal{G}^T\otimes\mathcal{G}^T\right) = \sum_{j=0}^{\infty}(k_j\otimes k_j)$$

is a principal submatrix of $\left(I_{(np)^2} - (\mathcal{A}\otimes\mathcal{A})\right)^{-1}$ with the same property and is therefore nonsingular. Thus, the set of zeros of this function is a proper algebraic set on Θ not depending on Σ_ν. On the complement of this proper algebraic set we have

$$\mathrm{vec}(\Sigma_\nu) = \left((\mathcal{G}\otimes\mathcal{G})\left(I_{(np)^2} - (\mathcal{A}\otimes\mathcal{A})\right)^{-1}(\mathcal{G}^T\otimes\mathcal{G}^T)\right)^{-1}\mathrm{vec}(\gamma(0)).$$

□

REMARK 3.2.9. From Theorems 3.2.6 and 3.2.8 we see that the system and noise parameters are g-identifiable, i.e. identifiable on the intersection of the set described in the proofs of Theorem 3.2.6 and Theorem 3.2.8. Note that the results shown in the proofs above are stronger than the genericity results because the set where Z_0 has not full row rank np is a proper algebraic set (see Wonham (1985) p. 28) and the same statement holds for the case of noise parameters.

Note that the property that Z_0 has full row rank np depends on $(A_1, \ldots, A_p)$ as well as on Σ_ν whereas the uniqueness of Σ_ν obtained via (3.2.21) only depends on $(A_1, \ldots, A_p)$.

If the system (2.1.11), (2.1.12) is not controllable, i.e. if Γ_p is singular, then we clearly have non-identifiability even for high-frequency data as the Yule-Walker equations then have no unique solution.

REMARK 3.2.10. We have not been able to give an explicit description of those elements in Θ which are not identifiable or those parameters where Z_0 is not of full row rank np. In Anderson et al. (2015a) another approach to identify the system and noise parameters which is based on blocking has been presented. The advantages of the blocking approach are on the one hand that we can specify sufficient conditions under which the parameters are identifiable and on the other hand all second moments which are directly observed are used.

An alternative method to identify the noise parameters is to reconstruct the missing autocovariances from the observed autocovariances (3.2.3) and then use equation (2.1.35) to identify the noise parameters. As will be shown in the next theorem, again this procedure only uses assumptions on the system parameters.

By using the Yule-Walker equations we observe that by choosing $k = \lceil \frac{p+1}{N} \rceil$, where $\lceil j \rceil$ leads to the smallest integer greater than or equal to j, it holds

$$\gamma(kN) = \mathcal{GA} \begin{pmatrix} \gamma(kN-1) \\ \gamma(kN-2) \\ \vdots \\ \gamma(kN-p) \end{pmatrix} = \mathcal{GA}^2 \begin{pmatrix} \gamma(kN-2) \\ \gamma(kN-3) \\ \vdots \\ \gamma(kN-p-1) \end{pmatrix} = \ldots = \mathcal{GA}^{kN-p} \begin{pmatrix} \gamma(p) \\ \vdots \\ \gamma(1) \end{pmatrix}.$$

Thereby, we can construct the following system of equations

$$\underbrace{\begin{pmatrix} \gamma(kN) \\ \gamma((k+1)N) \\ \vdots \\ \gamma((k+np-1)N) \end{pmatrix}}_{=\mathcal{O}_1} = \underbrace{\begin{pmatrix} \mathcal{GA}^{kN-p} \\ \mathcal{GA}^{(k+1)N-p} \\ \vdots \\ \mathcal{GA}^{(k+np-1)N-p} \end{pmatrix}}_{=\mathcal{O}} \begin{pmatrix} \gamma(p) \\ \vdots \\ \gamma(1) \end{pmatrix} \tag{3.2.23}$$

where the matrix on the left hand side of equation (3.2.23) can be directly observed from mixed-frequency data.

THEOREM 3.2.11. *For $(A_1, \dots, A_p) \in S$, if A_p is nonsingular and if for two eigenvalues $\lambda_i \neq \lambda_j$ of $\mathcal{A}$ it follows that $\lambda_i^N \neq \lambda_j^N$ holds, the missing autocovariances $\gamma^{ss}(p)$, $p \neq sN$, $s \in \mathbb{Z}$ can be reconstructed from the one which are directly observed.*

PROOF. We have to show that the matrix $\mathcal{O}$ has full column rank. Notice that $\mathcal{O}$ has an observability structure with the matrices $\left(\mathcal{G}\mathcal{A}^{kN-p}, \mathcal{A}^N\right)$. Since we assume that A_p is nonsingular, there exists no eigenvalue of $\mathcal{A}$ which is zero. Also note that, under our assumptions, if p_i is an eigenvector of $\mathcal{A}$ with $\lambda_i \neq 0$, then it is an eigenvector of $\mathcal{A}^N$ with $\lambda_i^N \neq 0$ and vice versa. To obtain a contradiction, suppose that $\mathcal{O}$ fails at having full column rank. Here we can use the PBH test of Theorem 2.1.6 to conclude that there exists an eigenvector p_i with corresponding eigenvalue $\lambda_i \neq 0$ of $\mathcal{A}$ so that

$$\begin{pmatrix} \mathcal{A}^N - I\lambda_i^N \\ \mathcal{G}\mathcal{A}^{kN-p} \end{pmatrix} p_i = \begin{pmatrix} 0 \\ (I_n, 0, \dots, 0)\, p_i \lambda_i^{kN-p} \end{pmatrix} = 0. \tag{3.2.24}$$

We can immediately conclude that $(I_n, 0, \dots, 0)\, p_i \lambda_i^{kN-p} = p_i^1 \lambda_i^{kN-p} = 0$, which is a contradiction to Lemma 2.1.7 because then p_i would be zero and therefore no eigenvector. This implies that $\mathcal{O}$ has full column rank. Therefore, we can reconstruct $\gamma(i)$, $i = 1, \dots, p$ with the Moore–Penrose pseudoinverse of $\mathcal{O}$, i.e.

$$\begin{pmatrix} \gamma(p) \\ \vdots \\ \gamma(1) \end{pmatrix} = \left(\mathcal{O}^T\mathcal{O}\right)^{-1} \mathcal{O}^T \mathcal{O}_1. \tag{3.2.25}$$

Now the non-observed autocovariances can be obtained via

$$\begin{aligned} \gamma(p+1) &= (A_1, \dots, A_p) \begin{pmatrix} \gamma(p) \\ \vdots \\ \gamma(1) \end{pmatrix} \\ \gamma(-p) &= \gamma(p)^T. \end{aligned} \tag{3.2.26}$$

□

By using Theorem 3.2.11 and equation (2.1.35) we directly get the next proposition:

PROPOSITION 3.2.12. *For $(A_1, \dots, A_p) \in S$, if A_p is nonsingular and if for two eigenvalues $\lambda_i \neq \lambda_j$ of $\mathcal{A}$ it follows that $\lambda_i^N \neq \lambda_j^N$ holds, the noise parameters Σ_ν are identifiable.*

REMARK 3.2.13. The assumptions on the system parameters in Proposition 3.2.12 are generic in the parameter space Θ (see Lemmas 3.1.3 and 3.1.6).

Theorem 3.2.11 can be modified to obtain g-identifiability for the MA part in the ARMA case under further assumptions. One of these assumptions is that the order of the moving average polynomial is smaller than or equal to p (see Anderson et al. (2015b)).

The next theorem states that the fourth moment of the inputs is g-identifiable, which is needed for the asymptotic behavior of the XYW estimator in Chapter 5.

THEOREM 3.2.14. *Let $(y_t)_{t\in\mathbb{Z}}$ be the output of system (2.1.6) with inputs $(\nu_t) \sim IID_n(0,\Sigma_\nu)$, assume that the fourth moment $\eta = \mathbb{E}\left(\nu_t\nu_t^T \otimes \nu_t\nu_t^T\right)$ of ν_t exists and let*

$$\kappa = \eta - vec\left(\Sigma_\nu\right) vec\left(\Sigma_\nu\right)^T - \left(\Sigma_\nu \otimes \Sigma_\nu\right) - K_{n,n}\left(\Sigma_\nu \otimes \Sigma_\nu\right).$$

Then κ is g-identifiable in Θ from

$$vec\left(\kappa\right) = \left(\mathcal{G}_2\left(I_{(np)^4} - \left(\mathcal{A}\otimes\mathcal{A}\otimes\mathcal{A}\otimes\mathcal{A}\right)\right)^{-1}\mathcal{G}_2^T\right)^{-1} vec\left(\psi\right) \tag{3.2.27}$$

where ψ is defined in the proof and $\mathcal{G}_2 = (\mathcal{G}\otimes\mathcal{G}\otimes\mathcal{G}\otimes\mathcal{G})$.

PROOF. It is easy to show that under our assumptions

$$\begin{aligned}\mathbb{E}\left(y_t y_t^T \otimes y_t y_t^T\right) &= \operatorname{vec}\left(\gamma\left(0\right)\right)\operatorname{vec}\left(\gamma\left(0\right)\right)^T + \left(\gamma\left(0\right)\otimes\gamma\left(0\right)\right) &(3.2.28)\\ &\quad + K_{n,n}\left(\gamma\left(0\right)\otimes\gamma\left(0\right)\right) + S_{0,0} &(3.2.29)\end{aligned}$$

where $S_{0,0} = \sum_{i=0}^{\infty}\left(k_i\otimes k_i\right)\kappa\left(k_i\otimes k_i\right)^T$ (compare Lemma 5.1.2 for the case $N=1$, $p=q=0$ and omit the superscript f and s). Reformulating and vectorizing equation (3.2.28) leads to

$$\operatorname{vec}\left(\psi\right) = \left(\sum_{i=0}^{\infty}\left(k_i\otimes k_i\otimes k_i\otimes k_i\right)\right)\operatorname{vec}\left(\kappa\right) \tag{3.2.30}$$

where $\psi = \mathbb{E}\left(y_t y_t^T \otimes y_t y_t^T\right) - \operatorname{vec}\left(\gamma\left(0\right)\right)\operatorname{vec}\left(\gamma\left(0\right)\right)^T - \left(\gamma\left(0\right)\otimes\gamma\left(0\right)\right) - K_{n,n}\left(\gamma\left(0\right)\otimes\gamma\left(0\right)\right)$. Note that, under the further assumption that the eighth moment $\mathbb{E}\left(\nu_t\nu_t^T\otimes\nu_t\nu_t^T\otimes\nu_t\nu_t^T\otimes\nu_t\nu_t^T\right)$ of ν_t exists, ψ can be consistently estimated from mixed-frequency data, see Theorem 5.3.2. Since $k_j = \mathcal{G}\mathcal{A}^j\mathcal{G}^T$ holds, the left term of the right hand side of (3.2.30) can be rewritten as

$$\begin{aligned}\sum_{i=0}^{\infty}\left(k_i\otimes k_i\otimes k_i\otimes k_i\right) &= \sum_{i=0}^{\infty}\mathcal{G}_2\left(\mathcal{A}^i\otimes\mathcal{A}^i\otimes\mathcal{A}^i\otimes\mathcal{A}^i\right)\mathcal{G}_2^T\\ &= \left(\mathcal{G}_2\left(I_{(np)^4} - \left(\mathcal{A}\otimes\mathcal{A}\otimes\mathcal{A}\otimes\mathcal{A}\right)\right)^{-1}\mathcal{G}_2^T\right). &(3.2.31)\end{aligned}$$

Following the same arguments as in the proof of Theorem 3.2.8 we can conclude that the right hand side of equation (3.2.31) is nonsingular on a generic subset and thus

$$\operatorname{vec}\left(\kappa\right) = \left(\mathcal{G}_2\left(I_{(np)^4} - \left(\mathcal{A}\otimes\mathcal{A}\otimes\mathcal{A}\otimes\mathcal{A}\right)\right)^{-1}\mathcal{G}_2^T\right)^{-1}\operatorname{vec}\left(\psi\right).$$

□

One big drawback of equation (3.2.27) is that the dimension of the matrix in the middle, i.e. $\left(I_{(np)^4} - \left(\mathcal{A}\otimes\mathcal{A}\otimes\mathcal{A}\otimes\mathcal{A}\right)\right)$, increases rapidly with np. For instance, for $n=4$ and $p=3$ the dimension of the matrix is 20736. In practice, of course, it is impossible to invert this matrix in a proper calculation time. One way out would be to use equation (3.2.30) and invert the first matrix on the right hand side, which has dimension n^4 or 256 in our example.

3.2.4. g-Identifiability of the System and Noise Parameters in the Case of Flow Variables. The first part of this section until Theorem 3.2.15 has been published in Anderson et al. (2015a).

In the previous sections only stock variables have been considered. Here we deal with the case where the process $(y_t^s)_{t\in\mathbb{Z}}$ consists of flow variables or variables aggregated by more general schemes.

For flow variables the aggregation to the corresponding observed process, $(w_t)_{t\in N\mathbb{Z}}$ say, is of the form

$$w_t = y_t^s + y_{t-1}^s + \cdots + y_{t-N+1}^s = (1 + z + \cdots + z^{N-1})y_t^s,\ t \in N\mathbb{Z}. \tag{3.2.32}$$

Note that the second moments required in the extended Yule-Walker equations are the autocovariances $\mathbb{E}\left(y_{t+h}^f \left(y_t^f\right)^T\right)$, $h \in \mathbb{Z}$ and the cross covariances $\mathbb{E}\left(y_{t+h}^s \left(y_t^f\right)^T\right)$, $h \in \mathbb{Z}$. We now show how these cross covariances can be retrieved from the cross covariances $\mathbb{E}\left(w_{t+h} \left(y_t^f\right)^T\right)$, $h \in \mathbb{Z}$ of the observations.

To show this, assume for the moment that w_t is available for all $t \in \mathbb{Z}$ and that the inverse of the linear transformation (3.2.32) exists for all $t \in \mathbb{Z}$, i.e.

$$y_t^s = \underset{M\to\infty}{\text{l.i.m.}} \sum_{j=0}^{M} h_j^{(M)} w_{t-j}, \quad h_j^{(M)} \in \mathbb{R}^{n_s\times n_s},\ t \in \mathbb{Z} \tag{3.2.33}$$

where again l.i.m. denotes the limit in mean square. Then

$$\gamma^{sf}(h) = \mathbb{E}\left(y_{t+h}^s \left(y_t^f\right)^T\right) = \lim_{M\to\infty} \sum_{j=0}^{M} h_j^{(M)} \underbrace{\mathbb{E}\left(w_{t+h-j} \left(y_t^f\right)^T\right)}_{\gamma^{wf}(h-j)}. \tag{3.2.34}$$

The transfer function $\left(1 + z + \cdots + z^{N-1}\right) I_{n_s}$ in (3.2.32) is inverted by inverting

$$\left(1 + e^{-i\lambda} + \cdots + e^{-i(N-1)\lambda}\right) = \prod_{k=1}^{N-1} \left(1 - \frac{e^{-i\lambda}}{e^{i2\pi\frac{k}{N}}}\right)$$

by using a partial fraction expansion of $\prod_{k=1}^{N-1} \left(1 - \frac{e^{-i\lambda}}{e^{i2\pi\frac{k}{N}}}\right)^{-1}$ and representing the inverse of each linear factor $\left(1 - \frac{e^{-i\lambda}}{e^{i2\pi\frac{k}{N}}}\right)$ as the limit of its Cesaro sum, i.e.

$$\left(1 - \frac{e^{-i\lambda}}{e^{i2\pi\frac{k}{N}}}\right)^{-1} = \lim_{M\to\infty} \sum_{j=0}^{M} \left(1 - \frac{j}{M}\right) e^{-i\left(\lambda + 2\pi\frac{k}{M}\right)j}.$$

Note that for our purposes the inverse of the linear transformation (3.2.33) only has to exist for the special input $(w_t)_{t\in\mathbb{Z}}$. In order to show the existence of the inverse transformation (3.2.33), it is more convenient to use the frequency domain rather than the time domain, see Rozanov (1967); Hannan (1970). Let

$$f_{y^s y^s}(\lambda) = (2\pi)^{-1} \sum_{h=-\infty}^{\infty} \gamma^{ss}(h) e^{-i\lambda h}$$

and

$$f_{ww}(\lambda) = (2\pi)^{-1} \sum_{h=-\infty}^{\infty} \mathbb{E}\left(w_{t+h} \left(w_t\right)^T\right) e^{-i\lambda h}$$

denote the spectral density of $(y_t^s)_{t\in\mathbb{Z}}$ and $(w_t)_{t\in\mathbb{Z}}$, respectively. As is well known, the spectral density $f_{ww}(\lambda)$ of $(w_t)_{t\in\mathbb{Z}}$ satisfies

$$f_{ww}(\lambda) = \left(1 + e^{-i\lambda} + \cdots + e^{-i(N-1)\lambda}\right) I_{n_s} f_{y^s y^s}(\lambda) I_{n_s} \left(1 + e^{i\lambda} + \cdots + e^{i(N-1)\lambda}\right)$$

and thus

$$\int \left(1 + e^{-i\lambda} + \cdots + e^{-i(N-1)\lambda}\right)^{-1} I_{n_s} f_{ww}(\lambda) I_{n_s} \left(1 + e^{i\lambda} + \cdots + e^{i(N-1)\lambda}\right)^{-1} d\lambda$$
$$= \int f_{y^s y^s}(\lambda) d\lambda < \infty.$$

Therefore, each row of $\left(1 + e^{-i\lambda} + \cdots + e^{-i(N-1)\lambda}\right)^{-1} I_{n_s}$ is an element of the frequency domain $\mathcal{L}_2\left(f_{ww} d\lambda\right)$ of f_{ww} and by the isomorphism between the frequency and the time domain the inverse transformation (3.2.33) is well defined. From (3.2.34) we then obtain

$$f_{y^s y^f}(\lambda) = (2\pi)^{-1} \sum_{h=-\infty}^{\infty} \gamma^{sf}(h) e^{-i\lambda h} = \left(1 + e^{-i\lambda} + \cdots + e^{-i(N-1)\lambda}\right)^{-1} I_{n_s} f_{wy^f}(\lambda) \tag{3.2.35}$$

and thus $\gamma^{sf}(h)$, $h \in \mathbb{Z}$. In this way, we get all covariances in the extended Yule-Walker equations. Note that, in contrast to stock variables, these covariances are the covariances of the observations, in the case considered here they have to be reconstructed as described above.

A completely analogous derivation holds if we replace (3.2.32) by the more general aggregation scheme

$$w_t = C_1 y_t^s + C_2 y_{t-1}^s + \cdots + C_N y_{t-N+1}^s \tag{3.2.36}$$

where $C_i \in \mathbb{R}^{n_s \times n_s}$ and C_1 is nonsingular. Thus, taking into account that generically Z_0 has row rank equal to np, we obtain the following result:

THEOREM 3.2.15. *Given the aggregation scheme (3.2.36) for the slow variables $(w_t)_{t\in N\mathbb{Z}}$, the system and noise parameters of the high-frequency system (2.1.6) are g-identifiable from $\gamma^{ff}(h)$ and $\gamma^{wf}(h)$, $h \in \mathbb{Z}$.*

Note that if we set $C_1 = I$ and $C_j = 0$, $j = 2, \ldots, N$, we have the case of stock variables. Thus, Theorems 3.2.6 and 3.2.8 are special cases of Theorem 3.2.15. As it can be immediately seen, Theorem 3.2.15 also covers the case where the slow variables are formed by a mixture of stock and flow variables.

In the following lines we introduce an alternative realization procedure where we assume that the slow component is observed by an aggregation scheme which is based on a simple structure. Simple structure means in this context that we assume aggregation schemes which are based, for example on the mean or the sum of the high-frequency observations. To be more precise, we assume that the aggregation scheme has the following representation:

$$w_t \quad = \quad c_1 y_t^s + c_2 y_{t-1}^s + \ldots + c_N y_{t-N+1}^s \tag{3.2.37}$$

where c_i, $i = 1, \ldots, N$, are real scalars and at least one c_i is unequal to zero. Examples of these scalar weights are $c_i = 1$, $i = 1, \ldots, N$, which represents the sum and $c_i = 1/N$, $i = 1, \ldots, N$ which represents the mean. For the case $n_s = 1$ the aggregation scheme in (3.2.36) and (3.2.37) coincide. We

can rewrite the cross autocovariance function of w_t and y_t^f as

$$\begin{aligned}\gamma^{wf}(h) &= \sum_{i=1}^{N} c_i \gamma^{sf}(h-i+1) \\ &= c_1\gamma^{sf}(h) + c_2\gamma^{sf}(h-1) + \ldots + c_N\gamma^{sf}(h-N+1).\end{aligned}$$

Furthermore, let

$$z_t = \begin{pmatrix} z_t^f \\ z_t^s \end{pmatrix} = \sum_{i=1}^{N} c_i y_{t-i+1} = \begin{pmatrix} \sum_{i=1}^{N} c_i y_{t-i+1}^f \\ w_t \end{pmatrix},$$

$\gamma_z(h) = \mathbb{E}\left(z_t z_{t-h}^T\right) = \sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j \gamma(j+h-i)$, $\gamma^{z^s f}(h) = \mathbb{E}\left(z_t^s \left(y_{t-h}^f\right)^T\right) = \gamma^{wf}(h)$ and $\gamma^{z^f f}(h) = \mathbb{E}\left(z_t^f \left(y_{t-h}^f\right)^T\right) = \sum_{i=1}^{N} c_i \gamma^{ff}(h-i+1)$. For simplicity we assume that $c_N \neq 0$. In this case we obtain another form of the XYW equations as follows:

$$\underbrace{\begin{pmatrix} \overbrace{c_N\gamma^{ff}(1) + \ldots + c_1\gamma^{ff}(N)}^{\gamma^{z^f f}(N)} & \gamma^{z^f f}(N+1) & \cdots & \overbrace{c_N\gamma^{ff}(np) + \ldots + c_1\gamma^{ff}(N+np-1)}^{\gamma^{z^f f}(N+np-1)} \\ \underbrace{c_N\gamma^{sf}(1) + \ldots + c_1\gamma^{sf}(N)}_{=\gamma^{wf}(N)} & \gamma^{wf}(N+1) & \cdots & \underbrace{c_N\gamma^{sf}(np) + \ldots + c_1\gamma^{sf}(N+np-1)}_{=\gamma^{wf}(N+np-1)} \end{pmatrix}}_{Z_1^g} =$$

$$(A_1, \ldots, A_p) \underbrace{\begin{pmatrix} \gamma^{z^f f}(N-1) & \gamma^{z^f f}(N) & \cdots & \gamma^{z^f f}(N+np-2) \\ \gamma^{wf}(N-1) & \gamma^{wf}(N) & \cdots & \gamma^{wf}(N+np-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma^{z^f f}(N-p) & \gamma^{z^f f}(N-p+1) & \cdots & \gamma^{z^f f}(N+(n-1)p-1) \\ \gamma^{wf}(N-p) & \gamma^{wf}(N-p+1) & \cdots & \gamma^{wf}(N+(n-1)p-1) \end{pmatrix}}_{Z_0^g}.$$

Again one can ask when the matrix Z_0^g has full row rank and thus the system parameters are identifiable. It is easy to see that the matrix Z_0^g can be rewritten as

$$\begin{aligned} Z_0^g &= \left(K, \mathcal{A}K, \mathcal{A}^2K, \ldots, \mathcal{A}^{np-1}K, \ldots, \mathcal{A}^{N+np-2}K\right)\left(I_{np} \otimes (c_N, \ldots, c_1)^T \otimes I_{n_f}\right) \\ &= \underbrace{\left(c_N I_{np} + c_{N-1}\mathcal{A} + \cdots + c_1\mathcal{A}^{N-1}\right)}_{F}\underbrace{\left(K, \mathcal{A}K, \mathcal{A}^2K, \ldots, \mathcal{A}^{np-1}K\right)}_{Z_0}. \end{aligned}$$

Since we already know that the matrix Z_0 has generically full row rank we have to check if the matrix F is at least generically nonsingular. For this purpose we assume that A_p is nonsingular (and thus $\mathcal{A}$ is nonsingular, too) and that the eigenvalues of $\mathcal{A}$ are distinct. Again, these two assumptions are fulfilled generically in the parameter space (see Lemmas 3.1.3 and 3.1.5). Thus, $\mathcal{A}$ has the representation $\mathcal{A} = P\Lambda P^{-1}$, where Λ is a diagonal matrix containing the eigenvalues of $\mathcal{A}$ and $P = (p_1, \ldots, p_{np})$ are the corresponding eigenvectors. Thus, we can write F in the following form

$$F = P\left(c_N I_{np} + c_{N-1}\Lambda + \cdots + c_1\Lambda^{N-1}\right)P^{-1}.$$

Since we want to show that F is generically nonsingular we have to look at each diagonal entry in $c_N I_{np} + c_{N-1}\Lambda + \ldots + c_1\Lambda^{N-1}$, which has to be generically unequal to zero. Thus, we have to show that for a fixed simple aggregation scheme $(c_1, \ldots, c_N)$ generically no eigenvalue of $\mathcal{A}$ is a root of $p(\lambda) = c_N + c_{N-1}\lambda + \ldots + c_1\lambda^{N-1}$. Let $(\lambda_1, \ldots, \lambda_{N-1})$ be the roots of $p(\lambda)$. Then Lemma 3.1.4 guarantees that generically no root of $p(\lambda)$ is an eigenvalue of $\mathcal{A}$ and for this reason F is generically nonsingular. Thus, we have shown:

THEOREM 3.2.16. *Given the aggregation scheme (3.2.37) for the slow variables $(w_t)_{t\in N\mathbb{Z}}$, the matrix Z_0^g has full row rank on a generic subset of the parameter space.*

The advantage of the XYW equations in this case case over the first approach, in which the missing autocovariances have to be reconstructed, is that the spectral density is not needed. Of course, one drawback is that only scalar weights are allowed. Note that if we set $c_i = 0$, $i = 1, \ldots, N-1$ and $c_N = 1$, we obtain that $Z_0^{\mathrm{g}} = Z_0$ and $Z_1^{\mathrm{g}} = Z_1$.

For the noise parameters we can arrange the following system of equations which rely on the equation set (3.2.23):

$$\underbrace{\begin{pmatrix} \gamma_z(N) \\ \gamma_z(2N) \\ \gamma_z(3N) \\ \vdots \\ \gamma_z((np-1)N) \end{pmatrix}}_{\mathcal{O}_1^{\mathrm{g}}} = \underbrace{\begin{pmatrix} \mathcal{G}\left(\sum_{l=1}^N \sum_{j=1}^N c_l c_j \mathcal{A}^{j-l+N}\right) \\ \mathcal{G}\left(\sum_{l=1}^N \sum_{j=1}^N c_l c_j \mathcal{A}^{j-l+N}\right)\mathcal{A}^N \\ \mathcal{G}\left(\sum_{l=1}^N \sum_{j=1}^N c_l c_j \mathcal{A}^{j-l+N}\right)\mathcal{A}^{2N} \\ \vdots \\ \mathcal{G}\left(\sum_{l=1}^N \sum_{j=1}^N c_l c_j \mathcal{A}^{j-l+N}\right)\mathcal{A}^{N(np-1)} \end{pmatrix}}_{\mathcal{O}^{\mathrm{g}}} \Gamma_p \mathcal{G}^T \tag{3.2.38}$$

where the left hand side of equation (3.2.38) is available.

THEOREM 3.2.17. *Let the aggregation scheme (3.2.37), $p(z) = c_1 + c_2 z + \ldots + c_N z^{N-1}$, for the slow variables $(w_t)_{t\in N\mathbb{Z}}$ be given and $(A_1, \ldots, A_p) \in S$. Assume that A_p is nonsingular, $\det(a(z))$ and $p(z)$ as well as $\det\left(a\left(z^{-1}\right)\right)$ and $p(z)$ are co-prime and if for two eigenvalues $\lambda_i \neq \lambda_j$ of $\mathcal{A}$ it follows that $\lambda_i^N \neq \lambda_j^N$ holds, then the autocovariances $\gamma(s)$, $s \in \mathbb{Z}$ can be reconstructed from the ones which are observed.*

PROOF. As in the proof of Theorem 3.2.11 we will use the PBH test to obtain our result. Because of our assumptions the following equalities hold

$$\begin{aligned} \begin{pmatrix} \mathcal{A}^N - I_{np}\lambda_i^N \\ \mathcal{G}\left(\sum_{l=1}^N \sum_{j=1}^N c_l c_j \mathcal{A}^{j-l+N}\right) \end{pmatrix} p_i &= \begin{pmatrix} 0 \\ \mathcal{G}\left(\sum_{l=1}^N \sum_{j=1}^N c_l c_j \mathcal{A}^{j-l+N}\right) p_i \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ \mathcal{G} p_i \left(\sum_{l=1}^N \sum_{j=1}^N c_l c_j \lambda_i^{j-l+N}\right) \end{pmatrix} \end{aligned}$$

where λ_i is an eigenvalue of $\mathcal{A}$ and p_i the corresponding eigenvector. Rewriting $\sum_{l=1}^N \sum_{j=1}^N c_l c_j \lambda^{j-l+N} = \lambda^N \left(\sum_{j=1}^N c_j \lambda^{j-1}\right)\left(\sum_{l=1}^N c_l \lambda^{-l+1}\right)$ and considering that we assume that $\det(a(z))$ and $p(z)$ and that

$\det\left(a\left(z^{-1}\right)\right)$ and $p(z)$ are left co-prime, we establish that the matrix $\mathcal{O}^{\mathrm{g}}$ has full column rank and thus

$$\Gamma_p \mathcal{G}^T = \left(\left(\mathcal{O}^{\mathrm{g}}\right)^T \mathcal{O}^{\mathrm{g}}\right)^{-1} \left(\mathcal{O}^{\mathrm{g}}\right)^T \mathcal{O}_1^{\mathrm{g}}.$$

□

Note that the assumptions of Theorem 3.2.17 are generically fulfilled on the parameter space Θ. If $c_i = c \neq 0$, $i = 1, \ldots, N$ for a $c \in \mathbb{R}$ then $p(z)$ has all its roots on the unit circle and thus the additional assumptions for the scheme (3.2.37), i.e. $\det(a(z))$ and $p(z)$ as well as $\det\left(a\left(z^{-1}\right)\right)$ and $p(z)$ are co-prime, are fulfilled.

If we want to extend the result of Theorem 3.2.8 to the simple aggregation scheme (3.2.37), we have to define an alternative (maybe non-minimal) state space system: Let

$$x_{t+1}^m = \mathcal{A}_m x_t^m + \mathcal{G}_m^T \nu_t \tag{3.2.39}$$

$$z_t = \mathcal{H}_m x_{t+1}^m, \tag{3.2.40}$$

where $m = \max(p, N)$, $x_t^m = \left(y_{t-1}^T, \ldots, y_{t-m}^T\right)^T$, $\mathcal{H}_m = \left(c_1 I_n, c_2 I_n, \ldots, c_N I_n, 0_{n \times n(m-N)}\right)$, $\mathcal{G}_m = \left(I_n, 0_{n \times n(m-1)}\right)$ and $\mathcal{A}_m$ be the companion form with system parameters

$$A_{m,i} = \begin{cases} A_i & i = 1, \ldots, p \\ 0 & i = p+1, \ldots, m \end{cases}.$$

It is easy to conclude that $\gamma_z(0) = \mathcal{H}_m \Gamma_m \mathcal{H}_m^T$ holds for $\Gamma_m = \mathbb{E}\left(x_t^m \left(x_t^m\right)^T\right)$ and that

$$\operatorname{vec}(\Gamma_m) = \left(\left(I_{(nm)^2} - (\mathcal{A}_m \otimes \mathcal{A}_m)\right)^{-1} \left(\mathcal{G}_m^T \otimes \mathcal{G}_m^T\right)\right) \operatorname{vec}(\Sigma_\nu).$$

Therefore, we have that

$$\operatorname{vec}(\gamma_z(0)) = \left((\mathcal{H}_m \otimes \mathcal{H}_m)\left(I_{(nm)^2} - (\mathcal{A}_m \otimes \mathcal{A}_m)\right)^{-1} \left(\mathcal{G}_m^T \otimes \mathcal{G}_m^T\right)\right) \operatorname{vec}(\Sigma_\nu).$$

THEOREM 3.2.18. *Given the aggregation scheme (3.2.37) for the slow variables $(w_t)_{t \in N\mathbb{Z}}$, the noise parameters Σ_ν can be generically reconstructed from*

$$vec(\Sigma_\nu) = \left((\mathcal{H}_m \otimes \mathcal{H}_m)\left(I_{(nm)^2} - (\mathcal{A}_m \otimes \mathcal{A}_m)\right)^{-1} \left(\mathcal{G}_m^T \otimes \mathcal{G}_m^T\right)\right)^{-1} vec(\gamma_z(0)). \tag{3.2.41}$$

PROOF. The proof is the same as the proof of Theorem 3.2.8 for the point $A_1 = \ldots = A_p = 0$. Indeed, it follows for this point that

$$(\mathcal{H}_m \otimes \mathcal{H}_m)\left(I_{(nm)^2} - (\mathcal{A}_m^* \otimes \mathcal{A}_m^*)\right)^{-1} \left(\mathcal{G}_m^T \otimes \mathcal{G}_m^T\right) = \sum_{i=1}^N c_i^2 I_n > 0$$

since at least one $c_i \neq 0$. Therefore this is also true for a generic subset of the parameter space. □

In an analogous way it is possible to generalize Theorem 3.2.14 to the case of a simple aggregation scheme (3.2.37).

CHAPTER 4

Estimators in the Mixed-Frequency Case

As has been shown in Chapter 3 the system and noise parameters in the mixed-frequency case are g-identifiable. We denote the generic subset of the parameter space obtained via the intersection of the two subsets from Theorem 3.2.6 and 3.2.8 by Θ_{XYW} and we restrict ourselves in this chapter to the stock case.

We want to present different estimators which are based on the one hand on Chapter 3, i.e. the extended Yule-Walker estimator and the generalized method of moments estimator, and on the other hand on (Gaussian) maximum likelihood type estimators, i.e. the maximum likelihood estimator as described in Hannan and Deistler (2012) and the EM algorithm for state space systems. In Chapter 7 these estimators are compared, where the main focus lies on asymptotic parameter estimation. This chapter is based on and further develops the theory of Koelbl et al. (2015).

