Estimating TyEL cash flow with VAR(p) model

Ville Päivinen

Master's thesis

University of Helsinki
Department of Mathematics and Statistics

June 12, 2020

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta/Osasto — Fakultet/Sektion — Faculty	Koulutusohjelma — Utbildningsprogram — Degree programme
Matemaattis-luonnontieteellinen	Matematiikan ja tilastotieteen maisteriohjelma

Tekijä — Författare — Author

Ville Päivinen

Työn nimi — Arbetets titel — Title

Estimating TyEL cash flow with VAR(p) model

Oppiaine — Läroämne — Subject

Matematiikka

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
Pro gradu -tutkielma	Kesäkuu 2020	77 p.

Tiivistelmä — Referat — Abstract

Efficient estimation and forecasting of the cash flow is an interest of pension insurance companies. At the turn of the year 2019 Finnish national Incomes Register was introduced and the payment cycle of TyEL (Employees Pensions Act) changed substantially. TyEL payments are calculated and paid monthly by all of the employers insured under TyEL after January 1st 2019.

Vector autoregressive (VAR) models are one of the most used and successful multivariate time series models. They are widely used with economic and financial data due to the good forecasting abilities and the possibility of analysing dynamic structures between the variables of the model.

The aim of this thesis is to determine whether a VAR model offers a good fit for predicting the incoming TyEL cash flow of a pension insurance company. With the monthly payment cycle arises a question of seasonality of the incoming TyEL cash flow, and thus the focus is on forecasting with seasonally varying data. The essential theory of VAR models is given. The forecast abilities are tested by building a VAR model for monthly, seasonally varying time series similar than the pension insurance companies would have and could use for the particular prediction problem.

Avainsanat — Nyckelord — Keywords

pension insurance, TyEL, forecasting, vector autoregression, VAR

Säilytyspaikka — Förvaringsställe — Where deposited

Kumpulan tiedekirjasto

Muita tietoja — Övriga uppgifter — Additional information

Contents

1 Introduction **3**

2 Background of TyEL insurance and TyEL contribution **5**

3 Mathematical background **8**

4 VAR(p) model **11**
4.1 Definition 11
4.1.1 Basic properties and assumptions 11
4.1.2 Stability condition and autocovariance of VAR(1) process 12
4.1.3 Extending the results to VAR(p) process and Yule-Walker equations 16
4.1.4 MA-representation 18
4.2 Forecasting 18
4.2.1 Point forecasts 19
4.2.2 Interval and region forecasts 21
4.3 Estimation of VAR(p) process 22
4.3.1 Multivariate least squares estimation 22
4.4 Modelling seasonality 27
4.4.1 A VAR representation with time invariant coefficients 28
4.5 Forecasting with estimated process 30

5 Simulation **33**
5.1 Order selection 33
5.1.1 Likelihood ratio tests 33
5.1.2 Information criteria 34
5.2 Checking the model adequacy 35
5.2.1 Residual autocorrelations 36
5.2.2 Multivariate Portmanteau statistics 38
5.3 Choosing parameters for the model and setting up the test data 39

5.3.1 Prediction problem of the pension insurance company 39
5.3.2 The test data 41
5.4 Building the model 43
5.5 Results and possible model improvements 49

6 Alternative models and conclusions **61**
6.1 Alternative models 61
6.2 Conclusions 63

A The data set **67**

B R codes **70**
B.1 Estimation of the VAR models 70
B.2 Normal approximation 72

Chapter 1

Introduction

Pension insurance companies have various reasons to forecast their insurance cash flow. Efficient forecasting of the cash flow helps the pension insurance company to allocate its funds more effectively in order to attain better incomes from its investments, for example. Larger investment incomes could lead to larger surplus, and thus help the company to reach the solvency regulations more easily.

In Finland there are several pension acts depending whether the insured person is working in the local government, state or in private sector. Our interest lies in the Employees Pensions Act, Työntekijän eläkelaki in Finnish and TyEL in short, which is the pension act for private sector workers in Finland. It is operated by 4 pension insurance companies (in 2020) competing against each other. The competition is based on operational efficiency, services and investment returns. Insurance premiums and pension benefits are regulated and cannot be subject to competition. In TyEL, client bonuses are paid once a year and they are based on the pension insurance company's solvency, efficiency and investment operations. Client bonuses are one of the largest subjects to the competition between the pension insurance companies, and an advantage could be gained trough the benefits attained by efficiently estimating the cash flow.

Vector autoregressive (VAR) models have been proven to be powerful and reliable tools for forecasting [19]. They are one of the most used and successful models for analysis of multivariate time series. Together with forecasting abilities VAR models offer the possibility of analysing the dynamic structure between a system of variables. Thus VAR models are often used with economic and financial time series.

The aim of this thesis is to determine whether a VAR model is a good fit for forecasting the incoming TyEL cash flow of a pension insurance company. The payment cycle of TyEL changed radically at the turn of the year 2019 when the Finnish national Incomes Register was introduced. Each TyEL insured employer has been obliged to pay its TyEL contributions monthly after the first of January 2019. Due to the monthly payment cycle some kind of seasonality is assumed to arise within the incoming TyEL cash flow of a pension insurance company. It is of interest whether the VAR models are able to capture the assumed seasonality, and thus we will be focusing on forecast abilities of VAR models with seasonally varying data.

In Chapter 2 a little bit more throughout, but brief backgrounds of TyEL insurance, TyEL contribution and the Incomes Register are given. The reader of the thesis is assumed to be familiar with probability theory and basic linear algebra. Some mathematical background is given in Chapter 3. The essential theory and basics of VAR models are given in Chapter 4. We are not going in the economic details behind the TyEL contribution and its possible seasonal variations or neither the theory of analysing the dynamic structures between the variables of the VAR model. The focus of the theory will be solely on forecasting.

In Chapter 5 we will build a VAR model with a seasonally varying data set and test its forecast abilities. The model building is supported by the theory of Chapter 4 and the first two subsections of Chapter 5. In Chapter 6 some suggestions for alternative models for forecasting are given and the results are compared to the ones attained with VAR models together with the conclusions of the thesis.

Chapter 2

Background of TyEL insurance and TyEL contribution

A brief overview of TyEL insurance

The earnings-related pension system in Finland was set up at the turn of the 1950s and 1960s. The first group to receive its own pension act were seafarers in 1956. The general Employees Pension Act TEL was set up in 1962, and it has been an integral part of Finnish social security system ever since. The Employees Pension Act TyEL in its current form was introduced in 2007, when TEL, Temporary Employees' Pensions Act LEL and Pensions Act for Performing Artists and Certain Groups of Employees TaEL were merged into one.

TyEL insurance is an earnings-related statutory insurance by which the employers in private sector ensure the pension cover of their employees. Every employee must be insured under TyEL by the employer, if the work is performed under an employment contract, the employee is between 17 and 67 years old and the monthly salary paid for the worker exceeds 60,56 euros. According to the statistics provided by The Finnish Pension Alliance TELA and Finnish Centre for Pensions, there were 1 659 114 persons insured under TyEL in 2018 [6]. They generated nearly 14 billions euros in insurance cash flow. Under all pension acts there were in total 2 664 746 insured persons, generating a cash flow of 22,2 billion euros in 2018. Measured in either numbers of the insured, or by the cash flow, TyEL is by far the largest pension act in Finland. The second largest pension act is Public Sector Pension Act JuEL with 535 000 insured persons and a cash flow of 5 billion euros in 2018.

TyEL contribution

TyEL contribution is an earnings-related payment, paid by both the employee and the employer. The contribution is paid in total to the pension insurance company by the employer, and the part of the employee is deduced from his or hers wages or salary directly. Thus the pension insurance companies get the whole cash flow generated by their TyEL insurance portfolios.

The basic TyEL contribution rate is confirmed for each year by the Finnish Ministry of Social Affairs and Health. The employers under TyEL receive various discounts to their TyEL contribution. Client bonuses and discounts for large payrolls are paid once every year. Every employer insured under TyEL is categorized either as a large or small employer for each year by the size of their payroll of the year before the last. Large employers receive a constant contribution loss discount as a percentage of their payrolls, and their history of disability pensions affects the size of the disability part component of the TyEL contribution trough the so-called contribution category. There are 11 contribution categories, with category 4 as a base category having no affect to the TyEL payment. Categories 1-3 reduce the TyEL payment, and categories 5-11 respectively increase the payment. The TyEL contribution of a single employer can be calculated as a difference between the basic TyEL contribution and all of the discounts received by the employer.

The exact actuarial principles of TyEL contribution can be found from `http://www.saadospalvelu.fi/fi/perusteet/index`, available in Finnish and Swedish.

Payment cycle and the Incomes Register

The Incomes Register is a national electronic database which includes all the information related to salaries and wages, pensions and benefits of individual citizens in Finland. It was introduced at the turn of the year 2019, and every payroll notification made after the first of January 2019 has to be made in real time and per payment to the Incomes Register by employers. Several operators, including pension insurance companies, receive payroll related informations from the Incomes Register.

As a payroll notification has been made to the Incomes Register, pension insurance companies will receive the needed information with a delay of couple of hours. Then, the TyEL contribution is calculated nearly instantly. This have had a significant effect on the payment cycle of TyEL payments. Before the Incomes Register employers made separate payroll notifications for the pension insurance companies, either once a year or monthly. An estimate for the current year's TyEL contribution was calculated and charged from the employer, and the final TyEL contribution was determined at the beginning of the next year. Employer could choose with how many advance payments the estimated TyEL contribution was paid, and after determining the final TyEL contribution the difference was either paid or charged from the employer by the pension insurance company.

After the introduction of the Incomes Register TyEL contributions are instantly calculated as final and charged monthly from the employers. The monthly TyEL payment of an employer consists of all TyEL contributions calculated from employer's payroll notifications made for the previous month. With small employers the monthly TyEL payment is the basic TyEL contribution, and with large employers the monthly TyEL payment includes the contribution loss discount and affect of the contribution category. Client bonuses and discounts for large payrolls are calculated once a year, and paid separately for the employers. Thus they do not affect the level of the monthly TyEL payments.

Chapter 3

Mathematical background

Definition 3.1. Let y_t be a stochastic process. If

$$\mathbb{E}(y_t) = \mu \quad \text{and} \quad \mathbb{E}\left[(y_t - \mu)(y_{t-h} - \mu)'\right] = \Gamma_y(h) = \Gamma_y(-h)$$

for all t and $h = 0, 1, 2, \ldots$ then the stochastic process y_t is (weakly) stationary.

Definition 3.2. Let $u_t = (u_{1t}, \ldots, u_{Nt})'$ be an N-dimensional stochastic process. If

1. $\mathbb{E}(u_t) = 0$ and $\mathbb{E}(u_t u_t') = \Sigma_u < \infty$ all t
2. $\mathbb{E}(u_t u_s') = 0$ all $t \neq s$,

u_t is called a (weak) *white noise* (process).

Definition 3.3. Let u_t be a white noise as in Definition 3.2. If u_t and u_s are independent for all $s \neq t$, u_t is called *independent* white noise.

Definition 3.4. Let u_t be an independent white noise as in Definition 3.3. If, in addition,

1. $\Sigma_u = \mathbb{E}(u_t u_t')$ is nonsingular
2. The fourth moment of u_t exists and is finite,

u_t is called a *standard* white noise.

Definition 3.5. Let A be an $(m \times n)$ matrix. The vectorization of the matrix A is the $(mn \times 1)$ column vector obtained by stacking the columns of the matrix A on top of each other and denoted by $\text{vec}(A)$. More precisely,

$$\text{vec}(A) = (a_{11}, \ldots, a_{m1}, a_{1,2}, \ldots, a_{m2}, \ldots, a_{1n}, \ldots, a_{mn})',$$

where a_{ij} represents the j'th entry of the row i of the matrix A.

Definition 3.6. Let A be an $(m \times n)$ matrix and B a $(p \times q)$ matrix. The *Kroenecker product* $A \otimes B$ is the $(pm \times qn)$ block matrix

$$A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix},$$

where a_{ij} represents the j'th entry of the row i of the matrix A.

Definition 3.7. Let $A = (a_{ij}) \in \mathbb{R}^{m \times n}$ be an $(m \times n)$ matrix. The *Frobenius norm* of the matrix A is denoted by $\|A\|_F$ and defined by

$$\|A\|_F := \left(\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2 \right)^{1/2}$$

Definition 3.8. Let $\{a_i\}$, $i = 0, \pm 1, \pm 2, \ldots$ be a doubly infinite sequence of real numbers. If

$$\lim_{n \to \infty} \sum_{i=-n}^{n} |a_i|$$

exists and is finite we say that the sequence $\{a_i\}$ is *absolutely summable.*

Definition 3.9. Let $\{A_i = (a_{mn,i}) \in \mathbb{R}^{N \times N}\}$, $i = 0, \pm 1, \pm 2, \ldots$ be a doubly infinite sequence of $(N \times N)$ matrices. If each sequence of real numbers

$$\{a_{mn,i}\}, \quad m, n = 1, \ldots, N, \quad i = 0, \pm 1, \pm 2, \ldots$$

is absolutely summable in the sense of Definition 3.8, we say that the sequence $\{A_i\}$ is absolutely summable.

Proposition 3.10. *Let $\{A_i\}$, $i = 0, \pm 1, \pm 2, \ldots$, be a sequence of real valued $(N \times N)$ matrices and $\{z_t\}$ a sequence of N-dimensional random variables such that*

$$\mathbb{E}(z_t' z_t) \leq c, \quad t \in \mathbb{Z},$$

where $c < \infty$ is a constant. Assume that $\{A_i\}$ is absolutely summable in the sense of Definition 3.9.

Then there exists a sequence of N-dimensional random variables $\{y_t\}$ such that

$$\sum_{i=-n}^{n} A_i z_{t-i} \to y_t,$$

in mean square as $n \to \infty$

Proof. See [8] pp. 29-31. Result follows by replacing absolute value by Frobenius norm. □

Definition 3.11. Let $M_{m,n}$ be the collection of all matrices A of order $m \times n$. A norm $\|\cdot\|$ on $M_{n,n}$ is said to be a *matrix norm* on $M_{n,n}$ if

$$\|AB\| \leq \|A\|\|B\| \tag{3.12}$$

for all $A, B \in M_{n,n}$.

Chapter 4

VAR(p) model

In this Chapter we will define the basic stationary finite order vector autoregressive (VAR) process. In Sections 4.1 and 4.2 we will assume that the process of interest is completely known. Estimation of the VAR process and modelling seasonality are dealt in Sections 4.3 and 4.4, respectively. In Section 4.5, we focus on the theory of forecasting using the estimated process.

4.1 Definition

4.1.1 Basic properties and assumptions

A multivariate time series is a data set of vector valued observations indexed by time points. That is

$$y_t = (y_{1t}, \ldots, y_{Nt}), \quad t = 1, \ldots, T, \quad N \in \{2, 3, \ldots\}$$

where t is the time point of a multivariate time series with N variables and T observations of the values of those variables. A very simple example of a multivariate time series is, say one hundred observations ($T = 100$) of the temperature and humidity ($N = 2$) of a room, or a data set of hourly development of one hundred chosen stock prices ($N = 100$) in New York Stock Exchange during a day ($T = 24$). Main difference to a univariate time series is that we are interested of the values of more than one variable varying over time, which is why N is required to be at least two, and dynamic interactions between the variables are allowed.

An observed multivariate time series can be interpreted as a realization of a stochastic process. Let $y_t = (y_{1t}, \ldots, y_{Nt})$ be a random vector. The set $\{y_1, \ldots, y_T\}$ contains the realized values of the stochastic process from time point 1 to the time point T. When we are talking about the process $\{y_t\}$, or with even shorter notation the process y_t, we will be talking about this underlying stochastic process behind the observed data. This stochastic process is sometimes called as the *data generation process* (DGP) of the observed time series. It will be always stated clearly whether we are talking about the process y_t or the random vector y_t, when necessary.

A *model* is used to describe the phenomena behind the observed time series. *Vector autoregression* is one way to model the phenomena. The idea of vector autoregression is that the values of the process depend linearly on the past values of itself and from an unobservable error term.

Definition 4.1. Let $y_t = (y_{1t}, \ldots, y_{Nt})'$ be a $(N \times 1)$ random vector. *Vector autoregressive process of order* p (VAR(p) process) is a stochastic process such that

$$y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t, \quad t \in \mathbb{Z}, \tag{4.2}$$

where the A_i are fixed $(N \times N)$ coefficient matrices, $\nu = (\nu_1, \ldots, \nu_N)'$ is a fixed $(N \times 1)$ vector of intercept and $u_t = (u_{1t}, \ldots, u_{Nt})'$ is a N-dimensional *white noise* process as in Definition 3.2.