4.1. Extended Yule-Walker Estimator

The autocovariances in equation (3.2.3) can be estimated by the mixed-frequency estimators

$$\hat{\gamma}^{ff}(h) = \frac{1}{T}\sum_{t=1}^{T-h}\left(y_{t+h}^f - \bar{y}_T^f\right)\left(y_t^f - \bar{y}_T^f\right)^T \quad h \geq 0 \tag{4.1.1}$$

$$\hat{\gamma}^{ff}(h) = \hat{\gamma}^{ff}(-h)^T \tag{4.1.2}$$

$$\hat{\gamma}^{sf}(h) = \frac{1}{T/N}\sum_{t=t_1}^{t_2}\left(y_{Nt}^s - \bar{y}_T^s\right)\left(y_{Nt-h}^f - \bar{y}_T^f\right)^T \tag{4.1.3}$$

where $\bar{y}_T^f = \frac{1}{T}\sum_{t=1}^{T} y_t^f$, $\bar{y}_T^s = \frac{1}{T/N}\sum_{t=1}^{\lfloor T/N \rfloor} y_{Nt}^s$ ($\lfloor j \rfloor$ leads to the largest integer smaller than or equal to j) and

$$t_1 = \begin{cases} 1 & N > h \\ \lfloor \frac{h}{N} \rfloor + 1 & N \leq h \end{cases}$$

$$t_2 = \begin{cases} \lfloor \frac{T}{N} \rfloor & h \geq 0 \\ \lfloor \frac{T+h}{N} \rfloor & h < 0 \end{cases}.$$

Due to the mixed-frequency structure the estimator of $\gamma^{sf}(h)$ has only (approximately) $1/N$ of the summands compared to the estimator of $\gamma^{ff}(h)$. Now we can replace the population second moments in (3.2.15) with their sample estimates (4.1.1) to (4.1.3) and we obtain

$$\hat{Z}_1 = A\hat{Z}_0 \tag{4.1.4}$$

where $\hat{Z}_1$ and $\hat{Z}_0$ are the sample estimates of Z_1 and Z_0, respectively. Note that we do not distinguish between the true system parameters in (2.1.6) and the variable A in (4.1.4). Note that for T large enough and $\theta \in \Theta_{\mathrm{XYW}}$, $\hat{Z}_0$ has full row rank and thus $\hat{Z}_0\hat{Z}_0^T$ is nonsingular. Therefore, we get the extended Yule-Walker estimator

$$\hat{A}_{\mathrm{XYW}} = \hat{Z}_1 \underbrace{\hat{Z}_0^T \left(\hat{Z}_0\hat{Z}_0^T\right)^{-1}}_{\hat{Z}_0^\dagger}. \tag{4.1.5}$$

In general, the estimated system parameters are not stable and thus have to be projected on Θ (see Chapter 6). The estimator for the covariance matrix Σ_ν is

$$\mathrm{vec}\left(\hat{\Sigma}_\nu\right) = \left((\mathcal{G}\otimes\mathcal{G})\left(I_{(np)^2} - \left(\hat{A}_{\mathrm{XYW}}\otimes\hat{A}_{\mathrm{XYW}}\right)\right)^{-1}\left(\mathcal{G}^T\otimes\mathcal{G}^T\right)\right)^{-1}\mathrm{vec}\left(\hat{\gamma}\left(0\right)\right). \tag{4.1.6}$$

Of course, this estimator depends on the estimated (projected) system parameters (and on the estimated autocovariance of the lag 0) and thus formula (4.1.6) can also be used when the system parameters are estimated, for instance, by the (projected) generalized method of moments estimator.

It may be helpful to use the information contained in the covariances corresponding to higher order lags in order to improve the quality of the estimator. Again we repeat that higher order lags do not increase the generic subset of identifiable parameters which was obtained in Theorem 3.2.6. For $k \geq 0$, we obtain

$$\underbrace{\mathbb{E}\left(y_t\left((y_{t-1}^f)^T,\ldots,(y_{t-np-k}^f)^T\right)\right)}_{Z_{1,k}} = (A_1,\ldots,A_p)Z_{0,k}$$

$$Z_{0,k} = \mathbb{E}\left(\begin{pmatrix} y_{t-1}\\ \vdots \\ y_{t-p}\end{pmatrix}\left((y_{t-1}^f)^T,\ldots,(y_{t-np-k}^f)^T\right)\right)$$

and thus we can define the XYW estimator for the lag $np+k$ as

$$\hat{A}_{\mathrm{XYW},k} = \hat{Z}_{1,k}\underbrace{\hat{Z}_{0,k}^T\left(\hat{Z}_{0,k}\hat{Z}_{0,k}^T\right)^{-1}}_{\hat{Z}_{0,k}^\dagger}. \tag{4.1.7}$$

Of course, there is no guarantee that this variation leads to a better estimator. One reason might be that the number of observations which can be used for estimating the autocovariances decreases with increasing lag order.

4.2. Generalized Method of Moments Estimator

In this section we want to propose the generalized method of moments (GMM) estimator, which was derived in Chen and Zadrozny (1998) and is based on Hansen (1982). Note that this estimator can only be applied in the „fat" case, i.e. where the matrix Z_0 in (3.2.15) has more columns than rows.

In the classic GMM setting, i.e. in the absence of missing data, we are typically interested in estimating finite dimensional parameters, say A_0. These parameters are defined via the moment

condition of $f_t(A)$ for which

$$\mathbb{E}(f_t(A)) = 0 \Leftrightarrow A = A_0 \tag{4.2.1}$$

holds. Reconsider the XYW equations (3.2.15), this function would be in our case

$$f_t(A) = \operatorname{vec}\left(\begin{pmatrix} y_t\left(y_{t-1}^f\right)^T & \cdots & y_t\left(y_{t-np}^f\right)^T \end{pmatrix} - A\begin{pmatrix} y_{t-1}\left(y_{t-1}^f\right)^T & \cdots & y_{t-1}\left(y_{t-np}^f\right)^T \\ \vdots & & \vdots \\ y_{t-p}\left(y_{t-1}^f\right)^T & \cdots & y_{t-p}\left(y_{t-np}^f\right)^T \end{pmatrix}\right).$$

The results in Section 3.2.2 imply that the moment condition (4.2.1) is fulfilled on the generic subset Θ_{XYW}. Furthermore, we have to introduce a sequence of weighting matrices $(Q_T) \in \mathbb{R}^{n^2pn_f \times n^2pn_f}$, which converges in probability to a symmetric, positive definite constant matrix Q_0. Then we can define the GMM estimator as the minimum of

$$g_T(A)^T Q_T g_T(A) \tag{4.2.2}$$

with respect to A, where $g_T(A) = \frac{1}{T}\sum_{t=np+1}^{T} f_t(A)$.

One drawback of equation (4.2.2) in the mixed-frequency case is that $f_t(A)$ is not feasible and thus also $g_T(A)$ is not feasible. In the last few years many papers have dealt with the GMM estimator in the case of missing data, see for example Abrevaya and Donald (2011); Muris (2014). One way to overcome this problem would be as described in Muris (2014), where a time depending selection matrix, say S_t, is introduced, so that $f_t(A)$ is replaced by the feasible term $S_t f_t(A)$. In general, for the mixed-frequency case this approach cannot be used since every entry in $f_t(A)$ contains missing data and thus S_t would be the zero matrix.

Therefore, we define our GMM estimator as

$$\hat{A}_{\mathrm{GMM}} = \underset{A\in\mathbb{R}^{n\times np}}{\arg\min} \operatorname{vec}\left(\hat{Z}_1 - A\hat{Z}_0\right)^T Q_T \operatorname{vec}\left(\hat{Z}_1 - A\hat{Z}_0\right). \tag{4.2.3}$$

The solution of (4.2.3) is

$$\operatorname{vec}\left(\hat{A}_{\mathrm{GMM}}\right) = \underbrace{\left(\left(\hat{Z}_0 \otimes I_n\right) Q_T \left(\hat{Z}_0^T \otimes I_n\right)\right)^{-1} \left(\hat{Z}_0 \otimes I_n\right) Q_T}_{\hat{Z}_{\mathrm{GMM}}^{\dagger}} \operatorname{vec}\left(\hat{Z}_1\right). \tag{4.2.4}$$

Here we assumed that $\hat{Z}_0$ has full row rank and that Q_T is nonsingular. If $n_f = 1$ (and $k = 0$), Z_0 is quadratic and therefore $\operatorname{vec}\left(\hat{A}_{\mathrm{GMM}}\right) = \operatorname{vec}\left(\hat{Z}_1\hat{Z}_0^{-1}\right) = \operatorname{vec}\left(\hat{A}_{\mathrm{XYW}}\right)$. An optimal weighting matrix Q_T for the GMM estimator will be derived in Chapter 5. Also note that for the case $Q_T = I_{n^2pn_f}$ it follows that $\operatorname{vec}\left(\hat{A}_{\mathrm{GMM}}\right) = \operatorname{vec}\left(\hat{A}_{\mathrm{XYW}}\right)$.

4.3. Maximum Likelihood Estimator

4.3.1. The AR(1) $N = 2$ Case. As has been shown in Section 3.2.1 the blocked process in the AR(1) case remains in the AR-class. Furthermore, the non-identifiable set has been described explicitly for the case $n = 2$ and $\Sigma_\nu > 0$ by Theorem 3.2.4. Let Θ_I be the identifiable set, which is generic in the parameter space. In this section we want to derive the maximum likelihood estimator

for the AR(1) case, where we assume that Σ_ν is nonsingular, $N = 2$ and T is even. For that reason we factorize the matrix $\tilde{\mathcal{A}}$ into $\tilde{\mathcal{A}} = a_1 a_2$ where

$$\tilde{\mathcal{A}} = \underbrace{\begin{pmatrix} 0 & a_{ff} & a_{fs} \\ 0 & a_{sf} & a_{ss} \\ I_{n_f} & 0 & 0 \end{pmatrix}}_{a_1} \underbrace{\begin{pmatrix} a_{ff} & a_{fs} & 0_{n_f \times n_f} \\ a_{ff} & a_{fs} & 0_{n_f \times n_f} \\ a_{sf} & a_{ss} & 0_{n_s \times n_f} \end{pmatrix}}_{a_2} \tag{4.3.1}$$

and rewrite $\Sigma_{\tilde{\nu}} = T_1 \Sigma_\nu T_1^T + T_2 \Sigma_\nu T_2^T$ where

$$T_1 = \begin{pmatrix} I_n \\ 0_{n_f \times n} \end{pmatrix}, \quad T_2 = \begin{pmatrix} a_{ff} & a_{fs} \\ a_{sf} & a_{ss} \\ I_{n_f} & 0 \end{pmatrix}. \tag{4.3.2}$$

We assume that the ν_t are independently normally distributed and thus $\tilde{\nu}_t$ are also independently normally distributed, i.e.

$$\tilde{v} = \text{vec}\left(\tilde{\nu}_2, \tilde{\nu}_4, \ldots, \tilde{\nu}_T\right) \sim \mathcal{N}_h\left(0, I_{T/2} \otimes \Sigma_{\tilde{\nu}}\right)$$

where $h = T/2\,(n + n_f)$. Thus, this variable has the density

$$f_{\tilde{v}}(v) = \frac{1}{(2\pi)^{(n+n_f)T/4}} \left|I_{T/2} \otimes \Sigma_{\tilde{\nu}}\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} v^T \left(I_{T/2} \otimes \Sigma_{\tilde{\nu}}^{-1}\right) v\right).$$

Defining $\tilde{y} = \text{vec}\left(\tilde{Y}\right)$ and using $\tilde{v} = \tilde{y} - (X^T \otimes I_{(n+n_f)})\text{vec}\left(\tilde{\mathcal{A}}\right)$, the density of $\tilde{y}$ is given by

$$\begin{aligned} f_{\tilde{y}}(y) &= \underbrace{\left|\frac{\partial \tilde{v}}{\partial \tilde{y}^T}\right|}_{1} f_{\tilde{v}}\left(v\right) = \frac{1}{(2\pi)^{(n+n_f)T/4}} \left|I_{T/2} \otimes \Sigma_{\tilde{\nu}}\right|^{-\frac{1}{2}} \\ &\exp\left(-\frac{1}{2}\left(y - (X^T \otimes I_{(n+n_f)})\text{vec}\left(\tilde{\mathcal{A}}\right)\right)^T \left(I_{T/2} \otimes \Sigma_{\tilde{\nu}}^{-1}\right)\left(y - (X^T \otimes I_{(n+n_f)})\text{vec}\left(\tilde{\mathcal{A}}\right)\right)\right) \end{aligned} \tag{4.3.3}$$

where $\tilde{Y} = (\tilde{y}_2, \tilde{y}_4, \ldots, \tilde{y}_T)$, $X = (\tilde{y}_0, \tilde{y}_2, \ldots, \tilde{y}_{T-2})$ and $\tilde{V} = (\tilde{\nu}_2, \tilde{\nu}_4, \ldots, \tilde{\nu}_T)$. For convenience we assume that the initial values, i.e. $\tilde{y}_0$, are known. Taking the logarithm of the above formula and taking into account that the last term in (4.3.3) can be rewritten into $\exp\left(\text{tr}\left(\left(\tilde{Y} - \tilde{\mathcal{A}}X\right)^T \Sigma_{\tilde{\nu}}^{-1}\left(\tilde{Y} - \tilde{\mathcal{A}}X\right)\right)\right)$ we obtain the likelihood for the AR(1) case:

$$\ln\left(l\left(A_1, \Sigma_\nu\right)\right) = c - \frac{T}{4}\ln|\Sigma_{\tilde{\nu}}| - \frac{1}{2}\text{tr}\left(\left(\tilde{Y} - \tilde{\mathcal{A}}X\right)^T \Sigma_{\tilde{\nu}}^{-1}\left(\tilde{Y} - \tilde{\mathcal{A}}X\right)\right) \tag{4.3.4}$$

In order to determine maximum likelihood estimators of A_1 and Σ_ν, the first order derivations are needed with respect to the high-frequency parameters:

THEOREM 4.3.1. *Assume that $\Sigma_\nu > 0$ and $\theta \in \Theta_I$. Then for the AR(1) case the score functions of (4.3.4) are given by*

$$\begin{aligned}\frac{\partial \ln\left(l\left(A_1,\Sigma_\nu\right)\right)}{\partial vec\left(A_1\right)} &= vec\left(T_2^T\Sigma_{\tilde{\nu}}^{-1}\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)X^TT_1\right)+vec\left(T_1^T\Sigma_{\tilde{\nu}}^{-1}\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)X^Ta_2^T\begin{pmatrix}0_{n_f\times n}\\ I_n\end{pmatrix}\right)\\ &\quad -vec\left(T_1^T\Sigma_{\tilde{\nu}}^{-1}\left(\frac{T}{2}\Sigma_{\tilde{\nu}}-\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)^T\right)\Sigma_{\tilde{\nu}}^{-1}T_2\Sigma_\nu\right)\end{aligned}$$

$$\begin{aligned}\frac{\partial \ln\left(l\left(A_1,\Sigma_\nu\right)\right)}{\partial vech\left(\Sigma_\nu\right)} &= -\frac{1}{2}D_n^Tvec\left(T_1^T\Sigma_{\tilde{\nu}}^{-1}\left[\frac{T}{2}\Sigma_{\tilde{\nu}}-\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)^T\right]\Sigma_{\tilde{\nu}}^{-1}T_1\right.\\ &\quad \left.+T_2^T\Sigma_{\tilde{\nu}}^{-1}\left[\frac{T}{2}\Sigma_{\tilde{\nu}}-\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)^T\right]\Sigma_{\tilde{\nu}}^{-1}T_2\right)\end{aligned}$$

where D_n is a duplication matrix (see Appendix C).

PROOF. The proof is given in Appendix A. □

Thus, by setting the score function equal to zero, i.e.

$$\begin{aligned}\frac{\partial \ln\left(l\left(A_1,\Sigma_\nu\right)\right)}{\partial \mathrm{vec}\left(A_1\right)} &= 0\\ \frac{\partial \ln\left(l\left(A_1,\Sigma_\nu\right)\right)}{\partial \mathrm{vech}\left(\Sigma_\nu\right)} &= 0,\end{aligned} \tag{4.3.5}$$

we can calculate the maximum likelihood estimator by a numerical algorithm such as the Newton-Raphson method. Of course, such an algorithm needs an initial value which can be obtained, for instance, via the (projected) XYW estimator.

4.3.2. The General AR(p) $p \geq 1$ Case. Throughout this section we assume that Σ_ν is non-singular. For a given N we define the blocked observed process

$$\tilde{y}_t = \begin{pmatrix} y_t \\ y_{t-1}^f \\ \vdots \\ y_{t-N+1}^f \end{pmatrix},\ t \in N\mathbb{Z}. \tag{4.3.6}$$

A Wold representation of this blocked process $(\tilde{y}_t)_{t\in N\mathbb{Z}}$ has been given in Filler (2010). Let T be a multiple of N. Then the likelihood of observations is given by (see Hannan and Deistler (2012))

$$\ln\left(\tilde{l}(\theta)\right) = c-\frac{T}{2}\ln\left(\det\left(\tilde{\Gamma}_T\left(\theta\right)\right)\right)-\frac{1}{2}\tilde{y}^T\left(\tilde{\Gamma}_T\left(\theta\right)\right)^{-1}\tilde{y} \tag{4.3.7}$$

where $\tilde{y} = \mathrm{vec}\left(\tilde{y}_N, \tilde{y}_{2N}, \ldots, \tilde{y}_T\right)$ contains the observed data and $\tilde{\Gamma}_T\left(\theta\right) = \mathbb{E}\left(\tilde{y}\tilde{y}^T\right)$ has block Toeplitz form. In Hannan and Deistler (2012) the consistency and the asymptotic normality of the maximum likelihood estimator under rather general conditions has been shown.

Note that the matrix $\tilde{\Gamma}_T(\theta)$ has dimension $(n+(N-1)n_f)T/N$ and is in general a very large matrix in block Toeplitz form. Accordingly its inversion may cause problems, even when taking advantage of its block Toeplitz structure (see Wax and Kailath (1983)).

Therefore, we establish the EM algorithm for state space models which was introduced in Shumway and Stoffer (1982) and is based on the work of Dempster et al. (1977). The main idea is to successively maximize (M-step) the conditional expectation (E-step) of the complete data likelihood given the observed data. In this context complete data likelihood stands for the high-frequency likelihood. As can be shown the likelihood does not decrease in each iteration step.

As in the high-frequency case, we assume that the innovations ν_t are independently normally distributed. First, we start with the complete data likelihood

$$\begin{aligned}\ln\left(l\left(A,\Sigma_\nu,\Gamma_p\right)\right) &= c-\frac{1}{2}\ln|V_1|-\frac{1}{2}x_1^TV_1^{-1}x_1 \\ &\quad -\frac{T}{2}\ln|\Sigma_\nu|-\frac{1}{2}\mathrm{tr}\left(\left(Y-AX\right)^T\Sigma_\nu^{-1}\left(Y-AX\right)\right) \end{aligned} \tag{4.3.8}$$

where again $Y=(y_1,y_2,\ldots,y_T)$, $X=(x_1,x_2,\ldots,x_T)$ and $A=(A_1,\ldots,A_p)$. It should be mentioned that in this case we assume that the initial values are normally distributed, i.e. $x_1=(y_0^T,\ldots y_{p-1}^T)^T\sim\mathcal{N}_{np}\left(0,V_1\right)$ and independent from $\nu_1,\ldots,\nu_T$, which is a different assumption than in Chapter 2, where we assumed that the initial values are given. Let $\tau^{(k)}=\left(A^{(k)},\Sigma_\nu^{(k)},V_1^{(k)}\right)$ be the parameters at the k-th iteration and define

$$Q\left(\tau|\tau^{(k)}\right)=\mathbb{E}_{\tau^{(k)}}\left(\ln\left(l(\tau)\right)|\tilde{y}\right) \tag{4.3.9}$$

where again $\tilde{y}=\mathrm{vec}\left(\tilde{y}_N,\tilde{y}_{2N},\ldots,\tilde{y}_T\right)$ and $\tau=(A,\Sigma_\nu,V_1)$. The function $Q\left(\tau|\tau^{(k)}\right)$ represents the E-step in this procedure. The initial values $\tau^{(0)}$ can be estimated, for instance, via the (projected) extended Yule-Walker estimator (see Chapter 6). In addition let

$$\begin{aligned} x_{t|T} &= \mathbb{E}_{\tau^{(k)}}\left(x_t|\tilde{y}\right) \\ P_{t|T} &= \mathbb{E}_{\tau^{(k)}}\left(x_t-x_{t|T}\right)\left(x_t-x_{t|T}\right)^T \\ P_{t,t-1|T} &= \mathbb{E}_{\tau^{(k)}}\left(x_t-x_{t|T}\right)\left(x_{t-1}-x_{t-1|T}\right)^T. \end{aligned} \tag{4.3.10}$$

In a next step we explicitly derive (4.3.9): Note that $x_1^TV_1^{-1}x_1$ can be rewritten into

$$\begin{aligned} x_1^TV_1^{-1}x_1 &= \mathrm{tr}\left(V_1^{-1}\left(x_{1|T}+x_1-x_{1|T}\right)\left(x_{1|T}+x_1-x_{1|T}\right)^T\right) \\ &= \mathrm{tr}\left(V_1^{-1}\left(x_{1|T}x_{1|T}^T+\left(x_1-x_{1|T}\right)\left(x_1-x_{1|T}\right)^T\right.\right. \\ &\quad \left.\left.+x_{1|T}\left(x_1-x_{1|T}\right)^T+\left(x_1-x_{1|T}\right)^Tx_{1|T}\right)\right). \end{aligned}$$

Taking the conditional expectation leads to

$$\mathbb{E}_{\tau^{(k)}}\left(x_1^TV_1^{-1}x_1|\tilde{y}\right)=\mathrm{tr}\left(V_1^{-1}\left(x_{1|T}x_{1|T}^T+P_{1|T}\right)\right) \tag{4.3.11}$$

since $\mathbb{E}_{\tau^{(k)}}\left(x_{1|T}\left(x_1-x_{1|T}\right)^T|\tilde{y}\right)=x_{1|T}\mathbb{E}_{\tau^{(k)}}\left(\left(x_1-x_{1|T}\right)^T|\tilde{y}\right)=0$. The second term, i.e.

$$\mathrm{tr}\left(\left(Y-AX\right)^T\Sigma_\nu^{-1}\left(Y-AX\right)\right),$$

can be rewritten in an analogous way:

$$\operatorname{tr}\left(\left(Y-AX\right)^{T}\Sigma_{\nu}^{-1}\left(Y-AX\right)\right) \;=\; \operatorname{tr}\left(\Sigma_{\nu}^{-1}\left(\sum_{i=1}^{T}\left(\mathcal{G}x_{t+1}-Ax_{t}\right)\left(\mathcal{G}x_{t+1}-Ax_{t}\right)^{T}\right)\right)$$

and with the help of $\left(\mathcal{G}x_{t+1}-Ax_{t}\right)=\left(\mathcal{G}x_{t+1|T}+\mathcal{G}\left(x_{t+1}-x_{t+1|T}\right)-Ax_{t|T}-A\left(x_{t}-x_{t|T}\right)\right)$ we obtain

$$\begin{aligned}\left(\mathcal{G}x_{t+1}-Ax_{t}\right)\left(\mathcal{G}x_{t+1}-Ax_{t}\right)^{T} &= \mathcal{G}x_{t+1|T}x_{t+1|T}^{T}\mathcal{G}^{T}+\mathcal{G}\left(x_{t+1}-x_{t+1|T}\right)\left(x_{t+1}-x_{t+1|T}\right)^{T}\mathcal{G}^{T}\\ &\quad +Ax_{t|T}x_{t|T}^{T}A^{T}+A\left(x_{t}-x_{t|T}\right)\left(x_{t}-x_{t|T}\right)^{T}A^{T}\\ &\quad -\mathcal{G}x_{t+1|T}x_{t|T}^{T}A^{T}-Ax_{t|T}x_{t+1|T}^{T}\mathcal{G}^{T}\\ &\quad -\mathcal{G}\left(x_{t+1}-x_{t+1|T}\right)\left(x_{t}-x_{t|T}\right)^{T}A^{T}\\ &\quad -A\left(x_{t}-x_{t|T}\right)\left(x_{t+1}-x_{t+1|T}\right)^{T}\mathcal{G}^{T}+\ldots\end{aligned}$$

where $\mathcal{G}=\left(I_{n},0,\ldots,0\right)$. Thus, with the same arguments as above, taking the conditional expectation we get

$$\begin{aligned}\mathbb{E}_{\tau^{(k)}}\left(\left(\mathcal{G}x_{t+1}-Ax_{t}\right)\left(\mathcal{G}x_{t+1}-Ax_{t}\right)^{T}|\tilde{y}\right) &= \mathcal{G}\left(x_{t+1|T}x_{t+1|T}^{T}+P_{t+1|T}\right)\mathcal{G}^{T}\\ &\quad +A\left(x_{t|T}x_{t|T}^{T}+P_{t|T}\right)A^{T}\\ &\quad -\mathcal{G}\left(x_{t+1|T}x_{t|T}^{T}+P_{t+1,t|T}\right)A^{T}\\ &\quad -A\left(x_{t|T}x_{t+1|T}^{T}+P_{t,t+1|T}\right)\mathcal{G}^{T}\end{aligned}$$

and thus

$$\begin{aligned}Q\left(\tau|\tau^{(k)}\right) &= -\frac{1}{2}\ln\left|V_{1}\right|-\frac{1}{2}\operatorname{tr}\left(V_{1}^{-1}\left(x_{1|T}x_{1|T}^{T}+P_{1|T}\right)\right)\\ &\quad -\frac{T}{2}\ln\left|\Sigma_{\nu}\right|-\frac{1}{2}\operatorname{tr}\left(\Sigma_{\nu}^{-1}\left(\mathcal{G}S_{11}\mathcal{G}^{T}+AS_{00}A^{T}-\mathcal{G}S_{10}A^{T}-AS_{10}^{T}\mathcal{G}^{T}\right)\right)\end{aligned} \tag{4.3.12}$$

where

$$\begin{aligned}S_{00} &= \sum_{t=1}^{T}\left(x_{t|T}x_{t|T}^{T}+P_{t|T}\right)\\ S_{11} &= \sum_{t=2}^{T+1}\left(x_{t|T}x_{t|T}^{T}+P_{t|T}\right)\\ S_{10} &= \sum_{t=2}^{T+1}\left(x_{t|T}x_{t-1|T}^{T}+P_{t,t-1|T}\right).\end{aligned} \tag{4.3.13}$$

The terms $x_{t|T}$, $P_{t|T}$ and $P_{t,t-1|T}$ can be calculated, for instance, by the Kalman smoothing algorithm for time varying models (see Shumway and Stoffer (1982)) or for simple cases by the interpolation formulas from Proposition 2.1.12. In the M-step we maximize the expected conditional likelihood

(4.3.12) with respect to the parameters τ. In Shumway and Stoffer (1982) it is argued that

$$\begin{aligned} A^{(k+1)} &= \mathcal{G}S_{10}S_{00}^{-1} \\ \Sigma^{(k+1)} &= \frac{1}{T}\left(\mathcal{G}S_{11}\mathcal{G}^T - \mathcal{G}S_{10}S_{00}^{-1}S_{10}^T\mathcal{G}^T\right) \\ V_1^{(k+1)} &= P_{1|T} + x_{1|T}x_{1|T}^T \end{aligned} \tag{4.3.14}$$

maximize the expected conditional likelihood function. Now we can construct the EM algorithm for our problem:

(1) Start with an initial estimator $\tau^{(0)}$ which can be derived, for instance, by the (projected) XYW estimator.
(2) Compute $x_{t|T}$, $P_{t|T}$ and $P_{t,t-1|T}$.
(3) Compute S_{00}, S_{11} and S_{10} from (4.3.13).
(4) Compute the estimates for the $(k+1)$-th step from (4.3.14).
(5) Repeat step 2-4 until the relative change of $\tau^{(k)}$ and $\tau^{(k+1)}$ is smaller than a predefined threshold.

For a fast implementation of the EM algorithm see for instance McLachlan and Krishnan (2008); Mader et al. (2014).

REMARK 4.3.2. As has been mentioned above, each iteration step of the EM algorithm does not decrease the likelihood function. Convergence of the EM algorithm is discussed in Wu (1983). One drawback of the EM algorithm is that, in general, there is a large dependence on the starting values $\tau^{(0)}$. One strategy to overcome this problem is to start with many different starting values and to choose the estimated parameters which maximize the likelihood.

Another way to estimate the system and noise parameters would be to estimate the high-frequency parameters via e.g. the extended Yule-Walker estimator, interpolate the missing slow variables, i.e. y_t^s $t \neq N, 2N, \ldots$ via the Kalman smoother and use the high-frequency maximum likelihood estimator to obtain an update of the system and noise parameters, e.g.

$$\mathring{A}^{(k+1)} = \mathcal{G}\left(\sum_{t=2}^{T+1}\left(x_{t|T}x_{t-1|T}^T\right)\right)\left(\sum_{t=1}^{T}\left(x_{t|T}x_{t|T}^T\right)\right)^{-1}. \tag{4.3.15}$$

Of course, this procedure can also be iterated. One reason why this procedure should not be used is that it is not guaranteed that the likelihood is non decreasing in each iteration step.

The main differences between the estimator (4.3.15) and the EM estimator, i.e.

$$A^{(k+1)} = \mathcal{G}\left(\sum_{t=2}^{T+1}\left(x_{t|T}x_{t-1|T}^T + P_{t,t-1|T}\right)\right)\left(\sum_{t=1}^{T}\left(x_{t|T}x_{t|T}^T + P_{t|T}\right)\right)^{-1}$$

are the terms $P_{t,t-1|T}$ and $P_{t|T}$. Loosely speaking, the EM algorithm takes the interpolation error into account whereas the estimator (4.3.15) does not. Note that, in general, both estimators do not lead to a stable AR polynomial.

Since interpolation of missing observations is a central point of the EM algorithm we state two examples of two systems, which will give us further insides into the problem of interpolation in the mixed-frequency setting:

EXAMPLE 4.3.3. Consider again the AR(1) case where Σ_ν is nonsingular as in Example 2.1.11 and let $N = 2$. Furthermore let $H_y^o(t)$ be the Hilbert space spanned by all y_t, $t \in 2\mathbb{Z}$ and y_t^f, $t \in 2\mathbb{Z} - 1$. Then the best linear interpolation of y_t^s is given by

$$P_{H_y^o(t)}\left(y_t^s\right) = y_{t,\text{int}}^s = \left(\rho_0^{ss}\right)^{-1}\left(0, I_{n_s}\right)\left(\Sigma_\nu^{-1}A_1 y_{t-1} + A_1^T\Sigma_\nu^{-1}y_{t+1} - \rho_0\begin{pmatrix} y_t^f \\ 0 \end{pmatrix}\right) \tag{4.3.16}$$

where $\rho_0 = \left(\Sigma_\nu^{-1} + A_1^T\Sigma_\nu^{-1}A_1\right)$ and $\rho_0^{ss} = \left(0, I_{n_s}\right)\rho_0\begin{pmatrix} 0 \\ I_{n_s} \end{pmatrix}$. The interpolation error covariance is

$$\mathbb{V}\left(y_t^s - y_{t,\text{int}}^s\right) = \left(\rho_0^{ss}\right)^{-1}$$

since the interpolation error is $y_t^s - y_{t,\text{int}}^s = \left(\rho_0^{ss}\right)^{-1}\left(0, I_{n_s}\right)\left(\Sigma_\nu^{-1}\nu_t - A_1^T\Sigma_\nu^{-1}\nu_{t+1}\right)$.

Thus, for a finite set of mixed-frequency observations we get

$$x_{t|T} = \begin{cases} y_t & t \in \{0, 2, \ldots, T\} \\ \begin{pmatrix} y_t^f \\ y_{t,\text{int}}^s \end{pmatrix} & t \in \{1, 3, \ldots, T-1\} \end{cases}, \tag{4.3.17}$$

$P_{t,t-1|T} = 0$ and

$$P_{t|T} = \begin{cases} 0 & t \in \{0, 2, \ldots, T\} \\ \begin{pmatrix} 0 & 0 \\ 0 & \left(\rho_0^{ss}\right)^{-1} \end{pmatrix} & t \in \{1, 3, \ldots, T-1\} \end{cases} \tag{4.3.18}$$

where we assume that y_0 is available and T is even. Note that each ρ_i depend on $\tau^{(k)}$, i.e. the k-th iteration of the EM algorithm. Thus, step 2 in the EM algorithm above can be replaced by (4.3.17) and (4.3.18).

EXAMPLE 4.3.4. Consider the AR(2) case where Σ_ν is nonsingular and $N = 2$. In this case the situation is more tricky since the interpolation of y_t^s in the high-frequency case depends on y_{t+1}, y_{t-1}, y_{t-2}, y_{t+2} and y_t^f, i.e.

$\cdots$	$t-3$	$t-2$	$t-1$	t	$t+1$	$t+2$	$t+3$	$\cdots$
$\cdots$	y_{t-3}^f	y_{t-2}^f	y_{t-1}^f	y_t^f	y_{t+1}^f	y_{t+2}^f	y_{t+3}^f	$\cdots$
$\cdots$	y_{t-3}^s	y_{t-2}^s	y_{t-1}^s	$y_{t,\text{int}}^s$	y_{t+1}^s	y_{t+2}^s	y_{t+3}^s	$\cdots$

(4.3.19)

As displayed above, the interpolation of y_t^s depends on the observed y_{t+1}, y_{t-1}, y_{t-2}^f, y_{t+2}^f and y_t^f in green and on the non-observed y_{t-2}^s and y_{t+2}^s in red. Thus, the best linear interpolation is given by

$$\begin{aligned} y_{t,\text{int}}^s &= (\rho_0^{ss})^{-1}(0, I_{n_s}) \left(-\rho_1 y_{t-1} - \rho_1^T y_{t+1} - \rho_2 \begin{pmatrix} y_{t-2}^f \\ y_{t-2,\text{int}}^s \end{pmatrix} \right. \\ &\quad \left. -\rho_2^T \begin{pmatrix} y_{t+2}^f \\ y_{t+2,\text{int}}^s \end{pmatrix} - \rho_0 \begin{pmatrix} y_t^f \\ 0 \end{pmatrix} \right) \end{aligned} \tag{4.3.20}$$

where $\rho_i = \begin{pmatrix} \rho_i^{ff} & \rho_i^{fs} \\ \rho_i^{sf} & \rho_i^{ss} \end{pmatrix} = \begin{pmatrix} \rho_i^f \\ \rho_i^s \end{pmatrix}$ is obtained from Proposition 2.1.12 and the interpolation error is

$$\begin{aligned} y_t^s - y_{t,\text{int}}^s &= (\rho_0^{ss})^{-1}(0, I_{n_s}) \left(\Sigma_\nu^{-1} \left(A_2 \begin{pmatrix} 0 \\ y_{t-2}^s - y_{t-2,\text{int}}^s \end{pmatrix} + \nu_t \right) - A_1^T \Sigma_\nu^{-1} \nu_{t+1} \right. \\ &\quad \left. + A_2^T \Sigma_\nu^{-1} \left(\begin{pmatrix} 0 \\ y_{t+2}^s - y_{t+2,\text{int}}^s \end{pmatrix} + \nu_{t+2} \right) \right). \end{aligned} \tag{4.3.21}$$

Equation (4.3.21) indicates that it is very hard to obtain an analogous formula for $P_{t|T}$ as in the AR(1) case.

Nevertheless for a set of observations $t = 1, \ldots, T$, where we assume that T is even and to simplify matters we assume that y_0, y_{-1} and y_{T+1} is observed, we can arrange the following system of equations:

$$\begin{pmatrix} \rho_0^{ss} & \rho_2^{ss} & & & & \\ \rho_2^{ss} & \rho_0^{ss} & \rho_2^{ss} & & & \\ & \rho_2^{ss} & \rho_0^{ss} & p_2^{ss} & & \\ & & \ddots & \ddots & \ddots & \\ & & & \rho_2^{ss} & \rho_0^{ss} & p_2^{ss} \\ & & & & \rho_2^{ss} & \rho_0^{ss} \end{pmatrix} \begin{pmatrix} y_{1,\text{int}}^s \\ y_{3,\text{int}}^s \\ \vdots \\ \vdots \\ y_{T-3,\text{int}}^s \\ y_{T-1,\text{int}}^s \end{pmatrix} = -X^{\text{ob}} \tag{4.3.22}$$

and

$$X^{\text{ob}} = \begin{pmatrix} \rho_2^s & \rho_1^s & \rho_0^s & \left(\rho_1^T\right)^s & \left(\rho_2^T\right)^s & & & & \\ & & \rho_2^s & \rho_1^s & \rho_0^s & \left(\rho_1^T\right)^s & \left(\rho_2^T\right)^s & & \\ & & & & \rho_2^s & \rho_1^s & \rho_0^s & \left(\rho_1^T\right)^s & \left(\rho_2^T\right)^s \\ & & & & & \ddots & \ddots & \ddots & \end{pmatrix} \begin{pmatrix} y_{-1} \\ y_0 \\ \begin{pmatrix} y_1^f \\ 0 \end{pmatrix} \\ y_2 \\ \vdots \\ \begin{pmatrix} y_{T-1}^f \\ 0 \end{pmatrix} \\ y_T \\ y_{T+1} \end{pmatrix}.$$

So, in a certain sense, the Kalman smoothing algorithm interpolates the missing observations by solving the system of equations (4.3.22) in a recursive way.

CHAPTER 5

Asymptotic Properties of the Mixed-Frequency Estimators

In this chapter we want to derive the asymptotic properties of the extended Yule-Walker and the GMM estimator for the stock case. Whereas under our assumptions in the high-frequency case the Yule-Walker estimator has the same asymptotic covariance as the maximum likelihood estimator and thus is asymptotically efficient, this is not the case for the XYW estimator in the mixed-frequency case. One reason might be that the XYW estimator does not include the information from the observed slow autocovariances. The asymptotic covariance of the XYW estimator is derived in two steps: First we derive the asymptotic covariance of the sample second moments of the observations and then we linearize the mapping attaching the parameters to the second moments of the observations. Furthermore, we derive the asymptotic covariance of the maximum likelihood estimator for the AR(1) case and the consistency of the second and fourth moment of the noise parameter estimators. Parts of the results in this chapter are contained in Koelbl et al. (2015).

5.1. Bartlett's Formula for the Mixed-Frequency Case

In Niebuhr and Kreiss (2013) Bartlett's formula was derived for the univariate low-frequency case. So the following results are on the one hand a generalization to the multivariate case and on the other hand a generalization to the mixed-frequency case.

In the following we will use the partition $k_j = \begin{pmatrix} k_j^f \\ k_j^s \end{pmatrix}$ where k_j^f denotes the first n_f and k_j^s the last n_s rows of k_j, respectively. Again let $\eta = \mathbb{E}\left(\nu_t \nu_t^T \otimes \nu_t \nu_t^T\right)$ and $\kappa = \eta - \operatorname{vec}(\Sigma_\nu)\operatorname{vec}(\Sigma_\nu)^T - (\Sigma_\nu \otimes \Sigma_\nu) - K_{n,n}(\Sigma_\nu \otimes \Sigma_\nu)$.

LEMMA 5.1.1. *Let $(y_t)_{t\in\mathbb{Z}}$ be a linear process and assume that the fourth moment of ν_t, $\eta = \mathbb{E}\left(\nu_t \nu_t^T \otimes \nu_t \nu_t^T\right)$, exists. Then*

$$\lim_{T\to\infty} T\,Cov\left(vec\left(\hat{\gamma}^{ff}(p)\right), vec\left(\hat{\gamma}^{ff}(q)\right)\right) = S_{p,q} + R_{p,q} \tag{5.1.1}$$

for $p, q \in \mathbb{Z}$ where

$$\begin{aligned} R_{p,q} &= \sum_{k=-\infty}^{\infty} \left(\gamma^{ff}(k+q-p) \otimes \gamma^{ff}(k)\right) + K_{n_f,n_f}\left(\gamma^{ff}(k+q) \otimes \gamma^{ff}(k-p)\right) \\ S_{p,q} &= \sum_{k=-\infty}^{\infty} \sum_{r=-\infty}^{\infty} \left(k_{k-p}^f \otimes k_k^f\right) \kappa \left(k_{r+k-q}^f \otimes k_{r+k}^f\right)^T \end{aligned}$$

and K_{n_f,n_f} and $K_{n,n}$ are commutation matrices.

PROOF. This is a consequence of Lemma 2.2.18. □

In the next few pages we want to introduce Bartlett's formula for the mixed-frequency setting. Thus, we need the following lemma whose proof is given in the Appendix.

LEMMA 5.1.2. *Let $(y_t)_{t\in\mathbb{Z}}$ be a linear process and assume that the fourth moment of ν_t, $\eta = \mathbb{E}\left(\nu_t\nu_t^T \otimes \nu_t\nu_t^T\right)$, exists. Then*

$$\begin{aligned}\mathbb{E}\left(\left(y_{t-p}^f\left(y_{Nu-q}^f\right)^T\right)\otimes\left(y_t^f\left(y_{Nu}^s\right)^T\right)\right) &= vec\left(\gamma^{ff}(p)\right)vec\left(\gamma^{sf}(q)\right)^T + R(p,q,Nu-t)\\ &\quad+\left(\gamma^{ff}(-Nu+t+q-p)\otimes\gamma^{fs}(-Nu+t)\right)\\ &\quad+K_{n_f,n_f}\left(\gamma^{ff}(-Nu+t+q)\otimes\gamma^{fs}(-Nu+t-p)\right)\end{aligned}$$

where p, q, t, $u\in\mathbb{Z}$, $1<N\in\mathbb{N}$ and

$$R(p,q,Nu-t)=\sum_{k=-\infty}^{\infty}\left(k_{k-p}^f\otimes k_k^f\right)\kappa\left(k_{k-q+Nu-t}^f\otimes k_{k+Nu-t}^s\right)^T.$$

PROOF. see Appendix B. □

The next lemma is a generalization of the well known formula

$$\sum_{u=1}^{T}\sum_{t=1}^{T}g(u-t)=\sum_{k=-(T-1)}^{T-1}(T-|k|)\,g(k). \tag{5.1.2}$$

LEMMA 5.1.3. *Let $N\in\mathbb{N}$, $g:\mathbb{Z}\to\mathbb{R}^{k\times q}$ and T be a multiple of N. Then $\sum_{u=1}^{T/N}\sum_{t=1}^{T}g(Nu-t)=\sum_{k=-(T/N-1)}^{T/N-1}\left(\frac{T}{N}-|k|\right)\sum_{i=0}^{N-1}g(Nk+i)$.*

PROOF. We prove this lemma by counting the individual occurrences of the terms $g(k)$, $k=-(T-1),...,T-1$: Let $u=1$ be fix and observe that in the sum above the following terms $g(N-T)$, ..., $g(2N-T-1)$, $g(2N-T)$, ..., $g(3N-T-1)$ till $g(0)$, ..., $g(N-1)$ occur. Now let $u=2$ and note that $g(2N-T)$, ..., $g(3N-T-1)$, ..., $g(0)$, ..., $g(N-1)$ and $g(N)$, ..., $g(2N-1)$ occur. It follows that there is a shift of N. For $u=T/N$ the terms $g(0),...,g(N-1)$ till $g(T-N),...,g(T-1)$ occur. To summarize, we can display the occurrences of the term $g(.)$ in Table 5.1.1, which proves this lemma.

□

From	To	Count
$g(0)$	$g(N-1)$	T/N
$g(N)$	$g(2N-1)$	$T/N-1$
⋮	⋮	⋮
$g(T-N)$	$g(T-1)$	1
$g(-N)$	$g(-1)$	$T/N-1$
⋮	⋮	⋮
$g(N-T)$	$g(2N-T-1)$	1

TABLE 5.1.1. Table of counts for the proof of Lemma 5.1.3

Note that if we set $N = 1$, we still obtain equation (5.1.2).

LEMMA 5.1.4. *Let $(y_t)_{t\in\mathbb{Z}}$ be a linear process and assume that the fourth moment of ν_t, $\eta = \mathbb{E}\left(\nu_t\nu_t^T \otimes \nu_t\nu_t^T\right)$, exists. Then we obtain*

$$\lim_{T\to\infty} TCov\left(vec\left(\hat{\gamma}^{ff}(p)\right), vec\left(\hat{\gamma}^{sf}(q)\right)\right) = S_{p,q} + R_{p,q} \tag{5.1.3}$$

where $p, q \in \mathbb{Z}$

$$\begin{aligned} R_{p,q} &= \sum_{k=-\infty}^{\infty} \left(\gamma^{ff}(k+q-p) \otimes \gamma^{fs}(k)\right) + K_{n_f,n_f}\left(\gamma^{ff}(k+q) \otimes \gamma^{fs}(k-p)\right) \\ S_{p,q} &= \sum_{k=-\infty}^{\infty}\sum_{r=-\infty}^{\infty} \left(k_{k-p}^f \otimes k_k^f\right) \kappa \left(k_{r+k-q}^f \otimes k_{r+k}^s\right)^T. \end{aligned}$$

PROOF. Let us assume that T is a multiple of N. We will prove this lemma with

$$\hat{\gamma}^{sf}(p) = \frac{N}{T}\sum_{t=1}^{T/N} y_{Nt}^s \left(y_{Nt-p}^f\right)^T, \quad \hat{\gamma}^{ff}(p) = \frac{1}{T}\sum_{t=1}^{T} y_t^f \left(y_{t-p}^f\right)^T$$

instead of $\hat{\gamma}^{sf}(p)$ and $\hat{\gamma}^{ff}(p)$, respectively, since, again as in the high-frequency case, it can be shown that this does not influence the asymptotic behavior. We start with the observation that $\text{vec}\left(y_{Nt}^s \left(y_{Nt-p}^f\right)^T\right) = \left(y_{Nt-p}^f \otimes y_{Nt}^s\right)$ and

$$\begin{aligned} \mathbb{E}\left(\text{vec}\left(\hat{\gamma}^{ff}(p)\right)\text{vec}\left(\hat{\gamma}^{sf}(q)\right)^T\right) &= \left(\frac{N}{T^2}\right)\sum_{u=1}^{T/N}\sum_{t=1}^{T}\mathbb{E}\left(\text{vec}\left(y_t^f\left(y_{t-p}^f\right)^T\right)\text{vec}\left(y_{Nu}^s\left(y_{Nu-q}^f\right)^T\right)^T\right) \\ &= \left(\frac{N}{T^2}\right)\sum_{u=1}^{T/N}\sum_{t=1}^{T}\mathbb{E}\left(y_{t-p}^f\left(y_{Nu-q}^f\right)^T \otimes y_t^f\left(y_{Nu}^s\right)^T\right). \end{aligned}$$

Now we can use Lemma 5.1.2 to obtain that the sum can be rewritten into

$$\begin{aligned} &\left(\frac{N}{T^2}\right)\sum_{u=1}^{T/N}\sum_{t=1}^{T}\text{vec}\left(\gamma^{ff}(p)\right)\text{vec}\left(\gamma^{sf}(q)\right)^T + \\ &+\left(\gamma^{ff}(-Nu+t+q-p)\otimes\gamma^{fs}(-Nu+t)\right) \\ &K_{n_f,n_f}\left(\gamma^{ff}(-Nu+t+q)\otimes\gamma^{fs}(-Nu+t-p)\right) + R(p,q,Nu-t) \\ = \;& \text{vec}\left(\gamma^{ff}(p)\right)\text{vec}\left(\gamma^{sf}(q)\right)^T + \left(\frac{N}{T^2}\right)\sum_{u=1}^{T/N}\sum_{t=1}^{T} g(Nu-t). \end{aligned}$$

Using the formula from Lemma 5.1.3 we can conclude that

$$\begin{aligned} = \;& \text{vec}\left(\gamma^{ff}(p)\right)\text{vec}\left(\gamma^{sf}(q)\right)^T + \\ &\left(\frac{N}{T}\right)\sum_{|k|<T/N}\left(\frac{1}{N}-\frac{|k|}{T}\right)\sum_{i=0}^{N-1}\left(\left(\gamma^{ff}(Nk-i+q-p)\otimes\gamma^{fs}(Nk-i)\right)+\right. \\ &\left.K_{n_f,n_f}\left(\gamma^{ff}(Nk-i+q)\otimes\gamma^{fs}(Nk-i-p)\right)+R(p,q,Nk+i)\right). \end{aligned}$$

It is worth mentioning that the summands of the above formula (without the scaling factor) are absolutely summable. Thus, dominated convergence leads to

$$\begin{aligned}\lim_{T/N\to\infty}\frac{T}{N}\mathrm{Cov}\left(\hat{\hat{\gamma}}^{ff}(p),\hat{\hat{\gamma}}^{sf}(q)\right) &= S_{p,q}+\frac{1}{N}\sum_{k=-\infty}^{\infty}\sum_{i=0}^{N-1}\left(\left(\gamma^{ff}(Nk-i+q-p)\otimes\gamma^{fs}(Nk-i)\right)+\right.\\ &\left.K_{n_f,n_f}\left(\gamma^{ff}(Nk-i+q)\otimes\gamma^{fs}(Nk-i-p)\right)\right)\end{aligned}$$

or

$$\begin{aligned}\lim_{T/N\to\infty}\frac{T}{N}\mathrm{Cov}\left(\hat{\hat{\gamma}}^{ff}(p),\hat{\hat{\gamma}}^{sf}(q)\right) &= S_{p,q}+\frac{1}{N}\sum_{k=-\infty}^{\infty}\left(\left(\gamma^{ff}(k+q-p)\otimes\gamma^{fs}(k)\right)\right.\\ &\left.+K_{n_f,n_f}\left(\gamma^{ff}(k+q)\otimes\gamma^{fs}(k-p)\right)\right)\end{aligned}$$

where

$$\begin{aligned}S_{p,q} &= \frac{1}{N}\sum_{i=0}^{N-1}\sum_{r=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}\left(k_{k-p}^{f}\otimes k_{k}^{f}\right)\kappa\left(k_{k-q+Nr+i}^{f}\otimes k_{k+Nr+i}^{s}\right)^{T}\\ &= \frac{1}{N}\sum_{r=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}\left(k_{k-p}^{f}\otimes k_{k}^{f}\right)\kappa\left(k_{k-q+r}^{f}\otimes k_{k+r}^{s}\right)^{T}.\end{aligned}$$

□

REMARK 5.1.5. Replacing $\hat{\gamma}^{sf}(q)$ in the lemma above by the "high-frequency autocovariance estimator", i.e. by $\frac{1}{T}\sum_{t=1}^{T}y_t^s\left(y_{t-q}^f\right)^T$, the result of Lemma 5.1.4 is still valid. This result is surprising since for the case $N=2$ we only use half of the y_t^s observations.

LEMMA 5.1.6. *Let $(y_t)_{t\in\mathbb{Z}}$ be a linear process and assume that the fourth moment of ν_t, $\eta=\mathbb{E}\left(\nu_t\nu_t^T\otimes\nu_t\nu_t^T\right)$, exists. Then*

$$\begin{aligned}\mathbb{E}\left(\left(y_{Nt-p}^{f}\left(y_{Nt+Nh-q}^{f}\right)^{T}\right)\otimes\left(y_{Nt}^{s}\left(y_{Nt+Nh}^{s}\right)^{T}\right)\right) &= vec\left(\gamma^{sf}(p)\right)vec\left(\gamma^{sf}(q)\right)^{T}+R(p,q,Nh)\\ &+\left(\gamma^{ff}(-Nh+q-p)\otimes\gamma^{ss}(-Nh)\right)\\ &+K_{n_f,n_s}\left(\gamma^{sf}(-Nh+q)\otimes\gamma^{fs}(-Nh-p)\right)\end{aligned}$$

where p, q, t, $h\in\mathbb{Z}$, $1<N\in\mathbb{N}$ and

$$R(p,q,Nh)=\sum_{k=-\infty}^{\infty}\left(k_{k-p}^{f}\otimes k_{k}^{s}\right)\kappa\left(k_{k-q+Nh}^{f}\otimes k_{k+Nh}^{s}\right)^{T}.$$

PROOF. see Appendix B. □

The following result is a generalization of the result of Niebuhr and Kreiss (2013) to the multivariate case.

LEMMA 5.1.7. *Let $(y_t)_{t\in\mathbb{Z}}$ be a linear process and assume that the fourth moment of ν_t, $\eta=\mathbb{E}\left(\nu_t\nu_t^T\otimes\nu_t\nu_t^T\right)$, exists. Then we obtain*

$$\lim_{T\to\infty}T\,Cov\left(vec\left(\hat{\gamma}^{sf}(p)\right),vec\left(\hat{\gamma}^{sf}(q)\right)\right)=N\left(S_{p,q}+R_{p,q}\right)\tag{5.1.4}$$

for $p, q \in \mathbb{Z}$ where

$$\begin{aligned} R_{p,q} &= \sum_{k=-\infty}^{\infty} \left(\gamma^{ff}(Nk+q-p) \otimes \gamma^{ss}(Nk)\right) + K_{n_f,n_s}\left(\gamma^{sf}(Nk+q) \otimes \gamma^{fs}(Nk-p)\right) \\ S_{p,q} &= \sum_{k=-\infty}^{\infty} \sum_{r=-\infty}^{\infty} \left(k^f_{k-p} \otimes k^s_k\right) \kappa \left(k^f_{Nr+k-q} \otimes k^s_{Nr+k}\right)^T. \end{aligned}$$

PROOF. Let us assume that T is a multiple of N. Again we will prove this lemma with $\hat{\hat{\gamma}}^{sf}(p) = \frac{N}{T}\sum_{t=1}^{T/N} y^s_{Nt}\left(y^f_{Nt-p}\right)^T$ instead of $\hat{\gamma}^{sf}(p)$ since it can be shown that this change does not influence the asymptotic behavior. Again we start with the observation that $\operatorname{vec}\left(y^s_{Nt}\left(y^f_{Nt-p}\right)^T\right) = \left(y^f_{Nt-p} \otimes y^s_{Nt}\right)$ and

$$\begin{aligned} \mathbb{E}\left(\operatorname{vec}\left(\hat{\hat{\gamma}}^{sf}(p)\right)\operatorname{vec}\left(\hat{\hat{\gamma}}^{sf}(q)\right)^T\right) &= \left(\frac{N}{T}\right)^2 \sum_{u=1}^{T/N}\sum_{t=1}^{T/N} \mathbb{E}\left(\operatorname{vec}\left(y^s_{Nt}\left(y^f_{Nt-p}\right)^T\right)\operatorname{vec}\left(y^s_{Nu}\left(y^f_{Nu-q}\right)^T\right)^T\right) \\ &= \left(\frac{N}{T}\right)^2 \sum_{u=1}^{T/N}\sum_{t=1}^{T/N} \mathbb{E}\left(y^f_{Nt-p}\left(y^f_{Nu-q}\right)^T \otimes y^s_{Nt}\left(y^s_{Nu}\right)^T\right). \end{aligned}$$

Now we can use Lemma 5.1.6 with $u = t+h$ to obtain that the sum can be rewritten into

$$\begin{aligned} &\left(\frac{N}{T}\right)^2 \sum_{u=1}^{T/N}\sum_{t=1}^{T/N} \operatorname{vec}\left(\gamma^{sf}(p)\right)\operatorname{vec}\left(\gamma^{sf}(q)\right)^T + \\ &\left(\gamma^{ff}(-N(u-t)+q-p) \otimes \gamma^{ss}(-N(u-t))\right) + \\ &K_{n_f,n_s}\left(\gamma^{sf}(-N(u-t)+q) \otimes \gamma^{fs}(-N(u-t)-p)\right) + R(p,q,N(u-t)) \\ = \;& \operatorname{vec}\left(\gamma^{sf}(p)\right)\operatorname{vec}\left(\gamma^{sf}(q)\right)^T + \left(\frac{N}{T}\right)^2 \sum_{u=1}^{T/N}\sum_{t=1}^{T/N} g(u-t). \end{aligned}$$

Using the formula $\sum_{u=1}^{n}\sum_{t=1}^{n} g(u-t) = \sum_{k=-(n-1)}^{n-1}(n-|k|)\,g(k)$ leads to

$$\begin{aligned} &= \operatorname{vec}\left(\gamma^{sf}(p)\right)\operatorname{vec}\left(\gamma^{sf}(q)\right)^T + \\ &\left(\frac{N}{T}\right) \sum_{|k|<T/N} \left(1 - \frac{N|k|}{T}\right)\left(\left(\gamma^{ff}(Nk+q-p) \otimes \gamma^{ss}(Nk)\right) + \right. \\ &\left. K_{n_f,n_s}\left(\gamma^{sf}(Nk+q) \otimes \gamma^{fs}(Nk-p)\right) + R(p,q,Nk)\right). \end{aligned}$$

Thus, dominated convergence leads to

$$\begin{aligned} \lim_{T/N\to\infty} \frac{T}{N}\operatorname{Cov}\left(\hat{\gamma}^{sf}(p), \hat{\gamma}^{sf}(q)\right) &= S_{p,q} + \sum_{k=-\infty}^{\infty}\left(\gamma^{ff}(Nk+q-p) \otimes \gamma^{ss}(Nk)\right) \\ &\quad + K_{n_f,n_s}\left(\gamma^{sf}(Nk+q) \otimes \gamma^{fs}(Nk-p)\right) \end{aligned}$$

where

$$S_{p,q} = \sum_{k=-\infty}^{\infty}\sum_{r=-\infty}^{\infty}\left(k^f_{k-p} \otimes k^s_k\right)\kappa\left(k^f_{Nr+k-q} \otimes k^s_{Nr+k}\right)^T.$$

□

Note that for the case $N = 1$ we still obtain Bartlett's formula for the high-frequency case. It is worth mentioning that for the case $N > 1$, regarding the superscripts, the right hand side of equation (5.1.4) is not a multiple of the right hand side of equation (5.1.1). Indeed, for fixed p and q the infinite sum $R_{p,q}$ of the right hand side of equation (5.1.4) contains only every N-th summand of the infinite sum of the right hand side of equation (5.1.1).

THEOREM 5.1.8. *Let $(y_t)_{t\in\mathbb{Z}}$ be a linear process, assume that the fourth moment of ν_t, $\eta = \mathbb{E}\left(\nu_t\nu_t^T \otimes \nu_t\nu_t^T\right)$ exists and $s \in \mathbb{N}$. Then we have asymptotic normality of the autocovariance estimators, i.e.*

$$\sqrt{T}\left(\begin{pmatrix} vec\left(\hat{\gamma}^{ff}(0)\right) \\ vec\left(\hat{\gamma}^{sf}(0)\right) \\ \vdots \\ vec\left(\hat{\gamma}^{ff}(s)\right) \\ vec\left(\hat{\gamma}^{sf}(s)\right) \end{pmatrix} - \begin{pmatrix} vec\left(\gamma^{ff}(0)\right) \\ vec\left(\gamma^{sf}(0)\right) \\ \vdots \\ vec\left(\gamma^{ff}(s)\right) \\ vec\left(\gamma^{sf}(s)\right) \end{pmatrix}\right) \xrightarrow{d} \mathcal{N}_h\left(0, \Sigma_\gamma\right)$$

where $h = (s+1)\,nn_f$ and Σ_γ is obtained from Lemma 5.1.1, Lemma 5.1.4 and Lemma 5.1.7.

PROOF. In order to apply Theorem 2.2.22 we define a particular blocked process

$$u_t = \begin{pmatrix} y_{Nt} \\ y_{Nt-1} \\ \vdots \\ y_{Nt-(N-1)} \end{pmatrix}.$$

Of course, this process is again a linear one as can be easily seen:

$$u_t = \sum_{j=-\infty}^{\infty} \begin{pmatrix} k_{Nj} & k_{Nj+1} & \dots & k_{Nj+N-1} \\ k_{Nj-1} & k_{Nj} & \dots & k_{Nj+N-2} \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \underbrace{\begin{pmatrix} \nu_{N(t-j)} \\ \vdots \\ \nu_{N(t-j)-(N-1)} \end{pmatrix}}_{\tilde{\nu}_{t-j}}$$

where $\tilde{\nu}_t \sim IID_{Nn}\left(0, I_N \otimes \Sigma_\nu\right)$. Now applying Theorem 2.2.22 leads to

$$\sqrt{\frac{T}{N}}\left(\begin{pmatrix} \text{vec}\left(\hat{\gamma}_u(0)\right) \\ \vdots \\ \text{vec}\left(\hat{\gamma}_u(s)\right) \end{pmatrix} - \begin{pmatrix} \text{vec}\left(\gamma_u(0)\right) \\ \vdots \\ \text{vec}\left(\gamma_u(s)\right) \end{pmatrix}\right) \xrightarrow{d} \mathcal{N}_{n^2(s+1)}\left(0, \Sigma_u\right)$$

where $\gamma_u(i)$ is the population autocovariance of the process $(u_t)_{t\in\mathbb{Z}}$ for lag i and $\hat{\gamma}_u(i)$ is its sample counterpart with T/N summands. In a last step we have to find a transformation matrix, say H, which transforms $\hat{\gamma}_u$ to the desired autocovariances. To obtain this transformation we define $H_1 = \begin{pmatrix} \frac{1}{N}I_{n_f^2} & \frac{1}{N}I_{n_f^2} & \cdots & \frac{1}{N}I_{n_f^2} \end{pmatrix}$ and observe that for $j = 0, \dots, s$

$$\begin{aligned} \text{vec}\left(\hat{\gamma}^{sf}(j)\right) &= \frac{N}{T}\sum_{i=1}^{T/N} \text{vec}\left(y_{Nt}^s\left(y_{Nt-j}^f\right)^T\right) \\ &= S_j^{sf}\left(\text{vec}\left(\hat{\gamma}_u(i)\right)\right)_{i=0,\dots,s} \end{aligned}$$

and

$$\begin{aligned}\operatorname{vec}\left(\hat{\hat{\gamma}}^{ff}(j)\right) &= H_1\begin{pmatrix}\frac{N}{T}\sum_{i=1}^{T/N}\operatorname{vec}\left(y^f_{Nt}\left(y^f_{Nt-j}\right)^T\right)\\ \frac{N}{T}\sum_{i=1}^{T/N}\operatorname{vec}\left(y^f_{Nt-1}\left(y^f_{Nt-1-j}\right)^T\right)\\ \vdots\\ \frac{N}{T}\sum_{i=1}^{T/N}\operatorname{vec}\left(y^f_{Nt-(N-1)}\left(y^f_{Nt-(N-1)-j}\right)^T\right)\end{pmatrix}\\ &= H_1 S_j^{ff}\left(\operatorname{vec}\left(\hat{\gamma}_u(i)\right)\right)_{i=0,\ldots,s}\end{aligned}$$

where S_j^{ff} and S_j^{sf} are selector matrices for lag j. Finally, we can construct our particular transformation matrix $H = \left(\left(H_1 S_0^{ff}\right)^T, \left(S_0^{sf}\right)^T, \ldots, \left(H_1 S_s^{ff}\right)^T, \left(S_s^{sf}\right)^T\right)^T$ to obtain the desired result

$$\begin{aligned}&\sqrt{\frac{T}{N}}\left(\begin{pmatrix}\operatorname{vec}\left(\hat{\gamma}^{ff}(i)\right)\\ \operatorname{vec}\left(\hat{\gamma}^{sf}(i)\right)\end{pmatrix} - \begin{pmatrix}\operatorname{vec}\left(\gamma^{ff}(i)\right)\\ \operatorname{vec}\left(\gamma^{sf}(i)\right)\end{pmatrix}\right)_{i=0,\ldots,s} =\\ &\qquad \sqrt{\frac{T}{N}} H\left(\operatorname{vec}\left(\hat{\gamma}_u(i)\right) - \operatorname{vec}\left(\gamma_u(i)\right)\right)_{i=0,\ldots,s} \xrightarrow{d} \mathcal{N}_h\left(0, \Sigma_\gamma\right).\end{aligned}$$

The asymptotic covariance $\Sigma_\gamma = H\Sigma_u H^T$ can be derived by using Lemmas 5.1.1, 5.1.4 and 5.1.7. □

REMARK 5.1.9. The last theorem can be extended to any set of lags including negative ones. Indeed we will use the set of lags $(-p+1, \ldots, np)$ for the XYW and GMM estimator. Also note, that the assumption that the innovations are i.i.d. can be relaxed, see e.g. Hall and Heyde (1980).

We do not distinguish between singular and nonsingular normal distributions. A singular normal distribution may occur since for example $\hat{\gamma}^{ff}(i) = \hat{\gamma}^{ff}(-i)^T$ holds.

5.2. Asymptotic Normality of the XYW/GMM Estimator

Before we consider the asymptotic normality of the XYW/GMM estimator, we first have a look at the consistency, which can be easily obtained with the aid of Theorem 2.2.17.

THEOREM 5.2.1. *Let $(y_t)_{t\in\mathbb{Z}}$ be the output of system (2.1.6) with inputs $(\nu_t)_{t\in\mathbb{Z}} \sim IID_n(0, \Sigma_\nu)$ and $\theta \in \Theta_{XYW}$. Then the XYW estimator for the system parameters is consistent, i.e.*

$$\hat{A}_{XYW} \xrightarrow{p} A.$$

PROOF. As mentioned above, Theorem 2.2.17 guarantees the consistency of the autocovariance estimators, i.e. $\hat{Z}_0 \xrightarrow{p} Z_0$ and $\hat{Z}_1 \xrightarrow{p} Z_1$. This together with Lemma 2.2.5, i.e. $\hat{Z}_0^\dagger \xrightarrow{p} Z_0^\dagger$, and Lemma 2.2.7, i.e. $\hat{Z}_1\hat{Z}_0^\dagger \xrightarrow{p} Z_1 Z_0^\dagger$, establishes our result. □

Having obtained the asymptotic distribution of the covariance estimators, we have to linearize the mapping attaching the system parameters to the second moments of the observations. The next theorem derives the asymptotic distribution of the XYW/GMM estimators and is related to Gingras (1985).

THEOREM 5.2.2. *Let $(y_t)_{t\in\mathbb{Z}}$ be the output of system (2.1.6) with inputs $(\nu_t)_{t\in\mathbb{Z}} \sim IID_n(0,\Sigma_\nu)$, $\theta \in \Theta_{XYW}$ and assume that $\eta = \mathbb{E}\left(\nu_t\nu_t^T \otimes \nu_t\nu_t^T\right)$ exists. Then the GMM estimator*

$$\begin{aligned} vec\left(\hat{A}_{GMM}\right) &= \left(\left(\hat{Z}_0 \otimes I_n\right) Q_T \left(\hat{Z}_0^T \otimes I_n\right)\right)^{-1} \left(\hat{Z}_0 \otimes I_n\right) Q_T vec\left(\hat{Z}_1\right) \\ &= \hat{G}^\dagger_{Q_T} vec\left(\hat{Z}_1\right) \end{aligned}$$

is asymptotically normal with zero mean and a covariance matrix given by

$$\Sigma_{GMM} = \left(G^\dagger_{Q_0} JP\right) \Sigma_\gamma \left(G^\dagger_{Q_0} JP\right)^T, \tag{5.2.1}$$

i.e.

$$\sqrt{T}\left(vec\left(\hat{A}_{GMM}\right) - vec\left(A\right)\right) \xrightarrow{d} \mathcal{N}_{n^2p}\left(0, \Sigma_{GMM}\right). \tag{5.2.2}$$

Here $Q_T \xrightarrow{p} Q_0$ where Q_0 is constant, symmetric and positive definite, Σ_γ is the asymptotic covariance of the mixed-frequency autocovariances described in Theorem 5.1.8 for the lags $(-p+1,\ldots,np)$. Furthermore,

$$\begin{aligned} G^\dagger_{Q_0} &= \left(\left(Z_0 \otimes I_n\right) Q_0 \left(Z_0^T \otimes I_n\right)\right)^{-1} \left(Z_0 \otimes I_n\right) Q_0 \\ J &= \begin{pmatrix} D & 0_{n\times n} & & 0_{n\times n} \\ 0_{n\times n} & D & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0_{n\times n} \\ 0_{n\times n} & \cdots & 0_{n\times n} & D \end{pmatrix} \in \mathbb{R}^{n^2 p n_f \times n(n+1) n_f p} \\ D &= \begin{pmatrix} -A_p & 0_{n\times(n_f-1)n} & -A_{p-1} & 0_{n\times(n_f-1)n} \\ \ldots & -A_1 & 0_{n\times(n_f-1)n} & I_n & 0_{n\times(n_f-1)n} \end{pmatrix} \end{aligned}$$

and the permutation matrix P is given as $P = \left(I_{(n+1)p} \otimes P_2\right)$ where

$$P_2 = \left(I_{n_f} \otimes \begin{pmatrix} I_{n_f} \\ 0_{n_s\times n_f} \end{pmatrix}, \quad I_{n_f} \otimes \begin{pmatrix} 0_{n_f\times n_s} \\ I_{n_s} \end{pmatrix}\right).$$

REMARK 5.2.3. It is easy to see that

$$\operatorname{vec}\begin{pmatrix} \hat{\gamma}^{ff}(h) \\ \hat{\gamma}^{sf}(h) \end{pmatrix} = P_2 \begin{pmatrix} \operatorname{vec}\left(\hat{\gamma}^{ff}(h)\right) \\ \operatorname{vec}\left(\hat{\gamma}^{sf}(h)\right) \end{pmatrix}$$

and thus P is a permutation matrix. In the case where only one fast component occurs, i.e. $n_f = 1$, the commutation matrix P is the identity matrix and

$$J = \begin{pmatrix} -A_p & \ldots & -A_1 & I_n & & \\ & -A_p & \ldots & -A_1 & I_n & \\ & & \ddots & & & \ddots \\ & & & -A_p & \ldots & -A_1 & I_n \end{pmatrix}.$$

Also note that in this case (and if $k = 0$ holds) the weighting matrix Q_T is not needed.