Remark. The vector of intercept ν allows the possibility of the process having a non-zero mean. The unobservable error term u_t is commonly referred as an *innovation* or a *shock* term.

In VAR(p) model the value of the process at time point t depends on its p previous values $y_{t-1}, \ldots, y_{t-p}$ and on the value of the error term u_t. We will start investigating the model further from VAR(1) processes. It is later seen that all the results considering VAR process of order one are easily generalized for processes of finite order $p > 1$.

In what follows we assume that the coefficient matrices A_i and the vector of intercept ν are known, if not otherwise stated. Equation (4.2) is referred as the data generation process of the VAR(p) process.

4.1.2 Stability condition and autocovariance of VAR(1) process

In VAR(1) model the value of the process depends of the previous value of the process, y_{t-1}, and of the value of error term u_t at time t. That is, by Definition 4.1

$$y_t = \nu + A_1 y_{t-1} + u_t, \tag{4.3}$$

where A_1 and ν are known constants. Assuming that this data generation method of the process y_t has started some time $t = 1$, we get the following equations

$$\begin{aligned}
y_1 &= \nu + A_0 y_0 + u_1 \\
y_2 &= \nu + A_1 y_1 + u_2 = \nu + A_1(\nu + A_0 y_0 + u_1) + u_2 \\
&= (I_N + A_1)\nu + A_1^2 y_0 + A_1 u_1 + u_2 \\
&\vdots \\
y_t &= (I_N + A_1 + \cdots + A_1^{t-1})\nu + A_1^t y_0 + \sum_{i=0}^{t-1} A_1^i u_{t-i} \\
&\vdots
\end{aligned}$$

where I_N is a $(N \times N)$ identity matrix. Since it is assumed that A_1 and ν are known constants, we see that the vectors $(y_0, \ldots, y_t)$ are uniquely determined by $y_0, u_1, \ldots, u_t$. Respectively the joint distribution of $y_0, \ldots, y_t$ is determined by the joint distribution of $y_0, u_1, \ldots, u_t$.

Without assuming a specified starting period of the process, we have the general form

$$\begin{aligned}
y_t &= \nu + A_1 y_{t-1} + u_t \\
&= (I_N + A_1 + \cdots + A_1^j)\nu + A_1^{j+1} y_{t-j-1} + \sum_{i=0}^{j} A_1^i u_{t-i},
\end{aligned} \tag{4.4}$$

where y_{t-j-1} is the initial value of the process starting at some time in infinite past (compare this to the y_0 in process starting at some given time $t = 1$). It is worth noting that Definition 4.1 makes no assumption of the process starting at some given time. The following result is needed to consider the process of (4.4) in the limit $j \to \infty$.

Theorem 4.5. *Suppose A is an $(n \times n)$ matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$ which are all strictly less than 1 in absolute value, that is $|\lambda_i| < 1$ all $i = 1, 2, \ldots, n$.*

Then

(i) $A^j \to 0$ *as* $j \to \infty$

(ii) The sequence $\{A^j\}$ *is absolutely summable*

(iii) $\sum_{j=0}^{\infty} A^j = (I_n - A)^{-1}$ *exists*

Proof. Define $\rho(A) = \max_{1\leq i\leq n} |\lambda_i|$ as the *spectral radius* of A. Given A and $\epsilon > 0$ it can be seen that there exists a matrix norm as defined in Definition 3.11 such that

$$\rho(A) \leq \|A\| < \rho(A) + \epsilon. \tag{4.6}$$

Suppose $\rho(A) < 1$. Now by (4.6) there exists a matrix norm $\|\cdot\|$ such that $\|A\| < 1$. By Definition 3.11 of the matrix norm $\|A^k\| \leq (\|A\|)^k$ which implies *(i)*.

For *(ii)*, let $\{a_j\}$ be a sequence of scalars. Now the series $\sum_{j=0}^{\infty} a_j A^j$ converges if the series $\sum_{j=0}^{\infty} |a_j| \|A\|^j$ converges. Taking $a_j = 1$ all j the latter series is a geometric series, which converges when $\|A\| < 1$.

For *(iii)* note that $(I_n - A)\sum_{i=0}^{m} A^i = I_n - A^{(m+1)}$, thus by letting $m \to \infty$ we have $(I_n - A)\sum_{i=0}^{\infty} A^i = I_n$ which gives the result.

For the proof of (4.6) see Chapter 11 of [18]. □

Assume that the coefficient matrix A_1 of VAR(1) process satisfies the conditions of Theorem 4.5. Then, by Theorem 4.5 the sequence $\{A_1^i\}$ is absolutely summable. Since u_t is a white noise process it has finite variance, and by Proposition 3.10 the sum

$$\sum_{i=0}^{j} A_1^i u_{t-i}$$

converges in mean square as $j \to \infty$. Furthermore by Theorem 4.5

$$(I_N + A_1 + \cdots + A_1^j)\nu \to (I_N - A_1)^{-1}\nu,$$
$$A_1^{j+1} y_{t-j-1} \to 0$$

as $j \to \infty$. Hence, if A_1 satisfies Theorem 4.5, that is all the eigenvalues of of A_1 have modulus strictly less than 1, VAR(1) process defined in (4.3) can be written as a well-defined stochastic process such that

$$y_t = \mu + \sum_{i=0}^{\infty} A_1^i u_{t-i}, \quad t \in \mathbb{Z}, \tag{4.7}$$

where $\mu := (I_N - A_1)^{-1}\nu$. Distributions and joint distributions of the y_t are uniquely determined by the distributions of the u_t. Expectation and autocovariance of the process are easily deduced from Equation (4.7). Since $\mathbb{E}(u_t) = 0$ for all t, we see that $\mathbb{E}(y_t) = \mu$ for all t. Autocovariance of the process is determined by

$$\begin{aligned}\Gamma_y(h) &:= \mathbb{E}\left[(y_t - \mu)(y_{t-h} - \mu)'\right] \\ &= \lim_{n\to\infty} \sum_{i=0}^{n}\sum_{j=0}^{n} A_1^i \mathbb{E}(u_{t-i}u'_{t-h-j})(A_1^j)' \\ &= \lim_{n\to\infty} \sum_{i=0}^{n} A_1^{h+i}\Sigma_u(A_1^i)' \\ &= \sum_{i=0}^{\infty} A_1^{h+i}\Sigma_u(A_1^i)',\end{aligned}$$

since $\mathbb{E}(u_t u'_s) = 0$ for all $s \neq t$ and $\mathbb{E}(u_t u'_t) = \Sigma_u$ for all t by Definition 3.2.

VAR(1) process satisfying Theorem 4.5 is called a *stable* process, and the condition regarding the absolute values of the eigenvalues the *stability condition.* Since eigenvalues are roots of the of the characteristic polynomial of the matrix, the condition of all the eigenvalues of A_1 being strictly less then 1 is equivalent to

$$det(I_N - A_1 z) \neq 0 \quad \text{for} \quad |z| \leq 1, \quad z \in \mathbb{C}. \tag{4.8}$$

Example 4.9. Assume that we have a time series of the temperature and humidity of a room with one hundred observations,

$$y_t = (y_{t,temp}, y_{t,hum}), \quad t = 1, \ldots, 100, \quad N = 2$$

where $y_{t,temp}$ corresponds to the observed temperature of the room at time point t and $y_{t,hum}$ to the observed humidity at time point t. The time series is known to be generated by a VAR(1) process. That is

$$\begin{aligned} y_t = \nu + A_1 y_{t-1} + u_t \iff & \begin{bmatrix} y_{t,temp} \\ y_{t,hum}\end{bmatrix} = \begin{bmatrix} \nu_{temp} \\ \nu_{hum}\end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22}\end{bmatrix}\begin{bmatrix} y_{t-1,temp} \\ y_{t-1,hum}\end{bmatrix} + \begin{bmatrix} u_{t-1,temp} \\ u_{t-1,hum}\end{bmatrix} \\ \iff & \begin{bmatrix} y_{t,temp} \\ y_{t,hum}\end{bmatrix} = \begin{bmatrix} \nu_{temp} \\ \nu_{hum}\end{bmatrix} + \begin{bmatrix} a_{11}y_{t-1,temp} + a_{12}y_{t-1,hum} \\ a_{21}y_{t-1,temp} + a_{22}y_{t-1,hum}\end{bmatrix} + \begin{bmatrix} u_{t-1,temp} \\ u_{t-1,hum}\end{bmatrix},\end{aligned}$$

where u_t is the 2-dimensional unobservable error term as in Definition 4.1.

If the mean temperature and humidity of the room are positive and differ from zero, then $\nu_{temp}, \nu_{hum} > 0$. Entries of the coefficient matrix A_1 represent the dynamic relations between the two variables.

4.1.3 Extending the results to VAR(p) process and Yule-Walker equations

Extending the results from VAR(1) process to a general VAR(p) process with $p > 1$ turns out to be rather straightforward, since any VAR(p) process can be written in a form of a VAR(1) process. Assuming that y_t is a VAR(p) process with $p > 1$ as in Definition 4.1 we can define

$$Y_t = \boldsymbol{\nu} + \mathbf{A}Y_{t-1} + U_t, \tag{4.10}$$

where

$$Y_t := \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix}, \quad \boldsymbol{\nu} := \begin{bmatrix} \nu \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{A} := \begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I_N & 0 & \cdots & 0 & 0 \\ 0 & I_N & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I_N & 0 \end{bmatrix}, \quad U_t := \begin{bmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Now $\mathbf{A}$ is a $(Np \times Np)$ matrix and $\boldsymbol{\nu}, Y_t$ and U_t are $(Np \times 1)$ matrices. It is easily seen by basic rules of matrix calculation that equations (4.2) and (4.10) correspond to each other in the sense that (4.2) is true if and only if (4.10) is true.

Definition of a stable VAR(p) process follows from replacing the coefficient matrix A_1 of VAR(1) process by the coefficient matrix $\mathbf{A}$ of representation (4.10). If all the eigenvalues of $\mathbf{A}$ are strictly less than 1, or equivalently

$$det(I_{Np} - \mathbf{A}z) \neq 0 \quad \text{for} \quad |z| \leq 1, \quad z \in \mathbb{C}, \tag{4.11}$$

then the VAR(p) process is called stable.

The determinant in (4.11) is not easily calculated in practice. This motivates the following formulation, which is obtained from (4.11) via determinant properties of block matrices, and is usually more efficient in practice.

Definition 4.12. Let y_t be a VAR(p) process as in Definition 4.1. Suppose $p > 1$. We call the process y_t *stable*, if

$$det(I_N - A_1 z - \cdots - A_p z^p) \neq 0 \quad \text{for} \quad |z| \leq 1, \quad z \in \mathbb{C}. \tag{4.13}$$

Assuming that the process y_t is stable, mean and auto-covariance can derived with similar arguments as for VAR(1) process previously. That is, the process has a mean-vector given by

$$\boldsymbol{\mu} := \mathbb{E}(Y_t) = (I_{Np} - \mathbf{A})^{-1}\boldsymbol{\nu} \tag{4.14}$$

and the auto-covariances are given by

$$\Gamma_Y(h) = \sum_{i=0}^{\infty} \mathbf{A}^{h+i}\Sigma_u(\mathbf{A}^i)'. \tag{4.15}$$

Auto-covariances of the process are often presented using the *Yule-Walker equations.* For this, we make the following observation.

Proposition 4.16. *A stable VAR(p) process is stationary.*

Proof. Properties of white noise process u_t defined in Definition 3.2 and linear representation of the VAR(p) process imply that a stable VAR(p) process is stationary in the sense of Definition 3.1. □

Without loss of generality we consider the mean-adjusted form of the VAR(p) process y_t.

Definition 4.17. Let y_t be a VAR(p) process as in Definition 4.1. We call

$$y_t - \mu = A_1(y_{t-1} - \mu) + \ldots + A_p(y_{t-p} - \mu) + u_t \tag{4.18}$$

the *mean-adjusted form* of the VAR(p) process.

Assuming that y_t is stable, the Yule-Walker equations are obtained by multiplying (4.18) from the right with $(y_{t-h} - \mu)'$ and taking expectation. By Proposition 4.16 $\Gamma_y(i) = \Gamma_y(-i)'$. Since $\mathbb{E}(u_t y_t') = \mathbb{E}(u_t u_t') = \Sigma_u$ and $\mathbb{E}(u_t y_{t+h}') = 0$ for $h > 0$, we have

$$\begin{aligned}\Gamma_y(0) &= A_1\Gamma_y(-1) + \cdots + A_p\Gamma_y(-p) + \Sigma_u \\ &= A_1\Gamma_y(1)' + \cdots + A_p\Gamma_y(p)' + \Sigma_u\end{aligned}$$

for $h = 0$ and for $h > 0$

$$\Gamma_y(h) = A_1\Gamma_y(h-1) + \cdots + A_p\Gamma_y(h-p).$$

These Yule-Walker equations may be used to compute the auto-covariance $\Gamma_y(h)$ recursively for $h \geq p$, when the coefficient matrices $A_1, \ldots, A_p$ and past auto-covariances $\Gamma_y(p-1), \ldots, \Gamma_y(0)$ are known.

It is worth of noting that all the previous results rely on the process y_t being stable. VAR(p) processes can be defined without the stability condition being satisfied. In fact, seasonally varying time series are an example of unstable processes. Though it is possible and rather simple to adjust seasonally varying time series to be stable, i.e to satisfy the stability condition. From this on without stating otherwise, we will assume that the process y_t satisfies the stability condition. Topic of modelling seasonality is handled later on the Chapter 4.4.

4.1.4 MA-representation

Considering the VAR(1) representation (4.10) of VAR(p) process, assuming that the process is stable, Y_t has a representation

$$Y_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \mathbf{A}^i U_{t-i}, \tag{4.19}$$

where $\boldsymbol{\mu} = (I_{Np} - \mathbf{A})^{-1}\boldsymbol{\nu}$. This representation in (4.19) is called the *moving avarage* (MA) representation of the VAR(p) process. MA representation of y_t can be found by multiplying (4.19) from the right with a $(N \times Np)$ matrix $J := [I_N : 0 : \cdots : 0]$. That is,

$$\begin{aligned} y_t = JY_t &= J\boldsymbol{\mu} + \sum_{i=0}^{\infty} J\mathbf{A}^i J' J U_{t-i} \\ &= \mu + \sum_{i=0}^{\infty} \phi_i u_{t-i}, \end{aligned} \tag{4.20}$$

where $\phi_i = J\mathbf{A}^i J'$ and $\mu = J\boldsymbol{\mu}$. This representation of y_t in (4.20) is often referred as the *canonical* MA representation of the VAR(p) process. Because the $\mathbf{A}^i$ are absolutely summable, so are the ϕ_i. Thus the equation (4.20) is well-defined.

4.2 Forecasting

As we have stated in the introduction of the thesis, forecasting with VAR models is the main interest of our study. Point forecasts and interval forecasts will be discussed in turn.

Generally speaking, the prediction problem of a forecaster goes as it follows: A forecaster needs to make statements about future values of variables $y_1, \ldots, y_N$ and has available a model for the data generation process and an information set Ω_t containing the available information at time t, that is Ω_t is a sigma-algebra generated by the random vectors $y_s = (y_{1s}, \ldots, y_{Ns})$, where $s \leq t$.

Time t when the forecasts are made is called the *forecast origin*, the number of periods into the future for which the forecast is made is called *forecast horizon* and the predictor h periods ahead a *h-step predictor*.

If forecasts are desired for a particular purpose, a specific cost function may be associated with the forecast errors. Forecast will be optimal, if it minimizes this associated cost. Of course, in practice one has to almost always consider the expected costs. In case of VAR models predictors that minimize the mean squared errors (MSEs) are the most widely used ones. Arguments in favor of using the MSE as a cost function can be found from [9] and [10]. For the clarity we will give formal definitions of mean square error and its minimization. Let $y_t = (y_{1t}, \ldots, y_{Nt})'$ be a N-dimensional stable VAR(p) process as in Definition 4.1.

Definition 4.21. Let $\bar{y}_t(h)$ be any h-step predictor of the process y_t. The mean square error (MSE) of $\bar{y}_t(h)$ is defined and denoted as

$$MSE\left[\bar{y}_t(h)\right] = \mathbb{E}\left[(y_{t+h} - \bar{y}_t(h))(y_{t+h} - \bar{y}_t(h))'\right]. \tag{4.22}$$

Definition 4.23. Let $\bar{y}_t(h)$ and $\tilde{y}_t(h)$ be h-step predictors at forecast origin t. If

$$MSE\left[\bar{y}_t(h)\right] \geq MSE\left[\tilde{y}_t(h)\right] \tag{4.24}$$

for all h-step predictors $\bar{y}_t(h)$, then $\tilde{y}_t(h)$ is called *the minimum mean square error predictor* for forecast horizon h at the forecast origin t.

Remark. The inequality sign in (4.24) means that the difference between the left- and right-hand side matrices is positive semi-definite.

In what follows, if not otherwise stated, we will assume that the information set available to the forecaster consists of the previous and present values of the process y_t.

Definition 4.25. Let t be the forecast origin. The sigma-algebra

$$\Omega_t \coloneqq \{y_s | s \leq t\} \tag{4.26}$$

is called the information set of the process y_t.

4.2.1 Point forecasts

It is a well-known fact from probability theory that conditional expectation of y_{t+h} conditioned on the information set Ω_t minimizes the mean square error of the h-step predictor $\bar{y}_t(h)$. That is,

$$MSE\left[\bar{y}_t(h)\right] \geq MSE\left[\mathbb{E}\left(y_{t+h}|\Omega_t\right)\right] \tag{4.27}$$

for all h-step predictors $\bar{y}_t(h)$. To shorten the notation we denote

$$\mathbb{E}_t\left(y_{t+h}\right) \coloneqq \mathbb{E}\left(y_{t+h}|\Omega_t\right)$$

and refer the conditional expectation $\mathbb{E}_t\left(y_{t+h}\right)$ as the optimal predictor meaning that it is optimal in the sense of Definition 4.23.

Optimality of the conditional expectation implies the following theorem.

Theorem 4.28. *Let $y_t = (y_{1t}, \dots, y_{Nt})'$ be a N-dimensional stable VAR(p) process where u_t is an independent white noise process as in Definition 3.3. Then*

$$\mathbb{E}_t\left(y_{t+h}\right) = \nu + A_1\mathbb{E}_t\left(y_{t+h-1}\right) + \cdots + A_p\mathbb{E}_t\left(y_{t+h-p}\right) \tag{4.29}$$

is the optimal h-step predictor of VAR(p)-process y_t.

From (4.29) we can calculate the h-step predictors recursively. Starting from $h = 1$ we get

$$\begin{aligned}
\mathbb{E}_t(y_{t+1}) &= \nu + A_1 y_t + \cdots + A_p y_{t-p+1} \\
\mathbb{E}_t(y_{t+2}) &= \nu + A_1\mathbb{E}_t(y_{t+1}) + A_2 y_t + \cdots + A_p y_{t-p+2} \\
&\vdots \\
\mathbb{E}_t(y_{t+h}) &= \nu + A_1\mathbb{E}_t(y_{t+h-1}) + \cdots + A_{h-1}\mathbb{E}_t(y_{t+1}) + A_h y_t + \cdots + A_p y_{t-p+h}
\end{aligned}$$

and so on.

It must be noted that the prediction formula in (4.29) and the recursive equations obtained via it rely on u_t being an independent white noise. If u_t is not an independent white noise, $\mathbb{E}_t(u_{t+h})$ will be non-zero in general and additional assumptions are usually required to obtain the conditional expectation of a VAR(p) process.

Indeed, assuming that u_t is not an independent white noise and without making any additional assumptions about the distribution of u_t, a less ambitious goal of finding the minimum MSE predictors among those that are linear functions of $y_t, y_{t-1}, \dots$ can be achieved. Letting $y_t(h) = B_0 y_t + B_1 y_{t-1} + \cdots$, where the B_i are $(N \times N)$ coefficient matrices, be any h-step predictor it can be seen that

$$y_t(h) = \nu + A_1 y_t(h-1) + \cdots + A_p y_t(h-p), \tag{4.30}$$

where $y_h(h-j) = y_{t+h-j}$, when $h \leq j$, is the optimal linear predictor for a VAR(p) process. The forecast error can be obtained by

$$y_{t+h} - y_t(h) = \sum_{i=0}^{h-1} \phi_i u_{t+h-i}, \tag{4.31}$$

where the ϕ_i are the coefficient matrices of the canonical MA representation of VAR(p) process y_t as in (4.20).

For the details of obtaining these results we refer the reader for [14], Section 2.2.

4.2.2 Interval and region forecasts

In order to set up interval forecasts or forecast intervals, it is necessary to make an assumption about the distribution of y_t. The most common one is to consider *Gaussian processes* where $y_t, y_{t+1}, \ldots, y_{t+h}$ follow a multivariate normal distribution for any t and h. An equivalent assumption is that the error terms u_t have a multivariate normal distribution with covariance matrix Σ_u and zero mean, with u_t and u_s being independent for $s \neq t$. If either of these equivalent assumptions is made, we say that y_t is a Gaussian process.

Let $y_t(h)$ be an h-step predictor. Assuming that y_t is Gaussian process and the covariance matrix Σ_u of the white noise u_t is positive definite, the forecast errors $y_{t+h} - y_t(h)$ are also normally distributed as linear transforms of normal vectors. That is,

$$y_{t+h} - y_t(h) \sim \mathcal{N}\left(0, \Sigma_y\left(h\right)\right).$$

Furthermore, since individual components of a normally distributed random vector are normally distributed, we have that

$$\frac{y_{k,t+h} - y_{k,t}\left(h\right)}{\sigma_k\left(h\right)} \sim \mathcal{N}\left(0,1\right),$$

where $y_{k,t}\left(h\right)$ is the k-th component of $y_t\left(h\right)$ and $\sigma_k\left(h\right)$ is the standard deviation of $y_{k,t+h} - y_{k,t}\left(h\right)$. Let z_α, $0 < \alpha < 1$, be the upper $100(1-\alpha)$ percentage point of the normal distribution. More precisely, given $Z \sim \mathcal{N}\left(0,1\right)$ z_α is such that $\mathbb{P}\left(Z > z_\alpha\right) = \alpha$. Then

$$\begin{aligned} 1 - \alpha &= \mathbb{P}\left\{-z_{\alpha/2} \leq \frac{y_{k,t+h} - y_{k,t}\left(h\right)}{\sigma_k\left(h\right)} \leq z_{\alpha/2}\right\} \\ &= \mathbb{P}\left\{y_{k,t}\left(h\right) - z_{\alpha/2}\sigma_k\left(h\right) \leq y_{k,t+h} \leq y_{k,t}\left(h\right) + z_{\alpha/2}\sigma_k\left(h\right)\right\}. \end{aligned}$$

Thus, a $100(1-\alpha)$ percentage interval forecast h-periods ahead, for the k-th component of y_t is

$$\left[y_{k,t}\left(h\right) - z_{\alpha/2}\sigma_k\left(h\right), y_{k,t}\left(h\right) + z_{\alpha/2}\sigma_k\left(h\right)\right].$$

When dealing with multivariate time series, it might of course be an object of interest to define a region that has at least the given probability of containing all of the N variables of the model under consideration. For this purpose, the so-called *Bonferroni's method* comes quite handy, especially if N is large.

The method is based on a fact that for any events $E_1, \ldots, E_N$

$$\mathbb{P}\left(E_1 \cup \cdots \cup E_N\right) \leq \mathbb{P}\left(E_1\right) + \cdots + \mathbb{P}\left(E_N\right).$$

Thus

$$\mathbb{P}\left(\bigcap_{i=1}^{N} E_i\right) \geq 1 - \sum_{i=1}^{N} \mathbb{P}\left(E_i^{\complement}\right), \tag{4.32}$$

for any events $E_1, \ldots, E_N$ by the basic rules of probability theory. Here $E_i^{\complement}$ denotes the complement of the set E_i.

Define a $(N \times K)$ matrix $F := [I_N : 0]$ and let E_i be the event that the component $y_{i,t+h}$ falls within an interval H_i. Then by (4.32),

$$\mathbb{P}\left(F y_{t+h} \in H_1 \times \cdots \times H_N\right) \geq 1 - \sum_{i=1}^{N} \mathbb{P}\left(E_i^{\complement}\right).$$

In other words, by choosing a $100\left(1 - \frac{\alpha}{N}\right)$ percent forecast interval for each of the N components of y_{t+h}, the resulting (joint) forecast region has at least a probability of $(1-\alpha)$ of containing all N variables jointly.

4.3 Estimation of VAR(p) process

Assume that a N-dimensional multiple time series $\{y_1, \ldots, y_T\}$ with $y_t = (y_{1t}, \ldots, y_{Nt})'$ is available, and it is known to be generated by a stable VAR(p) process

$$y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t, \tag{4.33}$$

with the coefficients ν and A_i's defined as in Definition 4.1 and u_t being white noise with positive definite covariance matrix Σ_u. Contrary to Sections 4.1 and 4.2, the coefficients $\nu, A_1, \ldots, A_p$ and Σ_u are assumed to be unknown, and the time series data will be used to estimate them. Generally speaking, our goal is to estimate the unknown parameters to best fit the observed data set. For this purpose, a method of least squares, more precisely a method of general least squares (GLS) is considered.

4.3.1 Multivariate least squares estimation

In addition to the T observed values for each of the N variables, we assume that p presample values for each variable, that is $y_{-p+1}, \ldots, y_0$, are available. This is convenient in order to simplify the notation.

Define

$$
\begin{aligned}
Y &:= (y_1, \dots, y_T) && (N \times T), \\
B &:= (\nu, A_1, \dots, A_p) && (N \times (Np+1)), \\
Z_t &:= \begin{bmatrix} 1 \\ y_t \\ \vdots \\ y_{t-p+1} \end{bmatrix} && ((Np+1) \times 1), \\
Z &:= (Z_0, \dots, Z_{T-1}) && ((Np+1) \times T), \\
U &:= (u_1, \dots, u_T) && (N \times T), \\
\boldsymbol{y} &:= \operatorname{vec}(Y) && (NT \times 1), \\
\boldsymbol{\beta} &:= \operatorname{vec}(B) && \big((N^2p+N) \times 1\big), \\
\boldsymbol{b} &:= \operatorname{vec}(B') && \big((N^2p+N) \times 1\big), \\
\boldsymbol{u} &:= \operatorname{vec}(U) && (NT \times 1),
\end{aligned}
\tag{4.34}
$$

where on the right side are the dimensions of the defined matrices and vec is the column stacking operator as in Definition 3.5.

Using this notation the VAR(p) model in (4.33) can be written as

$$Y = BZ + U \tag{4.35}$$

or

$$\operatorname{vec}(Y) = \operatorname{vec}(BZ) + \operatorname{vec}(U) \tag{4.36}$$

$$= (Z' \otimes I_N) \operatorname{vec}(B) + \operatorname{vec}(U) \tag{4.37}$$

which is equivalent to

$$\boldsymbol{y} = (Z' \otimes I_N)\, \boldsymbol{\beta} + \boldsymbol{u}. \tag{4.38}$$

Here $\otimes$ is the Kronecker product as in Definition 3.6. Now, our aim is to estimate the $\boldsymbol{\beta}$ in (4.38) that best fits the data.

Definition 4.39. Let $Y, B, Z, U, \boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{u}$ be as in (4.34) and $\Sigma_{\boldsymbol{u}}$ be the positive definite covariance matrix of $\boldsymbol{u}$. Define

$$\boldsymbol{u} = \boldsymbol{y} - (Z' \otimes I_N)\, \boldsymbol{\beta} \tag{4.40}$$

as the *residual vector* and

$$S(\boldsymbol{\beta}) = \boldsymbol{u}' \Sigma_{\boldsymbol{u}} \boldsymbol{u} \tag{4.41}$$

as the *objective function*.

An estimate $\hat{\boldsymbol{\beta}}$ which minimizes the objective function $S(\boldsymbol{\beta})$ is called the *(general) least squares estimator*. More precisely,

$$\hat{\boldsymbol{\beta}} = \operatorname*{argmin}_{\boldsymbol{\beta}} S(\boldsymbol{\beta}) \tag{4.42}$$

Remark 4.43. Note that $\Sigma_{\boldsymbol{u}} = I_T \otimes \Sigma_u$ when $\Sigma_{\boldsymbol{u}}$ is positive definite.

Theorem 4.44. *In addition to the assumptions of Definition 4.39 suppose that ZZ' is a non-singular matrix.*

Then

$$\hat{\boldsymbol{\beta}} = \Big((ZZ')^{-1} Z \otimes I_N \Big) \boldsymbol{y} \tag{4.45}$$

Proof. The least squares estimator is obtained by minimizing

$$\begin{aligned} S(\boldsymbol{\beta}) &= \boldsymbol{u}'(I_T \otimes \Sigma_u)^{-1}\boldsymbol{u} \\ &= [\boldsymbol{y} - (Z' \otimes I_N)\boldsymbol{\beta}]\,(I_T \otimes \Sigma_u^{-1})\,[\boldsymbol{y} - (Z' \otimes I_N)\boldsymbol{\beta}] \\ &= \operatorname{vec}(Y - BZ)'(I_T \otimes \Sigma_u^{-1})\operatorname{vec}(Y - BZ) \\ &= \operatorname{tr}\big[(Y - BZ)^{-1}\Sigma_u^{-1}(Y - BZ)\big], \end{aligned} \tag{4.46}$$

where the last equality holds, since Σ_u is a symmetric matrix and $\operatorname{tr}(ABC) = \operatorname{vec}(C')' \times (B' \otimes I)\operatorname{vec}(A)$ is used. Furthermore we have

$$\begin{aligned} S(\boldsymbol{\beta}) &= \boldsymbol{y}'(I_T \otimes \Sigma_u^{-1})\boldsymbol{y} + \boldsymbol{\beta}'(Z \otimes I_N)(I_T \otimes \Sigma_u^{-1})(Z' \otimes I_N)\boldsymbol{\beta} - 2\boldsymbol{\beta}(Z \otimes I_N)(I_T \otimes \Sigma_u^{-1})\boldsymbol{y} \\ &= \boldsymbol{y}'(I_T \otimes \Sigma_u^{-1})\boldsymbol{y} + \boldsymbol{\beta}'(ZZ' \otimes \Sigma_u^{-1})\boldsymbol{\beta}' - 2\boldsymbol{\beta}'(Z \otimes \Sigma_u^{-1})\boldsymbol{y}. \end{aligned}$$

Hence, taking partial derivatives w.r.t $\boldsymbol{\beta}$, we obtain

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 2(ZZ' \otimes \Sigma_u^{-1})\boldsymbol{\beta} - 2(Z \otimes \Sigma_u^{-1})\boldsymbol{y}.$$

Equating to zero gives the normal equations

$$(ZZ' \otimes \Sigma_u^{-1})\hat{\boldsymbol{\beta}} = (Z \otimes \Sigma_u^{-1})\boldsymbol{y}.$$

Consequently,

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= ((ZZ'Z^{-1} \otimes \Sigma_u^{-1})(Z \otimes \Sigma_u^{-1})\boldsymbol{y} \\ &= \Big((ZZ')^{-1} Z \otimes I_N \Big) \boldsymbol{y}. \end{aligned}$$

Checking that the Hessian of $S(\boldsymbol{\beta})$,

$$\frac{\partial^2 S}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = 2(ZZ' \otimes \Sigma_u^{-1}),$$

is positive definite confirms that $\hat{\boldsymbol{\beta}}$ is indeed a minimizing vector and thus the least squares estimator. □

Remark 4.47. For the result of Theorem 4.44 to hold, it is assumed that ZZ' is a non-singular matrix. This holds with probability 1 if y_t has a continuous distribution. Without stating otherwise, we will always assume a continuous distribution for y_t.

It can be seen that asymptotically $\hat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}$, that is as sample size $T \to \infty$, the resulting sequence of estimates $\hat{\boldsymbol{\beta}}$ converges in probability to the real parameter value $\boldsymbol{\beta}$. This, and the asymptotic distribution of $\hat{\boldsymbol{\beta}}$ are given by the following theorem.

Theorem 4.48. *Let y_t be a stable, N-dimensional VAR(p) process as in (4.33) with standard white noise residuals as in Definition 3.4. Then, as $T \to \infty$,*

$$\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$$

and

$$\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(0, \Gamma^{-1} \otimes \Sigma_u),$$

where $\Gamma^{-1} = plim ZZ'/T$. Moreover, $\xrightarrow{p}$ denotes the convergence in probability, and $\xrightarrow{d}$ convergence in distribution, as usual.