PROOF. We commence from the observation that

$$\begin{aligned}\sqrt{T}\left(\mathrm{vec}\left(\hat{A}_{\mathrm{GMM}}\right)-\mathrm{vec}\left(A\right)\right) &= \sqrt{T}\left(\hat{G}_{Q_T}^{\dagger}\mathrm{vec}\left(\hat{Z}_1\right)-\hat{G}_{Q_T}^{\dagger}\left(\hat{Z}_0^T\otimes I_n\right)\mathrm{vec}\left(A\right)\right)\\ &= \sqrt{T}\hat{G}_{Q_T}^{\dagger}\mathrm{vec}\left(\hat{Z}_1-A\hat{Z}_0-Z_1+AZ_0\right)\\ &= \sqrt{T}\hat{G}_{Q}^{\dagger}\mathrm{vec}\left(\begin{pmatrix}I_n & -A\end{pmatrix}\begin{pmatrix}\hat{Z}_1-Z_1\\ \hat{Z}_0-Z_0\end{pmatrix}\right)\\ &= \sqrt{T}\hat{G}_{Q_T}^{\dagger}\underbrace{\left(I_{pnn_f}\otimes\begin{pmatrix}I_n & -A\end{pmatrix}\right)}_{J_1}\mathrm{vec}\begin{pmatrix}\hat{Z}_1-Z_1\\ \hat{Z}_0-Z_0\end{pmatrix}.\end{aligned}$$

The last term $\mathrm{vec}\begin{pmatrix}\hat{Z}_1-Z_1\\ \hat{Z}_0-Z_0\end{pmatrix}$ has to be rewritten since there are many $\begin{pmatrix}\hat{\gamma}^{ff}(i)-\gamma^{ff}(i)\\ \hat{\gamma}^{sf}(i)-\gamma^{sf}(i)\end{pmatrix}$ which occur in $\hat{Z}_1-Z_1$ and in $\hat{Z}_0-Z_0$. Thus,

$$\begin{aligned}\begin{pmatrix}\hat{Z}_1\\ \hat{Z}_0\end{pmatrix} &= \underbrace{\begin{pmatrix}I_n\\ 0_{pn\times n}\end{pmatrix}}_{E_1}\begin{pmatrix}\hat{\gamma}^{ff}(-p+1) & \cdots & \hat{\gamma}^{ff}(np)\\ \hat{\gamma}^{sf}(-p+1) & \cdots & \hat{\gamma}^{sf}(np)\end{pmatrix}\underbrace{\begin{pmatrix}0_{pn_f\times npn_f}\\ I_{npn_f}\end{pmatrix}}_{F_1}\\ &+\underbrace{\begin{pmatrix}0_{n\times n}\\ I_n\\ 0_{n\times n}\\ \vdots\\ 0_{n\times n}\end{pmatrix}}_{E_2}\begin{pmatrix}\hat{\gamma}^{ff}(-p+1) & \cdots & \hat{\gamma}^{ff}(np)\\ \hat{\gamma}^{sf}(-p+1) & \cdots & \hat{\gamma}^{sf}(np)\end{pmatrix}\underbrace{\begin{pmatrix}0_{(p-1)n_f\times npn_f}\\ I_{npn_f}\\ 0_{n_f\times npn_f}\end{pmatrix}}_{F_2}+\cdots\\ &+\underbrace{\begin{pmatrix}0_{pn\times n}\\ I_n\end{pmatrix}}_{E_{p+1}}\begin{pmatrix}\hat{\gamma}^{ff}(-p+1) & \cdots & \hat{\gamma}^{ff}(np)\\ \hat{\gamma}^{sf}(-p+1) & \cdots & \hat{\gamma}^{sf}(np)\end{pmatrix}\underbrace{\begin{pmatrix}I_{npn_f}\\ 0_{pn_f\times npn_f}\end{pmatrix}}_{F_{p+1}}.\end{aligned}$$

Vectorizing the above formula leads to

$$\mathrm{vec}\begin{pmatrix}\hat{Z}_1\\ \hat{Z}_0\end{pmatrix}=\underbrace{\left(\sum_{i=1}^{p+1}\left(F_i^T\otimes E_i\right)\right)}_{J_2}\mathrm{vec}\begin{pmatrix}\hat{\gamma}^{ff}(-p+1) & \cdots & \hat{\gamma}^{ff}(np)\\ \hat{\gamma}^{sf}(-p+1) & \cdots & \hat{\gamma}^{sf}(np)\end{pmatrix}.$$

The last term in the last formula can be reformulated to

$$\mathrm{vec}\begin{pmatrix}\hat{\gamma}^{ff}(-p+1) & \cdots & \hat{\gamma}^{ff}(np)\\ \hat{\gamma}^{sf}(-p+1) & \cdots & \hat{\gamma}^{sf}(np)\end{pmatrix}=P\begin{pmatrix}\mathrm{vec}\left(\hat{\gamma}^{ff}(i)\right)\\ \mathrm{vec}\left(\hat{\gamma}^{sf}(i)\right)\end{pmatrix}_{i=-p+1,\ldots,np}.$$

To summarize, we can rewrite $\sqrt{T}\left(\mathrm{vec}\left(\hat{A}_{\mathrm{GMM}}\right)-\mathrm{vec}\left(A\right)\right)$ as

$$\sqrt{T}\left(\mathrm{vec}\left(\hat{A}_{\mathrm{GMM}}\right)-\mathrm{vec}\left(A\right)\right) = \sqrt{T}\hat{G}_{Q_T}^{\dagger}\underbrace{J_1J_2}_{J}P\left(\begin{pmatrix}\mathrm{vec}\left(\hat{\gamma}^{ff}(i)\right)\\ \mathrm{vec}\left(\hat{\gamma}^{sf}(i)\right)\end{pmatrix}-\begin{pmatrix}\mathrm{vec}\left(\gamma^{ff}(i)\right)\\ \mathrm{vec}\left(\gamma^{sf}(i)\right)\end{pmatrix}\right)_{i=-p+1,\ldots,np}$$

Theorem 5.1.8 implies that

$$\sqrt{T}\left(\begin{pmatrix}\operatorname{vec}\begin{pmatrix}\hat{\gamma}^{ff}(i)\end{pmatrix}\\ \operatorname{vec}\begin{pmatrix}\hat{\gamma}^{sf}(i)\end{pmatrix}\end{pmatrix}-\begin{pmatrix}\operatorname{vec}\begin{pmatrix}\gamma^{ff}(i)\end{pmatrix}\\ \operatorname{vec}\begin{pmatrix}\gamma^{sf}(i)\end{pmatrix}\end{pmatrix}\right)_{i=-p+1,\ldots,np} \xrightarrow{d} \mathcal{N}_h\left(0,\Sigma_\gamma\right).$$

Since the sample autocovariances are consistent estimators and $Q_T \xrightarrow{p} Q_0$, the same is true for $\hat{G}^\dagger_{Q_T}JP$, i.e. $\hat{G}^\dagger_{Q_T}JP \xrightarrow{p} G^\dagger_{Q_0}JP$. Now Slutsky's Lemma 2.2.7 directly leads to the result of the theorem. □

REMARK 5.2.4. As has been shown in Theorem 2.2.25 the asymptotic covariance for the standard high-frequency Yule-Walker estimator is of the form $\left(\Gamma_p^{-1}\otimes\Sigma_\nu\right)$ and thus, in this case, the fourth moment of the innovations does not influence the asymptotic covariance of the parameter estimates. In the case discussed here the fourth moment of the innovations does not vanish under linearization in general.

Having obtained the expression for the asymptotic covariance matrix we can determine the asymptotically optimal weighting matrix.

THEOREM 5.2.5. *Under the assumptions of Theorem 5.2.2, the optimal asymptotic weighting matrix for the GMM estimator is*

$$Q_0^* = \left(JP\Sigma_\gamma P^T J^T\right)^{-1} \tag{5.2.3}$$

and the corresponding asymptotic covariance is given by

$$\Sigma^*_{GMM} = \left(\left(Z_0\otimes I_n\right)Q_0^*\left(Z_0^T\otimes I_n\right)\right)^{-1}. \tag{5.2.4}$$

PROOF. The theorem directly follows from Theorem 3.2 in Hansen (1982). □

PROPOSITION 5.2.6. *Let $(y_t)_{t\in\mathbb{Z}}$ be the output of system (2.1.6) with inputs $(\nu_t)_{t\in\mathbb{Z}}\sim IID_n(0,\Sigma_\nu)$, $\theta\in\Theta_{XYW}$ and assume that $\eta=\mathbb{E}\left(\nu_t\nu_t^T\otimes\nu_t\nu_t^T\right)$ exists. Then the XYW estimator*

$$vec\left(\hat{A}_{XYW}\right)=\left(\left(\hat{Z}_0\hat{Z}_0^T\otimes I_n\right)\right)^{-1}\left(\hat{Z}_0\otimes I_n\right)vec\left(\hat{Z}_1\right)$$

is asymptotically normal with zero mean and a covariance matrix given by

$$\Sigma_{XYW}=\left(\left(Z_0^\dagger\right)^T\otimes I_n\right)JP\Sigma_\gamma P^TJ^T\left(Z_0^\dagger\otimes I_n\right) \tag{5.2.5}$$

i.e.

$$\sqrt{T}\left(vec\left(\hat{A}_{XYW}\right)-vec\left(A\right)\right)\xrightarrow{d}\mathcal{N}_{n^2p}\left(0,\Sigma_{XYW}\right) \tag{5.2.6}$$

PROOF. This is a special case of Theorem 5.2.2 with the weighting matrix $Q_T=I_{npn_f}$. □

To obtain a feasible GMM estimator we also have to estimate the optimal weighting matrix Q_0^* in a consistent way, say $\hat{Q}_T$. Note that Q_0^* depends on the system parameters, on the fourth moment of ν_t and on infinite sums of the autocovariances. Thus, in a first step we have to estimate the system parameters and the fourth moment of ν_t (for consistency see Theorem 5.3.2) which can be done, for instance, by the XYW estimator. In Robinson (1977); Berlinet and Francq (1997) it is stated that a consistent estimator for the infinite sums can be obtained by using, for instance, a truncation of these

sums where the truncation depends on the sample size T. As an example one can use the truncation at the $\lfloor K\left(T/N\right)^{\alpha}\rfloor$ summand where $K>0$ and $0<\alpha<1$.

With these results from a certain T_0 onwards, we can define our feasible (two-step) GMM estimator for the mixed-frequency case as

$$\operatorname{vec}\left(\hat{\hat{A}}_{\mathrm{GMM}}\right)=\left(\left(\hat{Z}_0\otimes I_n\right)\hat{Q}_T\left(\hat{Z}_0^T\otimes I_n\right)\right)^{-1}\left(\hat{Z}_0\otimes I_n\right)\hat{Q}_T\operatorname{vec}\left(\hat{Z}_1\right), \tag{5.2.7}$$

where the optimal weighting matrix is estimated in a first step.

REMARK 5.2.7. Until now, the asymptotic results obtained in this section are only valid for the stock case. Nevertheless, a generalization to the flow case with a simple aggregation scheme is straightforward: Using an obvious notation and following the same steps as in the proof of Theorem 5.2.2, we obtain

$$\sqrt{T}\left(\operatorname{vec}\left(\hat{A}^{\mathrm{g}}_{\mathrm{GMM}}\right)-\operatorname{vec}\left(A\right)\right) \quad = \quad \sqrt{T}\left(\hat{G}^{\mathrm{g}}_{Q_T}\right)^{\dagger}JP\left(\begin{pmatrix}\operatorname{vec}\left(\hat{\gamma}^{z^f f}\left(i\right)\right)\\ \operatorname{vec}\left(\hat{\gamma}^{wf}\left(i\right)\right)\end{pmatrix}-\begin{pmatrix}\operatorname{vec}\left(\gamma^{z^f f}\left(i\right)\right)\\ \operatorname{vec}\left(\gamma^{wf}\left(i\right)\right)\end{pmatrix}\right),$$

for $i=N-p,\ldots,N+np-1$. For the asymptotic covariance matrix of the autocovariance estimators we can modify Lemmas 5.1.1, 5.1.4 and 5.1.7 to

$$\lim_{T\to\infty}T\operatorname{Cov}\left(\operatorname{vec}\left(\hat{\gamma}^{z^f f}\left(p\right)\right),\operatorname{vec}\left(\hat{\gamma}^{z^f f}\left(q\right)\right)\right) = S^1_{p,q}+R^1_{p,q} \tag{5.2.8}$$

$$\lim_{T\to\infty}T\operatorname{Cov}\left(\operatorname{vec}\left(\hat{\gamma}^{z^f f}\left(p\right)\right),\operatorname{vec}\left(\hat{\gamma}^{wf}\left(q\right)\right)\right) = S^2_{p,q}+R^2_{p,q} \tag{5.2.9}$$

$$\lim_{T\to\infty}T\operatorname{Cov}\left(\operatorname{vec}\left(\hat{\gamma}^{wf}\left(p\right)\right),\operatorname{vec}\left(\hat{\gamma}^{wf}\left(q\right)\right)\right) = N\left(S^3_{p,q}+R^3_{p,q}\right), \tag{5.2.10}$$

where $p,q\in\mathbb{Z}$ and

$$\begin{aligned}
R^1_{p,q} &= \sum_{k=-\infty}^{\infty}\left(\gamma^{ff}\left(k+q-p\right)\otimes\gamma^{z^f z^f}\left(k\right)\right)+K_{n_f,n_f}\left(\gamma^{z^f f}\left(k+q\right)\otimes\gamma^{f z^f}\left(k-p\right)\right)\\
S^1_{p,q} &= \sum_{i,j=1}^{N}\sum_{k=-\infty}^{\infty}\sum_{r=-\infty}^{\infty}c_i c_j\left(k^f_{k-p}\otimes k^f_{k-i+1}\right)\kappa\left(k^f_{r+k-q}\otimes k^f_{r+k-j+1}\right)^T\\
R^2_{p,q} &= \sum_{k=-\infty}^{\infty}\left(\gamma^{ff}\left(k+q-p\right)\otimes\gamma^{z^f w}\left(k\right)\right)+K_{n_f,n_f}\left(\gamma^{z^f f}\left(k+q\right)\otimes\gamma^{fw}\left(k-p\right)\right)\\
S^2_{p,q} &= \sum_{i,j=1}^{N}\sum_{k=-\infty}^{\infty}\sum_{r=-\infty}^{\infty}c_i c_j\left(k^f_{k-p}\otimes k^f_{k-i+1}\right)\kappa\left(k^f_{r+k-q}\otimes k^s_{r+k-j+1}\right)^T\\
R^3_{p,q} &= \sum_{k=-\infty}^{\infty}\left(\gamma^{ff}\left(Nk+q-p\right)\otimes\gamma^{ww}\left(kN\right)\right)+K_{n_f,n_s}\left(\gamma^{wf}\left(Nk+q\right)\otimes\gamma^{fw}\left(Nk-p\right)\right)\\
S^3_{p,q} &= \sum_{i,j=1}^{N}\sum_{k=-\infty}^{\infty}\sum_{r=-\infty}^{\infty}c_i c_j\left(k^f_{k-p}\otimes k^s_{k-i+1}\right)\kappa\left(k^f_{Nr+k-q}\otimes k^s_{Nr+k-j+1}\right)^T.
\end{aligned}$$

In the next example we will have a closer look at the estimates of the mean and the autocovariance for the flow case.

EXAMPLE 5.2.8. For the case $N = 2$, consider the aggregation scheme

$$w_t = y_t^s + y_{t-1}^s$$

where we observe w_t at every second time point. Taking the expectation of w_t we get $\mathbb{E}(w_t) = \mathbb{E}(y_t^s) + \mathbb{E}\left(y_{t-1}^s\right) = 0$. The mean estimator in this case is

$$\begin{aligned} \bar{w}_T = \frac{2}{T}\sum_{t=1}^{T/2} w_{2t} &= \frac{2}{T}\sum_{t=1}^{T/2}\left(y_{2t}^s + y_{2t-1}^s\right) \\ &= \frac{2}{T}\sum_{i=1}^{T} y_t^s = 2\bar{y}_T^s. \end{aligned} \tag{5.2.11}$$

The cross-covariance of w_t and y_t is $\gamma^{wf}(i) = \mathbb{E}\left(w_t\left(y_{t-i}^f\right)^T\right) = \mathbb{E}\left(y_t^s\left(y_{t-i}^f\right)^T\right) + \mathbb{E}\left(y_{t-1}^s\left(y_{t-i}^f\right)^T\right) = \gamma^{sf}(i) + \gamma^{sf}(i-1)$ whereas the same relation is not true for the cross-covariance estimators:

$$\begin{aligned} \hat{\gamma}^{wf}(i) &= \frac{2}{T}\sum_{t=1}^{T/2}\left(w_{2t} - \bar{w}_T\right)\left(y_{2t-i}^f - \bar{y}_T^f\right)^T \\ &= \frac{2}{T}\sum_{t=1}^{T/2}\left(y_{2t}^s + y_{2t-1}^s - 2\bar{y}_T^s\right)\left(y_{2t-i}^f - \bar{y}_T^f\right)^T \\ &= \frac{2}{T}\sum_{t=1}^{T/2}\left(\left(y_{2t}^s - \bar{y}_T^s\right)\left(y_{2t-i}^f - \bar{y}_T^f\right)^T + \left(y_{2t-1}^s - \bar{y}_T^s\right)\left(y_{2t-i}^f - \bar{y}_T^f\right)^T\right) \\ &\neq \hat{\gamma}^{sf}(i) + \hat{\gamma}^{sf}(i-1) \end{aligned}$$

where $\hat{\gamma}^{sf}(i)$ is the autocovariance estimator in the stock case.

5.2.1. Upper and Lower Bounds for the Asymptotic Covariance. In Chen and Zadrozny (1998) on p. 56, upper and lower bounds for the asymptotic covariance of the XYW estimator are introduced. There are mainly three reasons why we are interested in such bounds:

- Firstly, as can be seen from Lemmas 5.1.1, 5.1.4 and 5.1.7 infinite sums occur in the formulas for the asymptotic variances. If we want to calculate these variances, we have to approximate them by finite ones. Of course, this approximation error can be made arbitrarily small (note that the $\gamma(i)$ decrease to zero in a geometric way) but on the other hand this will increase the computational load.
- Secondly, as mentioned in Remark 5.2.4, we have to calculate the fourth moment of ν_t, i.e. $\eta = \mathbb{E}\left(\nu_t\nu_t^T \otimes \nu_t\nu_t^T\right)$ or at least to assume that $(\nu_t)_{t\in\mathbb{Z}}$ are normally distributed so that the fourth moment does not occur.
- Thirdly, as will become clear later on, we want to split the estimation loss from high-frequency Yule-Walker to mixed-frequency extended Yule-Walker in several parts. The lower bound will be an appropriate intermediate step for that analysis.

For the lower bound we will use the XYW equations, as described in (3.2.12), but instead of using the mixed-frequency autocovariance estimators we will use the high-frequency autocovariance estimators.

THEOREM 5.2.9. *Let $(y_t)_{t\in\mathbb{Z}}$ be the output of system (2.1.6) with inputs $(\nu_t)_{t\in\mathbb{Z}} \sim IID_n(0,\Sigma_\nu)$ and $\theta \in \Theta_{XYW}$. Then the XYW estimator obtained via the high-frequency autocovariance estimators, say $\hat{A}^{full}_{XYW}$, is asymptotically normal with a zero mean and covariance matrix given by*

$$\Sigma^{full}_{XYW} = \left(\left(\left(Z_0^\dagger\right)^T \Gamma^{ff}_{np} Z_0^\dagger\right) \otimes \Sigma_\nu\right) \tag{5.2.12}$$

i.e.

$$\sqrt{T}\left(vec\left(\hat{A}^{full}_{XYW}\right) - vec\left(A\right)\right) \xrightarrow{d} \mathcal{N}_{n^2p}\left(0, \Sigma^{full}_{XYW}\right) \tag{5.2.13}$$

where

$$\Gamma^{ff}_{np} = \begin{pmatrix} \gamma^{ff}(0) & \cdots & \gamma^{ff}(np-1) \\ \vdots & \ddots & \vdots \\ \gamma^{ff}(1-np) & \cdots & \gamma^{ff}(0) \end{pmatrix}. \tag{5.2.14}$$

PROOF. The proof follows the same idea as the proof of Theorem 2.2.25. Consider

$$\sqrt{T}\hat{Z}_1 = \sqrt{T}\left(A_1,\ldots,A_p\right)\hat{Z}_0 + \sqrt{T}\underbrace{\left(\hat{e}(1),\ldots,\hat{e}(np)\right)}_{\hat{e}} + \sqrt{T}c_T \tag{5.2.15}$$

where

$$\hat{e}\left(i\right) = \frac{1}{T}\sum_{t=1}^{T} \nu_t \left(y^f_{t-i}\right)^T \quad i = 1,\ldots,np$$

and $\sqrt{T}c_T \xrightarrow{p} 0$ and thus will be neglected. Columnwise vectorization of equation (5.2.15) leads to

$$\sqrt{T}\text{vec}\left(\hat{Z}_1\right) = \sqrt{T}\left(\left(\hat{Z}_0\right)^T \otimes I_n\right)\text{vec}\left(A\right) + \sqrt{T}\text{vec}\left(\hat{e}\right).$$

Thus, premultiplying the above equation by $\left(\left(Z_0^\dagger\right)^T \otimes I_n\right)$ we directly get

$$\sqrt{T}\text{vec}\left(A\right) + \sqrt{T}\left(\left(Z_0^\dagger\right)^T \otimes I_n\right)\text{vec}\left(\hat{e}\right) = \sqrt{T}\left(\left(Z_0^\dagger\right)^T \otimes I_n\right)\text{vec}\left(\hat{Z}_1\right) = \sqrt{T}\text{vec}\left(\hat{A}^{\text{full}}_{\text{XYW}}\right)$$

or in another form:

$$\sqrt{T}\left(\text{vec}\left(\hat{A}^{\text{full}}_{\text{XYW}}\right) - \text{vec}\left(A\right)\right) = \sqrt{T}\left(\left(Z_0^\dagger\right)^T \otimes I_n\right)\text{vec}\left(\hat{e}\right). \tag{5.2.16}$$

It is easy to see that $\sqrt{T}\text{vec}\left(\hat{e}\right)$ has a zero mean and covariance $(\Gamma^{ff}_{np} \otimes \Sigma_\nu)$ since for a particular element of $\mathbb{V}\left(\text{vec}\left(\hat{e}\right)\right)$ it follows that

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{T^2}\left(\sum_{t=1}^{T} y^f_{t-p}\otimes\nu_t\right)\left(\sum_{n=1}^{T} y^f_{n-q}\otimes\nu_n\right)^T\right) &= \frac{1}{T^2}\mathbb{E}\left(\sum_{t=1}^{T} y^f_{t-p}\left(y^f_{t-q}\right)^T\otimes\nu_t\nu_t^T\right) \\
&\quad + \underbrace{\frac{1}{T^2}\mathbb{E}\left(\sum_{t=1}^{T}\sum_{n=1,t\neq n}^{T} y^f_{t-p}\left(y^f_{n-q}\right)^T\otimes\nu_t\nu_n^T\right)}_{=0} \\
&= \frac{1}{T}\left(\gamma^{ff}(p-q)\otimes\Sigma_\nu\right).
\end{aligned}$$

The remainder of the proof is the same as in Theorem 2.2.25. □

Now we are interested in the upper bound as described in Chen and Zadrozny (1998). The idea is to use a modification of the autocovariance estimator for $\gamma^{ff}(h)$ which uses only approximately half of the observations for $N = 2$, i.e. $\hat{\gamma}^{ff}(h) = \frac{2}{T}\sum_{t=2,4,\dots}^{T} y_t^f \left(y_{t-1}^f\right)^T$ and the same $\hat{\gamma}^{sf}(h)$ estimator as in the mixed-frequency setting. The authors of Chen and Zadrozny (1998) argue that the upper bound is the lower bound times N since for $N = 2$ we are using half of the observations for the autocovariance estimators. The next example shows that this is a false conclusion.

EXAMPLE 5.2.10. Consider an AR(1) model ($n = 2$, $N = 2$) with inputs $(\nu_t)_{t\in\mathbb{Z}} \sim IID_2(0, \Sigma_\nu)$ and assume that every even data point $(y_t)_{t\in 2\mathbb{Z}}$ is fully observed. Then we can arrange the following system of equations:

$$\begin{pmatrix} \frac{2}{T}\sum_{t=2,4,\dots}^{T} y_t y_{t-1}^f & \frac{2}{T}\sum_{t=2,4,\dots}^{T} y_t y_{t-2}^f \end{pmatrix} = A_1 \begin{pmatrix} \frac{2}{T}\sum_{t=3,5,\dots}^{T-1} y_{t-1} y_{t-1}^f & \frac{2}{T}\sum_{t=3,5,\dots}^{T-1} y_{t-1} y_{t-2}^f \end{pmatrix} + \hat{e}_2 + A_1 \hat{e}_3$$

where

$$\hat{e}_2 = \begin{pmatrix} \frac{2}{T}\sum_{t=2,4,\dots}^{T} \nu_t y_{t-1}^f & \frac{2}{T}\sum_{t=2,4,\dots}^{T} \nu_t y_{t-2}^f \end{pmatrix}$$

and

$$\begin{aligned} \hat{e}_3 = & \begin{pmatrix} \frac{2}{T}\sum_{t=2,4,..}^{T} y_{t-1} y_{t-1}^f & \frac{2}{T}\sum_{t=2,4,..}^{T} y_{t-1} y_{t-2}^f \end{pmatrix} \\ & - \begin{pmatrix} \frac{2}{T}\sum_{t=3,5,\dots}^{T-1} y_{t-1} y_{t-1}^f & \frac{2}{T}\sum_{t=3,5,\dots}^{T-1} y_{t-1} y_{t-2}^f \end{pmatrix}. \end{aligned} \tag{5.2.17}$$

In a next step we have to derive, as in the proof of Theorem 5.2.9, the covariance of $\operatorname{vec}(\hat{e}_2 + A_1\hat{e}_3)$, which is, of course, not straightforward. Nevertheless, we can argue that the covariance of $\operatorname{vec}(\hat{e}_2)$, i.e.

$$\mathbb{V}(\operatorname{vec}(\hat{e}_2)) = \frac{2}{T}\left(\gamma^{ff}(p-q) \otimes \Sigma_\nu\right)_{p,q=1,2}, \tag{5.2.18}$$

is two times the covariance of the lower bound, i.e.

$$\mathbb{V}(\operatorname{vec}(\hat{e})) = \frac{1}{T}\left(\gamma^{ff}(p-q) \otimes \Sigma_\nu\right)_{p,q=1,2}. \tag{5.2.19}$$

Now we can conclude that, in general, the covariance of $\operatorname{vec}(\hat{e}_2 + A_1\hat{e}_3)$ cannot be two times the covariance of the lower bound, too. Since the exact covariance of the upper bound is more complicated to derive than the asymptotic covariance of the XYW estimator and does not give us further insights, we pass the upper bound for our asymptotic analysis.

5.2.2. Asymptotic Behavior of the Maximum Likelihood Estimator for the AR(1) Case. In Theorem 4.3.1 we have derived the score function of the AR(1) likelihood. Now we are interested in the asymptotic behavior of this estimator.

THEOREM 5.2.11. *Let $(y_t)_{t\in\mathbb{Z}}$ be the output of system (2.1.6) with independent inputs $(\nu_t)_{t\in\mathbb{Z}} \sim \mathcal{N}_n(0, \Sigma_\nu)$, assume $\Sigma_\nu > 0$, $N = 2$ and $\theta \in \Theta_I$. Then the maximum likelihood estimator for the AR(1)*

case is asymptotically normally distributed, i.e.

$$\sqrt{T}\left(\begin{pmatrix} vec\left(\hat{A}_1^{ML}\right) \\ vech\left(\hat{\Sigma}_\nu^{ML}\right)\end{pmatrix} - \begin{pmatrix} vec\left(A_1\right) \\ vech\left(\Sigma_\nu\right)\end{pmatrix}\right) \xrightarrow{d} \mathcal{N}_{(3n^2+n)/2}\left(0, \Sigma_{AR\ 1}\right) \tag{5.2.20}$$

where $\Sigma_{AR\ 1} = \begin{pmatrix} \Sigma_{A_1} & \Sigma_{A_1,\Sigma_\nu} \\ \Sigma_{A_1,\Sigma_\nu}^T & \Sigma_{\Sigma_\nu}\end{pmatrix}$ *can be constructed with the formulas*

$$\begin{aligned}
\Sigma_{A_1} &= 2H_2\left(\left(\Sigma_{\tilde{\nu}} \otimes \Sigma_{\tilde{\nu}}\right) + K_{n+n_f,n+n_f}\left(\Sigma_{\tilde{\nu}} \otimes \Sigma_{\tilde{\nu}}\right)\right) H_2^T \\
&\quad + 2H_3\left(\tilde{\gamma}\left(0\right) \otimes \Sigma_{\tilde{\nu}}\right) H_3^T \\
\Sigma_{A_1,\Sigma_\nu} &= 2H_2\left(\left(\Sigma_{\tilde{\nu}} \otimes \Sigma_{\tilde{\nu}}\right) + K_{n+n_f,n+n_f}\left(\Sigma_{\tilde{\nu}} \otimes \Sigma_{\tilde{\nu}}\right)\right) H_1^T \\
\Sigma_{\Sigma_\nu} &= 2H_1\left(\left(\Sigma_{\tilde{\nu}} \otimes \Sigma_{\tilde{\nu}}\right) + K_{n+n_f,n+n_f}\left(\Sigma_{\tilde{\nu}} \otimes \Sigma_{\tilde{\nu}}\right)\right) H_1^T
\end{aligned}$$

where H_1, H_2 *and* H_3 *are given in Appendix A.*

While the asymptotic covariance between the system and noise parameters in the high-frequency case is zero, this is not true for the mixed-frequency case. Furthermore, as will be shown in the next example, the mixed-frequency XYW estimator is, in general, not asymptotically equivalent to the mixed-frequency maximum likelihood estimator.

EXAMPLE 5.2.12. Consider the following AR(1) model with dimension $n = 2$ and $N = 2$:

$$y_t = \begin{pmatrix} -1.1665 & -0.1865 \\ 1.4113 & 0.1063 \end{pmatrix} y_{t-1} + \nu_t \tag{5.2.21}$$

where we assume that the innovations are normally distributed with a zero mean and covariance matrix $\Sigma_\nu = I_2$. The roots of $\det\left(a\left(z\right)\right)$ are

$$\begin{aligned} z_0 &= -1.1029 \\ z_1 &= -6.5120. \end{aligned}$$

Tables 5.2.1 and 5.2.2 show the asymptotic covariance of the XYW estimator in the mixed-frequency and the maximum likelihood estimator in the high- and mixed-frequency case corresponding to Theorem 2.2.24, Proposition 5.2.6 and Theorem 5.2.11.

$$\Sigma_{\text{XYW}} = \begin{array}{c} \\ a^{ff} \\ a^{sf} \\ a^{fs} \\ a^{ss} \end{array} \begin{array}{c} \begin{array}{cccc} a^{ff} & a^{sf} & a^{fs} & a^{ss} \end{array} \\ \begin{pmatrix} 1.38 & 1.31 & 1.04 & 0.94 \\ & 10.99 & 1.12 & 8.11 \\ & & 0.85 & 0.79 \\ & & & 6.14 \end{pmatrix} \end{array}$$

TABLE 5.2.1. Asymptotic covariance of the XYW estimator for the system parameters for model (5.2.21)

While the asymptotic variance of the XYW estimator of the a^{ff} and of the a^{fs} part is approximately twice the variance of the maximum likelihood estimator in the mixed-frequency case, the two

remaining variances are much larger. Indeed, the variance of the a^{sf} part of the XYW estimator is six times higher than the variance of the maximum likelihood estimator. In this case the asymptotic variances of the mixed-frequency maximum likelihood estimator of the a^{ff} and a^{fs} part are very close to the one of the high-frequency estimator. Nevertheless, the a^{sf} and a^{ss} part are rather high.

$$\Sigma_{\text{AR 1}} = \begin{array}{c} \\ a^{ff} \\ a^{sf} \\ a^{fs} \\ a^{ss} \end{array} \begin{array}{c} \begin{array}{cccc} a^{ff} & a^{sf} & a^{fs} & a^{ss} \end{array} \\ \begin{pmatrix} 0.78 & -0.04 & 0.55 & -0.03 \\ & 1.89 & -0.01 & 1.36 \\ & & 0.46 & -0.02 \\ & & & 1.13 \end{pmatrix} \end{array} \qquad \Sigma_{\text{YW}} = \begin{array}{c} \\ a^{ff} \\ a^{sf} \\ a^{fs} \\ a^{ss} \end{array} \begin{array}{c} \begin{array}{cccc} a^{ff} & a^{sf} & a^{fs} & a^{ss} \end{array} \\ \begin{pmatrix} 0.65 & & 0.45 & \\ & 0.65 & & 0.45 \\ & & 0.37 & \\ & & & 0.37 \end{pmatrix} \end{array}$$

TABLE 5.2.2. Asymptotic covariance of the ML estimator for the system parameters for model (5.2.21) in the mixed-frequency case (left) and in the high-frequency case (right)

5.3. Consistency of the Noise Covariance Estimators

The next theorem shows that the two different estimators of the noise parameters from Section 3.2.3 are both consistent.