Proof. Omitted. See [14] Section 3.2.2 and Theorem 8.2.3 of [8]. □

In Theorem 4.44 we have seen that the least squares estimator $\hat{\boldsymbol{\beta}}$ does not interestingly depend on Σ_u. One might be interested in estimating the covariance matrix Σ_u for different purposes. Since $\Sigma_u = \mathbb{E}(u_t u_t')$ a plausible estimator is given by

$$\tilde{\Sigma}_u = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t \hat{u}_t' = \frac{1}{T}\hat{U}\hat{U}' = \frac{1}{T}(Y - \hat{B}Z)(Y - \hat{B}Z)' = \frac{1}{T}Y(I_T - Z'(ZZ')^{-1}Z)Y'. \tag{4.49}$$

Often a degrees of freedom adjustment is desired, because it leads to an unbiased estimator of the covariance matrix. Thus, an estimator

$$\hat{\Sigma}_u = \frac{T}{T - Np - 1}\tilde{\Sigma}_u$$

might be considered. Both $\tilde{\Sigma}_u$ and $\hat{\Sigma}_u$ are consistent estimators of Σ_u under the conditions of Theorem 4.48, see Proposition 3.2 of [14].

Lastly, we will consider the estimation of the VAR(p) model in a mean-adjusted form. That is, assume that a VAR(p) process in 4.33 is given as

$$(y_t - \mu) = A_1 (y_{t-1} - \mu) + \cdots + A_p (y_{t-p} - \mu) + u_t. \tag{4.50}$$

Assume that the mean vector μ is known. We define

$$\begin{aligned} Y^0 &:= (y_1 - \mu, \dots, y_T - \mu) && (N \times T), \\ A &:= (A_1, \dots, A_p) && (N \times Np, \\ Y_t^0 &:= \begin{bmatrix} y_t - \mu \\ \vdots \\ y_{t-p+1} - \mu \end{bmatrix} && (Np \times 1) \\ X &:= \left(Y_0^0, \dots, Y_{T-1}^0\right) && (Np \times T), \\ \boldsymbol{y}^0 &:= \operatorname{vec}(Y^0) && (NT \times 1) \\ \boldsymbol{\alpha} &:= \operatorname{vec}(A) && (N^2 p \times 1). \end{aligned} \tag{4.51}$$

Then

$$Y^0 = AX + U \tag{4.52}$$

or

$$\boldsymbol{y}^0 = (X' \otimes I_N)\, \boldsymbol{\alpha} + \boldsymbol{u}, \tag{4.53}$$

where U and $\boldsymbol{u}$ are defined as in (4.34). The least squares estimator of (4.50) is

$$\hat{\boldsymbol{\alpha}} = \left((XX')^{-1} X \otimes I_N \right) \boldsymbol{y}^0. \tag{4.54}$$

This can be seen by similar arguments than in the proof of Theorem 4.44.

Remark 4.55. In some literature *ordinary* least squares (OLS) estimates are discussed. The method of OLS differs from GLS only by the objective function in (4.41). In OLS, the covariance matrix $\Sigma_{\boldsymbol{u}}$ in (4.41) is replaced with identity matrix I_N. Both of the methods lead to similar results, thus it makes no difference if the estimation is done by either OLS or GLS. For details see e.g [23].

4.4 Modelling seasonality

Assume that the time series under consideration shows a seasonal pattern. That is, the obtained values of the time series are seen to be dependent of the season of the year, for example. One option to model the seasonality is a VAR(p) model with different intercept term for each season. That is

$$y_t = \nu_i + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t, \tag{4.56}$$

where ν_i is a $(N \times 1)$ intercept vector associated with the i-th season. In (4.56) a time point t is assumed to be within i'th season. Time variant vector of intercept allows the procecss to have a different mean for each season.

We can write (4.56) more precisely with the use of the so-called *dummy variable.*

Definition 4.57. Let the obtained time series have k seasons denoted by $s_1, \dots, s_k$. If

$$n_{it} = \begin{cases} 1 & \text{if} \quad t \in s_i \\ 0 & \text{otherwise} \end{cases}, \quad i = 1, \dots, k.$$

and $\sum_{i=1}^{k} n_{it} = 1$ for all t, then n_{it} is called a *seasonal dummy variable.*

Remark. The requirement of the sum of the dummy variables to be one for each t guarantees that every time point t belongs to one season only. Equivalently we could require that the seasons are distinct time intervals.

With the use of the seasonal dummy variable n_{it}, we can write (4.56) as

$$y_t = n_{1t}\nu_1 + \cdots + n_{kt}\nu_k + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t, \tag{4.58}$$

assuming that the process has k seasons as in Definition 4.57.

It is possible in the seasonal context that also the other coefficients than the vector of intercept ν vary seasonally. In that case, one might need a more general model. We can write

$$y_t = \nu_t + A_{1t} y_{t-1} + \cdots + A_{pt} y_{t-p} + u_t, \tag{4.59}$$

where u_t is an innovation process with zero mean and covariance matrices $\mathbb{E}\left(u_t u_t'\right) = \Sigma_u$. We also assume that u_t and u_s are independent all $s \neq t$. Since the covariances of u_t are allowed to vary, the innovation terms are not generally identically distributed.

It is worth noting that the VAR(p) model with constant coefficients defined in Definition 4.1 is a special case of this general form in (4.59). In (4.59) we allow the coefficients of the process vary with each time point. Seasonality is not as explicitly expressed in (4.59) as in (4.58), but this very general form allows the possibility of all kind of seasonality. Although it has to be noted, this form is more of a theoretical than a practical one.

4.4.1 A VAR representation with time invariant coefficients

Instead of going in to the properties and estimation of the general representation (4.59) (for that see Chapter 17 of [14]) we will focus on the representation which allows us to simplify the analysis of seasonally varying time series with time varying coefficients. Our goal is to present the seasonal (or periodic) time series in (4.59) as constant, non-periodic VAR process as defined in Definition 4.1.

For this purpose suppose that the process considered is a quarterly varying process, that is it has 4 periods during a year. We may define an annual process with

$$\varphi_1 := \begin{bmatrix} y_4 \\ y_3 \\ y_2 \\ y_1 \end{bmatrix}, \varphi_2 := \begin{bmatrix} y_8 \\ y_7 \\ y_6 \\ y_5 \end{bmatrix}, \dots, \varphi_\tau := \begin{bmatrix} y_{4\tau} \\ y_{4\tau-3} \\ y_{4\tau-2} \\ y_{4\tau-1} \end{bmatrix}, \dots,$$

where φ_τ is the annual process of year τ and each process $y_{4(\tau-1)-(4-i)}$ corresponds to the i-th period of the year τ.

Denote the i-th quarter of year τ by $s_{\tau,i}$. Assuming that each quarterly process $y_{4(\tau-1)-(4-i)}$ is a VAR(1) process such that

$$\begin{aligned} y_t &= \nu_t + A_{1,t} y_{t-1} + u_t \\ &= \nu_i + A_{1,i} y_{t-1} + u_t, \quad \text{if } t \in s_{\tau,i}, \quad i = 1,2,3,4, \quad \tau = 1,2,\dots, \end{aligned}$$

then we can write the annual process φ_τ as

$$\begin{bmatrix} I_N & -A_{1,4} & 0 & 0 \\ 0 & I_N & -A_{1,3} & 0 \\ 0 & 0 & I_N & -A_{1,2} \\ 0 & 0 & 0 & I_N \end{bmatrix} \begin{bmatrix} y_{4\tau} \\ y_{4\tau-3} \\ y_{4\tau-2} \\ y_{4\tau-1} \end{bmatrix} = \begin{bmatrix} \nu_4 \\ \nu_3 \\ \nu_2 \\ \nu_1 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ A_{1,1} & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_{4\tau} \\ y_{4\tau-3} \\ y_{4\tau-2} \\ y_{4\tau-1} \end{bmatrix} + \begin{bmatrix} u_{4\tau} \\ u_{4\tau-1} \\ u_{4\tau-2} \\ u_{4\tau-3} \end{bmatrix}.$$

More generally, assume that the process has k different seasons per year, with each process $y_{k(\tau-1)-(k-i)}$ corresponding to the i-th season of the year. Assuming that the processes $y_{k(\tau-1)-(k-i)}$ have constant parameters within each season $i = 1, \dots, k$ and that y_1 belongs to the first season, we may define the kN-dimensional annual process as

$$\varphi_\tau := \begin{bmatrix} y_{k\tau} \\ y_{k\tau-1} \\ \cdots \\ y_{k\tau+1} \end{bmatrix}, \quad \tau = 0, \pm 1, \pm 2, \dots$$

This annual process can be presented as a VAR(P) process, where P is the smallest integer greater than or equal to p/s. More precisely, we define

$$\mathbf{\Lambda}_0 \varphi_\tau = \boldsymbol{\nu} + \mathbf{\Lambda}_1 \varphi_{\tau-1} + \cdots + \mathbf{\Lambda}_P \varphi_{\tau-P} + \boldsymbol{u}_\tau, \tag{4.60}$$

where

$$\mathbf{\Lambda}_0 := \begin{bmatrix} I_N & -A_{1,k} & -A_{2,k} & \dots & -A_{k-1,k} \\ 0 & I_N & -A_{1,k-1} & \dots & -A_{1,k-1} \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \dots & I_N \end{bmatrix},$$

$$\boldsymbol{\nu} := \begin{bmatrix} \nu_k \\ \nu_{k-1} \\ \vdots \\ \nu_1 \end{bmatrix},$$

$$\mathbf{\Lambda}_i := \begin{bmatrix} A_{ik,k} & A_{ik+1,k} & \dots & A_{(i+1)k-1,k} \\ A_{ik-1,k-1} & A_{ik,k-1} & \dots & A_{(i+1)k-2,k-1} \\ \vdots & \vdots & \ddots & \vdots \\ A_{ik-k+1,1} & A_{ik-k+2,1} & \dots & A_{ik,1} \end{bmatrix}, \quad i = 1, \dots, P,$$

$$\boldsymbol{u}_\tau := \begin{bmatrix} u_{k\tau} \\ u_{k\tau-1} \\ \vdots \\ u_{k\tau-k+1} \end{bmatrix}.$$

Here $\mathbf{\Lambda}_0$ and $\mathbf{\Lambda}_i$ are $(kN \times kN)$ matrices, and $\boldsymbol{\nu}$ and $\boldsymbol{u}_\tau$ are $(kN \times 1)$ matrices. Thus the equation (4.60) is well-defined.

The stability condition of the process in (4.60) is as for VAR(p) process in Definition 4.12. That is, φ_τ is stable, if

$$\begin{aligned} & det(\mathbf{\Lambda}_0 - \mathbf{\Lambda}_1 z - \cdots - \mathbf{\Lambda}_P z^P) \\ & = det(I_{kN} - \mathbf{\Lambda}_0^{-1}\mathbf{\Lambda}_1 z - \cdots - \mathbf{\Lambda}_0^{-1}\mathbf{\Lambda}_P z^P) \neq 0 \quad \text{for} \quad |z| \leq 1, \quad z \in \mathbb{C}. \end{aligned} \tag{4.61}$$

If in addition the y_t's of φ_τ have bounded first and second moments, the process φ_τ is stationary.

Remark. Stationarity of φ_τ does not imply stationarity of the original process y_t. Assume for example that a stationary quarterly process φ_τ has a time-invariant mean vector $\boldsymbol{\mu} = (\mu_4, \mu_3, \mu_2, \mu_1)'$. Clearly the mean vectors for different quarters may be different.

If the stability condition (4.61) is satisfied, we can treat the process φ_τ as a normal stable VAR(P) process. In practice though, the representation with dummy variables in (4.58) is often sufficient in order to achieve stability. Usually other coefficients than the vector of intercept are not assumed to be varying over time. If stability cannot be achieved via dummy variables or there is a suspicion about the other coefficients varying over time, the method presented in this Section could be used in order to achieve a stable process.

4.5 Forecasting with estimated process

In Section 4.2 we have seen that the optimal linear h-step forecast of the VAR(p) process is

$$y_t(h) = \nu + A_1 y_t(h-1) + \cdots + A_p y_t(h-p),$$

where $y_t(j) = y_{t+j}$ for $j \leq 0$. By replacing the true coefficients $B = (\nu, A_1, \ldots, A_p)$ with their estimates $\hat{B} = (\hat{\nu}, \hat{A}_1, \ldots, \hat{A}_p)$ we obtain a forecast

$$\hat{y}_t(h) = \hat{\nu} + \hat{A}_1 \hat{y}_t(h-1) + \cdots + \hat{A}_p \hat{y}_t(h-p).$$

The forecast error is

$$\begin{aligned} y_{t+h} - \hat{y}_t(h) &= [y_{t+h} - y_t(h)] + [y_t(h) - \hat{y}_t(h)] \\ &= \sum_{i=0}^{h-1} \phi_i u_{t+h-i} + [y_t(h) - \hat{y}_t(h)], \end{aligned} \tag{4.62}$$

where the ϕ_i are the coefficient matrices of the canonical MA representation of VAR(p) process y_t as in Definition 4.20. Thus the forecast error with estimated coefficients is the forecast error with true coefficients plus the difference between the predictions with true and estimated coefficients respectively. It can be seen that under quite general conditions for the process y_t that the forecasts are unbiased, i.e forecast errors have zero mean, even with estimated coefficients (see e.g [3]).

In order to measure precision of the forecasts with estimated parameters we will need an expression for the MSE matrix of the forecasts. By noting that $[y_{t+h} - y_t(h)]$ is a function of $\{u_{t+1}, \ldots, u_{t+h}\}$ and $[y_t(h) - \hat{y}_t(h)]$ is a function of the past and present values of the process y_t at the forecast origin t, we see that the two terms in RHS of equality (4.62) are uncorrelated and

$$\begin{aligned} \Sigma_{\hat{y}}(h) \coloneqq MSE[\hat{y}_t(h)] &= \mathbb{E}\left([y_{t+h} - \hat{y}_t(h)]\,[y_{t+h} - \hat{y}_t(h)]'\right) \\ &= \Sigma_y(h) + MSE[y_t(h) - \hat{y}_t(h)], \end{aligned} \tag{4.63}$$

where $\Sigma_y(h) = \sum_{i=0}^{h-1} \phi_i \Sigma_u \phi_i'$. The distribution of the estimator $\hat{B}$ is needed in order to evaluate the last term in (4.63). In order to facilitate the following results there are two alternative assumptions that can be made:

1. Only data up to the forecast origin are used for estimation.
2. Estimation is done using a realization of a stochastic process that is independent of the process used for prediction. It is also assumed that these processes have the same stochastic structure (for instance, the process is Gaussian and has the same first and second moments as the process used for prediction).

The first assumption is often a realistic one from a practical point of view because the estimation and forecasting are usually based on the same data set. Though it can be seen that asymptotically the first assumption implies the same results as the second one, and thus the second assumption can alternatively be made in order to derive the following results. In addition to either one, it is assumed that we have for $\boldsymbol{\beta} = \text{vec}(B)$ and $\hat{\boldsymbol{\beta}} = \text{vec}(\hat{B})$

$$\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to \mathcal{N}(0, \Sigma_{\hat{\boldsymbol{\beta}}}),$$

asymptotically in distribution.

Under these assumptions we get the asymptotic approximation $\Omega(h)/T$, where

$$\Omega(h) = \mathbb{E}\left[\frac{\partial y_t(h)}{\partial \boldsymbol{\beta}'} \Sigma_{\hat{\boldsymbol{\beta}}} \frac{\partial y_t(h)'}{\partial \boldsymbol{\beta}}\right] \tag{4.64}$$

for the $MSE\left[\hat{y}_t(h) - y_t(h)\right]$. For the MSE matrix of $\hat{y}_t(h)$ we get an asymptotic approximation by

$$\Sigma_{\hat{y}}(h) = \Sigma_y(h) + \frac{1}{T}\Omega(h),$$

when y_t is Gaussian. An explicit expression for $\Omega(h)$ can be derived. It depends on the forecast horizon h, for $h = 1$ we have for example

$$\Omega(1) = (Np + 1)\Sigma_u,$$

and thus the approximation

$$\Sigma_{\hat{y}}(1) = \Sigma_u + \frac{Np+1}{T}\Sigma_u = \frac{T + Np + 1}{T}\Sigma_u \tag{4.65}$$

of the MSE matrix of the 1-step forecast with estimated coefficients is obtained. For $h > 1$ it is not possible to evaluate $\Omega(h)$ without knowing the summarized coefficients in matrix B. Although a consistent estimator $\hat{\Omega}(h)$ can be obtained by replacing all the unknown parameters by their least squares estimates.

We won't go into further details of forecast MSE matrix approximations, but rather satisfy with the results stated here. It is worth noting that from (4.64) we see that efficient estimation of $\hat{\boldsymbol{\beta}}$ reduces the forecast uncertainty. What is maybe even more interesting, is that from (4.65) we see that the MSE of 1-step forecast increases as the order p of the VAR model increases. Thus it might be in interest to fit a model of lower order, especially when forecasting is the main objective of the VAR model. Order selection will be discussed more thoroughly in Section 5.1.

For the details of the results stated in this Section we refer the reader to Section 3.5 of [14].

Chapter 5

Simulation

5.1 Order selection

In Chapter 4 we have assumed that VAR(p) process as in Definition 4.1 is behind the observed time series. We have not made an assumption that all the coefficient matrices $A_1, \ldots, A_p$ are non-zero. In other words, p is assumed as the upper bound for the order of the VAR process. Usually the real order of the data generation process is unknown, and statistical methods are needed to determine the right order. As seen in Section 4.5, choosing an order too high might reduce the forecast accuracy of the estimated VAR model. Thus, different statistical tools might be considered depending whether the interest is in fitting the correct order model corresponding to the data generation process, or in forecasting with best possible precision.

5.1.1 Likelihood ratio tests

The likelihood ratio tests are based on the likelihood functions. Assuming that y_t is Gaussian VAR(p) process, maximum likelihood estimation leads to similar results than GLS estimation in Section 4.3. For more information see e.g [14] or [21].

Assume that we have in some way or another chosen P as the upper bound for our VAR model order. The idea of likelihood ratio test is to compare this model of order P with a VAR($P-1$) model, or more precisely to determine whether the coefficient matrix A_P is non-zero. Statistically speaking, null-hypothesis $H_0^1 : A_P = 0$ is compared with alternative hypothesis $H_a^1 : A_P \neq 0$. If the null-hypothesis is rejected, we will choose P to be the order of the model. If the null-hypothesis is accepted, we will move on and compare the model of order $P-1$ with the model with lower order $P-2$, with similar hypothesis. This procedure is continued until the null-hypothesis is rejected with some order $P-i$, $0 \leq i \leq P$, and we will choose $P-i$ to be the estimated order of our model.

Formally the hypothesis for this series of tests can be written as

$$
\begin{aligned}
H_0^1 : A_P = 0 \quad &\text{versus} \quad H_a^1 : A_P \neq 0 \\
H_0^2 : A_{P-1} = 0 \quad &\text{versus} \quad H_a^2 : A_{P-1} \neq 0 | H_0^1 \\
&\vdots \\
H_0^i : A_{P-i+1} = 0 \quad &\text{versus} \quad H_a^i : A_{P-i+1} \neq 0 | H_0^{i-1} \\
&\vdots \\
H_0^P : A_1 = 0 \quad &\text{versus} \quad H_a^P : A_1 \neq 0 | H_0^{P-1},
\end{aligned}
$$

where conditioning on the null-hypothesis H_0^{i-1} is equivalent on conditioning to $A_P = \cdots = A_{P-i+2} = 0$.

The likelihood ratio test statistic is

$$
\lambda_{LR}(i) = T\left[\log\det\left(\tilde{\Sigma}_u(P-i)\right)\log\det\left(\tilde{\Sigma}_u(P-i+1)\right)\right], \quad i = 1, \ldots, P, \tag{5.1}
$$

for the i'th null-hypothesis H_0^i, when y_t is Gaussian. Here $\tilde{\Sigma}_u(a)$ is the estimator as in (4.49) for Σ_u, when VAR(a) model is fitted to a time series of length T by the method of GLS. It can be seen that the likelihood ratio test statistic in (5.1) follows asymptotically a $\chi^2_{N^2}$-distribution, where N is the number of variables in time series under consideration and lower index N^2 denotes the degrees of freedom of the chi-squared distribution (for this result see e.g [14], Proposition 4.1, or Result 7.11 of [12]). Thus the p-value of the likelihood ratio test statistics is calculated from the $\chi^2_{N^2}$-distribution, and the null-hypothesis considered is accepted or rejected according to the chosen significance level.

Although performing this kind of statistical testing scheme is a common strategy for detecting non-zero parameters, the approach might not be completely satisfactory if the VAR model is intended for a specific purpose, for instance if the main objective is forecasting. In such a case we might not be so interested in finding the correct order for the underlying data generation process by detecting non-zero parameters but finding a good model for prediction.

5.1.2 Information criteria

We have seen in Section 4.5 that the mean square error of 1-step forecast increases as the VAR order p increases. When forecasting is the main objective, it makes sense to choose an order p such that the precision of the forecast is maximized. In other words, one might want to choose an order p such that the theoretical MSE matrix of forecast errors is minimized. For this purpose several different criteria, commonly referred as information criteria, have been proposed. The most common ones are

$$
\begin{aligned}
\mathrm{FPE}(m) &= \left[\frac{T+Nm+1}{T-Nm-1}\right]\det\tilde{\Sigma}_u,\\
\mathrm{AIC}(m) &= \log\det(\tilde{\Sigma}_u) + \frac{2mN^2}{T},\\
\mathrm{BIC}(m) &= \log\det(\tilde{\Sigma}_u) + \frac{\log(T)}{T}mN^2,\\
\mathrm{HQ}(m) &= \log\det(\tilde{\Sigma}_u) + \frac{2\log[\log(T)]}{T}mN^2.
\end{aligned}
$$

Here FPE stands for *Final prediction error*, AIC for *Akaike's information criterion*, BIC for *Bayesian information criterion*, which is also commonly referred by the notation SQ(m), and HQ(m) stands for *Hannan-Quinn criterion*. Each of the criteria is used in the same way, the aim is to find an order $m = 0, \ldots, M$ such that the value of the selected information criteria is minimized and choose the order of the model accordingly.

In AIC, BIC and HQ the term $\log\det(\tilde{\Sigma}_u)$ measures the goodness of the fit of the model to the data, and the additive terms are used to penalize the more complicated models. It might be worth of noting that HQ and BIC are consistent estimators for the VAR order p when FPE and AIC are designed for minimizing the forecast error variance. Thus, if forecasting is the main objective, models based on AIC and FPE might produce superior forecasts, but they may not estimate the orders correctly. In practice it is recommendable to use information criteria as one of the methods choosing the model order, and not to choose the order by mechanically minimizing the information criteria. One way to further study the goodness of the estimated model is residual autocorrelations, which will be discussed next.

For more discussion about the topics of Section 5.1 see Sections 4.2 and 4.3 of [14].

5.2 Checking the model adequacy

Model diagnostics and checking the adequacy of the estimated model are essential part of time series analysis. We will assume that a time series of T observations generated by VAR(p) process with N variables is available. Estimation of the VAR(p) model is done using this particular time series.

5.2.1 Residual autocorrelations

Assuming that a VAR(p) model has been estimated, the residuals of an adequate model should behave like a white noise series. Thus checking the autocorrelation of the residuals becomes an integral part of the model diagnostics. Assuming further that the parameters of the model have been estimated using the method of (general) least squares as presented in Section 4.3 and using the notations from Equation (4.34), we have the estimated coefficient vector $\hat{B}$ and the corresponding residual matrix defined by $\hat{U} = (\hat{u}_1, \dots, \hat{u}_T) \coloneqq Y - \hat{B}Z$. The autocovariance matrices of the residuals are defined as

$$\hat{C}_i \coloneqq \frac{1}{T} \sum_{t=i+1}^{T} \hat{u}_t \hat{u}'_{t-i} = \frac{1}{T} \hat{U} F_i \hat{U}', \quad i = 0, 1, \dots, h < T, \tag{5.2}$$

where F_i is a $(T \times T)$ matrix such that $\hat{U} F_i \hat{U}' = \sum_{t=i+1}^{T} \hat{u}_t \hat{u}'_{t-i}$. The precise expression for F_i is not important neither interesting here. Furthermore we define

$$\hat{\boldsymbol{C}}_h \coloneqq \left(\hat{C}_1, \dots, \hat{C}_h\right) = \hat{U} F \left(I_h \otimes \hat{U}'\right),$$

where $F \coloneqq (F_1, \dots, F_h)$ is a $(T \times hT)$ matrix depending explicitly on h and T.

The autocorrelation matrices of the residuals are defined correspondingly as

$$\hat{R}_i = \hat{D}^{-1} \hat{C}_i \hat{D}^{-1}, \quad \hat{\boldsymbol{R}}_h \coloneqq \left(\hat{R}_1, \dots, \hat{R}_h\right),$$

where $\hat{D}$ is a $(N \times N)$ diagonal matrix of the standard errors of residual series, i.e $\hat{D}_{mm} = \sqrt{\hat{C}_{0,mm}}$ for $m = 1, \dots, N$.

The main theoretical results considering whiteness of the residuals are the asymptotic distributions of autocovariances and autocorrelations of an estimated process. We will state the results and the notations needed, and refer the reader to other sources for the proofs of those results.

Theorem 5.3. *Let y_t be a (stationary) stable VAR(p) process as in Definition 4.1 with identically distributed standard white noise u_t as in Definition 3.4 and let the coefficients be estimated by multivariate least squares as in Section 4.3.1 (or an asymptotically equivalent procedure). Then*

$$\sqrt{T} \operatorname{vec}(\hat{\boldsymbol{C}}_h) \xrightarrow{d} \mathcal{N}(0, \Sigma_c(h)),$$

where

$$\begin{aligned}\Sigma_c(h) &= (I_h \otimes \Sigma_u - \tilde{G}'\Gamma^{-1}\tilde{G}) \otimes \Sigma_u \\ &= (I_h \otimes \Sigma_u \otimes \Sigma_u) - \hat{G}\left[\Gamma_Y(0)^{-1} \otimes \Sigma_u\right]\hat{G}'\end{aligned}$$

and

$$\begin{aligned}\tilde{G} &:= \begin{bmatrix} 0 & 0 & \dots & 0 \\ \Sigma_u & \phi_1\Sigma_u & \dots & \phi_{h-1}\Sigma_u \\ 0 & \Sigma_u & \dots & \phi_{h-2}\Sigma_u \\ \vdots & & & \vdots \\ 0 & 0 & \dots & \phi_{h-p}\Sigma_u \end{bmatrix} \\ \Gamma &:= plim ZZ'/T\end{aligned}$$

and $\Gamma_Y(0)$ *is the covariance matrix of* $Y_t = (y_t', \dots, y_{t-p+1}')$ *and* $\hat{G} := \check{G} \otimes I_N$*, where* $\check{G}$ *is a* $(Np \times Nh)$ *submatrix of* $\tilde{G}$ *with the first row of zeros eliminated.*

Proof. See [14], Section 4.4. □

Theorem 5.4. *Let* D *be the* $(N \times N)$ *diagonal matrix with the square roots of* Σ_u *on the diagonal and define* $G_0 := \tilde{G}(I_h \otimes D^{-1})$*. Then, under the conditions of Theorem 5.3,*

$$\sqrt{T}\,\mathrm{vec}(\hat{\boldsymbol{r}}_h) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\boldsymbol{r}}(h)),$$

where $\Sigma_{\boldsymbol{r}}(h) = [(I_h \otimes R_u) - G_0'\Gamma^{-1}G_o] \otimes R_u$ *and* R_u *is the zero-lag autocorrelation matrix of* u_t*.*

Proof. See [14], Section 4.4. □

In practice, all unknown quantities will be replaced by estimates. It might be worth of noting that if Γ is estimated by ZZ'/T, estimator $\tilde{\Sigma}_u$ must be used for Σ_u in order to ensure positive variances.

In applied work, a popular way to test a white noise hypothesis of residuals is to plot the estimated autocorrelations and $\pm 2/\sqrt{T}$ upper- and lower bounds around zero. The white noise hypothesis is rejected, if any of the estimated correlation coefficients reach out the area between those bounds. One has to keep in mind that considering the individual correlation coefficients does not provide a picture of their overall significance. We present a test for the overall significance in Section 5.2.2.

5.2.2 Multivariate Portmanteau statistics

A *Portmanteau test* is a popular test for the overall significance of the residual autocorrelations up to the lag h. Let $\boldsymbol{R}_h$ be the lag h theoretical autocorrelation matrix of the residuals. Then the test hypothesis is

$$H_0 : \boldsymbol{R}_h = (R_1, \ldots, R_h) = 0 \quad \text{versus} \quad H_1 : \boldsymbol{R}_h \neq 0.$$

The test statistic is

$$\begin{aligned}
Q_h &:= T \sum_{i=1}^{h} \mathrm{tr}(\hat{R}_i' \hat{R}_u^{-1} \hat{R}_i \hat{R}_u^{-1}) \\
&= T \sum_{i=1}^{h} \mathrm{tr}(\hat{R}_i' \hat{R}_u^{-1} \hat{R}_i \hat{R}_u^{-1} \hat{D}^{-1} \hat{D}) \\
&= T \sum_{i=1}^{h} \mathrm{tr}(\hat{D} \hat{R}_i' \hat{D} \hat{D}^{-1} \hat{R}_u^{-1} \hat{D}^{-1} \hat{D} \hat{R}_i \hat{D} \hat{D}^{-1} \hat{R}_u^{-1} \hat{D}^{-1}) \\
&= T \sum_{i=1}^{h} \mathrm{tr}(\hat{C}_i' \hat{C}_0^{-1} \hat{C}_i \hat{C}_0^{-1}).
\end{aligned}$$

This test statistic has by Theorem 5.3 an approximate asymptotic χ^2-distribution.

Theorem 5.5. *Assume that the conditions of Theorem 5.3 hold. Then, approximately, for large T and h*

$$\begin{aligned}
Q_h &= T \sum_{i=1}^{h} tr(\hat{C}_i' \hat{C}_0^{-1} \hat{C}_i \hat{C}_0^{-1}) \\
&= T \,\mathrm{vec}(\hat{\boldsymbol{C}})' (I_h \otimes \hat{C}_0^{-1} \otimes \hat{C}_0^{-1}) \,\mathrm{vec}(\hat{\boldsymbol{C}}) \approx \chi^2(N^2(h-p)).
\end{aligned}$$

Proof. See Section 4.4 of [14] for a sketch of the proof and [1] for more details. □

As with residual autocovariances and -correlations the result in Theorem 5.5 is asymptotic. In practice it has been suggested to use modified test statistic

$$\overline{Q}_h := T^2 \sum_{i=1}^{h} (T-i)^{-1} \mathrm{tr}(\hat{C}_i' \hat{C}_0^{-1} \hat{C}_i \hat{C}_0^{-1}), \tag{5.6}$$

which, approximately in large samples and for large h has the same asymptotic χ^2-distribution with $N^2(h-p)$ degrees of freedom as Q_h.

Remark. The use of modified test statistic is a consequence of founding out that in small samples the nominal size of Portmanteau test tends to be lower than the significance level chosen. For more information see e.g [2], [11] and [13].

In some literature Lagrange multiplier tests have also been suggested for checking the adequacy of a fitted VAR model. Tsay and Lütkepohl both note in [21] and [14] that it has been found out that the asymptotic chi-square distribution of the Lagrange multiplier test is found to be a poor approximation, in [4] for instance. Thus we ignore the Lagrange multiplier tests, an interested reader might see [14] for the definitions of Lagrange multiplier test and further discussion.

5.3 Choosing parameters for the model and setting up the test data

In the following Sections our wish is to fit a VAR model for some seasonally varying time series and test its prediction accuracy. Naturally, we hope that the data used for the simulation would somehow resemble the time series pension insurance companies got. As mentioned in Chapter 2, before the Incomes Register there was an option for the employers to provide monthly payroll notifications for the pension insurance companies. If this option was chosen, the employers paid their TyEL payments also every month. The option of monthly payroll notifications was available from January 2007 to December 2018. Thus, it is a natural assumption that the pension insurance company would have a time series of the employers which have chosen the monthly payroll notification cycle from January 2007 until December 2018. This data set could be used in order to estimate the cycle of monthly payments together with the data from January 2019 until the forecast origin.

In this Section we will take a closer look to the prediction problem of the pension insurance company and set up the data set for our model building.

5.3.1 Prediction problem of the pension insurance company

As argued in Chapter 1 it is natural that the pension insurance company wants to forecast its incoming cash flow. With accurate predictions of the incoming cash flow the pension insurance company might be able to allocate its funds more efficiently, for instance, and gain advantage to the competitors.

As presented in Chapter 2 the TyEL contribution of one employer consists of the difference between the basic TyEL contribution and discounts received by the employer. Different discounts are given to the insured depending whether they are either large or small employers. The client bonus and the discount for large payrolls are discounts which are calculated and paid fully once in a year, making them only outgoing cashflow. Thus, the monthly incoming TyEL cashflow of the pension insurance company consists of three parts, that is from the basic TyEL contribution, affects of the contribution categories and contribution loss discounts paid within their insurance portfolio.