THEOREM 5.3.1. *Let $(y_t)_{t\in\mathbb{Z}}$ be the output of system (2.1.6) with inputs $(\nu_t)_{t\in\mathbb{Z}} \sim IID_n(0,\Sigma_\nu)$ and $\theta \in \Theta_{XYW}$. Then the estimator of the noise parameters from equation (3.2.21), i.e.*

$$vec\left(\hat{\Sigma}_\nu\right) = \left((\mathcal{G}\otimes\mathcal{G})\left(I_{(np)^2} - \left(\hat{\mathcal{A}}\otimes\hat{\mathcal{A}}\right)\right)^{-1}(\mathcal{G}^T\otimes\mathcal{G}^T)\right)^{-1} vec\left(\hat{\gamma}(0)\right) \tag{5.3.1}$$

$$= \hat{L}^{-1} vec\left(\hat{\gamma}(0)\right) \tag{5.3.2}$$

where the system parameters are estimated consistently, is also consistent. Under the further assumptions of Theorem 3.2.11 the same is true for the estimator of the noise parameters given by (3.2.25) and (2.1.35), i.e.

$$\begin{pmatrix} \hat{\gamma}(p) \\ \vdots \\ \hat{\gamma}(1) \end{pmatrix} = \left(\hat{\mathcal{O}}^T\hat{\mathcal{O}}\right)^{-1}\hat{\mathcal{O}}^T\hat{\mathcal{O}}_1. \tag{5.3.3}$$

$$\hat{\Sigma}_\nu = \hat{\gamma}(0) - \left(\hat{A}_1\cdots\hat{A}_p\right)\begin{pmatrix} \hat{\gamma}(-1) \\ \vdots \\ \hat{\gamma}(-p) \end{pmatrix}. \tag{5.3.4}$$

PROOF. Under our assumptions Theorem 2.2.17 implies that the mixed-frequency estimator of the autocovariance at lag $i \in N\mathbb{Z}$ is consistent, i.e. $\hat{\gamma}(i) \xrightarrow{p} \gamma(i)$. Furthermore, it follows from our assumptions that $\hat{\mathcal{A}} \xrightarrow{p} \mathcal{A}$ and therefore with the aid of Lemma 2.2.5 it follows that $\hat{L}^{-1} \xrightarrow{p} L^{-1}$, which implies $\text{vec}\left(\hat{\Sigma}_\nu\right) = \hat{L}^{-1}\text{vec}\left(\hat{\gamma}(0)\right) \xrightarrow{p} L^{-1}\text{vec}\left(\gamma(0)\right) = \text{vec}\left(\Sigma_\nu\right)$. In a same way one can argue that the second estimator is also consistent. □

THEOREM 5.3.2. *Let $(y_t)_{t\in\mathbb{Z}}$ be the output of system (2.1.6) with inputs $(\nu_t)_{t\in\mathbb{Z}} \sim IID_n(0,\Sigma_\nu)$, $\theta \in \Theta_{XYW}$, assume that the eighth moment $\mathbb{E}\left(\nu_t\nu_t^T \otimes \nu_t\nu_t^T \otimes \nu_t\nu_t^T \otimes \nu_t\nu_t^T\right)$ exists, let $\eta = \mathbb{E}\left(\nu_t\nu_t^T \otimes \nu_t\nu_t^T\right)$ and let the system and noise parameters be consistently estimated. Then the estimator of η, i.e.*

$$\begin{aligned} vec\left(\hat{\eta}\right) &= \left(\mathcal{G}_2\left(I_{(np)^4} - \left(\hat{\mathcal{A}}\otimes\hat{\mathcal{A}}\otimes\hat{\mathcal{A}}\otimes\hat{\mathcal{A}}\right)\right)^{-1}\mathcal{G}_2^T\right)^{-1} vec\left(\hat{\psi}\right) + \\ &\quad vec\left(\hat{\Sigma}_\nu\right) vec\left(\hat{\Sigma}_\nu\right)^T + \left(\hat{\Sigma}_\nu\otimes\hat{\Sigma}_\nu\right) + K_{n,n}\left(\hat{\Sigma}_\nu\otimes\hat{\Sigma}_\nu\right) \end{aligned} \tag{5.3.5}$$

is consistent where $\mathcal{G}_2 = (\mathcal{G}\otimes\mathcal{G}\otimes\mathcal{G}\otimes\mathcal{G})$,

$$\hat{\psi} = \hat{\chi} - vec\left(\hat{\gamma}\left(0\right)\right) vec\left(\hat{\gamma}\left(0\right)\right)^T - \left(\hat{\gamma}\left(0\right)\otimes\hat{\gamma}\left(0\right)\right) - K_{n,n}\left(\hat{\gamma}\left(0\right)\otimes\hat{\gamma}\left(0\right)\right)$$

and $\hat{\chi}$ is the sample mixed-frequency estimator of $\mathbb{E}\left(y_t y_t^T \otimes y_t y_t^T\right)$.

PROOF. Under our assumptions the consistency of the estimators of the autocovariance of lag 0, of the system parameters and of the noise parameters are guaranteed. Thus, it remains to prove that $\hat{\chi}$ is a consistent estimator for $\mathbb{E}\left(y_t y_t^T \otimes y_t y_t^T\right)$. This has been shown in Lomnicki (1961) for the high-frequency case. Again using Lemma 2.2.5 establishes our result. □

CHAPTER 6

Projecting Estimators on the Parameter Space

It is well known that in the single-frequency case the Yule-Walker estimator always leads to a stable AR polynomial, provided $\hat{\Gamma}_p > 0$ holds, and the estimated covariance matrix of the noise is non-negative definite. In general, the XYW estimator does not fulfill these desirable properties. So, in a second step, one has to check whether the estimated parameters, say $\hat{\theta}$, lie in the parameter space Θ. If $\hat{\theta}$ is not contained in this space, the question of finding a $\hat{\theta}_P \in \Theta$ which is sufficiently close to $\hat{\theta}$ arises. In this chapter we are separating this problem in two sub-problems: The first problem is to find a stable polynomial matrix close to an unstable estimator of $a(z)$. The second problem is to find a positive (semi)-definite covariance matrix of rank q, which is close to an indefinite (symmetric) estimator of Σ_ν. Parts of the results of this chapter are contained in Koelbl et al. (2015).

6.1. Stabilization of the Estimated System Parameters

In this section we assume that we have an estimate for the system parameters, say $\hat{A}_{\text{un}} \in \mathbb{R}^{n \times np}$, corresponding to at least one unstable root, say $z_0 \in \mathbb{C}$, so that $|z_0| \leq 1$ and $\det(\hat{a}_{\text{un}}(z_0)) = 0$. Note that the parameter space S (as described in Section 2.1) is an open set and thus there exists no best approximation of such an $\hat{A}_{\text{un}}$, for instance in Frobenius norm, by an element of S. In addition, S is even for the multivariate AR(1) case non-convex as the following example shows:

EXAMPLE 6.1.1. Consider the two stable AR(1) system matrices

$$A = \begin{pmatrix} 0.8 & 10 \\ 0 & 0.8 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0.8 & 0 \\ 10 & 0.8 \end{pmatrix}.$$

Then the linear combination

$$\frac{1}{2}A + \frac{1}{2}B = \begin{pmatrix} 0.8 & 5 \\ 5 & 0.8 \end{pmatrix}$$

is not stable and thus not contained in S. Indeed, the eigenvalues of this linear combination are $\lambda_0 = 5.8$ and $\lambda_1 = -4.2$. A picture of the non-convex stability set for the univariate AR(3) case has been given in Combettes and Trussell (1992).

We consider the problem of finding

$$\inf_{A \in S} \left\| A - \hat{A}_{\text{un}} \right\|_F^2 \tag{6.1.1}$$

where S is the parameter space and $\|.\|_F$ denotes the Frobenius norm. There exists a huge range of literature about finding the nearest stable polynomial in the univariate case in an iterative way. In Moses and Liu (1991); Stoica and Moses (1992); Combettes and Trussell (1992) the "stabilization" of a

univariate polynomial is discussed where the reflection coefficients are used to guarantee the stability. In D'haene et al. (2006); Balogh and Pintelon (2008) some special minimization methods are used to stabilize the multivariate transfer function. In this case the step size of the optimization algorithm is chosen so that the obtained transfer function is still stable and causal. A more interesting way to solve the univariate stabilization problem is proposed in Orbandexivry et al. (2013) using the so-called Dikin Ellipsoid. We will repeat the most important steps from this procedure and generalize it to the multivariate case, which can be easily done. We point out that all these different methods need a stable initial value.

Problem (6.1.1) can be reformulated by the Lyapunov Theorem (see Theorem 1 in Orbandexivry et al. (2013)) as

$$\inf_{A,P} \frac{1}{2} \left\| A - \hat{A}_{\text{un}} \right\|_F^2 \tag{6.1.2}$$

where minimization with respect to P runs over $P = P^T > 0$, $P - \mathcal{A}P\mathcal{A}^T > 0$ where $\mathcal{A}$ is the companion form of A. For a fixed $P = P^T > 0$, we can define the set

$$S_P = \left\{ A \in \mathbb{R}^{n \times np} : \; P - \mathcal{A}P\mathcal{A}^T > 0, \, \mathcal{A} \text{ is the companion form of } A \right\} \subset S$$

and the function $b_P(A) = -\ln\left(\det\left(P - \mathcal{A}P\mathcal{A}^T\right)\right)$, which is a barrier function. It follows from Theorem 5 in Orbandexivry et al. (2013) that for $A \in S$, $P = P^T > 0$ so that $P - \mathcal{A}P\mathcal{A}^T > 0$ and any $0 \leq \alpha < 1$, the so-called Dikin Ellipsoid

$$\mathcal{E}\left(P, A; \alpha\right) = A + \left\{ H \in \mathbb{R}^{n \times np} : \left\langle b_P''(A)H, H \right\rangle \leq \alpha \right\}$$

is a subset of S_P where $\langle A, B \rangle = \text{tr}\left(AB^T\right)$ and $\left\langle b_P''(A)H, H \right\rangle$ is the second derivative of $b_P(A)$ in a given direction H. Now, for given $A \in S$ and α, the question arises which P should be chosen so that $\mathcal{E}\left(P, A; \alpha\right)$ is maximized. In Orbandexivry et al. (2013) the authors argue that a good choice is given by solving the problem

$$\min_{P \in P_A} b_P(A) \tag{6.1.3}$$

where $P_A = \left\{ P \in \mathbb{R}^{np \times np} : \; P = P^T > 0, \, P - \mathcal{A}P\mathcal{A}^T > 0, \, \text{tr}\left(P\right) = 1, \, \mathcal{A} \text{ is the companion form of } A \right\}$. The solution, say P^*, of problem (6.1.3) satisfies

$$\begin{aligned} Q^{-1} - \mathcal{A}^T Q^{-1} \mathcal{A} &= npI_{np} \\ P^* - \mathcal{A}P^*\mathcal{A}^T &= Q > 0 \end{aligned}$$

and can be derived via

$$\text{vec}\left(Q^{-1}\right) = \left(I_{(np)^2} - \left(\mathcal{A}^T \otimes \mathcal{A}^T \right) \right)^{-1} \text{vec}\left(npI_{np}\right) \tag{6.1.4}$$

$$\text{vec}\left(P^*\right) = \left(I_{(np)^2} - \left(\mathcal{A} \otimes \mathcal{A} \right) \right)^{-1} \text{vec}\left(Q\right). \tag{6.1.5}$$

Note that $A \in S$ and that the absolute value of all eigenvalues of $\mathcal{A}$ (and obviously $\mathcal{A}^T$) is smaller than one. Now we are in a position to formulate a new, restricted optimization problem for a given

$0 \leq \alpha < 1$, $A \in S$ and a corresponding P^*:

$$\min_H \frac{1}{2} \left\| A + H - \hat{A}_{\text{un}} \right\|_F^2 \tag{6.1.6}$$

where H is so that $\left\langle b''_{P^*}(A)H, H \right\rangle \leq \alpha$. Note that we now have a convex optimization problem. It can be shown that $\left\langle b''_{P^*}(A)H, H \right\rangle \leq \alpha$ can be rewritten as $\text{vec}(H)^T B \text{vec}(H) \leq \alpha$ where

$$\begin{aligned} \frac{1}{2}B &= \left(P^* \otimes \mathcal{G}Q^{-1}\mathcal{G}^T\right) + \\ &\quad \left(P^*\mathcal{A}^T Q^{-1}\mathcal{A}P^* \otimes \mathcal{G}Q^{-1}\mathcal{G}^T\right) + \\ &\quad \left(P^*\mathcal{A}^T Q^{-1}\mathcal{G}^T \otimes \mathcal{G}Q^{-1}\mathcal{A}P^*\right) K_{n,np} \end{aligned} \tag{6.1.7}$$

and $K_{n,np}$ is a commutation matrix.

LEMMA 6.1.2. *The matrix $B \in \mathbb{R}^{n^2p \times n^2p}$ is symmetric positive definite and thus can be factorized as $B = UDU^T$ where D is a diagonal matrix with positive entries d_i, $i = 1, \ldots, n^2p$ and U is an orthonormal matrix.*

PROOF. Since P^* and Q are positive definite matrices the first term of the right hand side of (6.1.7) is positive definite. Thus it remains to show that the sum of the second and third term is at least positive semi-definite. These two terms can be rewritten as

$$\left(P^*\mathcal{A}^T Q^{-\frac{1}{2}} \otimes \mathcal{G}Q^{-\frac{1}{2}}\right)(I_{np} + K_{np,np})\left(Q^{-\frac{1}{2}}\mathcal{A}P^* \otimes Q^{-\frac{1}{2}}\mathcal{G}^T\right).$$

Now it is easy to conclude that B is positive definite since $K_{np,np}$ is a symmetric matrix which has eigenvalues ± 1 (see Magnus and Neudecker (1979)) and therefore $(I_{np} + K_{np,np})$ has eigenvalues 0 and 2. □

The solution of (6.1.6) fulfills the following equations:

$$\begin{aligned} \left(I_{n^2p} + \lambda B\right)\text{vec}(H) &= \text{vec}\left(\hat{A}_{\text{un}} - A\right) \\ \text{vec}(H)^T B \text{vec}(H) &= \alpha. \end{aligned} \tag{6.1.8}$$

The solution of (6.1.8) can be derived by finding the root of the following function (see Orbandexivry et al. (2013), p. 1199)

$$\psi(\lambda) = \sum_{i=1}^{n^2p} \frac{d_i \left(e_i^T U^T \text{vec}\left(\hat{A}_{\text{un}} - A\right)\right)^2}{(1 + \lambda d_i)^2} - \alpha = 0 \tag{6.1.9}$$

with respect to $\lambda \in (0, \infty)$ where e_i is the i-th unit vector and then substituting into

$$\text{vec}(H) = \left(I_{n^2p} + \lambda B\right)^{-1} \text{vec}\left(\hat{A}_{\text{un}} - A\right). \tag{6.1.10}$$

It is worth mentioning that there exists a unique λ and therefore a unique H. In a last step we can derive our new approximation by $A + H$.

The whole procedure has to be iterated:

(1) Start with a stable polynomial $A^{(0)}$ which can be derived by an initial method $a-e$ below and choose an $\eta > 0$.
(2) Compute Q and P^* from (6.1.4) and (6.1.5).
(3) Compute λ and H from (6.1.9) and (6.1.10).
(4) $A^{(i)} = A^{(i-1)} + H$
(5) Repeat steps 2-4 until $\|H\|_F^2 < \eta$

We consider five different initializations for the algorithm described above:

a: Reflecting the unstable roots of $\hat{a}_{\text{un}}(z)$ on the unit circle: Method 1

In Hannan (1970); Lippi and Reichlin (1994) a procedure was proposed to obtain a causal (fundamental) representation of a transfer function from a non-causal (non-fundamental) one via the so-called Blaschke matrices. We say that $B(z)$ is a Blaschke matrix if

$$B(z)\left(\overline{B}\left(z^{-1}\right)\right)^T = I_n. \tag{6.1.11}$$

We will adapt this procedure to our problem for the generic case that all roots of the polynomial $\hat{a}_{\text{un}}(z)$ are different and do not lie on the unit circle. Let us assume that there exists an unstable root $z_0 \in \mathbb{C}$ so that $|z_0| < 1$ and $\det\left(\hat{a}_{\text{un}}\left(z_0\right)\right) = 0$. Then there exists a vector $0 \neq g \in \mathbb{C}^n$ with $g^T\bar{g} = 1$, which lies in the left kernel of $\hat{a}_{\text{un}}\left(z_0\right)$, i.e.

$$g^T\hat{a}_{\text{un}}\left(z_0\right) = 0.$$

Now let K be a unitary matrix, i.e. $K\bar{K}^T = I_n$, where g^T is in the first row of K. For instance, such a K can be obtained by defining $K_2 = \begin{pmatrix} g^T \\ (0, I_{n-1}) \end{pmatrix}$ and applying the Gram–Schmidt process (note that K_2 is generically nonsingular). If we premultiply $\hat{a}_{\text{un}}\left(z\right)$ by K, we can observe that all entries in the first row of $K\hat{a}_{\text{un}}\left(z\right)$ contain the factor $z - z_0$. Let $B_{z_0}\left(z\right)$ be the matrix

$$\begin{pmatrix} \frac{1-\bar{z}_0 z}{z-z_0} & 0 \\ 0 & I_{n-1} \end{pmatrix}.$$

Then premultiplying $K\hat{a}_{\text{un}}\left(z\right)$ by $B_{z_0}\left(z\right)$ we obtain that $\tilde{a}(z) = B_{z_0}\left(z\right)K\hat{a}_{\text{un}}\left(z\right)$ has not a root at z_0 anymore. Furthermore, it is easy to see that $B_{z_0}\left(z\right)K$ is a Blaschke matrix. It is worth mentioning that $\tilde{a}(z)$ is still an AR(p) polynomial. One can repeat this procedure until no root is inside the unit circle. In a last step one has to premultiply the obtained polynomial with $\tilde{A}_0^{-1}$ so that the stable polynomial fulfills the restriction $A_0 = I_n$.

One nice property of this procedure is that if we apply this procedure to a regular AR(p) process, i.e. $b^{-1}a(z)$, the spectral density will be preserved.

b: Reflecting the unstable roots of $\hat{a}_{\text{un}}(z)$ on the unit circle: Method 2

As mentioned after Lemma 3.1.5, generically, the unstable polynomial has the following representation

$$\mathcal{A} = P\Lambda P^{-1}$$

where Λ is a diagonal matrix containing the eigenvalues $\lambda_i \neq 0$ and P contains the corresponding eigenvectors of $\mathcal{A}$. Furthermore, we assume that $|\lambda_i| \neq 1$. As has been shown in

Lemma 2.1.7 P has a special structure:

$$P = \begin{pmatrix} P_1 \\ P_1\Lambda^{-1} \\ \vdots \\ P_1\Lambda^{-p+1} \end{pmatrix}$$

where $P_1 \in \mathbb{R}^{n\times np}$. We can now define our new companion form $\mathcal{A}^{(0)} = P^{(0)}\Lambda^{(0)}\left(P^{(0)}\right)^{-1}$ where $\Lambda^{(0)}$ is a diagonal matrix with the entries

$$\lambda_i^0 = \begin{cases} \lambda_i & |\lambda_i| < 1 \\ 1/\bar{\lambda}_i & \text{else} \end{cases}$$

and

$$P^{(0)} = \begin{pmatrix} P_1 \\ P_1\left(\Lambda^{(0)}\right)^{-1} \\ \vdots \\ P_1\left(\Lambda^{(0)}\right)^{-p+1} \end{pmatrix},$$

which is, of course, stable.

c: Setting the magnitude of the unstable zeros to $1+\epsilon$

This method is analogous to Method b, but now we define

$$\lambda_i^0 = \begin{cases} \lambda_i & |\lambda_i| < 1 \\ \frac{(1-\epsilon)}{|\lambda_i|}\lambda_i & \text{else} \end{cases}$$

where $0 < \epsilon < 1$ and ϵ is close to 0.

d: Scaling of the polynomial: Method 1

We start with $0 < \mu_0 < 1$ and define $\mu_0(A_1, \ldots, A_p) = B_{\mu_0}$. If B_{μ_i} is not stable, set $\mu_{i+1} = \mu_i/2$ and repeat until B_{μ_i} is stable. Set $B_{\mu_i} = A^{(0)}$.

e: Scaling of the polynomial: Method 2

For $0 < \epsilon < 1$ define $\mu = (1-\epsilon)\min_i(|z_i|) < 1$. Now define the initial starting point as $A^{(0)} = (A_1\mu, \ldots, A_p\mu^p)$, which is stable.

EXAMPLE 6.1.3. The question arises which of these initialization methods should be chosen for the algorithm. In a simulation framework we would choose the method which leads to a stable polynomial which is as close as possible to the true system parameters (and where the computation time is appropriate small), which is, of course, infeasible in practice.

We compare the stabilization algorithm described above initialized with the initialization methods $a-e$ for an AR(2) polynomial of dimension 3 for different points in the parameter space S. To get an unstable polynomial we add to each component a perturbation, say $\psi_i^{j,k}$ $i = 1, 2$, $j, k = 1, 2, 3$, where $\psi_i^{j,k} \sim \mathcal{N}_1(0, 0.1)$. We repeat this simulation for every point 10^3 times. As a performance criteria we consider the norm $\left\|A_{\text{st}} - \hat{A}_{\text{un}}\right\|_F$ where A_{st} is the stabilized polynomial, the computation time in seconds and the count of the iteration steps of the stabilization procedure. It turns out that in most

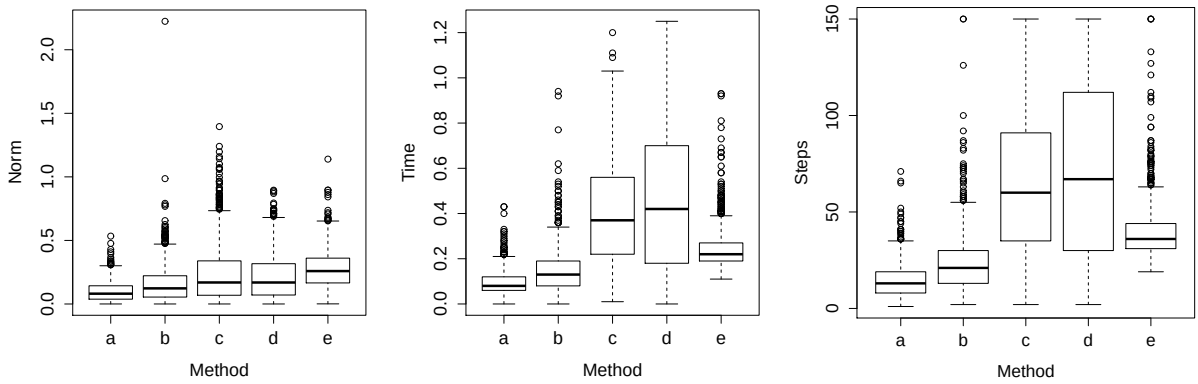

FIGURE 6.1.1. Comparison of the five initialization methods for an AR(2) $n = 3$ polynomial

of the cases the initialization method a performs best. Thus, we only want to present a point of the parameter space S:

$$A_1 = \begin{pmatrix} -0.844 & 0.658 & -1.182 \\ 0.428 & 0.471 & -1.209 \\ 1.401 & -1.760 & 0.005 \end{pmatrix}, A_2 = \begin{pmatrix} -1.127 & 0.971 & -0.815 \\ -0.049 & 0.127 & 0.102 \\ 1.536 & 0.077 & -0.855 \end{pmatrix} \tag{6.1.12}$$

which has the roots:

$$\begin{aligned} z_{0,1} &= -0.398 \pm 0.962i \\ z_2 &= -1.149 \\ z_{3,4} &= 0.376 \pm 1.095i \\ z_5 &= 1.481. \end{aligned} \tag{6.1.13}$$

In Figure 6.1.1 there are boxplots of the performance of the five different initialization methods $a - e$ displayed with respect to the three criteria. As can be seen in the left picture of Figure 6.1.1 the medians of the calculated norms are $(0.07, 0.08, 0.10, 0.11, 0.11)$ which shows that Method a performs best. Furthermore, the medians of the computational times in seconds are $(0.05, 0.07, 0.22, 0.25, 0.13)$, which are displayed in the picture in the middle of Figure 6.1.1. The medians of the iteration steps of the five methods are $(8, 12, 35, 40, 20)$, which are displayed in the picture in the right of Figure 6.1.1.

To summarize, the stabilization algorithm works quite fast for "small" polynomials, where "small" means depending on n and p and gives us stable polynomials which are near to the unstable ones. Of course, this algorithm is an iterative algorithm and thus different starting values should be considered.

One drawback of the stabilization procedure is that if we are close to the border of the stability region, the matrix Q, which is in a certain sense a measure for the distance to the border, will be near to singularity. Thus, the inverse of it will cause numerical troubles, for instance the matrix B will not be symmetric anymore. This problem can be fixed by increasing the threshold η for $\|H\|_F^2$ or to use

$B_2 = \left(B + B^T\right)/2$ instead of B. Note that B_2 is the nearest symmetric matrix to B (see Higham (1989)).

Reconsidering problem (6.1.1), one can ask the question of finding

$$\inf_{A \in S} \left\| \hat{Z}_1 - A\hat{Z}_0 \right\|_F^2 \tag{6.1.14}$$

where S is again the stability set and the true system parameters are identifiable via the XYW equations. With this problem we expect to find a stable matrix A, which will adapt to the estimates $\hat{Z}_0$ and $\hat{Z}_1$ more directly. Following the same ideas as above we get our new, restricted optimization problem

$$\min_H \frac{1}{2} \left\| \hat{Z}_1 - A\hat{Z}_0 - H\hat{Z}_0 \right\|_F^2 \tag{6.1.15}$$

with respect to $\left\langle b''_{P^*}(A)H, H \right\rangle \leq \alpha$. The solution of problem (6.1.15) has to satisfy the equations

$$\begin{aligned} \operatorname{vec}(H) &= \left(I_{n^2p} + \lambda \left(\left(\hat{Z}_0 \hat{Z}_0^T \right)^{-1} \otimes I_n \right) B \right)^{-1} \operatorname{vec}\left(\hat{A}_{\mathrm{XYW}} - A \right) \\ \operatorname{vec}(H)^T B \operatorname{vec}(H) &= \alpha. \end{aligned} \tag{6.1.16}$$

Note that the equations above and (6.1.8) only differ in the term $\left(\hat{Z}_0 \hat{Z}_0^T \right)^{-1}$, which has the drawback that an analogue of (6.1.9) cannot be achieved. Thus, we have to find the root of

$$\operatorname{vec}(H)^T B \operatorname{vec}(H) - \alpha = 0$$

with respect to the variable $\lambda > 0$ where we have to insert equation (6.1.16) into it. Of course, the computational time will increase in general as can be seen in Example 6.1.4.

EXAMPLE 6.1.4. We compare the two different stabilization procedures proposed in (6.1.6), say Method 1, and (6.1.14), say Method 2, for the AR(2) process with the system parameters (6.1.12) and $(\nu_t)_{t \in \mathbb{Z}} \sim \mathcal{N}_3(0, I_3)$. Again we repeat the simulation 10^3 times and we initialize the procedures with Method a. As can be seen in Figure 6.1.2, it turns out that the medians of $\left\| A_{\mathrm{st}} - \hat{A}_{\mathrm{un}} \right\|_F$ of the two methods are $(0.04, 0.11)$, respectively. Thus, Method 2 does not seem to perform quite well. The calculation time of the second procedure also increases. Furthermore, the medians of $\|A_{\mathrm{st}} - A\|_F$ for the two methods, where A are the true system parameters, are $(2.90, 2.88)$, which do not differ very much.

6.2. Positive (Semi)-Definiteness of the Estimated Noise Covariance Matrix

Under specific assumptions the estimator for the noise covariance matrix Σ_ν from equation (3.2.21) is consistent. Furthermore, this estimate is symmetric since

$$\begin{aligned} \operatorname{vec}\left(\hat{\Sigma}_\nu^T \right) &= K_{n,n} \operatorname{vec}\left(\hat{\Sigma}_\nu \right) \\ &= \left((\mathcal{G} \otimes \mathcal{G}) \left(I_{(np)^2} - \left(\hat{\mathcal{A}} \otimes \hat{\mathcal{A}} \right) \right)^{-1} (\mathcal{G}^T \otimes \mathcal{G}^T) \right)^{-1} K_{n,n} \operatorname{vec}\left(\hat{\gamma}(0) \right) \\ &= \operatorname{vec}\left(\hat{\Sigma}_\nu \right) \end{aligned}$$

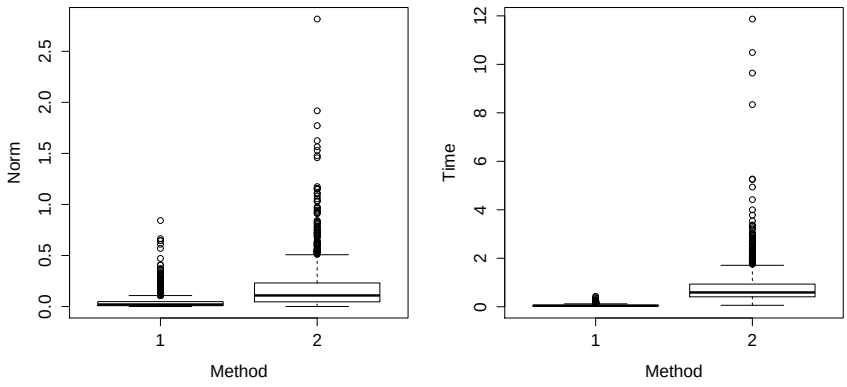

FIGURE 6.1.2. Comparison of the new stabilization method for an AR(2) $n = 3$ polynomial

but may not be positive (semi)-definite and of rank q and thus not contained in the set D. Note that the estimate for the covariance matrix $\hat{\gamma}(0)$ is always positive (semi)-definite in the single-frequency setting, however, this is not the case in the mixed-frequency setting.

There exists an enormous amount of literature about finding the nearest member of a class of matrices to a given matrix, see e.g. Eckart and Young (1936); Higham (1989, 2002). In Higham (1989) a survey of nearness problems is given, where the focus lies, among other properties, on positive definiteness or rank deficiency.

Before we get further insides into our main problem, we first consider the problem of finding the nearest positive (semi)-definite diagonal matrix of rank q, when $\hat{\Sigma}_\nu$ is also a diagonal matrix where the entries are ordered in a descending way. Thus, we are faced with the problem of finding

$$\inf_{\Lambda_{\mathrm{ps}} \in D_d} \left\| \Lambda_{\mathrm{ps}} - \hat{\Sigma}_\nu \right\|_F^2 \tag{6.2.1}$$

where $D_d = \{\Lambda_\nu \in \mathbb{R}^{n \times n} | \, \Lambda_\nu \text{ is diagonal}, \Lambda_\nu \geq 0, \mathrm{rk}(\Lambda_\nu) = q\}$. We can rewrite the norm of the difference as $\left\| \Lambda_{\mathrm{ps}} - \hat{\Sigma}_\nu \right\|_F^2 = \sum_{i=1}^n (\lambda_i^{ps} - \lambda_i)^2$, where λ_i^{ps}, λ_i are the diagonal entries of Λ_{ps} and $\hat{\Sigma}_\nu$, respectively. Thus one arbitrary near solution of (6.2.1) is the diagonal matrix Λ_+ with the entries

$$\lambda_i^+ = \begin{cases} \max(\lambda_i, \epsilon) & i = 1, \ldots, q \\ 0 & i = q+1, \ldots, n \end{cases}$$

for a sufficiently small $\epsilon > 0$. Of course, since the set of positive definite matrices is open, one cannot assume that the solution of (6.2.1) is in the set D_d. This issue is presented in the next example:

EXAMPLE 6.2.1. Let Σ_ν be the indefinite diagonal matrix

$$\Sigma_\nu = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Then, for a sufficiently small $\epsilon > 0$, a near solution of problem (6.2.1) for the case $q = 2$ is given by

$$\Lambda_+ = \begin{pmatrix} 1 & 0 \\ 0 & \epsilon \end{pmatrix}$$

whereas the infimum of (6.2.1) is

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \underset{\Sigma_{\text{ps}} \in D_d}{\arg\inf} \left\| \Sigma_{\text{ps}} - \hat{\Sigma}_\nu \right\|_F^2,$$

which is not contained in D_d.

Now we consider our main problem, namely to find a near solution to

$$\inf_{\Sigma_{\text{ps}} \in D} \left\| \Sigma_{\text{ps}} - \hat{\Sigma}_\nu \right\|_F^2 \tag{6.2.2}$$

where $D = \left\{ \Sigma_\nu \in \mathbb{R}^{n \times n} \middle| \, \Sigma_\nu = \Sigma_\nu^T, \Sigma_\nu \geq 0, \text{rk}\,(\Sigma_\nu) = q \right\}$ and $\hat{\Sigma}_\nu$ is a symmetric matrix. The matrix $\hat{\Sigma}_\nu$ can be represented as $\hat{\Sigma}_\nu = Q \Lambda Q^T$ where Λ is the diagonal matrix containing the eigenvalues λ_i in a descending order and Q is a matrix containing the appropriate eigenvectors. For simplicity, we assume that the q-th and the $(q+1)$-th eigenvalue are distinct.

To obtain an arbitrarily close solution to the problem (6.2.2) we define the following matrix $\hat{\Sigma}_{\text{ps}} = Q \Lambda_+ Q^T$ where again Λ_+ is a diagonal matrix with the entries

$$\lambda_i^+ = \begin{cases} \max(\lambda_i, \epsilon) & i = 1, \ldots, q \\ 0 & i = q+1, \ldots, n \end{cases}$$

for a sufficiently small $\epsilon > 0$. Note that by the so-called Wielandt-Hoffman Theorem (see Hoffman and Wielandt (1953); Wilkinson (1979)) $\sum_{i=1}^n \left(\lambda_i^A - \lambda_i^B \right)^2 \leq \|A - B\|_F^2$ holds for symmetric matrices A, $B \in \mathbb{R}^{n \times n}$, where λ_i^A and λ_i^B are the corresponding eigenvalues in a descending order, respectively. Thus, $\hat{\Sigma}_{\text{ps}}$ gives an arbitrarily close solution.

EXAMPLE 6.2.2. Let $q = 2$ and

$$\hat{\Sigma}_\nu = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}.$$

Then the infimum of (6.2.2) with rank one is

$$\Sigma_s = \underset{\Sigma_{\text{ps}} \in D_d}{\arg\inf} \left\| \Sigma_{\text{ps}} - \hat{\Sigma}_\nu \right\|_F^2 = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

and $\left\| \Sigma_{\text{s}} - \hat{\Sigma}_\nu \right\|_F^2 = 1$. In contrast, for an $\epsilon = 10^{-6}$, $\hat{\Sigma}_{\text{ps}}$ with rank $q = 2$ is

$$\hat{\Sigma}_{\text{ps}} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \epsilon \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} + 10^{-6} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

and $\left\| \Sigma_{\text{s}} - \hat{\Sigma}_\nu \right\|_F^2 = 1 + \epsilon$.

CHAPTER 7

Monte Carlo Simulations

In this chapter the emphasis is on Monte Carlo simulations of the estimation procedures presented in Chapter 5. We compare the finite sample and the asymptotic behavior of the estimators for the case of stock variables. We do this by simulating the finite sample covariance matrix of the autocovariance estimator, the extended Yule-Walker estimator and the generalized method of moments estimator and compare them with the exact asymptotic ones. In order to do this, we choose two different AR(2) models where one model has a long and the other a short memory. Furthermore, we simulate the finite sample covariance matrix of the maximum likelihood estimator for the AR(1) case for one model. Many of the simulation results presented in this chapter are based on the software in R written by Alexander Braumann.

7.1. Mixed-Frequency Bartlett's Formula

In this section we compare the analytical asymptotic mixed-frequency Bartlett's formula, which was derived in Theorem 2.2.22, with the finite sample mean squared error for two different AR(2) models.