The basic TyEL contribution is always a pre-determined constant percentage of the TyEL payroll for a given year. Contribution categories are calculated for the upcoming year before the start of it. Thus, the pension insurance company knows the contribution categories of its insurance portfolio for the upcoming year at some point of the current year, and is able to estimate the affects as a percentage of the TyEL payroll for each insured large employer. The contribution loss discount is a known pre-determined constant percentage of the TyEL payroll when the size of the employer is determined. Since the sizes of the employers are determined for the calculations of the contribution categories, it is also known which of the insured employers receive contribution loss discounts in the upcoming year.

Thus, it seems that the prediction problem of the pension insurance company really comes down to predicting the monthly TyEL payrolls of its insured employers. The affects of the contribution categories are dependant from the age-distribution of the employer. The pension insurance company is able to calculate a single estimate as a percentage of the TyEL payroll for the effects of contribution category for each insured company, using e.g the age-distribution of the respective employer's from the last year. On a level of a large employer, and especially on the level of the whole insurance portfolio of the pension insurance company, there are usually a sufficiently large number of insured persons making the age-wage distributions pretty much stable over the years.

That being said, use of the out of the model -estimate for the effect of contribution category (or categories) seems to make sense. Similarly, knowing the sizes of the employers the contribution loss discounts paid can be taken into account by a single estimate outside of the model. Then, it makes no difference whether the insurance company wants to estimate incoming TyEL cashflow from its entire insurance portfolio, a certain subset of it or even from a single employer. The out of the model -estimates can always be adjusted accordingly. Although if estimating cash flow generated by more than one insured employer, some kind of estimates of wages are needed in order to derive single estimates for the effect of the contribution categories and paid contribution loss discounts on the level of the whole insurance portfolio.

Example 5.7. Assume that the pension insurance company has a portfolio of 4 clients - clients A, B, C and D. Clients A and B are large employers, clients C and D are small employers. Moreover, client A has a contribution category 7 and client B has a contribution category 1 for the upcoming year. Based on adequate statistics, the pension insurance company has estimated the following figures:

Client	Effect of contribution category, % of TyEL payroll	Contribution loss discount, % of TyEL payroll
A	+0.975%	-0.187%
B	-1.275%	-0.187%
C	0.00%	0.00%
D	0.00%	0.00%

Wanting single estimates for the upcoming year for the whole insurance portfolio, the insurance company uses the TyEL payrolls of the previous year and calculates the estimates as a weighted average:

Client	TyEL payroll of previous year %	Proportion of the insurance portfolio
A	22 000 000 EUR	50%
B	18 000 000 EUR	41%
C	2 000 000 EUR	4,5%
D	2 000 000 EUR	4,5%

The single estimates are

$$\lambda_C = \frac{0.975 * 22000000 - 1.275 * 18000000 + 0.00 * 2000000 + 0.00 * 2000000}{44000000} \approx -0.034\%$$
$$\lambda_M = \frac{-0.187 * 22000000 - 0.187 * 18000000 + 0.00 * 2000000 + 0.00 * 2000000}{44000000} = -0.17\%,$$

of the TyEL payroll of the whole insurance portfolio, λ_C corresponding to the effects of contribution categories and λ_M to the contribution loss discounts paid. Of course, predicted payrolls can be used in determining the proportions of the insurance portfolio, but in that case the pension insurance company would need to estimate TyEL payrolls of each employer for the upcoming year. With large number of insured employers this could not be practical from a computational viewpoint.

5.3.2 The test data

By the arguments in Section 5.3.1 we are interested in predicting the monthly TyEL payrolls of the insured employers. It is somewhat natural assumption that the payrolls have some kind of a relation with the number of insured persons within the insured company or the whole insurance portfolio. Keeping in mind, although forecasting being main objective of the thesis, that VAR models can also be used to structural analysis between the variables of the model, it seems quite a good choice to choose monthly TyEL payrolls and monthly numbers of insured persons for our variables. We might hope also that with these variables the model could capture some patterns such as the effect of bonuses paid for holiday seasons every year at the same time together with other seasonal factors and thus make the predictions more accurate. An economic variable could also be added to the model, but in the lack of economical theory and probable precision loss by choosing a poor economic variable, we decide to leave any economic variables out of the model.

Unfortunately any type of the data we would be exactly looking for, that is data of TyEL payrolls and numbers of insured persons per month, is not publicly available. Although Finnish Centre of Pensions provides yearly figures of TyEL payrolls and persons insured under TyEL per industry [7]. Keeping in mind that we are looking data with seasonal variation, construction seems a good industry to choose for our test data. Construction is known to be quite highly seasonal business in Finland, due to the extreme weather conditions for instance. Summers are usually the high seasons, with winters substantially the low seasons.

In order to derive the yearly data to the monthly level, we make use of the monthly wage and salary indices by industry provided by Statistics Finland [5]. These indices describe the monthly development of wages and salaries sum per industry. By combining these two data sets, we get estimated monthly figures of the TyEL payrolls and numbers of insured persons under TyEL of the construction industry, from January 2008 till December 2018. Although the figures are not exact, this data set satisfies our purposes of testing the VAR model with seasonal, monthly data, and is very much similar to the actual time series which might be used by the pension insurance company. Seasonality of the obtained data is clearly visible from Figure 5.1.

Remark. The test data set concerning the estimated monthly TyEL payrolls and numbers of insured persons of the construction industry in Finland starts from January 2008, not from January 2007 which would be assumed as the first observation available for the pension insurance company. This is due to the reform of coding of the industries in Finland at the turn of the year 2008.

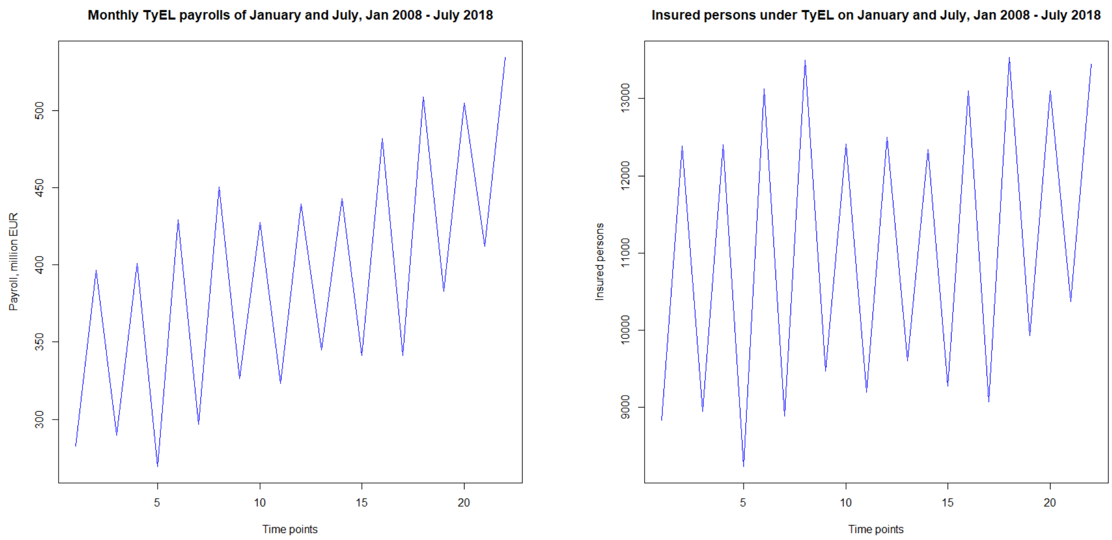

Figure 5.1: Estimated TyEL payrolls and numbers of insured persons under TyEL of the construction industry in Finland. Every odd time point corresponds to the values of Januaries, and even time points to the values of Julys from the year 2008 to the year 2018.

5.4 Building the model

Now that we have a data set for the simulated prediction problem we can start building the VAR model. Our time series consists of two variables, that is $N = 2$. Observation at the time point t will be denoted by $y_t = (y_{t,p}, y_{t,i})'$, $y_{t,p}$ corresponding to the monthly TyEL payroll and $y_{t,i}$ to the monthly number of insured persons under TyEL. Both variables concern specifically the construction industry, as mentioned before. The test data set is available in whole in Tables A.1 and A.2 of Appendix A.

Since our data set consists of values from January 2008 to December 2018, it seems natural to choose the year 2018 to be our forecast horizon with the forecast origin being theoretically the start of the year 2018, such that the values from December 2017 have been observed and are in our use, but no values from the year 2018 have been observed yet. That is, we will build our model with the information set consisting of the 120 observed values from January 2008 to December 2017, moreover $\Omega_{120} = \{y_s | 1 \leq s \leq 120\}$, forecast the values from January 2018 to December 2018 and compare them to the real, observed, values of the year 2018 to see if our model forecasts well. The forecast horizon of 12 months seems also natural given the discussion regarding the calculation of contribution

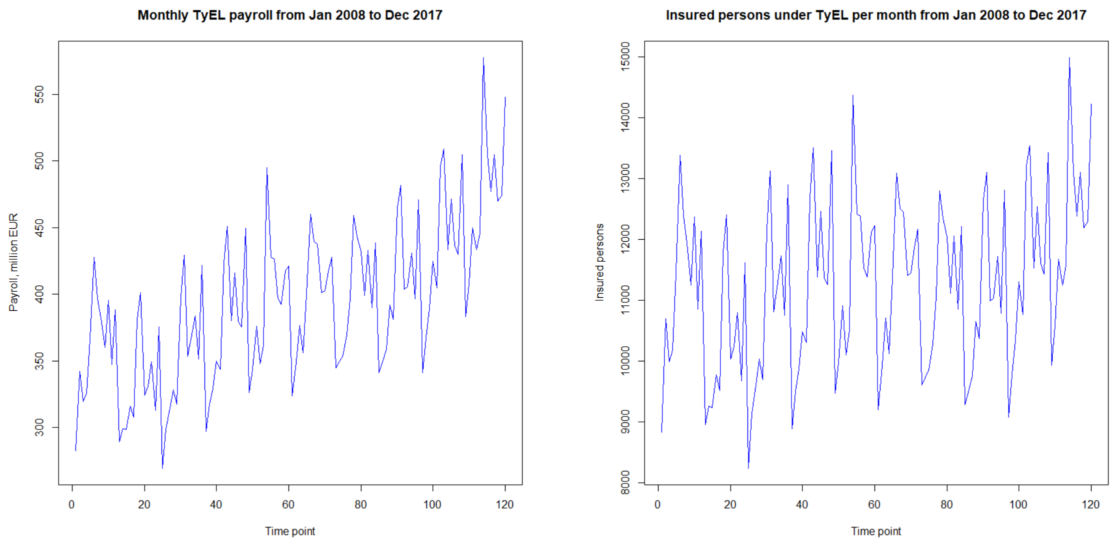

Figure 5.2: Information set Ω_{120} for the model estimation. First observations are from January 2008 and the last ones from December 2017.

categories in Section 5.3.1. From the computational viewpoint we will be using R [17] version 3.6.3 with the package 'vars' made for VAR modelling by Bernhard Pfaff for the model estimation. See [16] and [15] for the implementation of the package 'vars'. The R code used for our model estimation is presented in Appendix B.

The data of the information set Ω_{120} is visualised on Figure 5.2. The underlying seasonality is again clearly visible. There seems to be no large outliers on the data, so we will be satisfied with the data set as it is. One thing to be considered though is the visible trend in both the payrolls and numbers of insured persons. Clearly the overall levels of payrolls and insured persons have increased as a function of time. As discussed before, we were not keen on using any economic variable in the model. We will take two different approaches on the trend component of the data - one is to take it into account with built-in functions of the 'vars' package, and the other is to try and detrend the economic trend from the observed time series.

The starting point of the model building is to choose some candidates for the VAR model order p. Forecasting being the main objective, we might be not so interested in fitting the correct order model for the underlying data generation process, so we will prefer the method of minimizing the information criterion presented in Section 5.1 over the likelihood ratio tests to get us some candidates. Higher order models may not be preferred by the theory presented in Section 4.5, but we may test what model orders are suggested by the information criteria with higher maximum orders P. Results with maximum orders $P = 20, 13$ and 6 are presented in Table 5.1.

Max. order	AIC	HQ	SC	FPE
$P = 20$	15	13	13	15
$P = 13$	13	13	13	13
$P = 6$	6	3	3	6

Table 5.1: Suggested model orders with given maximum model orders P by each information criterion. Results are attained via VARSelect -function of the 'vars' package by setting the argument 'season' as 12 and argument 'type' as trend.

By suggestions in Table 5.1 we will try models with orders $p = 15, 13, 6$ and 3. Estimation of the VAR models is done by the function VAR of the 'vars' package. Seasonality is taken into account with argument 'season', which is set to be 12 since our data is monthly, and the underlying trend component is handled with argument 'type' which is set to be trend. With these arguments the function VAR estimates a VAR(p) model with seasonal dummy variables as in (4.58) and a constant intercept term which includes the deterministic trend-component, and notably captures the seasonality of one category, which here will be January. Thus the resulting models are of the form

$$y_t = \nu_s + n_{1t}\nu_1 + \cdots + n_{11t}\nu_{11} + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t, \quad p \in \{15, 13, 6, 3\}, \tag{5.8}$$

where ν_s is the estimated intercept term for January together with the effect of the trend. The seasonal dummy variables are as Definition 4.57 with $k = 11$ and $s_1 = (1, 2], s_2 = (2, 3] \ldots, s_{11} = (11, 12]$ corresponding to the each month of the year of the time point t. The function VAR estimates the parameters $\nu_s, \nu_1, \ldots, \nu_{11}, A_1, \ldots, A_p$ in (5.8) with the method of ordinary least squares, but as we have noted in Section 4.3 the OLS leads to similar results as with the general least squares method presented in Section 4.3.

After estimating the VAR models, we wish to determine whether the models are stationary or not. This is achieved via roots -function of the 'vars' package, which calculates and returns the eigenvalues of the coefficient matrix $\mathbf{A}$ of the VAR(1) representation in (4.10) of the estimated VAR(p) model. Remembering that the stability condition in Definition 4.12 is equivalent to the eigenvalues of $\mathbf{A}$ being all strictly less than one, stability of each model is easily seen from the Table 5.2.

VAR(15)	1.0184391	1.0184391	0.9974798	0.9937347
VAR(13)	1.0248404	1.0248404	0.9915615	0.9896849
VAR(6)	0.9871302	0.9871302	0.8874787	0.8874787
VAR(3)	0.9959825	0.9363501	0.8368390	0.8368390

Table 5.2: Four largest eigenvalues of the VAR(1) representations' coefficient matrices **A** of the estimated VAR(p) models with $p = 15, 13, 6$ and 3.

The method presented in Section 4.4.1 could be used for non-stationary VAR(15) and VAR(13) models to try obtain stationary models, but as we are not especially keen on fitting these higher order models anyway, we will continue with the VAR(6) and VAR(3) models as they are seen to be stationary. It has to be also noted, that even if the higher order models would be of an interest, the method presented in Section 4.4.1 would decrease our sample size from $T = 120$ to $T = 30$, if we would fit a quarterly process. Thus, it might not be a reasonable option without a couple of more years of data in our information set even if our interest would be on models with higher orders.

As we have chosen the VAR(3) and VAR(6) models for our contenders, we would like to investigate the adequacy of these models. As presented in Section 5.2, the residuals of an adequate model should behave like white noise series and have no significant auto- or crosscorrelations. Residual auto- and crosscorrelations of the estimated VAR(3) and VAR(6) models are plotted in Figure 5.3 together with the $\pm 2/\sqrt{T} = \pm 2/\sqrt{120}$ -significance bounds. There are significant single autocorrelations with lags 5,9 and 13 with the VAR(3) model, and similar significant crosscorrelations with lags $\pm 5, \pm 9$ and ± 13. We see that the same single auto- and crosscorrelations arise also with the model of order 6.

In Section 4.5 we have seen that increasing the model order reduces the forecast accuracy. Thus, when nearly similar auto- and cross-correlations arise with VAR(6) model and VAR(3) model, we could prefer the model with lower order to continue with. Though, the auto- and crosscorrelations at lags 5 and 9 could be something to worry about, which gives us enough reason to fit a model with higher order than 6. We will estimate a VAR(9) model in the hope of getting rid of the auto- and crosscorrelations at the lags 5 and 9.

VAR(9)	0.9914393	0.9914393	0.9793456	0.9793456

Table 5.3: Four largest eigenvalues of the coefficient matrix **A** of VAR(1) representation of the estimated VAR(9) model.

From Table 5.3 it is seen that the estimated VAR(9) is stable. From Figure 5.4 we see that the single auto- and crosscorrelations with lags 5 and 9 vanished, as we hoped, but also the $\pm 2/\sqrt{120}$ -significant correlation at lag 13 vanished. The VAR(9) model seems to capture the underlying data generation process better than than the previous lower order models.