EXAMPLE 7.1.1. We call the following model "Model 1":

$$y_t = \underbrace{\begin{pmatrix} -0.844 & 0.658 & -1.182 \\ 0.428 & 0.471 & -1.209 \\ 1.401 & -1.760 & 0.005 \end{pmatrix}}_{A_1} y_{t-1} + \underbrace{\begin{pmatrix} -1.127 & 0.971 & -0.815 \\ -0.049 & 0.127 & 0.102 \\ 1.536 & 0.077 & -0.855 \end{pmatrix}}_{A_2} y_{t-2} + \nu_t, \tag{7.1.1}$$

where we assume that the innovations are normally distributed with a zero mean and covariance matrix $\Sigma_\nu = I_3$. The roots of $\det(a(z))$ are

$$\begin{aligned} z_{0,1} &= -0.398 \pm 0.962i \\ z_2 &= -1.149 \\ z_{3,4} &= 0.376 \pm 1.095i \\ z_5 &= 1.481 \end{aligned}$$

and thus $z_{0,1}$, where $|z_{0,1}| = 1.0411$, is very close to the unit circle. The covariance matrix of this model is

$$\gamma(0) = \begin{pmatrix} 32.13 & 20.38 & -8.54 \\ & 45.46 & 19.10 \\ & & 27.94 \end{pmatrix}. \tag{7.1.2}$$

EXAMPLE 7.1.2. We call the following model "Model 2":

$$y_t = \underbrace{\begin{pmatrix} 1.030 & 4.475 & 5.972 \\ -0.314 & -1.201 & -1.533 \\ -0.094 & -0.224 & -0.303 \end{pmatrix}}_{A_1} y_{t-1} + \underbrace{\begin{pmatrix} -1.562 & -5.656 & -6.511 \\ 0.524 & 2.021 & 1.695 \\ -0.004 & 0.049 & 0.628 \end{pmatrix}}_{A_2} y_{t-2} + \nu_t, \tag{7.1.3}$$

where we again assume that the innovations are normally distributed with a zero mean and covariance matrix $\Sigma_\nu = I_3$. The roots of $\det(a(z))$ are

$$\begin{array}{rclrcl} z_0 & = & 1.250 & z_3 & = & 1.399 \\ z_1 & = & -1.279 & z_{4,5} & = & -0.431 \pm 1.334i \\ z_2 & = & -1.302 & & & \end{array} .$$

The covariance matrix of this model is

$$\gamma(0) = \begin{pmatrix} 265.43 & -68.88 & -8.95 \\ & 19.35 & 2.14 \\ & & 1.59 \end{pmatrix} . \tag{7.1.4}$$

In Figure 7.1.1 the positions of the roots of Model 1 and Model 2 are shown. The blue points are the roots of Model 1 and the green crosses are the roots of Model 2. The black circle represents the unit circle. We assume that the first two components of each model are fast ones, i.e. observed at each time point, and the third component is the slow one observed at each third time point. Thus, we have the following situation: $N = 3$, $n = 3$, $q = 3$, $n_f = 2$ and $n_s = 1$.

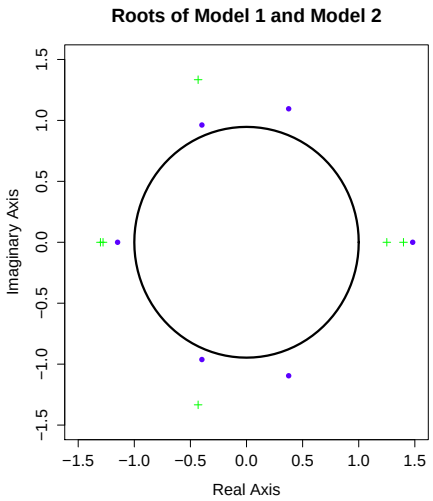

FIGURE 7.1.1. Positions of the roots of Model 1 (blue, points) and Model 2 (green, cross)

In Figure 7.1.2 the autocorrelations of Model 1 and Model 2 are displayed. To be more precise, on the main diagonal the autocorrelations and on the off-diagonal the cross-autocorrelations from lag

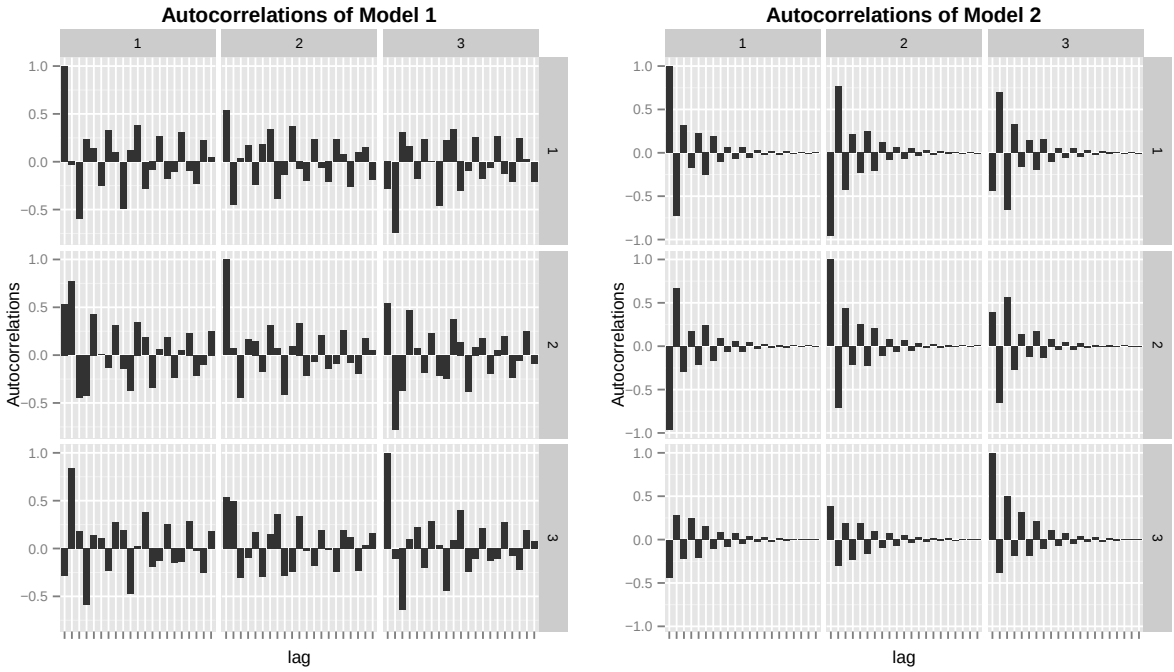

FIGURE 7.1.2. Autocorrelations of Model 1 and Model 2 of the lags 0 to 20

0 to lag 20 are shown. As can be seen, Model 1 has a longer memory than Model 2. Also note that the variance of the fast part of Model 2 is much higher than of the slow one (see (7.1.4)).

The sample mean squared error (SMSE) is defined as

$$\text{SMSE}\left(\hat{\theta},\theta\right) = \frac{1}{M}\sum_{i=1}^{M}\left(\hat{\theta}_i-\theta\right)\left(\hat{\theta}_i-\theta\right)^T \tag{7.1.5}$$

where θ is the vector of true values, $\hat{\theta}_i$ are the estimated values at the i-th repetition and M is the total number of repetitions. Note that this sample error can be split into the sample covariance, $S\left(\hat{\theta}\right)$, and into the squared bias, $\text{Bias}^2\left(\bar{\hat{\theta}}_M,\theta\right)$, i.e.

$$\begin{aligned}\text{SMSE}\left(\hat{\theta},\theta\right) &= \frac{1}{M}\sum_{i=1}^{M}\left(\hat{\theta}_i-\bar{\hat{\theta}}_M\right)\left(\hat{\theta}_i-\bar{\hat{\theta}}_M\right)^T+\left(\bar{\hat{\theta}}_M-\theta\right)\left(\bar{\hat{\theta}}_M-\theta\right)^T\\ &= S\left(\hat{\theta}\right)+\text{Bias}^2\left(\bar{\hat{\theta}}_M,\theta\right)\end{aligned} \tag{7.1.6}$$

where $\bar{\hat{\theta}}_M = \frac{1}{M}\sum_{i=1}^{M}\hat{\theta}_i$.

For our first simulation we generate $T = 10^5$ observations of Model 1 with Gaussian innovations and delete observations of the slow components to obtain our mixed-frequency pattern for $N = 3$. Thereafter, we calculate $\hat{\gamma}^{ff}(0)$, $\hat{\gamma}^{sf}(0)$, $\hat{\gamma}^{ff}(1)$ and $\hat{\gamma}^{sf}(1)$ and repeat these steps $M = 10^5$ times. In Table 7.1.1 the sample mean squared errors multiplied by $\sqrt{T}$ for these autocovariance estimators,

	$\hat{\gamma}^{ff}(0)$			$\hat{\gamma}^{sf}(0)$		$\hat{\gamma}^{ff}(1)$				$\hat{\gamma}^{sf}(1)$	
$\hat{\gamma}^{ff}(0)$	9414	2907	11323	-3784 (-3776)	6421 (6433)	-2584	8568	-8657	-2796	8244 (8239)	3951 (3926)
		1849	3890	-822 (-815)	1661 (1669)	-45	2636	-1677	205	2511 (2513)	2010 (2008)
			13924	-4421 (-4410)	7619 (7634)	-2799	10360	-9953	-2859	9847 (9844)	4947 (4920)
$\hat{\gamma}^{sf}(0)$				2728 (1670)	-1584 (-2678)	1297 (1298)	-3478 (-3470)	3822 (3818)	1470 (1473)	-2996 (-3332)	-1072 (-1309)
					6523 (4609)	-2049 (-2049)	5819 (5830)	-6355 (-6361)	-2380 (-2378)	6718 (5666)	3189 (2399)
$\hat{\gamma}^{ff}(1)$						1326	-2336	3205	1661	-2282 (-2279)	-432 (-421)
							7863	-7829	-2457	7501 (7497)	3573 (3550)
								9203	3864	-7691 (-7685)	-2878 (-2850)
									2255	-2535 (-2531)	-338 (-325)
$\hat{\gamma}^{sf}(1)$										8213 (7311)	4378 (3552)
											3968 (2575)

TABLE 7.1.1. Rounded SMSE multiplied by $\sqrt{T}$ of the autocovariance estimators of Model 1 in the mixed- (high-) frequency case for $T = 10^5$ and $M = 10^5$

say $\hat{\Sigma}_\gamma$, are displayed for the mixed-frequency case. Furthermore, the SMSEs multiplied by $\sqrt{T}$ for the high-frequency autocovariance estimators, say $\hat{\Sigma}_\gamma^{\text{high}}$, i.e. when all observations are available, are shown in brackets. Of course, the autocovariance estimators for the fast components in the high- and in the mixed-frequency case, e.g. $\hat{\gamma}^{ff}(0)$, are the same and so the SMSEs do not differ. It should be mentioned that the bias in this simulation is very small and will not be presented.

As has been shown in Chapter 5 the asymptotic cross-covariance of $\hat{\gamma}^{ff}(i)$ and $\hat{\gamma}^{sf}(j)$ is the same in the high- and in the mixed-frequency case. Indeed, as can be seen in Table 7.1.1, the SMSEs do not differ very much.

In Table 7.1.2 the exact asymptotic covariance matrix for the autocovariance estimators for the mixed-frequency case, say Σ_γ, is displayed, where, again, the high-frequency covariance matrix, say $\Sigma_\gamma^{\text{high}}$, is shown in brackets. Comparing Table 7.1.1 and Table 7.1.2 the sample mean squared error multiplied by $\sqrt{T}$ seems to converge to the exact asymptotic one.

Let us define the relative Frobenius norm of two matrices as

$$\varpi(A, B) = \|A - B\|_F / \|B\|_F \tag{7.1.7}$$

where $\|B\|_F > 0$. The relative Frobenius norms of the SMSEs multiplied by $\sqrt{T}$ and the exact asymptotic covariance matrix for different sample sizes $T = 10^2$, 10^3, 10^4 and 10^5 for the high- and the mixed-frequency case for Model 1 and Model 2 are given in Table 7.1.3, where the count of repetitions

	$\hat{\gamma}^{ff}(0)$			$\hat{\gamma}^{sf}(0)$		$\hat{\gamma}^{ff}(1)$				$\hat{\gamma}^{sf}(1)$	
$\hat{\gamma}^{ff}(0)$	9457	2913	11372	-3797	6463	-2601	8608	-8704	-2819	8276	3938
		1849	3898	-819	1673	-48	2641	-1683	200	2517	2007
			13981	-4434	7668	-2819	10405	-10005	-2884	9886	4934
$\hat{\gamma}^{sf}(0)$				2744	-1588	1307	-3490	3841	1483	-3001	-1055
				(1680)	(-2692)					(-3350)	(-1316)
					6560	-2060	5857	-6393	-2393	6748	3191
					(4629)					(5692)	(2408)
$\hat{\gamma}^{ff}(1)$						1332	-2353	3224	1669	-2294	-429
							7900	-7872	-2478	7531	3561
								9255	3889	-7725	-2865
									2265	-2551	-334
$\hat{\gamma}^{sf}(1)$										8231	4363
										(7342)	(3561)
											3959
											(2576)

TABLE 7.1.2. Rounded asymptotic covariance matrix of the autocovariance estimators of Model 1 in the mixed- (high-) frequency case

is $M = 10^5$. In accordance with the theory from Chapter 5 the relative norms seem to converge to zero.

		$T = 10^2$	$T = 10^3$	$T = 10^4$	$T = 10^5$
Model 1	MF $\hat{\Sigma}_\gamma$	0.1184	0.0103	0.0030	0.0048
	HF $\hat{\Sigma}_\gamma^{\text{high}}$	0.1119	0.0097	0.0027	0.0048
Model 2	MF $\hat{\Sigma}_\gamma$	0.0266	0.0055	0.0043	0.0019
	HF $\hat{\Sigma}_\gamma^{\text{high}}$	0.0266	0.0056	0.0043	0.0018

TABLE 7.1.3. Relative Frobenius norms of the SMSE multiplied by $\sqrt{T}$ and of the asymptotic covariance matrix of the autocovariance estimators for different sample sizes T for Model 1 and 2

7.2. Extended Yule-Walker Estimator

In this section we compare the asymptotic covariance matrix of the extended Yule-Walker estimator with the finite sample mean squared error multiplied by $\sqrt{T}$ for Model 1 and Model 2. Again, we choose the count of repetitions as $M = 10^5$. In Table 7.2.1 the relative Frobenius norms of the SMSE multiplied by $\sqrt{T}$ and the exact asymptotic covariance matrices are displayed for the two models with different increasing sample sizes T. Here MF corresponds to the asymptotic covariance in Proposition 5.2.6 and HF to the asymptotic covariance in Theorem 5.2.9. It seems that the relative norms converge to zero, as they are supposed to in theory. It is worth mentioning that the convergence of Model 1 is faster than the convergence of Model 2.

In Table C.0.1 the asymptotic covariance and in Table C.0.2 the SMSE multiplied by $\sqrt{T}$ for $T = 10^5$ for the extended Yule-Walker estimator in the mixed-frequency case for Model 1 are shown.

		$T = 10^2$	$T = 10^3$	$T = 10^4$	$T = 10^5$
Model 1	MF $\hat{A}_{\mathrm{XYW}}$	0.8999	0.2449	0.0169	0.0062
	HF $\hat{A}^{\mathrm{full}}_{\mathrm{XYW}}$	2.5622	0.2468	0.0253	0.0074
Model 2	MF $\hat{A}_{\mathrm{XYW}}$	0.8865	0.3736	0.1832	0.0217
	HF $\hat{A}^{\mathrm{full}}_{\mathrm{XYW}}$	2.6586	0.6384	0.0655	0.0053

TABLE 7.2.1. Relative Frobenius norms of the SMSE multiplied by $\sqrt{T}$ and of the asymptotic covariance matrix of the XYW estimator for different sample sizes T for Model 1 and 2

In addition, in Table C.0.3 the asymptotic covariance of the extended Yule-Walker estimator using high-frequency autocovariance estimators and in Table C.0.4 the asymptotic covariance of the standard high-frequency Yule-Walker estimator for Model 1 are displayed. By comparing these tables we immediately see that there is a large gap between the high- and mixed-frequency estimators, i.e. the covariance matrix of the mixed-frequency extended Yule-Walker estimator in Table C.0.1 is much larger than the covariance matrix shown in Table C.0.3, where the autocovariances are estimated via the high-frequency estimators. Furthermore, we can observe that there is also a gap between the asymptotic covariance matrices displayed in Table C.0.3 and Table C.0.4, which correspond to the two different high-frequency estimators.

7.3. Generalized Method of Moments Estimator

In an analogous way as in the section before, we compare the asymptotic covariance matrix of the feasible generalized method of moments estimator with the finite sample mean squared error multiplied by $\sqrt{T}$ for Model 1 and Model 2. In Table 7.3.1 the relative Frobenius norms of the SMSE multiplied by $\sqrt{T}$ and the exact asymptotic covariance are displayed for the two models, where the relative bias is shown in brackets, i.e.

$$\varpi\left(\bar{\hat{A}}, A\right) = \left\| \left(\bar{\hat{A}}_1, \bar{\hat{A}}_2\right) - (A_1, A_2) \right\|_F / \left\| (A_1, A_2) \right\|_F$$

where $\bar{\hat{A}}_j = \frac{1}{M}\sum_{i=1}^{M} \hat{A}_j^i$, $j = 1, 2$. Here we choose $M = 10^4$ repetitions and we truncate the infinite sum at the 100-th summand (for the case $T = 10^2$ we truncate the sum at the 50-th summand). This has mainly two reasons: Firstly, the average calculation period decreases (which is nevertheless very high) and secondly, the fractions of the norms of the autocovariance at lag $i \geq 100$ and at lag 0 of both models are rather small so that omitting these lags does not influence our results a lot.

The simulation shows that the relative SMSE of Model 1 for $T = 10^2$ is 22, which is very high and it increases at $T = 10^3$ to 72. Also note that the relative bias is very high. Nevertheless, by increasing sample size T they seem to converge to zero. Thus, we can conclude that in the case where T is small, the finite covariance matrix of the GMM estimator may differ a lot from the asymptotic one.

As can be seen in Table 7.3.1, the same conclusion follows for Model 2, where the relative SMSE as well as the relative bias seem to converge to zero, too.

In Table C.0.5 the asymptotic covariance matrix of the GMM estimator for Model 1 is presented. In contrast to Table C.0.1, i.e. the asymptotic covariance matrix of the XYW estimator, this covariance

	$T = 10^2$	$T = 10^3$	$T = 10^4$	$T = 10^5$	$T = 10^6$
Model 1	22 (0.4801)	72 (0.0581)	15 (0.0045)	0.2338 (0.0001)	0.0101 (0.0001)
Model 2	42 (0.3154)	27 (0.0571)	22 (0.0357)	5 (0.0033)	0.0112 (0.0001)

TABLE 7.3.1. Relative Frobenius norms of the SMSE multiplied by $\sqrt{T}$ and of the asymptotic covariance matrix of the GMM estimator for different sample sizes T for Model 1 and 2

matrix is smaller, i.e. the difference between these two models is positive (semi)-definite. Indeed, the relative norm between these two covariance matrices is 10.65. In Table C.0.6 the SMSE multiplied by $\sqrt{T}$ for $T = 10^6$ and $M = 10^5$ is displayed, which does not differ very much from the exact asymptotic variance displayed in Table C.0.5.

In the following lines we want to discuss the information loss measured by the asymptotic variance of different estimators which are based on the XYW equations. This information loss can be split into mainly two parts: Firstly, the loss of information which occurs due to the structure of the XYW equations, i.e. the missing observable slow autocovariances $\gamma^{ss}(h)$, $h \in N\mathbb{Z}$. Secondly, the information loss that occurs through the mixed-frequency setting, i.e. the mixed-frequency estimators for the autocovariances. In Table 7.3.2 a comparison of the asymptotic covariance matrices of the following estimators relative to the asymptotically efficient high-frequency YW estimator, say Σ_{YW}, is shown:

- XYW based on high-frequency observations, say $\Sigma^{\mathrm{full}}_{\mathrm{XYW}}$
- GMM based on mixed-frequency observations, say Σ_{GMM}
- XYW based on mixed-frequency observations, say Σ_{XYW}
- GMM based on mixed-frequency observations with a further lag $k = 1$, say $\Sigma_{\mathrm{GMM},1}$
- XYW based on mixed-frequency observations with a further lag $k = 1$, say $\Sigma_{\mathrm{XYW},1}$

As can be seen in the first column, where the autocovariance estimators are obtained using high-frequency observations, the relative norms for the two models are 3 and 62, which reflects that the use of the XYW equations instead of using standard YW equations can lead to a significant rise in the asymptotic covariances depending on the specific point in the parameter space. This effect increases when using the mixed-frequency autocovariance estimators as can be seen in the third column, i.e. 1352 and 8103. If we add a further lag to the XYW estimator the relative difference decreases for Model 1, i.e. 1124, and increases for Model 2, i.e. 9180, which is displayed in the last column. Nevertheless, the GMM estimator diminishes this effects as can be seen in the second and in the fourth column.

	$\varpi\left(\Sigma^{\mathrm{full}}_{\mathrm{XYW}}, \Sigma_{\mathrm{YW}}\right)$	$\varpi\left(\Sigma_{\mathrm{GMM}}, \Sigma_{\mathrm{YW}}\right)$	$\varpi\left(\Sigma_{\mathrm{XYW}}, \Sigma_{\mathrm{YW}}\right)$	$\varpi\left(\Sigma_{\mathrm{GMM},1}, \Sigma_{\mathrm{YW}}\right)$	$\varpi\left(\Sigma_{\mathrm{XYW},1}, \Sigma_{\mathrm{YW}}\right)$
1	3	117	1352	113	1124
2	62	650	8103	404	9180

TABLE 7.3.2. Comparison of the asymptotic covariance matrices for different estimators for Model 1 and 2

Tables C.0.7 and C.0.8 report the asymptotic covariance matrices of the XYW and the GMM estimator for Model 1 in the case of flow variables, where $w_t = y^s_t + y^s_{t-1} + y^s_{t-2}$. Let $\Sigma^{\mathrm{flow}}_{\mathrm{XYW}}$ and $\Sigma^{\mathrm{flow}}_{\mathrm{GMM}}$ denote these covariance matrices. The difference between the asymptotic covariance matrices of the

XYW estimator for the stock and the flow case, i.e. $\Sigma_{\text{XYW}} - \Sigma^{\text{flow}}_{\text{XYW}}$, yields in an indefinite matrix. Thus, using flow rather than stock data, does not seem to result in a better estimator in general. The same conclusion holds for the GMM estimators.

7.4. Maximum Likelihood AR(1) Estimator

In this section we consider the asymptotic covariance matrix of the maximum likelihood estimator for the special AR(1) case proposed in Section 5.2.2 and compare it to the SMSE multiplied by $\sqrt{T}$ of the roots of the score functions of Theorem 4.3.1 and to the SMSE multiplied by $\sqrt{T}$ of the EM algorithm introduced in Section 4.3.2. Since we are interested in asymptotic considerations we initialize the EM algorithm as well as the Newton-Raphson algorithm, which we use to find the roots of the score functions, with the true parameters. Note that this is in practice, of course, not possible.

In Table 7.4.1 the relative Frobenius norms of the SMSE multiplied by $\sqrt{T}$ and of the asymptotic covariance matrix are presented for Model (5.2.21), which we call "Model 3", where $N = 2$, $n = 2$, $\Sigma_\nu = I_2$ and $M = 5000$ repetitions. According to the theory, the relative Frobenius norms of the roots of the score functions seem to converge to zero. The relative Frobenius norms corresponding to the EM algorithm also converge to zero. It is worth noting that the calculation time of the EM algorithm is much higher than the time of finding the roots of the score functions. This points out the importance of more structure theory for the AR(p) $p > 1$ case.

Model 3	$T = 10^2$	$T = 10^3$	$T = 10^4$	$T = 10^5$
Roots of the score function	0.1009	0.0172	0.0091	0.0056
EM algorithm	0.1143	0.0207	0.0175	0.0152

TABLE 7.4.1. Relative Frobenius norms of the SMSE multiplied by $\sqrt{T}$ and of the asymptotic covariance matrix for different sample sizes T for Model 3

As in the section before the same comparison of the asymptotic covariance matrices, compared to the high-frequency YW covariance matrix, are displayed in Table 7.4.2 with the additional asymptotic covariance matrix of the ML estimator in the mixed-frequency case, say $\Sigma_{\text{AR }1}$. Note that in the AR(1) case, for $k = 0$, the GMM and the XYW estimator are the same. For this model, it follows that the asymptotic covariance matrix of the ML estimator in the mixed-frequency case is at least three times smaller than the asymptotic covariance matrices of the other mixed-frequency estimators.

$\varpi\left(\Sigma^{\text{full}}_{\text{XYW}}, \Sigma_{\text{YW}}\right)$	$\varpi\left(\Sigma_{\text{AR }1}, \Sigma_{\text{YW}}\right)$	$\varpi\left(\Sigma_{\text{XYW}}, \Sigma_{\text{YW}}\right)$	$\varpi\left(\Sigma_{\text{XYW},1}, \Sigma_{\text{YW}}\right)$	$\varpi\left(\Sigma_{\text{GMM},1}, \Sigma_{\text{YW}}\right)$
0.5730	1.4157	11.808	5.7571	4.736

TABLE 7.4.2. Comparison of the asymptotic covariance matrices for different estimators for Model 3

CHAPTER 8

Summary and Outlook

This thesis deals with identifiability and the asymptotic behavior of estimators of vector autoregressive systems in the (single and) mixed-frequency case. The asymptotic distribution of the autocovariance estimator, the extended Yule-Walker estimator, a generalized method of moments estimator and the maximum likelihood estimator (in the latter case only for the AR(1) case) are considered in detail.

In Chapter 2 the main focus is on autoregressive systems in the single-frequency case. In particular, the problem of forecasting and interpolation of observations and the estimation of the system and noise parameters where the AR system could be regular as well as singular are considered. In addition, the asymptotic behavior of two estimators is analyzed. Furthermore, system theoretical concepts, which are needed for the theory of identifiability and estimation of AR systems in the mixed-frequency case, are explained. Chapter 3 deals with the problem of identifiability of the system and noise parameters from mixed-frequency observations. For that reason the genericity property on a subset of the parameter space is introduced. It is shown that the parameters are g-identifiable in the parameter space. The last part of this chapter treats the case of flow rather than stock variables where g-identifiability is also established. In the fourth chapter different estimation procedures which are based on the XYW equations and on the (Gaussian) maximum likelihood theory are presented. One of the main contributions of this thesis is contained in Chapter 5, where the asymptotic distribution of the autocovariance, the XYW, the GMM and the ML estimator for the AR(1) case in the mixed-frequency case are derived. It turns out that the XYW and the GMM estimator are both not asymptotically efficient. Furthermore, upper and lower bounds for the asymptotic covariance of the XYW estimator are derived. In Chapter 6 the problems of projecting unstable AR polynomials and indefinite symmetric covariance matrices on the parameter space are considered. An adaption of a univariate stabilization algorithm for the multivariate case is presented and the problem of the initialization of this algorithm is treated. In Chapter 7 some finite sample effects are investigated through simulations.

There are still a number of open questions for further research which are connected with the topic of this thesis. The next few bullet points should give a small overview:

- Throughout this thesis it is assumed that the order of the AR polynomial p and the rank of the innovation covariance matrix q are known. Of course, in practice these two parameters have to be estimated, too.
- In Anderson et al. (2015a) an alternative realization procedure which is based on the idea of blocking the observed outputs is presented. Indeed, this procedure, which uses a particular state space representation, can be used to estimate the system and noise parameters of the

high-frequency system. Nevertheless, further research is needed since estimating this state space system yields in general an ARMA rather than an AR system. Thus, in a second step, one has to project this estimate back on the AR class. Furthermore, a comparison of the asymptotic behavior of this estimation procedure and the XYW estimator has to be done.

- As mentioned in Chapter 1 there are other ways to overcome the problem of forecasting from mixed-frequency data, which are not necessarily connected with a high-frequency system. One could think about a comparison of different procedures by comparing the one-step-ahead prediction error covariance matrix.
- In Anderson et al. (2015b) the mixed-frequency ARMA case is considered where g-identifiability of the system parameters is based on a modification of the XYW equations. The asymptotic results of Chapter 5 can be adapted to this ARMA case.
- One motivation of singular AR systems is that they can be used for modeling generalized dynamic factor models. In the single-frequency case there exists a wide range of literature about denoising, for instance Stock and Watson (2002); Doz et al. (2006, 2011); Choi (2012). Nevertheless, in the mixed-frequency case only a procedure proposed in Felsenstein (2014), which is based on the work of Hallin and Liska (2011), is known.

APPENDIX A

Derivation of the Maximum Likelihood AR(1) Estimator

The underlying representation of the AR(1) maximum likelihood with which we work is

$$\ln\left(l\left(A_1,\Sigma_\nu\right)\right) = c-\frac{T}{4}\ln|\Sigma_{\tilde{\nu}}|-\frac{1}{2}\mathrm{tr}\left(\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)^T\Sigma_{\tilde{\nu}}^{-1}\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)\right).$$

In order to obtain the score function we have to derive the first derivations with respect to the underlying high-frequency parameters, i.e.

$$\begin{aligned}\mathrm{vech}\left(s_{\Sigma_\nu}\right) &= \frac{\partial\ln\left(l\left(A_1,\Sigma_\nu\right)\right)}{\partial\mathrm{vech}\left(\Sigma_\nu\right)} \\ \mathrm{vec}\left(s_{A_1}\right) &= \frac{\partial\ln\left(l\left(A_1,\Sigma_\nu\right)\right)}{\partial\mathrm{vec}\left(A_1\right)}.\end{aligned} \tag{A.0.1}$$

In a first step we derive the first term $\ln|\Sigma_{\tilde{\nu}}|$ with respect to the noise parameters. All derivation rules can be found in Appendix C.

$$\begin{aligned}\frac{\partial\ln|\Sigma_{\tilde{\nu}}|}{\partial\mathrm{vech}\left(\Sigma_\nu\right)^T} &= \frac{\partial\ln|\Sigma_{\tilde{\nu}}|}{\partial\mathrm{vec}\left(\Sigma_{\tilde{\nu}}\right)^T}\frac{\partial\mathrm{vec}\left(\Sigma_{\tilde{\nu}}\right)}{\partial\mathrm{vech}\left(\Sigma_\nu\right)^T} \\ &= \mathrm{vec}\left(\Sigma_{\tilde{\nu}}^{-1}\right)^T\frac{\partial\mathrm{vec}\left(T_1\Sigma_\nu T_1^T+T_2\Sigma_\nu T_2^T\right)}{\partial\mathrm{vech}\left(\Sigma_\nu\right)^T} \\ &= \mathrm{vec}\left(\Sigma_{\tilde{\nu}}^{-1}\right)^T\left(\left(T_1\otimes T_1\right)+\left(T_2\otimes T_2\right)\right)D_n \\ &= \left(D_n^T\mathrm{vec}\left(T_1^T\Sigma_\nu^{-1}T_1+T_2^T\Sigma_\nu^{-1}T_2\right)\right)^T\end{aligned}$$

where we used $\frac{\partial\mathrm{vec}(\Sigma_\nu)}{\partial\mathrm{vech}(\Sigma_\nu)^T}=D_n\frac{\partial\mathrm{vech}(\Sigma_\nu)}{\partial\mathrm{vech}(\Sigma_\nu)^T}=D_n$ and Lemma C.0.3. Now the second term can be derived as follows:

$$\begin{aligned}\frac{\partial\mathrm{tr}\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)^T\Sigma_{\tilde{\nu}}^{-1}\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)}{\partial\mathrm{vech}\left(\Sigma_\nu\right)^T} &= \mathrm{vec}\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)^T\left(\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)^T\otimes I\right)\frac{\partial\mathrm{vec}\left(\Sigma_{\tilde{\nu}}^{-1}\right)}{\partial\mathrm{vech}\left(\Sigma_\nu\right)^T} \\ &= \mathrm{vec}\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)^T\left(\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)^T\otimes I\right)\frac{\partial\mathrm{vec}\left(\Sigma_{\tilde{\nu}}^{-1}\right)}{\partial\mathrm{vec}\left(\Sigma_{\tilde{\nu}}\right)^T}\frac{\mathrm{vec}\left(\Sigma_{\tilde{\nu}}\right)}{\partial\mathrm{vech}\left(\Sigma_\nu\right)^T} \\ &= -\Bigg(D_n^T\mathrm{vec}\Bigg(T_1^T\Sigma_{\tilde{\nu}}^{-1}\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)^T\Sigma_{\tilde{\nu}}^{-1}T_1 \\ &\quad +T_2^T\Sigma_{\tilde{\nu}}^{-1}\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)^T\Sigma_{\tilde{\nu}}^{-1}T_2\Bigg)\Bigg)^T.\end{aligned}$$

To summarize, the score function of the noise parameters is

$$\begin{aligned}\operatorname{vech}\left(s_{\Sigma_\nu}\right) &= -\frac{1}{2}D_n^T\operatorname{vec}\left(T_1^T\Sigma_{\tilde{\nu}}^{-1}\left(\frac{T}{2}\Sigma_{\tilde{\nu}}-\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)^T\right)\Sigma_{\tilde{\nu}}^{-1}T_1\right.\\ &\left.+T_2^T\Sigma_{\tilde{\nu}}^{-1}\left(\frac{T}{2}\Sigma_{\tilde{\nu}}-\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)^T\right)\Sigma_{\tilde{\nu}}^{-1}T_2\right).\end{aligned}$$

For the derivation of the second score function we will use

$$\begin{aligned}\frac{\partial\operatorname{vec}\left(T_2\right)}{\partial\operatorname{vec}\left(A_1\right)^T} &= \left(I_n\otimes T_1\right)\\ \frac{\partial\operatorname{vec}\left(a_1\right)}{\partial\operatorname{vec}\left(A_1\right)^T} &= \begin{pmatrix}0_{n_f\times n}\\ I_n\end{pmatrix}\otimes T_1\\ \frac{\partial\operatorname{vec}\left(a_2\right)}{\partial\operatorname{vec}\left(A_1\right)^T} &= T_1\otimes\left(\begin{pmatrix}0_{n_f\times n}\\ I_n\end{pmatrix}+\begin{pmatrix}I_{n_f} & 0_{n_f\times n_s}\\ 0_{n\times n_f} & 0_{n\times n_s}\end{pmatrix}\right)\end{aligned}$$

and

$$\begin{aligned}\frac{\partial\operatorname{vec}\left(\Sigma_{\tilde{\nu}}\right)}{\partial\operatorname{vec}\left(A_1\right)^T} &= \frac{\partial\operatorname{vec}\left(T_2\Sigma_\nu T_2^T\right)}{\partial\operatorname{vec}\left(A_1\right)^T}\\ &= \left(\left(I_{n+n_f}\otimes T_2\Sigma_\nu\right)K_{n+n_f,n}+\left(T_2\Sigma_\nu\otimes I_{n+n_f}\right)\right)\frac{\partial\operatorname{vec}\left(T_2\right)}{\partial\operatorname{vec}\left(A_1\right)^T}\\ &= K_{n+n_f,n+n_f}\left(T_2\Sigma_\nu\otimes T_1\right)+\left(T_2\Sigma_\nu\otimes T_1\right).\end{aligned}$$