For testing the overall significance of the correlation coefficients we are using the modified Portmanteau test statistic in (5.6) as suggested in Section 5.2.2. The serial.test -function of 'vars' package calculates the modified test statistic with argument 'type' set as PT.Adjusted and returns the p-value for the null-hypothesis $H_0 : \boldsymbol{R}_h = (R_1, \ldots, R_h) = 0$. Obviously, we expect the null-hypothesis to be rejected with VAR(3) model, but as we see from Table 5.4 it's rejected also with the VAR(9) model. This is something of what we are not too concerned, if the models forecast well. The interpretation of overall correlation with higher h is that the higher lags still contain information, which should be taken into account in the VAR model. This is consistent with the higher model order suggestions given by information criteria in Table 5.1. It is still highly reasoned to consider the VAR(9) model, since it reduced all the single auto- and crosscorrelations. The residual means of the estimated models are 14116.27 for y_{tp} and 1.022356 for y_{ti} with the VAR(3) model, and 19008.21 for y_{tp} and 0.9161942 for y_{ti} with the VAR(9) model. They are still considerably close to zero, so that we could be somewhat satisfied with the residuals acting like white noise series and the models being adequate enough.

Given these results, we might be wanting to test the forecast abilities of VAR(9) and VAR(3) models. If either of the models forecasts well, we could satisfy ourselves with the results or maybe test a model of lower order to see if it increases the forecast precision. Obtained forecasts and some possibilities of model improvement are considered next in Section 5.5.

Estimated VAR model	h	p-value	H_0 accepted/rejected
VAR(3)	24	1.053e-09	Rejected
VAR(3)	36	1.335e-07	Rejected
VAR(3)	48	2.755e-08	Rejected
VAR(9)	24	7.458e-06	Rejected
VAR(9)	36	0.0002656	Rejected
VAR(9)	48	0.000555	Rejected

Table 5.4: Results of the modified Portmanteau test with test statistic $\overline{Q}_h$ as in (5.6) and null-hypothesis $H_0 : \boldsymbol{R}_h = (R_1, \ldots, R_h) = 0$.

Remark. The choice of the h in the Portmanteau test is a source of some debate. Tsay has suggested the use of $h \approx \log T$ with non-seasonal data set, and with h multiples of seasons with seasonal data in [20]. Thus we are testing with h two, three and four times the seasons of our data.

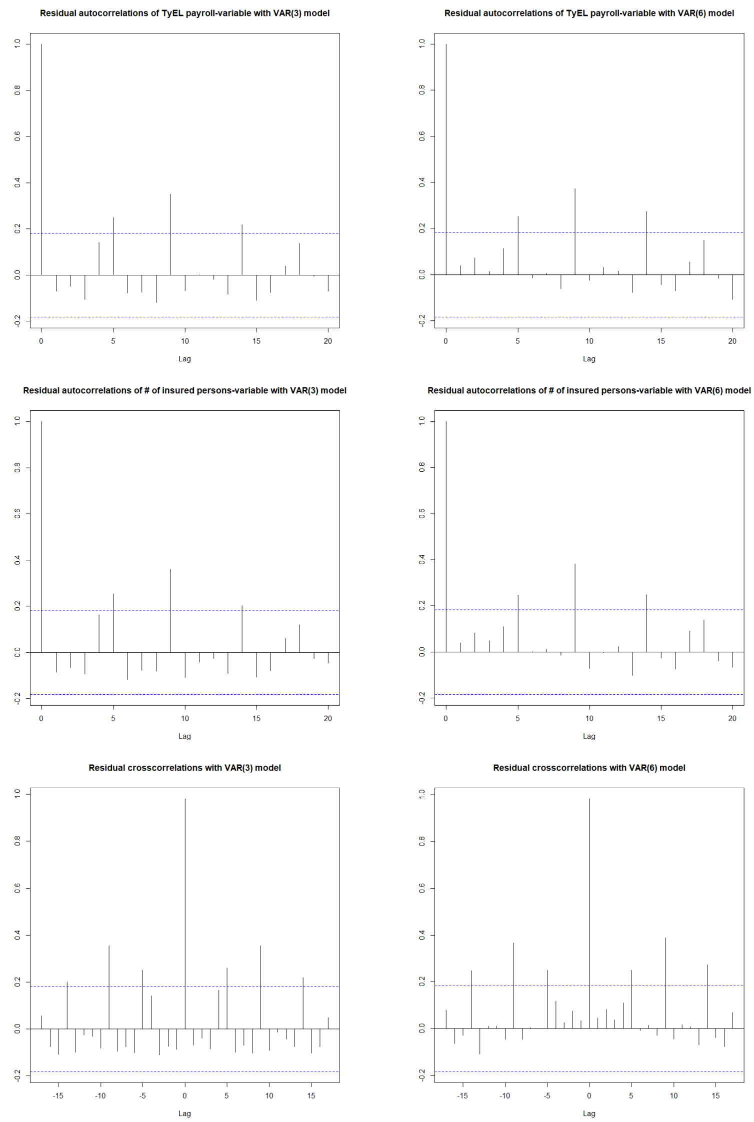

Figure 5.3: Residual auto- and crosscorrelations of the estimated VAR(3) and VAR(6) models with $\pm 2/\sqrt{T}$ -significance bounds. Blue lines are the $\pm 2/\sqrt{T}$ -significance bounds, and single auto- and crosscorrelations at given lags reaching out of the area between them are considered significant.

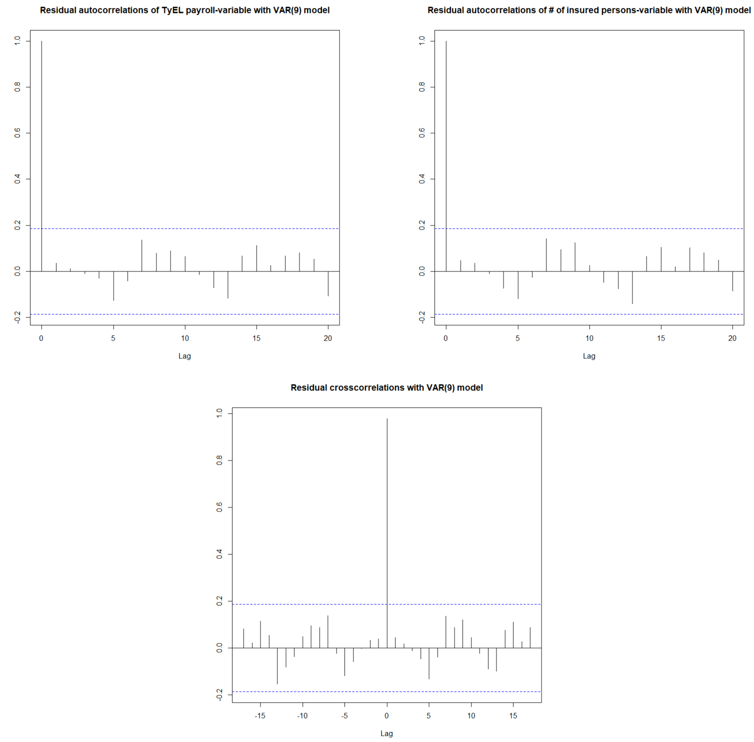

Figure 5.4: Residual auto- and crosscorrelations of the estimated VAR(9) model with $\pm 2/\sqrt{T}$ -significance bounds. No single auto- or crosscorrelations reach out the area between the blue $\pm 2/\sqrt{T}$ -significance bounds.

5.5 Results and possible model improvements

As mentioned in the start of the Section 5.4, our forecast horizon is 12 months from the hypothetical forecast origin in the start of the year 2018. The obtained forecast results with the VAR(3) and VAR(9) models estimated in Section 5.4 are presented in Tables 5.5, 5.6, 5.7 and 5.8. Interval forecasts are omitted, since we have made no assumptions about the distribution of y_t. More generally, from our viewpoint the interest lies mostly on point forecasts anyway.

Time t	Real $y_{t,p}$ EUR	Pred. $\hat{y}_{t,p}$ EUR	Difference EUR	Difference %
2018M01	412209073	394733236	-17475837.6	-4.24
2018M02	437740504	437553338	-187166.4	-0.04
2018M03	477273043	458250463	-19022580.3	-3.99
2018M04	454212395	442586131	-11626264.3	-2.56
2018M05	475625854	468487925	-7137929.3	-1.50
2018M06	607400982	549910705	-57490277.4	-9.46
2018M07	534512864	530008394	-4504470.6	-0.84
2018M08	541101621	498009336	-43092284.6	-7.96
2018M09	495392123	491125978	-4266145.5	-0.86
2018M10	497039312	490617840	-6421471.8	-1.29
2018M11	537807242	480104030	-57703212.3	-10.73
2018M12	546454985	527277905	-19177080.8	-3.51

Table 5.5: Forecast results of the estimated VAR(3) model, TyEL payroll-variable.

Time t	Real $y_{t,i}$ pers.	Pred. $\hat{y}_{t,i}$ pers.	Difference pers.	Difference %
2018M01	10367	9701	-667	-6.43
2018M02	11010	10944	-65	-0.59
2018M03	12004	11469	-535	-4.46
2018M04	11424	11047	-377	-3.30
2018M05	11962	11795	-167	-1.40
2018M06	15277	14100	-1177	-7.70
2018M07	13444	13560	116	0.87
2018M08	13609	12630	-980	-7.20
2018M09	12460	12414	-45	-0.36
2018M10	12501	12427	-74	-0.60
2018M11	13526	12102	-1425	-10.53
2018M12	13744	13447	-296	-2.16

Table 5.6: Forecast results of the estimated VAR(3) model, number of insured persons under TyEL-variable.

Time t	Real $y_{t,p}$ EUR	Pred. $\hat{y}_{t,p}$ EUR	Difference EUR	Difference %
2018M01	412209073	384954770	-27254303.5	-6.61
2018M02	437740504	431806452	-5934052.2	-1.36
2018M03	477273043	481447747	4174704.5	0.87
2018M04	454212395	428508948	-25703447.3	-5.66
2018M05	475625854	474886307	-739547.2	-0.16
2018M06	607400982	565835153	-41565828.8	-6.84
2018M07	534512864	509555362	-24957502.0	-4.67
2018M08	541101621	510243472	-30858148.3	-5.70
2018M09	495392123	511708262	16316138.7	3.29
2018M10	497039312	467307057	-29732254.9	-5.98
2018M11	537807242	494068691	-43738551.0	-8.13
2018M12	546454985	546840428	385442.9	0.07

Table 5.7: Forecast results of the estimated VAR(9) model, TyEL payroll-variable.

Time t	Real $y_{t,i}$ pers.	Pred. $\hat{y}_{t,i}$ pers.	Difference pers.	Difference %
2018M01	10367	9488	-880	-8.48
2018M02	11010	10827	-183	-1.66
2018M03	12004	12044	40	0.33
2018M04	11424	10615	-809	-7.08
2018M05	11962	11930	-33	-0.27
2018M06	15277	14367	-910	-5.95
2018M07	13444	12942	-502	-3.73
2018M08	13609	12930	-679	-4.99
2018M09	12460	12810	350	2.81
2018M10	12501	11696	-805	-6.44
2018M11	13526	12412	-1114	-8.24
2018M12	13744	13796	52	0.38

Table 5.8: Forecast results of the estimated VAR(9) model, number of insured persons under TyEL-variable.

The VAR(3) model is able to predict the values of monthly TyEL payrolls with mean absolute difference ($|y_{t,p} - \hat{y}_{t,p}|$) of 3.915% to the observed, real payrolls, and the number of insured persons with mean absolute difference ($|y_{t,i} - \hat{y}_{t,i}|$) of 3.8% to the real numbers. Respectively, the VAR(9) model predicts the TyEL payrolls with mean absolute difference of 4.111667%, and the number of insured persons with mean absolute difference of 4.196667% to the real values. On a yearly level the VAR(3) predicts the TyEL payroll with -4.123553% difference to the real payroll, as the VAR(9) model predicts the yearly payroll with -3.483719% difference to the real value. It could be noted that the VAR(3) model predicts all the payrolls as less than observed values, which could be taken into account with a deterministic trend component.

Based on these results one could consider of fitting a lowest-possible order model, which reduces the auto- and crosscorrelations obtained with lags 3 and 5 to see whether fitting a model slightly better than VAR(3) increases the forecast accuracy enough together with lowering the model order to obtain more accurate forecasts than with VAR(9) model. As we have seen, with VAR(6) model there were significant auto- and cross-correlations with the lags 3 and 5, so we could try fitting a VAR(7) model. Instead of doing this, we will first consider a different approach with the trend component.

As mentioned in the beginning of Section 5.4 we wish to test whether detrending the economic trend out of the observed data set would better the obtained forecasts. The intercept term approach for the trend component is rather crude method, which could reduce the accuracy of our forecasts. The underlying economic trend can cause some distortion as well to the estimation of the parameters of our models. There are several ways to detrend observed trends from the time series data. As we again wish to make no assumptions about the underlying economic trend, we'll use a quite simple method of differencing the observed values. That is, we will transform every observation at time point t to the difference between observed values at time points t and $t-1$. It is worth of noting that differencing the time series treats the underlying trend as a stochastic one, i.e, the trend would be generated by a random walk. From the economic viewpoint one could argue that increase in salaries would be better understood as a deterministic trend.

The resulting detrended data set is visualised in Figure 5.5. The trend has vanished, as we hoped. It might be noted that the detrending the data set by differencing reduces the sample size by one to $T = 119$. The information set consisting of the detrended values is referred by Ω_{119} and precisely defined as $\Omega_{119} \coloneqq \{y_s - y_{s-1} | 2 \leq s \leq 120\}$.

For building the VAR models with this detrended data set we will use similar procedure than in Section 5.4. We will start by obtaining some candidates for our model order using the method of minimizing the information criteria. Estimation of the model is done using the same 'vars' package functions is in Section 5.4, but without the argument 'type' being set as trend, since the data is now in detrended form to start with. The resulting models will be as in (5.8) with the intercept ν_s being a general intercept term capturing the

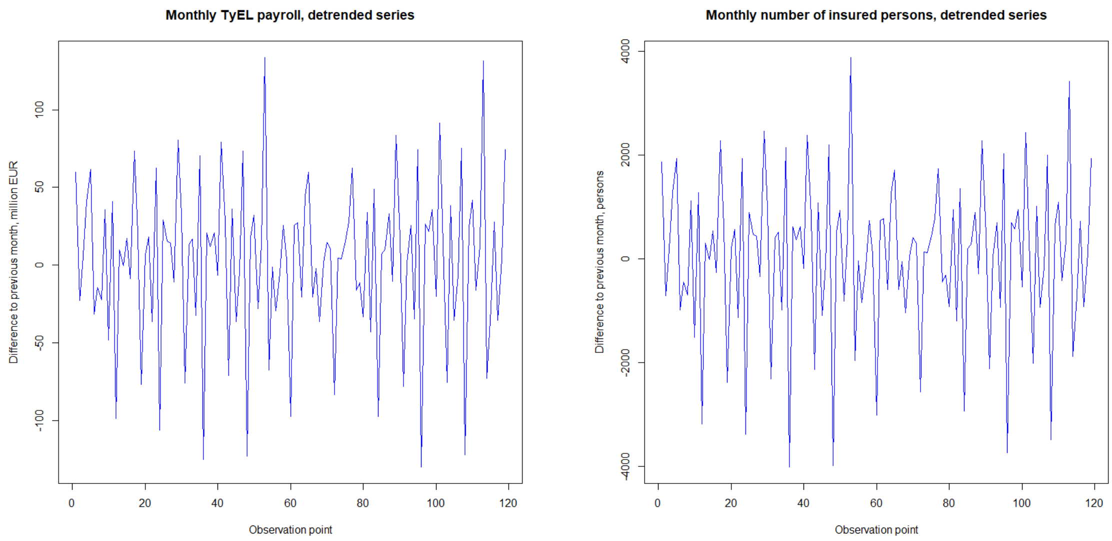

Figure 5.5: Detrended time series, information set Ω_{119}.

Remark. First observation point corresponds to the difference between observed values for February 2008 and January 2008. The last observation is the difference between December 2017 and November 2017.

behaviour of January in the model. Argument 'season' is set again to be 12 to obtain seasonal dummies.

From Table 5.9 we see that the suggestions based on information criteria differ a little bit from the suggestions in Table 5.1. Seen from the Table 5.10, estimated VAR(14), VAR(13) and VAR(12) models are not stable, and we will continue the model building with VAR(2) and VAR(6) models.