Thus, with the formulas above it is easy to conclude that

$$\begin{aligned}\frac{\partial\ln\left|\Sigma_{\tilde{\nu}}\right|}{\partial\operatorname{vec}\left(A_1\right)^T} &= \operatorname{vec}\left(\Sigma_{\tilde{\nu}}^{-1}\right)^T\frac{\partial\operatorname{vec}\left(\Sigma_{\tilde{\nu}}\right)}{\partial\operatorname{vec}\left(A_1\right)^T}\\ &= \left(\left(K_{n,n}\left(T_1^T\otimes\Sigma_\nu T_2^T\right)+\left(\Sigma_\nu T_2^T\otimes T_1^T\right)\right)\operatorname{vec}\left(\Sigma_{\tilde{\nu}}^{-1}\right)\right)^T\\ &= \left(K_{n,n}\operatorname{vec}\left(\Sigma_\nu T_2^T\Sigma_{\tilde{\nu}}^{-1}T_1\right)+\operatorname{vec}\left(T_1^T\Sigma_{\tilde{\nu}}^{-1}T_2\Sigma_\nu\right)\right)^T\\ &= 2\operatorname{vec}\left(T_1^T\Sigma_{\tilde{\nu}}^{-1}T_2\Sigma_\nu\right)^T.\end{aligned}$$

As can be shown the second term yields

$$\begin{aligned}\frac{\partial\operatorname{tr}\left(\Sigma_{\tilde{\nu}}^{-1}\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)^T\right)}{\partial\operatorname{vec}\left(A_1\right)^T} &= 2\operatorname{vec}\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)^T\left(I\otimes\Sigma_{\tilde{\nu}}^{-1}\right)\frac{\partial\operatorname{vec}\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)}{\partial\operatorname{vec}\left(A_1\right)^T}+\\ &\operatorname{vec}\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)^T\left(\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)^T\otimes I\right)\frac{\partial\operatorname{vec}\left(\Sigma_{\tilde{\nu}}^{-1}\right)}{\partial\operatorname{vec}\left(A_1\right)^T}\end{aligned}$$

where

$$\begin{aligned}
\operatorname{vec}\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)^T\left(\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)^T\otimes I\right)\frac{\partial\operatorname{vec}\left(\Sigma_{\tilde{\nu}}^{-1}\right)}{\partial\operatorname{vec}\left(A_1\right)^T} &= \operatorname{vec}\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)^T\left(\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)^T\otimes I\right)\frac{\partial\operatorname{vec}\left(\Sigma_{\tilde{\nu}}^{-1}\right)}{\partial\operatorname{vec}\left(\Sigma_{\tilde{\nu}}\right)^T}\frac{\partial\operatorname{vec}\left(\Sigma_{\tilde{\nu}}\right)}{\partial\operatorname{vec}\left(A_1\right)^T}\\
&= -\operatorname{vec}\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)^T\left(\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)^T\Sigma_{\tilde{\nu}}^{-1}\otimes\Sigma_{\tilde{\nu}}^{-1}\right)\\
&\quad \left(K_{n+n_f,n+n_f}\left(T_2\Sigma_\nu\otimes T_1\right)+\left(T_2\Sigma_\nu\otimes T_1\right)\right)\\
&= -\left(K_{n,n}\operatorname{vec}\left(\Sigma_\nu T_2^T\Sigma_{\tilde{\nu}}^{-1}\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)^T\Sigma_{\tilde{\nu}}^{-1}T_1\right)\right)^T\\
&\quad -\operatorname{vec}\left(T_1^T\Sigma_{\tilde{\nu}}^{-1}\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)^T\Sigma_{\tilde{\nu}}^{-1}T_2\Sigma_\nu\right)^T\\
&= -2\operatorname{vec}\left(T_1^T\Sigma_{\tilde{\nu}}^{-1}\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)^T\Sigma_{\tilde{\nu}}^{-1}T_2\Sigma_\nu\right)^T.
\end{aligned}$$

To obtain the second term we observe that

$$\begin{aligned}
\frac{\partial\operatorname{vec}\left(\tilde{\mathcal{A}}X\right)}{\partial\operatorname{vec}\left(A_1\right)^T} &= \left(X^T\otimes I\right)\frac{\partial\operatorname{vec}\left(\tilde{\mathcal{A}}\right)}{\partial\operatorname{vec}\left(A_1\right)^T}\\
&= \left(X^T\otimes I\right)\left(\left(I\otimes a_1\right)\frac{\partial\operatorname{vec}\left(a_2\right)}{\partial\operatorname{vec}\left(A_1\right)^T}+\left(a_2^T\otimes I\right)\frac{\partial\operatorname{vec}\left(a_1\right)}{\partial\operatorname{vec}\left(A_1\right)^T}\right)\\
&= \left(X^T\otimes I\right)\left(\left(T_1\otimes T_2\right)+\left(a_2^T\begin{pmatrix}0_{n_f\times n}\\ I_n\end{pmatrix}\otimes T_1\right)\right)
\end{aligned}$$

where we used

$$a_1\left(\begin{pmatrix}0_{n_f\times n}\\ I_n\end{pmatrix}+\begin{pmatrix}I_{n_f} & 0_{n_f\times n_s}\\ 0_{n\times n_f} & 0_{n\times n_s}\end{pmatrix}\right)=T_2.$$

Now it follows

$$\begin{aligned}
\operatorname{vec}\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)^T\left(I\otimes\Sigma_{\tilde{\nu}}^{-1}\right)\frac{\partial\operatorname{vec}\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)}{\partial\operatorname{vec}\left(A_1\right)^T} &= -\operatorname{vec}\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)^T\left(X^T\otimes\Sigma_{\tilde{\nu}}^{-1}\right)\left(\left(T_1\otimes T_2\right)+\right.\\
&\quad \left.\left(a_2^T\begin{pmatrix}0_{n_f\times n}\\ I_n\end{pmatrix}\otimes T_1\right)\right)\\
&= -\operatorname{vec}\left(T_2^T\Sigma_{\tilde{\nu}}^{-1}\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)X^TT_1\right)\\
&\quad -\operatorname{vec}\left(T_1^T\Sigma_{\tilde{\nu}}^{-1}\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)X^Ta_2^T\begin{pmatrix}0_{n_f\times n}\\ I_n\end{pmatrix}\right)
\end{aligned}$$

which leads to

$$\begin{aligned}
\operatorname{vec}\left(s_{A_1}\right) &= \operatorname{vec}\left(T_2^T\Sigma_{\tilde{\nu}}^{-1}\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)X^TT_1\right)+\operatorname{vec}\left(T_1^T\Sigma_{\tilde{\nu}}^{-1}\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)X^Ta_2^T\begin{pmatrix}0_{n_f\times n}\\ I_n\end{pmatrix}\right)\\
&\quad -\operatorname{vec}\left(T_1^T\Sigma_{\tilde{\nu}}^{-1}\left(\frac{T}{2}\Sigma_{\tilde{\nu}}-\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)\left(\tilde{Y}-\tilde{\mathcal{A}}X\right)^T\right)\Sigma_{\tilde{\nu}}^{-1}T_2\Sigma_\nu\right).
\end{aligned}$$

Under our assumptions the asymptotic covariance matrix of the ML estimator is given as

$$\Sigma_{\text{AR 1}}=\lim_{\frac{T}{2}\to\infty}\frac{T}{2}\mathbb{I}_T\left(\theta\right)^{-1} \tag{A.0.2}$$

where

$$\mathbb{I}_T(\theta) = \mathbb{E}\left(\frac{\partial \ln\left(l\left(A_1, \Sigma_\nu\right)\right)}{\partial \mathrm{vec}\left(\theta\right)}\left(\frac{\partial \ln\left(l\left(A_1, \Sigma_\nu\right)\right)}{\partial \mathrm{vec}\left(\theta\right)}\right)^T\right). \tag{A.0.3}$$

To obtain the information matrix we have to rewrite the score functions into

$$\begin{aligned}
\mathrm{vech}\left(s_{\Sigma_\nu}\right) &= \underbrace{-\frac{1}{2}D_n^T\left(\left(T_1^T\Sigma_{\tilde{\nu}}^{-1}\otimes T_1^T\Sigma_{\tilde{\nu}}^{-1}\right)+\left(T_2^T\Sigma_{\tilde{\nu}}^{-1}\otimes T_2^T\Sigma_{\tilde{\nu}}^{-1}\right)\right)}_{=H_1}\mathrm{vec}\left(\frac{T}{2}\Sigma_{\tilde{\nu}}-\tilde{V}\tilde{V}^T\right)\\
\mathrm{vec}\left(s_{A_1}\right) &= \underbrace{-\left(\Sigma_\nu T_2^T\Sigma_{\tilde{\nu}}^{-1}\otimes T_1^T\Sigma_{\tilde{\nu}}^{-1}\right)}_{H_2}\mathrm{vec}\left(\frac{T}{2}\Sigma_{\tilde{\nu}}-\tilde{V}\tilde{V}^T\right)\\
&\quad+\underbrace{\left(\left(T_1^T\otimes T_2^T\Sigma_{\tilde{\nu}}^{-1}\right)+\left(\left(0_{n\times n_f}\quad I_n\right)a_2\otimes T_1^T\Sigma_{\tilde{\nu}}^{-1}\right)\right)}_{H_3}\mathrm{vec}\left(\tilde{V}X^T\right).
\end{aligned}$$

The structure of X and $\tilde{V}$ implies that $\mathbb{E}\left(\tilde{V}X^T\right)=0$, $\mathbb{E}\left(\mathrm{vec}\left(\tilde{V}X^T\right)\mathrm{vec}\left(\tilde{V}\tilde{V}^T\right)^T\right)=0$ since $\mathbb{E}\left(\tilde{y}_t\tilde{\nu}_t^T\right)=0$ and

$$\mathbb{E}\left(\mathrm{vec}\left(\tilde{V}X^T\right)\mathrm{vec}\left(\tilde{V}X^T\right)^T\right)=\frac{T}{2}\left(\tilde{\gamma}\left(0\right)\otimes\Sigma_{\tilde{\nu}}\right)$$

where $\tilde{\gamma}\left(0\right)=\mathbb{E}\left(\tilde{y}_t\tilde{y}_t^T\right)$. Let $\tilde{\eta}=\mathbb{E}\left(\tilde{V}_t\tilde{V}_t^T\otimes\tilde{V}_t\tilde{V}_t^T\right)$ and observe that $\tilde{\eta}=\mathrm{vec}\left(\Sigma_{\tilde{\nu}}\right)\mathrm{vec}\left(\Sigma_{\tilde{\nu}}\right)^T+\left(\Sigma_{\tilde{\nu}}\otimes\Sigma_{\tilde{\nu}}\right)+K_{n+n_f,n+n_f}\left(\Sigma_{\tilde{\nu}}\otimes\Sigma_{\tilde{\nu}}\right)$. It is straightforward to show that $\mathbb{E}\mathrm{vec}\left(\tilde{V}\tilde{V}^T\right)=\frac{T}{2}\mathrm{vec}\left(\Sigma_{\tilde{\nu}}\right)$ and

$$\mathbb{E}\left(\mathrm{vec}\left(\tilde{V}\tilde{V}^T\right)\mathrm{vec}\left(\tilde{V}\tilde{V}^T\right)^T\right)=\left(\frac{T^2}{4}-\frac{T}{2}\right)\mathrm{vec}\left(\Sigma_{\tilde{\nu}}\right)\mathrm{vec}\left(\Sigma_{\tilde{\nu}}\right)^T+\frac{T}{2}\tilde{\eta}.$$

Thus,

$$\begin{aligned}
\mathbb{E}\left(\mathrm{vec}\left(\frac{T}{2}\Sigma_{\tilde{\nu}}-\tilde{V}\tilde{V}^T\right)\mathrm{vec}\left(\frac{T}{2}\Sigma_{\tilde{\nu}}-\tilde{V}\tilde{V}^T\right)^T\right) &= \frac{T}{2}\left(\tilde{\eta}-\mathrm{vec}\left(\Sigma_{\tilde{\nu}}\right)\mathrm{vec}\left(\Sigma_{\tilde{\nu}}\right)^T\right)\\
&= \frac{T}{2}\left(\Sigma_{\tilde{\nu}}\otimes\Sigma_{\tilde{\nu}}\right)+K_{n+n_f,n+n_f}\left(\Sigma_{\tilde{\nu}}\otimes\Sigma_{\tilde{\nu}}\right).
\end{aligned}$$

Now we have everything to construct our information matrix for the AR(1) case:

$$\begin{aligned}
\mathbb{E}\left(\mathrm{vech}\left(s_{\Sigma_\nu}\right)\mathrm{vech}\left(s_{\Sigma_\nu}\right)^T\right) &= \frac{T}{2}H_1\left(\left(\Sigma_{\tilde{\nu}}\otimes\Sigma_{\tilde{\nu}}\right)+K_{n+n_f,n+n_f}\left(\Sigma_{\tilde{\nu}}\otimes\Sigma_{\tilde{\nu}}\right)\right)H_1^T\\
\mathbb{E}\left(\mathrm{vec}\left(s_{A_1}\right)\mathrm{vec}\left(s_{A_1}\right)^T\right) &= \frac{T}{2}H_2\left(\left(\Sigma_{\tilde{\nu}}\otimes\Sigma_{\tilde{\nu}}\right)+K_{n+n_f,n+n_f}\left(\Sigma_{\tilde{\nu}}\otimes\Sigma_{\tilde{\nu}}\right)\right)H_2^T\\
&\quad+\frac{T}{2}H_3\left(\tilde{\gamma}\left(0\right)\otimes\Sigma_{\tilde{\nu}}\right)H_3^T\\
\mathbb{E}\left(\mathrm{vec}\left(s_{A_1}\right)\mathrm{vech}\left(s_{\Sigma_\nu}\right)^T\right) &= \frac{T}{2}H_2\left(\left(\Sigma_{\tilde{\nu}}\otimes\Sigma_{\tilde{\nu}}\right)+K_{n+n_f,n+n_f}\left(\Sigma_{\tilde{\nu}}\otimes\Sigma_{\tilde{\nu}}\right)\right)H_1^T\\
&= \left(\mathbb{E}\left(\mathrm{vec}\left(s_{\Sigma_\nu}\right)\mathrm{vec}\left(s_{A_1}\right)^T\right)\right)^T.
\end{aligned}$$

APPENDIX B

Proofs of Lemma 5.1.2 and Lemma 5.1.6

PROOF OF LEMMA 5.1.2. Expanding y_t^f and y_t^s into $\sum_{i=-\infty}^{\infty} k_i^f \nu_{t-i}$ and $\sum_{i=-\infty}^{\infty} k_i^s \nu_{t-i}$, respectively, yields

$$\mathbb{E}\left(\left(y_{t-p}^f \left(y_{Nu-q}^f\right)^T\right) \otimes \left(y_t^f \left(y_{Nu}^s\right)^T\right)\right) =$$

$$\mathbb{E}\left(\sum_h \sum_i \sum_j \sum_k \left(k_i^f \nu_{t-p-i} \otimes k_k^f \nu_{t-k}\right)\left(\nu_{Nu-q-j}^T \left(k_j^f\right)^T \otimes \nu_{Nu-h}^T \left(k_h^s\right)^T\right)\right). \tag{B.0.1}$$

In this context we use the notation $\sum_i$ instead of $\sum_{i=-\infty}^{\infty}$. The term in (B.0.1) can be separated into four different cases since in all other cases there occur terms such as $\mathbb{E}\left(\nu_{t-i}\nu_{t-j}^T \otimes \nu_{t-k}\nu_{t-l}^T\right)$, which are zero for $i \neq j$, $i \neq k$ and $i \neq l$. The following formulas will be used in the proof:

$$\begin{aligned}
\operatorname{vec}\left(\gamma^{ff}(h)\right) &= \sum_i \left(k_{i-h}^f \otimes k_i^f\right) \operatorname{vec}\left(\Sigma_\nu\right) \\
\operatorname{vec}\left(\gamma^{sf}(h)\right) &= \sum_i \left(k_{i-h}^f \otimes k_i^s\right) \operatorname{vec}\left(\Sigma_\nu\right).
\end{aligned}$$

In the first case we assume that all ν_t occur at the same time point, i.e. $p + i = k$, $j + q = h$ and $t - k = Nu - q - j$. Thus, we obtain

$$\begin{aligned}
\mathrm{I} &= \sum_k \mathbb{E}\left(k_{k-p}^f \nu_{t-k} \otimes k_k^f \nu_{t-k}\right)\left(\nu_{t-k}^T \left(k_{k-q+Nu-t}^f\right)^T \otimes \nu_{t-k}^T \left(k_{k+Nu-t}^s\right)^T\right) \\
&= \sum_k \left(k_{k-p}^f \otimes k_k^f\right) \underbrace{\mathbb{E}\left(\nu_{t-k}\nu_{t-k}^T \otimes \nu_{t-k}\nu_{t-k}^T\right)}_{=\eta} \left(k_{k-q+Nu-t}^f \otimes k_{k+Nu-t}^s\right)^T \\
&= \sum_k \left(k_{k-p}^f \otimes k_k^f\right) \eta \left(k_{k-q+Nu-t}^f \otimes k_{k+Nu-t}^s\right)^T.
\end{aligned}$$

In the second case we assume that the first two ν_t and the last two ν_t in equation (B.0.1) occur at the same time point but not all at the same time, i.e. $p+i=k$, $q+j=h$ but $t-k\neq Nu-q-j$:

$$\begin{aligned}
\text{II} &= \sum_k \sum_{j,t-k\neq Nu-q-j} \mathbb{E}\left(\left(k^f_{k-p}\nu_{t-k}\otimes k^f_k\nu_{t-k}\right)\left((\nu^T_{Nu-q-j}\left(k^f_j\right)^T\otimes\nu^T_{Nu-q-j}\left(k^s_{q+j}\right)^T\right)\right)\\
&= \sum_k \sum_{j,t-k\neq Nu-q-j}\left(k^f_{k-p}\otimes k^f_k\right)\mathbb{E}\left(\nu_{t-k}\otimes\nu_{t-k}\right)\mathbb{E}\left(\nu^T_{Nu-q-j}\otimes\nu^T_{Nu-q-j}\right)\left(k^f_j\otimes k^s_{q+j}\right)^T\\
&= \sum_k \sum_{j,t-k\neq Nu-q-j}\left(k^f_{k-p}\otimes k^f_k\right)\mathrm{vec}\left(\Sigma_\nu\right)\mathrm{vec}\left(\Sigma_\nu\right)^T\left(k^f_j\otimes k^s_{q+j}\right)^T\\
&= \sum_k\sum_j\left(k^f_{k-p}\otimes k^f_k\right)\mathrm{vec}\left(\Sigma_\nu\right)\mathrm{vec}\left(\Sigma_\nu\right)^T\left(k^f_j\otimes k^s_{q+j}\right)^T\\
&\quad -\sum_k\left(k^f_{k-p}\otimes k^f_k\right)\mathrm{vec}\left(\Sigma_\nu\right)\mathrm{vec}\left(\Sigma_\nu\right)^T\left(k^f_{k-q+Nu-t}\otimes k^s_{k+Nu-t}\right)^T\\
&= \mathrm{vec}\left(\gamma^{ff}\left(p\right)\right)\mathrm{vec}\left(\gamma^{sf}\left(q\right)\right)^T-\sum_k\left(k^f_{k-p}\otimes k^f_k\right)\mathrm{vec}\left(\Sigma_\nu\right)\mathrm{vec}\left(\Sigma_\nu\right)^T\left(k^f_{k-q+Nu-t}\otimes k^s_{k+Nu-t}\right)^T.
\end{aligned}$$

In the third case we assume that the first and the third ν_t and the second and the fourth ν_t in equation (B.0.1) occur at the same time point, but not all at the same time, i.e. $t-p-i=Nu-q-j$, $t-k=Nu-h$ but $p+i\neq k$:

$$\begin{aligned}
\text{III} &= \sum_i\sum_{k,k\neq p+i}\mathbb{E}\left(\left(k^f_i\nu_{t-p-i}\otimes k^f_k\nu_{t-k}\right)\left(\nu^T_{t-p-i}\left(k^f_{p+i-q+Nu-t}\right)^T\otimes\nu^T_{t-k}\left(k^s_{Nu-t+k}\right)^T\right)\right)\\
&= \sum_i\sum_{k,k\neq p+i}\left(k^f_i\otimes k^f_k\right)\mathbb{E}\left(\nu_{t-p-i}\nu^T_{t-p-i}\otimes\nu_{t-k}\nu^T_{t-k}\right)\left(\left(k^f_{p+i-q+Nu-t}\right)^T\otimes\left(k^s_{Nu-t+k}\right)^T\right)\\
&= \sum_i\sum_{k,k\neq p+i}\left(k^f_i\otimes k^f_k\right)\left(\Sigma_\nu\otimes\Sigma_\nu\right)\left(\left(k^f_{p+i-q+Nu-t}\right)^T\otimes\left(k^s_{Nu-t+k}\right)^T\right)\\
&= \sum_i\sum_k\left(k^f_i\otimes k^f_k\right)\left(\Sigma_\nu\otimes\Sigma_\nu\right)\left(\left(k^f_{p+i-q+Nu-t}\right)^T\otimes\left(k^s_{Nu-t+k}\right)^T\right)\\
&\quad -\sum_k\left(k^f_{k-p}\otimes k^f_k\right)\left(\Sigma_\nu\otimes\Sigma_\nu\right)\left(k^f_{k-q+Nu-t}\otimes k^s_{Nu-t+k}\right)^T\\
&= \sum_i\sum_k\left(k^f_i\Sigma_\nu\left(k^f_{p+i-q+Nu-t}\right)^T\otimes k^f_k\Sigma_\nu\left(k^s_{Nu-t+k}\right)^T\right)\\
&\quad -\sum_k\left(k^f_{k-p}\otimes k^f_k\right)\left(\Sigma_\nu\otimes\Sigma_\nu\right)\left(k^f_{k-q+Nu-t}\otimes k^s_{Nu-t+k}\right)^T\\
&= \left(\gamma^{ff}(-Nu+t+q-p)\otimes\gamma^{fs}\left(-Nu+t\right)\right)\\
&\quad -\sum_k\left(k^f_{k-p}\otimes k^f_k\right)\left(\Sigma_\nu\otimes\Sigma_\nu\right)\left(k^f_{k-q+Nu-t}\otimes k^s_{Nu-t+k}\right)^T.
\end{aligned}$$

In the fourth case, where we assume that $t-p-i=Nu-h$, $t-k=Nu-q-j$ but $p+i\neq k$, we will use the result that for $p+i\neq k$

$$\begin{aligned}
\mathbb{E}\left(\nu_{t-p-i}\nu_{t-k}^T \otimes \nu_{t-k}\nu_{t-p-i}^T\right) &= \mathbb{E}\left(\mathbb{E}\left(\nu_{t-p-i}\nu_{t-k}^T \otimes \nu_{t-k}\nu_{t-p-i}^T\right)|\nu_{t-k}\right) \\
&= \mathbb{E}\left(\mathbb{E}\left(\left(I_n \otimes \nu_{t-k}\right)\left(\nu_{t-p-i} \otimes \nu_{t-p-i}^T\right)\left(\nu_{t-k}^T \otimes I_n\right)\right)|\nu_{t-k}\right) \\
&= \mathbb{E}\left(\left(I_n \otimes \nu_{t-k}\right)\left(\mathbb{E}\left(\nu_{t-p-i} \otimes \nu_{t-p-i}^T\right)\right)\left(\nu_{t-k}^T \otimes I_n\right)\right) \qquad \text{(B.0.2)} \\
&= \mathbb{E}\left(\left(I_n \otimes \nu_{t-k}\right)\Sigma_\nu\left(\nu_{t-k}^T \otimes I_n\right)\right) \qquad \text{(B.0.3)}
\end{aligned}$$

holds. Then it follows:

$$\begin{aligned}
\text{IV} &= \sum_k \sum_{i,p+i\neq k} \mathbb{E}\left(\left(k_i^f \nu_{t-p-i} \otimes k_k^f \nu_{t-k}\right)\left(\nu_{t-k}^T \left(k_{k-q+Nu-t}^f\right)^T \otimes \nu_{t-p-i}^T \left(k_{i+p+Nu-t}^s\right)^T\right)\right) \\
&= \sum_k \sum_{i,p+i\neq k} \left(k_i^f \otimes k_k^f\right)\mathbb{E}\left(\nu_{t-p-i}\nu_{t-k}^T \otimes \nu_{t-k}\nu_{t-p-i}^T\right)\left(k_{k-q+Nu-t}^f \otimes k_{i+p+Nu-t}^s\right)^T.
\end{aligned}$$

Rewriting the above equation with equation (B.0.2) results in

$$\begin{aligned}
\text{IV} &= \sum_k \sum_{i,p+i\neq k} \left(k_i^f \otimes k_k^f\right)\mathbb{E}\left(\left(I_n \otimes \nu_{t-k}\right)\Sigma_\nu\left(\nu_{t-k}^T \otimes I_n\right)\right)\left(k_{k-q+Nu-t}^f \otimes k_{i+p+Nu-t}^s\right)^T \\
&= \sum_k \sum_{i,p+i\neq k} \mathbb{E}\left(k_i^f \otimes k_k^f \nu_{t-k}\right)\Sigma_\nu\left(k_{k-q+Nu-t}^f \nu_{t-k} \otimes k_{i+p+Nu-t}^s\right)^T.
\end{aligned}$$

Now we use the well known fact that $\left(k_i^f \otimes k_k^f \nu_{t-k}\right) = K_{n_f,n_f}\left(k_k^f \nu_{t-k} \otimes k_i^f\right)$, compare Lemma C.0.2,

$$\begin{aligned}
\text{IV} &= \sum_k \sum_{i,p+i\neq k} K_{n_f,n_f}\mathbb{E}\left(k_k^f \nu_{t-k} \otimes k_i^f\right)\Sigma_\nu\left(k_{k-q+Nu-t}^f \nu_{t-k} \otimes k_{i+p+Nu-t}^s\right)^T \\
&= K_{n_f,n_f}\sum_k \sum_{i,p+i\neq k} \mathbb{E}\left(k_k^f \nu_{t-k} \otimes k_i^f \Sigma_\nu\right)\left(k_{k-q+Nu-t}^f \nu_{t-k} \otimes k_{i+p+Nu-t}^s\right)^T \\
&= K_{n_f,n_f}\sum_k \sum_{i,p+i\neq k} \mathbb{E}\left(k_k^f \nu_{t-k}\nu_{t-k}^T \otimes k_i^f \Sigma_\nu\right)\left(k_{k-q+Nu-t}^f \otimes k_{i+p+Nu-t}^s\right)^T \\
&= K_{n_f,n_f}\sum_k \sum_{i,p+i\neq k} \left(k_k^f \Sigma_\nu \otimes k_i^f \Sigma_\nu\right)\left(k_{k-q+Nu-t}^f \otimes k_{i+p+Nu-t}^s\right)^T \\
&= K_{n_f,n_f}\sum_k \sum_i \left(k_k^f \Sigma_\nu\left(k_{k-q+Nu-t}^f\right)^T \otimes k_i^f \Sigma_\nu\left(k_{i+p+Nu-t}^s\right)^T\right) \\
&\quad -K_{n_f,n_f}\sum_k \left(k_k^f \Sigma_\nu\left(k_{k-q+Nu-t}^f\right)^T \otimes k_{k-p}^f \Sigma_\nu\left(k_{k+Nu-t}^s\right)^T\right) \\
&= K_{n_f,n_f}\left(\gamma^{ff}(-Nu+t+q) \otimes \gamma^{fs}(-Nu+t-p)\right) \\
&\quad -K_{n_f,n_f}\sum_k \left(k_k^f \otimes k_{k-p}^f\right)\left(\Sigma_\nu \otimes \Sigma_\nu\right)\left(k_{k-q+Nu-t}^f \otimes k_{k+Nu-t}^s\right)^T.
\end{aligned}$$

Now combining all four cases and rearranging the according terms where we use $K_{n_f,n_f}\left(k_k^f \otimes k_{k-p}^f\right) = \left(k_{k-p}^f \otimes k_k^f\right)K_{n,n}$ yields our desired result. □

PROOF OF LEMMA 5.1.6. Expanding y_t^f and y_t^s into $\sum_{i=-\infty}^{\infty} k_i^f \nu_{t-i}$ and $\sum_{i=-\infty}^{\infty} k_i^s \nu_{t-i}$, respectively, yields

$$\mathbb{E}\left(\left(y_{Nt-p}^f \left(y_{Nt+Nh-q}^f\right)^T\right) \otimes \left(y_{Nt}^s \left(y_{Nt+Nh}^s\right)^T\right)\right) =$$

$$\mathbb{E}\left(\sum_u \sum_i \sum_j \sum_k \left(k_i^f \nu_{Nt-p-i} \otimes k_k^s \nu_{Nt-k}\right)\left(\nu_{Nt+Nh-q-j}^T \left(k_j^f\right)^T \otimes \nu_{Nt+Nh-u}^T \left(k_u^s\right)^T\right)\right). \quad \text{(B.0.4)}$$

In this context we, again, use the notation $\sum_i$ instead of $\sum_{i=-\infty}^{\infty}$. As in the proof of Lemma 5.1.2 we split (B.0.4) into four different cases, again in all other cases the expectation is zero:

In the first case we assume that all ν_t occur at the same time point, i.e. $p+i=k$, $j+q=u$ and $k=j+q-Nh$. Thus, we obtain

$$\begin{aligned}
\text{I} &= \sum_k \mathbb{E}\left(k_{k-p}^f \nu_{Nt-k} \otimes k_k^s \nu_{Nt-k}\right)\left(\nu_{Nt-k}^T \left(k_{k-q+Nh}^f\right)^T \otimes \nu_{Nt-k}^T \left(k_{k+Nh}^s\right)^T\right) \\
&= \sum_k \left(k_{k-p}^f \otimes k_k^s\right) \underbrace{\mathbb{E}\left(\nu_{Nt-k}\nu_{Nt-k}^T \otimes \nu_{Nt-k}\nu_{Nt-k}^T\right)}_{=\eta} \left(k_{k-q+Nh}^f \otimes k_{k+Nh}^s\right)^T \\
&= \sum_k \left(k_{k-p}^f \otimes k_k^s\right) \eta \left(k_{k-q+Nh}^f \otimes k_{k+Nh}^s\right)^T.
\end{aligned}$$

In the second case we assume that the first two ν_t and the last two ν_t in equation (B.0.4) occur at the same time point but not all at the same time, i.e. $p+i=k$, $q+j=u$ but $k \neq j+q-Nh$:

$$\begin{aligned}
\text{II} &= \sum_k \sum_{j,k\neq j+q-Nh} \mathbb{E}\left(\left(k_{k-p}^f \nu_{Nt-k} \otimes k_k^s \nu_{Nt-k}\right)\left(\left(\nu_{Nt+Nh-q-j}^T \left(k_j^f\right)^T \otimes \nu_{Nt+Nh-q-j}^T \left(k_{q+j}^s\right)^T\right)\right)\right) \\
&= \sum_k \sum_{j,k\neq j+q-Nh} \left(k_{k-p}^f \otimes k_k^s\right) \mathbb{E}\left(\nu_{Nt-k} \otimes \nu_{Nt-k}\right) \mathbb{E}\left(\nu_{Nt+Nh-q-j}^T \otimes \nu_{Nt+Nh-q-j}^T\right)\left(k_j^f \otimes k_{q+j}^s\right)^T \\
&= \sum_k \sum_{j,k\neq j+q-Nh} \left(k_{k-p}^f \otimes k_k^s\right) \operatorname{vec}\left(\Sigma_\nu\right) \operatorname{vec}\left(\Sigma_\nu\right)^T \left(k_j^f \otimes k_{q+j}^s\right)^T \\
&= \sum_k \sum_j \left(k_{k-p}^f \otimes k_k^s\right) \operatorname{vec}\left(\Sigma_\nu\right) \operatorname{vec}\left(\Sigma_\nu\right)^T \left(k_j^f \otimes k_{q+j}^s\right)^T \\
&\quad - \sum_k \left(k_{k-p}^f \otimes k_k^s\right) \operatorname{vec}\left(\Sigma_\nu\right) \operatorname{vec}\left(\Sigma_\nu\right)^T \left(k_{k-q+Nh}^f \otimes k_{k+Nh}^s\right)^T \\
&= \operatorname{vec}\left(\gamma^{sf}\left(p\right)\right) \operatorname{vec}\left(\gamma^{sf}\left(q\right)\right)^T - \sum_k \left(k_{k-p}^f \otimes k_k^s\right) \operatorname{vec}\left(\Sigma_\nu\right) \operatorname{vec}\left(\Sigma_\nu\right)^T \left(k_{k-q+Nh}^f \otimes k_{k+Nh}^s\right)^T.
\end{aligned}$$

In the third case we assume that the first and the third ν_t and the second and the fourth ν_t in equation (B.0.4) occur at the same time point but not all at the same time, i.e. $p+i=q+j-Nh$, $k=u-Nh$ but $k \neq p+i$:

$$
\begin{aligned}
\text{III} &= \sum_i \sum_{k,k\neq p+i} \mathbb{E}\left(\left(k_i^f \nu_{Nt-p-i} \otimes k_k^s \nu_{Nt-k}\right)\left(\nu_{Nt-p-i}^T \left(k_{p+i-q+Nh}^f\right)^T \otimes \nu_{Nt-k}^T \left(k_{Nh+k}^s\right)^T\right)\right) \\
&= \sum_i \sum_{k,k\neq p+i} \left(k_i^f \otimes k_k^s\right) \mathbb{E}\left(\nu_{Nt-p-i}\nu_{Nt-p-i}^T \otimes \nu_{Nt-k}\nu_{Nt-k}^T\right)\left(\left(k_{p+i-q+Nh}^f\right)^T \otimes \left(k_{Nh+k}^s\right)^T\right) \\
&= \sum_i \sum_{k,k\neq p+i} \left(k_i^f \otimes k_k^s\right) \left(\Sigma_\nu \otimes \Sigma_\nu\right)\left(\left(k_{p+i-q+Nh}^f\right)^T \otimes \left(k_{Nh+k}^s\right)^T\right) \\
&= \sum_i \sum_k \left(k_i^f \otimes k_k^s\right) \left(\Sigma_\nu \otimes \Sigma_\nu\right)\left(\left(k_{p+i-q+Nh}^f\right)^T \otimes \left(k_{Nh+k}^s\right)^T\right) \\
&\quad - \sum_k \left(k_{k-p}^f \otimes k_k^s\right)\left(\Sigma_\nu \otimes \Sigma_\nu\right)\left(k_{k-q+Nh}^f \otimes k_{Nh+k}^s\right)^T \\
&= \sum_i \sum_k \left(k_i^f \Sigma_\nu \left(k_{p+i-q+Nh}^f\right)^T \otimes k_k^s \Sigma_\nu \left(k_{Nh+k}^s\right)^T\right) \\
&\quad - \sum_k \left(k_{k-p}^f \otimes k_k^s\right)\left(\Sigma_\nu \otimes \Sigma_\nu\right)\left(k_{k-q+Nh}^f \otimes k_{Nh+k}^s\right)^T \\
&= \left(\gamma^{ff}(-Nh+q-p) \otimes \gamma^{ss}(-Nh)\right) - \sum_k \left(k_{k-p}^f \otimes k_k^s\right)\left(\Sigma_\nu \otimes \Sigma_\nu\right)\left(k_{k-q+Nh}^f \otimes k_{Nh+k}^s\right)^T.
\end{aligned}
$$

In the fourth case, where we assume that $p+i = u - Nh$, $k-q+Nh = j$ but $p+i \neq k$, we will use the result that for $p+i \neq k$

$$
\begin{aligned}
\mathbb{E}\left(\nu_{Nt-p-i}\nu_{Nt-k}^T \otimes \nu_{Nt-k}\nu_{Nt-p-i}^T\right) &= \mathbb{E}\left(\mathbb{E}\left(\nu_{Nt-p-i}\nu_{Nt-k}^T \otimes \nu_{Nt-k}\nu_{Nt-p-i}^T\right)|\nu_{Nt-k}\right) \\
&= \mathbb{E}\left(\mathbb{E}\left(\left(I_n \otimes \nu_{Nt-k}\right)\left(\nu_{Nt-p-i} \otimes \nu_{Nt-p-i}^T\right)\left(\nu_{Nt-k}^T \otimes I_n\right)\right)|\nu_{Nt-k}\right) \\
&= \mathbb{E}\left(\left(I_n \otimes \nu_{Nt-k}\right)\Sigma_\nu\left(\nu_{Nt-k}^T \otimes I_n\right)\right) \qquad \text{(B.0.5)}
\end{aligned}
$$

holds. It follows:

$$
\begin{aligned}
\text{IV} &= \sum_k \sum_{i,p+i\neq k} \mathbb{E}\left(\left(k_i^f \nu_{Nt-p-i} \otimes k_k^s \nu_{Nt-k}\right)\left(\nu_{Nt-k}^T \left(k_{k-q+Nh}^f\right)^T \otimes \nu_{Nt-p-i}^T \left(k_{i+p+Nh}^s\right)^T\right)\right) \\
&= \sum_k \sum_{i,p+i\neq k} \left(k_i^f \otimes k_k^s\right) \mathbb{E}\left(\nu_{Nt-p-i}\nu_{Nt-k}^T \otimes \nu_{Nt-k}\nu_{Nt-p-i}^T\right)\left(k_{k-q+Nh}^f \otimes k_{i+p+Nh}^s\right)^T.
\end{aligned}
$$

Rewriting the above equation with equation (B.0.5) results in

$$
\begin{aligned}
\text{IV} &= \sum_k \sum_{i,p+i\neq k} \left(k_i^f \otimes k_k^s\right) \mathbb{E}\left(\left(I_n \otimes \nu_{Nt-k}\right)\Sigma_\nu\left(\nu_{Nt-k}^T \otimes I_n\right)\right)\left(k_{k-q+Nh}^f \otimes k_{i+p+Nh}^s\right)^T \\
&= \sum_k \sum_{i,p+i\neq k} \mathbb{E}\left(k_i^f \otimes k_k^s \nu_{Nt-k}\right)\Sigma_\nu\left(k_{k-q+Nh}^f \nu_{Nt-k} \otimes k_{i+p+Nh}^s\right)^T.
\end{aligned}
$$

Now we will use the fact that $\left(k_i^f \otimes k_k^s \nu_{Nt-k}\right) = K_{n_f,n_s}\left(k_k^s \nu_{Nt-k} \otimes k_i^f\right)$, compare Lemma (C.0.2),

$$\begin{aligned}
\text{IV} &= \sum_k \sum_{i,p+i\neq k} K_{n_f,n_s}\mathbb{E}\left(k_k^s \nu_{Nt-k} \otimes k_i^f\right)\Sigma_\nu\left(k_{k-q+Nh}^f \nu_{Nt-k} \otimes k_{i+p+Nh}^s\right)^T \\
&= K_{n_f,n_s}\sum_k \sum_{i,p+i\neq k} \mathbb{E}\left(k_k^s \nu_{Nt-k} \otimes k_i^f \Sigma_\nu\right)\left(k_{k-q+Nh}^f \nu_{Nt-k} \otimes k_{i+p+Nh}^s\right)^T \\
&= K_{n_f,n_s}\sum_k \sum_{i,p+i\neq k} \mathbb{E}\left(k_k^s \nu_{Nt-k}\nu_{Nt-k}^T \otimes k_i^f \Sigma_\nu\right)\left(k_{k-q+Nh}^f \otimes k_{i+p+Nh}^s\right)^T \\
&= K_{n_f,n_s}\sum_k \sum_{i,p+i\neq k} \left(k_k^s \Sigma_\nu \otimes k_i^f \Sigma_\nu\right)\left(k_{k-q+Nh}^f \otimes k_{i+p+Nh}^s\right)^T \\
&= K_{n_f,n_s}\sum_k \sum_i \left(k_k^s \Sigma_\nu \left(k_{k-q+Nh}^f\right)^T \otimes k_i^f \Sigma_\nu \left(k_{i+p+Nh}^s\right)^T\right) \\
&\quad -K_{n_f,n_s}\sum_k \left(k_k^s \Sigma_\nu \left(k_{k-q+Nh}^f\right)^T \otimes k_{k-p}^f \Sigma_\nu \left(k_{k+Nh}^s\right)^T\right) \\
&= K_{n_f,n_s}\left(\gamma^{sf}(-Nh+q) \otimes \gamma^{fs}(-Nh-p)\right) \\
&\quad -K_{n_f,n_s}\sum_k \left(k_k^s \otimes k_{k-p}^f\right)(\Sigma_\nu \otimes \Sigma_\nu)\left(k_{k-q+Nh}^f \otimes k_{k+Nh}^s\right)^T.
\end{aligned}$$

Again, combining all four cases and rearranging the according terms where we use $K_{n_f,n_s}\left(k_k^s \otimes k_{k-p}^f\right) = \left(k_{k-p}^f \otimes k_k^s\right)K_{n,n}$ yields our desired result. □

APPENDIX C

Vectorization, Kronecker Product and Tables of Simulation

The following two lemmas repeat useful properties of the columnwise vectorization vec (.) and the Kronecker operator.

LEMMA C.0.1. *Let A, B and C matrices from which the dimension should be clear out of the context. Then:*

(1) $vec(A+B) = vec(A) + vec(B)$.
(2) $vec(ABC) = \left(C^T \otimes A\right) vec(B)$.
(3) $vec(AB) = (I \otimes A)\, vec(B) = \left(B^T \otimes I\right) vec(A)$.

(4) $\mathrm{tr}(ABC) = \mathrm{vec}\left(A^T\right)^T \left(C^T \otimes I\right) \mathrm{vec}(B)$.
(5) $\mathrm{tr}(ABC) = \mathrm{vec}\left(B^T\right)^T \left(A^T \otimes I\right) \mathrm{vec}(C)$.
(6) $A \otimes B \neq B \otimes A$, $(A \otimes B)^T = \left(A^T \otimes B^T\right)$.
(7) $(A \otimes B)(C \otimes D) = (AC \otimes BD)$.
(8) $(A \otimes B)^{-1} = \left(A^{-1} \otimes B^{-1}\right)$ *where A and B are nonsingular.*
(9) $(A \otimes B)^{\dagger} = \left(A^{\dagger} \otimes B^{\dagger}\right)$ *where $A^{\dagger}$ is the Moore–Penrose pseudoinverse of A.*
(10) *If A and B are square matrices with eigenvalues λ_i^A and λ_j^B, respectively, then $(A \otimes B)$ has the eigenvalues $\lambda_i^A \lambda_j^B$.*
(11) $|A \otimes B| = |A|^n |B|^m$ *where $A \in \mathbb{R}^{m \times m}$ and $B \in \mathbb{R}^{n \times n}$.*

PROOF. see Lütkepohl (2005), Appendix A. □

LEMMA C.0.2. *Let $A \in \mathbb{R}^{n \times s}$, $B \in \mathbb{R}^{m \times t}$, $x \in \mathbb{R}^k$ and $K_{i,j} \in \mathbb{R}^{ij \times ij}$ be a commutation matrix, i.e. $vec\left(A^T\right) = K_{n,s} vec(A)$. Furthermore let $D_n \in \mathbb{R}^{n^2 \times \frac{n(n+1)}{2}}$ be a duplication matrix, i.e. $vec(C) = D_n vech(C)$ for any symmetric matrix $C \in \mathbb{R}^{n \times n}$, and $L_n \in \mathbb{R}^{\frac{n(n+1)}{2} \times n^2}$ be an elimination matrix, i.e. $vech(D) = L_n vec(D)$ where $D \in \mathbb{R}^{n \times n}$. Then:*

(1) $K_{m,n}(A \otimes B) K_{s,t} = (B \otimes A)$.
(2) $K_{k,n}(A \otimes x) = (x \otimes A)$.
(3) $K_{m,n}^T = K_{n,m}$.
(4) $K_{m,n}^{-1} = K_{n,m}$.
(5) *If C is symmetric, then $D_n^{\dagger} vec(C) = L_n vec(C)$ where $D_n^{\dagger}$ is the Moore–Penrose pseudoinverse of D_n.*

PROOF. see Magnus and Neudecker (1979). □

The next lemma is useful for deriving the score and the information matrix in Appendix A.

LEMMA C.0.3. *Let $A \in \mathbb{R}^{n\times n}$ and $\theta \in \mathbb{R}^n$. Then:*

(1) $\frac{\partial A\theta}{\partial \theta^T} = A$.
(2) $\frac{\partial \ln|A|}{\partial A} = \left(A^T\right)^{-1}$ where A is nonsingular.
(3) $\frac{\partial \mathrm{vec}\left(A^{-1}\right)}{\partial \mathrm{vec}(A)^T} = -\left(\left(A^{-1}\right)^T \otimes A^{-1}\right)$ where A is nonsingular.
(4) Let $A(\theta) \in \mathbb{R}^{m\times p}$, $C(\theta) \in \mathbb{R}^{q\times r}$ and $B \in \mathbb{R}^{p\times q}$. Then $\frac{\partial \mathrm{vec}(ABC)}{\partial \theta^T} = (I_r \otimes AB)\frac{\partial \mathrm{vec}(C)}{\partial \theta^T} + \left(C^T B^T \otimes I_m\right)\frac{\partial \mathrm{vec}(A)}{\partial \theta^T}$.

PROOF. see Lütkepohl (2005), Appendix A. □

$\mathrm{vec}(A_1, A_2)$																	
133	72	100	-339	-203	-299	177	97	129	86	44	89	57	45	51	-234	-158	-219
	47	56	-166	-111	-156	99	60	77	44	27	42	19	19	25	-100	-76	-110
		92	-254	-150	-262	148	91	112	58	23	88	36	28	36	-168	-111	-177
			1038	580	842	-449	-214	-358	-274	-158	-220	-202	-126	-155	794	473	655
				356	463	-268	-128	-232	-156	-113	-95	-103	-71	-87	423	273	361
					820	-432	-257	-308	-203	-73	-288	-136	-93	-124	599	363	580
						251	149	186	108	50	138	66	52	66	-299	-200	-302
							107	93	46	2	109	19	22	31	-121	-88	-161
								195	93	81	43	49	33	47	-244	-158	-225
									80	54	41	52	31	42	-210	-124	-167
										73	-38	31	15	25	-126	-78	-81
											169	32	29	33	-145	-87	-178
												55	31	32	-178	-101	-120
													23	20	-101	-67	-75
														27	-121	-73	-101
															652	369	491
																232	291
																	436

$\mathrm{vec}(A_1, A_2)$

TABLE C.0.1. Rounded asymptotic covariance matrix of the XYW estimator of Model 1 in the mixed-frequency case

$\mathrm{vec}(A_1, A_2)$																	
132	72	100	-337	-202	-298	176	97	129	85	44	88	57	45	51	-233	-158	-218
	46	56	-165	-111	-156	99	60	77	43	27	42	19	19	25	-99	-75	-109
		92	-253	-149	-261	148	91	112	57	23	88	35	28	36	-167	-110	-176
			1031	576	837	-447	-214	-356	-271	-156	-219	-200	-126	-154	789	471	651
				354	460	-267	-128	-231	-155	-112	-94	-103	-71	-86	420	272	358
					817	-430	-256	-307	-202	-72	-287	-135	-93	-124	595	361	577
						251	149	186	107	50	136	65	52	65	-298	-199	-301
							107	93	45	2	108	19	22	31	-120	-88	-161
								195	92	80	43	48	33	47	-242	-157	-224
									79	54	41	52	31	41	-208	-122	-165
										72	-38	30	15	24	-124	-77	-80
											169	32	28	33	-144	-86	-178
												55	31	32	-177	-101	-119
													23	20	-101	-67	-75
														26	-120	-72	-100
															648	367	487
																231	289
																	433

$\mathrm{vec}(A_1, A_2)$

TABLE C.0.2. Rounded SMSE multiplied by $\sqrt{T}$ of the XYW estimator of Model 1 in the mixed-frequency case for $T = 10^5$ and $M = 10^5$

vec (A_1, A_2) / vec $(A_1, A_2)^T$					
0.7 0.7 0.7	-1.4 -1.4 -1.4	0.7 0.7 0.7	0.5 0.5 0.5	0.2 0.2 0.2	-0.8 -0.8 -0.8
	4.0 4.0 4.0	-1.3 -1.3 -1.3	-1.4 -1.4 -1.4	-0.9 -0.9 -0.9	3.0 3.0 3.0
		1.0 1.0 1.0	0.2 0.2 0.2	0.1 0.1 0.1	-0.7 -0.7 -0.7
			0.8 0.8 0.8	0.2 0.2 0.2	-0.8 -0.8 -0.8
				0.5 0.5 0.5	-1.0 -1.0 -1.0
					2.8 2.8 2.8

TABLE C.0.3. Rounded asymptotic covariance matrix of the XYW estimator of Model 1 in the high-frequency case

vec (A_1, A_2) / vec $(A_1, A_2)^T$					
0.2 0.2 0.2	-0.2 -0.2 -0.2	0.1 0.1 0.1	0.1 0.1 0.1	0.0 0.0 0.0	0.0 0.0 0.0
	0.7 0.7 0.7	-0.1 -0.1 -0.1	-0.3 -0.3 -0.3	-0.2 -0.2 -0.2	0.6 0.6 0.6
		0.2 0.2 0.2	-0.1 -0.1 -0.1	0.0 0.0 0.0	0.0 0.0 0.0
			0.4 0.4 0.4	-0.1 -0.1 -0.1	-0.1 -0.1 -0.1
				0.2 0.2 0.2	-0.4 -0.4 -0.4
					0.9 0.9 0.9

TABLE C.0.4. Rounded asymptotic covariance matrix of the YW estimator of Model 1 in the high-frequency case

vec $(A_1, A_2)^T$																	
vec (A_1, A_2)	11	7	4	-17	-12	-12	10	7	1	7	1	8	1	5	1	-5	-9	-7
		6	2	-8	-8	-7	7	6	1	4	1	5	-2	2	1	2	-3	-4
			6	-16	-5	-22	8	7	2	5	-2	14	1	1	2	-9	-3	-13
				91	49	63	-26	-7	-32	-33	-27	-10	-13	-6	-9	68	34	48
					38	16	-13	3	-24	-20	-27	13	-7	-6	-4	35	25	18
						96	-30	-32	-5	-22	11	-64	2	2	-7	34	7	56
							14	12	3	10	-1	20	-0	3	2	-10	-8	-17
								19	-10	3	-16	34	-6	0	0	7	4	-13
									23	13	26	-22	6	-0	3	-29	-14	-12
										13	12	2	3	1	3	-22	-11	-17
											33	-36	7	0	3	-28	-15	-5
												71	-8	-0	1	9	9	-26
													8	3	1	-17	-10	-2
														5	0	-4	-8	1
															2	-7	-3	-6
																64	28	32
																	22	9
																		37

TABLE C.0.5. Rounded asymptotic covariance matrix of the GMM estimator of Model 1 in the mixed-frequency case

vec $(A_1, A_2)^T$																	
vec (A_1, A_2)	11	7	4	-17	-12	-12	10	7	1	7	1	8	1	5	1	-5	-9	-7
		6	2	-7	-7	-7	7	6	1	4	1	5	-2	2	1	2	-3	-4
			6	-17	-5	-22	7	7	2	5	-2	14	1	1	2	-10	-4	-13
				92	49	63	-25	-6	-32	-33	-28	-10	-13	-6	-9	69	34	49
					38	16	-13	3	-24	-20	-27	14	-7	-6	-4	35	25	18
						96	-29	-32	-4	-22	11	-64	1	2	-7	34	7	56
							14	12	3	10	-1	19	-0	3	2	-10	-8	-17
								19	-10	2	-16	35	-6	0	0	8	4	-13
									23	13	26	-22	6	-0	3	-30	-14	-12
										13	12	1	3	1	3	-22	-11	-17
											33	-37	7	0	3	-28	-15	-4
												71	-8	-0	1	10	9	-26
													8	4	1	-17	-10	-2
														5	0	-4	-8	1
															2	-7	-3	-6
																66	28	33
																	23	10
																		38

TABLE C.0.6. Rounded SMSE multiplied by $\sqrt{T}$ of the GMM estimator of Model 1 in the mixed-frequency case for $T = 10^6$ and $M = 10^5$

vec (A_1, A_2) / vec $(A_1, A_2)^T$																	
126	74	48	-191	-125	-93	31	0	-8	86	41	39	41	49	35	-102	-91	-74
	66	27	-93	-94	-45	30	7	15	63	57	-1	-5	12	23	-19	-37	-41
		57	-79	-32	-116	33	29	-24	29	-16	102	14	17	11	-43	-31	-51
			354	221	125	-35	38	-18	-140	-100	-5	-109	-92	-62	248	187	119
				233	-18	-33	55	-101	-121	-185	149	-37	-45	-41	117	134	39
					336	-93	-129	129	-19	166	-374	-19	-40	-15	64	33	145
						75	62	0	6	-29	94	-16	14	-5	8	-25	-36
							95	-59	-31	-114	185	-33	-2	-15	57	29	-35
								122	34	152	-211	-10	-11	4	-4	-33	43
									88	102	-61	20	17	37	-70	-59	-43
										260	-313	4	-12	35	-47	-60	31
											519	-1	33	-20	10	29	-123
												68	41	15	-119	-69	-27
													42	7	-75	-64	-27
														22	-40	-22	-28
															237	145	75
																128	43
																	89

TABLE C.0.7. Rounded asymptotic covariance matrix of the XYW estimator of Model 1 in the mixed-frequency flow case ($c_i = 1$, $i = 1, 2, 3$)

vec (A_1, A_2) / vec $(A_1, A_2)^T$																	
15	12	-3	-5	-14	6	9	6	-0	4	-0	1	-5	7	-2	13	-7	5
	19	-8	7	-12	10	7	7	-0	7	7	-7	-16	-3	1	28	7	2
		9	-2	12	-21	3	7	-8	-5	-16	25	6	2	-1	-6	-0	-5
			30	28	-24	1	20	-23	-3	-23	35	-22	-16	4	37	34	-12
				65	-69	0	33	-48	-15	-65	92	-7	-14	3	15	39	-28
					97	-12	-51	63	19	90	-131	8	6	-1	-22	-35	35
						9	13	-9	-1	-13	21	-5	4	-2	14	1	-1
							39	-40	-8	-53	79	-16	-4	-1	36	25	-16
								49	12	66	-94	12	6	-1	-28	-30	23
									7	20	-27	-3	-2	2	3	-3	4
										96	-134	9	0	3	-27	-31	29
											193	-16	-3	-3	44	47	-42
												23	11	-2	-40	-24	5
													14	-5	-14	-21	7
														4	1	6	-5
															74	39	-8
																45	-19
																	19

TABLE C.0.8. Rounded asymptotic covariance matrix of the GMM estimator of Model 1 in the mixed-frequency flow case ($c_i = 1$, $i = 1, 2, 3$)

Bibliography

Abrevaya, J. and Donald, S. G. (2011). A GMM Approach for Dealing with Missing Data on Regressors and Instruments. Mimeo.

Anderson, B., Deistler, M., Felsenstein, E., Funovits, B., Koelbl, L., and Zamani, M. (2015a). Multivariate AR Systems and Mixed Frequency Data: g-Identifiability and Estimation. *Econometric Theory*, doi:10.1017/S0266466615000043:1–34, reproduced with permission.

Anderson, B., Deistler, M., Felsenstein, E., and Koelbl, L. (2015b). The Structure of Multivariate AR and ARMA Systems: Regular and Singular Systems; the Single and the Mixed Frequency Case. *submitted to Annals of Journal of Econometrics, Resubmitted after Revision*.

Anderson, B. D. O., Deistler, M., Felsenstein, E., Funovits, B., Zadrozny, P. A., Eichler, M., Chen, W., and Zamani, M. (2012). Identifiability of Regular and Singular Multivariate Autoregressive Models. *Proceedings of the 51th IEEE Conference on Decision and Control (CDC)*, 184-189.

Anderson, T. (1958). *An Introduction to Multivariate Statistics*. John Wiley & Sons, New York.

Anderson, T. (1994). *The Statistical Analysis of Time Series*. John Wiley & Sons.

Balogh, L. and Pintelon, R. (2008). Stable Approximation of Unstable Transer Function Models. *IEEE Transactions on Instrumentation and Measurement*, 57(12):2720–2726.

Berlinet, A. and Francq, C. (1997). On Bartlett's Formula for Non-Linear Processes. *Journal of Time Series Analysis*, 18(6):535–552.

Bochnak, J., Coste, M., and Roy, M.-F. (1998). *Real Algebraic Geometry*. Springer Verlag, Berlin.

Brockwell, P. J. and Davis, R. A. (1987). *Time Series: Data Theory and Methods*. Springer, New York.

Brubacher, S. and Tunnicliffe Wilson, G. (1976). Interpolating Time Series with Application to the Estimation of Holiday Effects on Electricity Demand. *Journal of the Royal Statistical Society*, 25(2):107–116.

Burg, J. P. (1975). *Maximum Entropy Spectral Analysis*. PhD thesis, Stanford University.

Chen, B. and Zadrozny, P. A. (1998). An Extended Yule-Walker Method for Estimating a Vector Autoregressive Model with Mixed-Frequency Data. *Advances in Econometrics*, 13:47–73.

Chen, W., Anderson, B. D. O., Deistler, M., and Filler, A. (2011). Solutions of Yule-Walker Equations for Singular AR Processes. *Journal of Time Series Analysis*, 32:531–538.

Choi, I. (2012). Efficient Estimation of Factor Models. *Econometric Theory*, 2:274–308.

Combettes, P. L. and Trussell, H. J. (1992). Best Stable and Invertible Approximations for ARMA Systems. *IEEE Transactions on Signal Processing*, 40(12):3066–3069.

De Hoon, M., van der Hagen, T., Schoonewelle, H., and van Dam, H. (1996). Why Yule-Walker should not be used for Autoregressive Modelling. *Annals of Nuclear Energy*, 1523(15):1219–1228.

Deistler, M., Anderson, B. D. O., Filler, A., Zinner, C., and Chen, W. (2010). Generalized Linear Dynamic Factor Models - An Approach via Singular Autoregressions. *European Journal of Control*, 16(3):211–224.

Deistler, M., Filler, A., and Funovits, B. (2011). AR Systems and AR Processes: The Singular Case. *Communications in Information and Systems*, 11(3):225–236.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society*, 39(1):1–38.

D'haene, T., Pintelon, R., and Vandersteen, G. (2006). An Iterative Method to Stabilize a Transfer Function in the s- and z- Domains. *IEEE Transactions on Instrumentation and Measurement*, 55(4):1192–1196.

Doz, C., Giannone, D., and Reichlin, L. (2006). A Quasi Maximum Likelihood Approach for Large Approximate Dynamic Factor Models. *ECB Working Paper No. 674*.

Doz, C., Giannone, D., and Reichlin, L. (2011). A Two-Step Estimator for Large Approximate Dynamic Factor Models based on Kalman Filtering. *Journal of Econometrics*, 164(1):188–205.

Eckart, C. and Young, G. (1936). The Approximation of One Matrix by Another of Lower Rank. *Psychometrika*, 1(3):211–218.

Felsenstein, E. (2014). *Regular and Singular AR and ARMA Models: The Single and The Mixed Frequency Case*. PhD thesis, Vienna University of Technology.

Filler, A. (2010). *Generalized Dynamic Factor Models - Structure Theory and Estimation for Single Frequency and Mixed Frequency Data*. PhD thesis, Vienna University of Technology.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The Generalized Dynamic Factor Model: Identification and Estimation. *Review of Economics and Statistics*, 82(4):540–554.

Forni, M., Hallin, M., Lippi, M., and Zaffaroni, P. (2015). Dynamic Factor Models with Infinite-Dimensional Factor Spaces: One-Sided Representations. *Journal of Econometrics*, 185:359–371.

Friedman, M. (1962). The Interpolation of Time Series by Related Series. *Journal of the American Statistical Association*, 57(300):729–757.

Ghysels, E., Santa-Clara, P., and Valkanov, R. (2004). The MIDAS touch: MIxed DAta Sampling models. *mimeo, Chapel Hill, N.C.*

Ghysels, E., Sinko, A., and Valkanov, R. (2007). MIDAS Regressions: Further Results and New Directions. *Econometric Reviews*, 26(1):53–90.

Gingras, D. F. (1985). Asymptotic Properties of High-Order Yule-Walker Estimates of the AR Parameters of an ARMA Time Series. *IEEE Transactions on Acoustics Speech and Signal Processing*, 33(4):1095–1101.

Golub, G. and Van Loan, C. (1996). *Matrix Computations*. Johns Hopkins Univ Pr.

Gomez, V., Maravall, A., and Pena, D. (1999). Missing Observations in ARIMA Models: Skipping Approach Versus Additive Outlier Approach. *Journal of Econometrics*, 88:341–363.

Grenander, U. and Rosenblatt, M. (1957). *Statistical Analysis of Stationary Time Series*. Wiley publications in statistics: Mathematical statistics.

Hall, P. and Heyde, C. (1980). *Martingal Limit Theory and Its Application*. Academic Press, New York.

Hallin, M. and Liska, R. (2011). Dynamic Factors in the Presence of Blocks. *Journal of Econometrics*, 163:29–41.

Hannan, E. J. (1970). *Multiple Time Series*. John Wiley & Sons, New York.

Hannan, E. J. and Deistler, M. (2012). *The Statistical Theory of Linear Systems*. SIAM Classics in Applied Mathematics, Philadelphia.

Hannan, E. J. and Heyde, C. (1972). On Limit Theorems for Quadratic Function of Discrete Time Series. *The Annals of Mathematical Statistics*, 43(6):2058–2066.

Hansen, L. P. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, 50(4):1029–1054.

Harvey, A. C. and Pierse, R. G. (1984). Estimating Missing Observations in Economic Time Series. *Journal of the American Statistical Association*, 79(385):125–131.

Higham, J. N. (1989). Matrix Nearness Problems and Applications. In Gover, M. and Barnett, S., editors, *Applications of Matrix Theory*, pages 1–27. Oxford University Press.

Higham, J. N. (2002). Computing the Nearest Correlation Matrix - A Problem from Finance. *IMA Journal of Numerical Analysis*, 22(3):329–343.

Hoffman, A. and Wielandt, H. (1953). The Variation of the Spectrum of a Normal Matrix. *Duke Mathematical Journal*, 20:37–39.

Kailath, T. (1980). *Linear Systems*. Prentice Hall, Englewood Cliffs, New Jersey.

Khatri, C. G. (1961). Some Results for the Singular Normal Multivariate Regression Models. *The Indian Journal of Statistics*, 30:267–280.

Koelbl, L., Braumann, A., Felsenstein, E., and Deistler, M. (2015). Estimation of VAR Systems from Mixed Frequency Data: The Stock and the Flow Case. *to appear in Advances in Econometrics*, 35.

Kohn, R. and Ansley, C. F. (1986). Estimation, Prediction and Interpolation for ARIMA Models with Missing Data. *Journal of the American Statistical Association*, 81(395):751–761.

Kuzin, V., Marcellino, M., and Schumacher, C. (2011). MIDAS versus Mixed-Frequency VAR: Nowcasting GDP in the Euro Area. *International Journal of Forecasting*, 27:529–542.

Lee, E. B. and Markus, L. (1967). *Foundations of Optimal Control Theory*. John Wiley & Sons.

Lippi, M. and Reichlin, L. (1994). VAR Analysis, Nonfundamental Representations, Blaschke Matrices. *Journal of Econometrics*, 63:307–325.

Lomnicki, Z. (1961). Tests for Departure from Normality in the Case of Linear Stochastic Processes. *Metrika*, 4(1):37–62.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer Berlin.

Mader, W., Linke, Y., Mader, M., Sommerlade, L., Timmer, J., and Schelter, B. (2014). A Numerically Efficient Implementation of the Expectation Maximization Algorithm for State Space Models. *Applied Mathematics and Computation*, 241:222–232.

Magnus, J. R. and Neudecker, H. (1979). The Commutation Matrix: Some Properties and Applications. *Annals of Statistics*, 7(2):381–394.

Mann, H. B. and Wald, A. (1943). On the Statistical Treatment of Linear Stochastic Difference Equations. *Econometrica*, 11(3/4):173–220.

Mariano, R. S. and Murasawa, Y. (2003). A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series. *Journal of Applied Econometrics*, 18(4):427–443.

McLachlan, G. and Krishnan, T. (2008). *The EM Algorithm and Extensions*. John Wiley & Sons.

Moses, L. R. and Liu, D. (1991). Determining the Closest Stable Polynomial to an Unstable One. *IEEE Transactions on Signal Processing*, 39(4):901–906.

Muris, C. (2014). Efficient GMM Estimation with General Missing Data Patterns. Technical report, Simon Frasier University.

Niebuhr, T. and Kreiss, J.-P. (2013). Asymptotics for Autocovariances and Integrated Periodograms for Linear Processes Observed at Lower Frequencies. *International Statistical Review*, (1):123–140.

Orbandexivry, F.-X., Nesterov, Y., and Van Dooren, P. (2013). Nearest Stable System Using Successive Convex Approximations. *Automatica*, 49:1195–1203.

Pourahmadi, M. (1988). Estimation and Interpolation of Missing Values of a Stationary Time Series. *Journal of Time Series Analysis*, 10(2):149–169.

Rao, C. R. (1972). *Linear Statistical Inference and Its Applications*. John Wiley & Sons.

Rao, C. R. and Mitra, S. (1971). Generalized Inverse of a Matrix and its Applications. *Wiley, New York*.

Reinsel, G. (1993). *Elements of Multivariate Time Series Analysis*. Springer Series in Statistics, New York.

Robinson, P. (1977). Estimating Variances and Covariances of Sample Autocorrelations and Autocovariances. *Australian Journal of Statistics*, 19(3):236–240.

Rozanov, Y. A. (1967). *Stationary Random Processes*. Holden-Day, San Francisco.

Shumway, R. and Stoffer, D. (1982). An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm. *Journal of Time Series Analysis*, 3(4):253–264.

Srivastava, M. S. and von Rosen, D. (2002). Regression Models with Unknown Singular Covariance Matrix. *Linear Algebra and its Application*, 354:255–273.

Stock, J. H. and Watson, M. W. (2002). Forecasting Using Principal Components From a Large Number of Predictors. *Journal of the American Statistical Association*, 97(460):1167–1179.

Stock, J. H. and Watson, M. W. (2008). Forecasting in Dynamic Factor Models Subject to Structural Instability. In Castle, J. and Shephard, N., editors, *The Methodology and Practice of Econometrics*. Oxford: Oxford University Press.

Stoica, P. and Moses, L. R. (1992). On the Unit Circle Problem: The Schur-Cohn Procedure Revisited. *Signal Processing*, 26:95–118.

Su, N. and Lund, R. (2011). Multivariate Versions of Bartlett's Formula. *Journal of Multivariate Analysis*, 105:18–31.

van der Vaart, A. (2000). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics.

Wax, M. and Kailath, T. (1983). Efficient Inversion of Toeplitz-Block Toeplitz Matrix. *IEEE Transactions on Acoustics Speech and Signal Processing*, 31(5):1218–1221.

Wilkinson, J. (1979). Elemementary Proof of the Wielandt-Hoffman Theorem and of its Generalization. Technical report, Computer Science Department, Standford University.

Wohlrabe, K. (2008). *Forecasting with Mixed-frequency Time Series Models*. PhD thesis, Ludwig-Maximilians-Universität München.

Wold, H. (1938). *A Study in the Analysis of Stationary Time Series*. PhD thesis, Almqvistand Wiksell, Sweden.

Wonham, W. M. (1985). *Linear Multivariable Control - A Geometric Approach*. Springer Verlag, third edition.

Wu, C. F. J. (1983). On the Convergence Properties of the EM Algorithm. *The Annals of Statistics*, 11(1):95–103.

Zadrozny, P. A. (1988). Gaussian Likelihood of Continuous-Time ARMAX Models When Data Are Stocks and Flows at Different Frequencies. *Econometric Theory*, 4(1):108–124.

Curriculum Vitae

Lukas Kölbl

Date of Birth	28 September 1986
Place of Birth	Mistelbach
Citizenship	Austrian

Education

11/2013	Research visit at the RSISE, ANU Canberra, Australia
Since 9/2012	Ph.D. student in Technical Mathematics at the Vienna University of Technology
2010-2012	Dipl.-Ing. (MSc) with distinction in Technical Mathematics, with focus on Economic Mathematics Vienna University of Technology
2007-2010	BSc with distinction in Technical Mathematics, with focus on Economic Mathematics Vienna University of Technology
2001-2006	Matura at HTL Donaustadt, Vienna

Professional career

2012-2015	(FWF) Project Assistant at the Institute for Econometrics and System Theory
2012	Tutor at the Institute of Statistics and Mathematical Methods in Economics
2011, 2012	Tutor at the Institute of Numerics and Simulation of Differential Equations
7/2011-8/2011	Internship at Raiffeisenlandesbank Niederösterreich-Wien