Residual auto- and cross-correlations of the estimated VAR(2) and VAR(6) models are visualized in Figure 5.6. We see that $\pm 2/\sqrt{119}$ -significant single correlations occur again at lags 5, 9 and 13 with the VAR(2) model and at lags 9 and 13, with lag 5 being just on the edge of the significance bound with the VAR(6) model. These correlations suggest again fitting a VAR(9) model in order to get rid of the significant correlations at lag 9. From Table 5.11 we see that the estimated VAR(9) model is again stable. Residual auto- and crosscorrelations in Figure 5.7 show us that all the single auto- and crosscorrelations have vanished with estimated VAR(9) model.

By the modifed Portmanteau test results in Table 5.12 we have to reject the null-hypothesis on non-correlation of the residuals with all models. This, again, thus not concern us too much if the models forecast well. We would like to test first the forecast abilities of the VAR(2) and VAR(9) models in order to compare the results attained with models estimated with original information set. The obtained results can be seen from Tables 5.13, 5.14, 5.15 and 5.16.

Max. order	AIC	HQ	SC	FPE
$P = 20$	14	12	12	14
$P = 13$	13	13	12	13
$P = 6$	6	2	2	6

Table 5.9: Suggested model orders with given maximum model orders P by each information criterion with detrended data. Results are attained via vars::VARSelect-function by setting the argument 'season' as 12.

VAR(14)	1.0180855	1.0180855	0.9945052	0.9945052
VAR(13)	1.0268931	1.0268931	0.9922615	0.9922615
VAR(12)	1.0260925	1.0260925	0.9927337	0.9850482
VAR(6)	0.9105159	0.9105159	0.8695096	0.8695096
VAR(2)	0.8275305	0.8275305	0.5333085	0.5333085

Table 5.10: Four largest eigenvalues of the coefficient matrix **A** of VAR(1) representations of the with detrended data estimated VAR(p) models with $p = 14, 13, 12, 6$ and 2 .

VAR(9)	0.9770823	0.9770823	0.9366831	0.9366831

Table 5.11: Four largest eigenvalues of the coefficient matrix **A** of VAR(1) representation of the with detrended data estimated VAR(9) model.

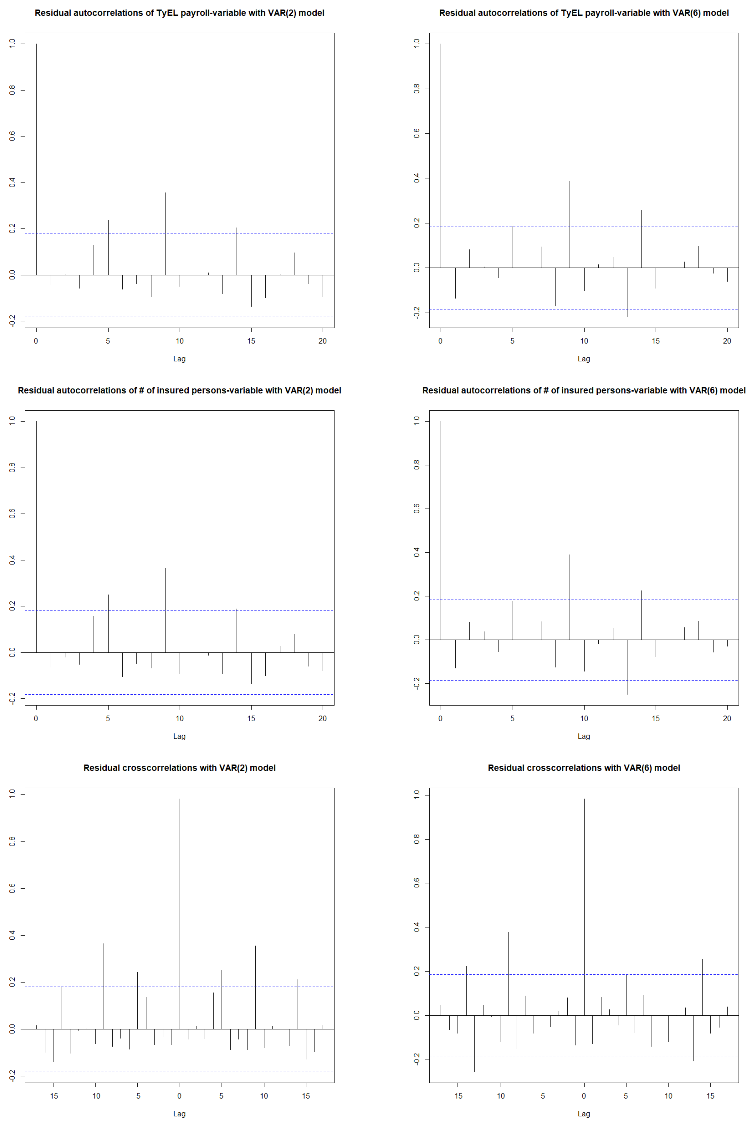

Figure 5.6: Residual auto- and crosscorrelations of the with detrended data estimated VAR(2) and VAR(6) models with $\pm 2/\sqrt{T}$-significance bounds.

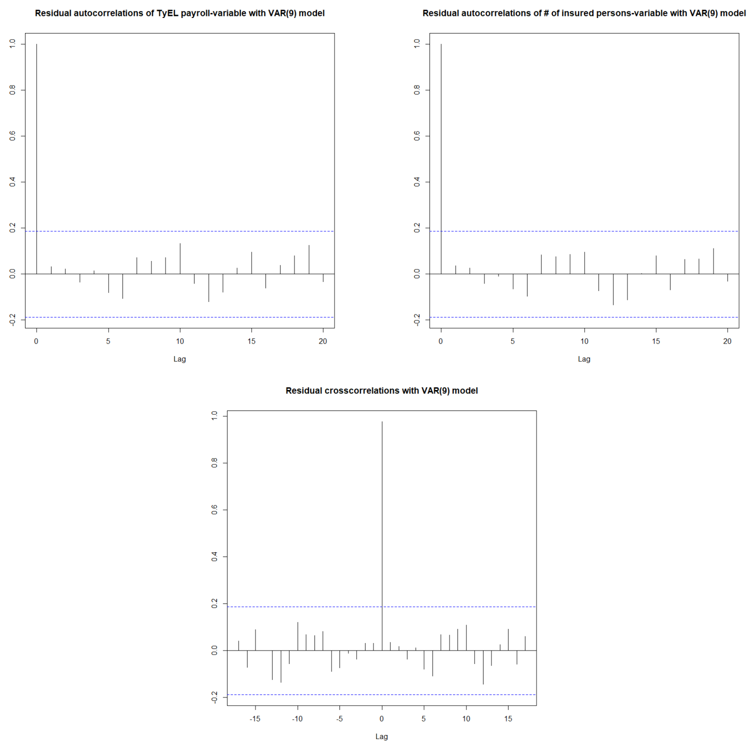

Figure 5.7: Residual auto- and crosscorrelations of the with detrended data estimated VAR(9) model with $\pm 2/\sqrt{T}$-significance bounds.

Estimated VAR model	h	p-value	H_0 accepted/rejected
VAR(3)	24	4.142e-09	Rejected
VAR(3)	36	2.807e-07	Rejected
VAR(3)	48	1.466e-07	Rejected
VAR(6)	24	6.204e-12	Rejected
VAR(6)	36	3.26e-12	Rejected
VAR(6)	48	2.442e-15	Rejected
VAR(9)	24	5.972e-05	Rejected
VAR(9)	36	0.008064	Rejected
VAR(9)	48	0.009304	Rejected

Table 5.12: Results of the modified Portmanteau test with test statistic $\overline{Q}_h$ as in 5.6 and null-hypothesis $H_0 : \boldsymbol{R}_h = (R_1, \ldots, R_h) = 0$ with models estimated with detrended information set Ω_{131}.

Time t	Real $y_{t,p}$ EUR	Pred. $\hat{y}_{tp}$ EUR	Difference EUR	Difference %
2018M01	412209073	399332949	-12876124	-3.12
2018M02	437740504	441080947	3340443	0.76
2018M03	477273043	463218701	-14054342	-2.94
2018M04	454212395	449433447	-4778949	-1.05
2018M05	475625854	474337536	-1288318	-0.27
2018M06	607400982	557693705	-49707277	-8.18
2018M07	534512864	538836302	4323437	0.81
2018M08	541101621	506347210	-34754411	-6.42
2018M09	495392123	501453822	6061699	1.22
2018M10	497039312	501371154	4331842	0.87
2018M11	537807242	490991732	-46815511	-8.70
2018M12	546454985	539872446	-6582539	-1.20

Table 5.13: Forecast results of the VAR(2) model estimated with detrended data, TyEL payroll-variable.

Time t	Real $y_{t,i}$ pers.	Pred. $\hat{y}_{t,i}$ pers.	Difference pers.	Difference %
2018M01	10367	9825	-542	-5.23
2018M02	11010	11039	29	0.26
2018M03	12004	11607	-396	-3.30
2018M04	11424	11239	-185	-1.62
2018M05	11962	11963	1	0.00
2018M06	15277	14325	-951	-6.23
2018M07	13444	13815	372	2.76
2018M08	13609	12875	-734	-5.40
2018M09	12460	12718	259	2.08
2018M10	12501	12744	243	1.94
2018M11	13526	12426	-1100	-8.13
2018M12	13744	13822	78	0.57

Table 5.14: Forecast results of the VAR(2) model estimated with detrended data, number of insured persons under TyEL-variable.

Time t	Real $y_{t,p}$ EUR	Pred. $\hat{y}_{t,p}$ EUR	Difference EUR	Difference %
2018M01	412209073	384730404	-27478669.7	-6.67
2018M02	437740504	429584272	-8156232.5	-1.86
2018M03	477273043	477037041	-236001.8	-0.05
2018M04	454212395	427590177	-26622218.3	-5.86
2018M05	475625854	475325480	-300373.7	-0.06
2018M06	607400982	567091271	-40309711.6	-6.64
2018M07	534512864	510870681	-23642183.4	-4.42
2018M08	541101621	511797592	-29304028.7	-5.42
2018M09	495392123	512888114	17495991.4	3.53
2018M10	497039312	470240054	-26799258.5	-5.39
2018M11	537807242	499536195	-38271047.1	-7.12
2018M12	546454985	548254691	1799705.9	0.33

Table 5.15: Forecast results of the VAR(9) model estimated with detrended data, TyEL payroll-variable.

Time t	Real $y_{t,i}$ pers.	Pred. $\hat{y}_{t,i}$ pers.	Difference pers.	Difference %
2018M01	10367	9514	-854	-8.24
2018M02	11010	10825	-185	-1.68
2018M03	12004	12018	15	0.12
2018M04	11424	10703	-721	-6.31
2018M05	11962	12057	95	0.79
2018M06	15277	14528	-749	-4.90
2018M07	13444	13114	-330	-2.45
2018M08	13609	13120	-489	-3.59
2018M09	12460	13012	552	4.43
2018M10	12501	11938	-564	-4.51
2018M11	13526	12740	-786	-5.81
2018M12	13744	14037	293	2.13

Table 5.16: Forecast results of the VAR(9) model estimated with detrended data, number of insured persons under TyEL-variable.

The VAR(2) model is able to predict the values of monthly TyEL payrolls with mean absolute difference of 2.961667% to the real payrolls, and the number of insured persons with mean absolute difference of 3.126667% in to the real numbers. The VAR(9) model predicts the TyEL payrolls with mean absolute difference of 3.945833% and the number of insured persons with mean absolute difference of 3.746667% to the real values. Comparing these results to the values obtained with the original estimation done by the non-differenced data (information set Ω_{120}), the accuracy has increased especially with the lower order model. Improvement is seen also on the yearly level, as the VAR(2) model predicts the TyEL payroll of 2018 with -2.539569% accuracy and the number of insured persons with -1.936192% accuracy. Correspondingly the VAR(9) model predicts the yearly payroll with -3.354358% accuracy and the numbers of insured persons with -2.459558% accuracy, making the VAR(2) model more accurate on yearly level too contrary to the previous results. Detrending the data seems to make the estimation of the VAR parameters more efficient, which gives the advantage in forecasting to the lower order models as supported by the theory in Section 4.5. It is worth noting that the improvements in the VAR(9) model are much smaller than in the lower order model, which is probably due the fact that the VAR(9) model corresponds more accurately to the underlying data generation process and captures it with the original, non-detrended data, also.

Given these results, we might want to test a VAR(1) model to see whether any more forecast precision is attained by lowering the model order. The forecasts obtained with estimated VAR(1) model are given in Tables 5.17 and 5.18.

Time t	Real $y_{t,p}$ EUR	Pred. $\hat{y}_{t,p}$ EUR	Difference EUR	Difference %
2018M01	412209073	421991492	9782419	2.37
2018M02	437740504	451721771	13981266	3.19
2018M03	477273043	461031110	-16241933	-3.40
2018M04	454212395	471920579	17708183	3.90
2018M05	475625854	478188899	2563045	0.54
2018M06	607400982	565173994	-42226988	-6.95
2018M07	534512864	554479151	19966287	3.74
2018M08	541101621	511581606	-29520015	-5.46
2018M09	495392123	511577427	16185304	3.27
2018M10	497039312	513730393	16691080	3.36
2018M11	537807242	497572888	-40234354	-7.48
2018M12	546454985	550991959	4536973	0.83

Table 5.17: Forecast results of the VAR(1) model estimated with detrended data, TyEL payroll-variable.

Time t	Real $y_{t,i}$ pers.	Pred. $\hat{y}_{t,i}$ pers.	Difference pers.	Difference %
2018M01	10367	10444	77	0.74
2018M02	11010	11269	259	2.35
2018M03	12004	11530	-474	-3.95
2018M04	11424	11840	416	3.64
2018M05	11962	12025	63	0.52
2018M06	15277	14508	-769	-5.03
2018M07	13444	14221	777	5.78
2018M08	13609	12983	-626	-4.60
2018M09	12460	12972	512	4.11
2018M10	12501	13056	555	4.44
2018M11	13526	12575	-952	-7.03
2018M12	13744	14103	359	2.61

Table 5.18: Forecast results of the VAR(1) model estimated with detrended data, number of insured persons under TyEL-variable.

With the estimated VAR(1) model the mean absolute difference is 3.7075% for $y_{t,p}$ and 3.733333% for $y_{t,i}$ to the real values. On a yearly level though the estimated VAR(1) model outperforms all of the previous models, with only -0.4455668% difference in the predicted TyEL payrolls and 0.1308416% difference in the predicted number of insured persons for year 2018. It seems that in general differencing of the time series improved accuracy of our forecasts. Thus, it might be preferred to work with detrended information sets. It has to be noted also that we have not made any assumptions about the economic trend of the year 2018. If one has any, the obtained forecasts could easily to be adjusted outside of the model to take the expected economic trend into account.

We will summarize the attained results together with our conclusions in Section 6.2, after comparing the VAR forecasts to some alternative approaches.

Remark 5.9. It is mentioned in Section 5.4 that a VAR(7) model could be estimated (with the information set Ω_{120}). Also, the VAR(6) model is considered in this Section. The forecasts with these models are a bit more accurate than with the respective VAR(9) models estimated with the information sets Ω_{120} and Ω_{119}, but lose clearly on precision to the lower order VAR(3), VAR(2) and VAR(1) models and are thus omitted.

Chapter 6

Alternative models and conclusions

6.1 Alternative models

In order to determine the efficiency of the VAR forecasts we would like to test the results against some alternative, easily formed forecasts to see whether taking the multivariate time series approach to the forecasting problem shows clear benefits. We will not be testing the VAR results against other multivariate time series models, since the choice of the VAR model is already argued in the introduction of the thesis. Comparing the VAR results against univariate autoregressive model with only the TyEL payroll as variable seems also unnecessary, since, as mentioned before, VAR offers together with forecasts an option to investigate the dynamic relations between the variables. It is also an intuitive assumption that the numbers of the insured persons and the payrolls paid have some kind of correlation between them. That is, the use of two variables instead of one should only offer additional benefits for the prediction problem of the pension insurance company.

One alternative approach to the prediction problem could be due normality assumption. Especially when working with the detrended data, we could try to fit normal distributions for each month with monthly means and variances. This would take the seasonality into account, as every month would have different means, and the assumption of normality is quite general. Of course, the assumption could be a poor one, but as said it is quite general and a simple approach. The number of insured persons variable is omitted in this approach, since the main interest lies after all in the TyEL payrolls.

Willing to test the normal approximation, we have fitted normal distributions for TyEL payroll data for each month using the detrended data set. That is, we have for each month mean μ_i for the difference from the observed value at previous month and estimated standard deviation σ_i, $i = 1, 2 \ldots, 12$. Index 1 corresponds to the difference between TyEL payrolls at February and January, month 2 to the difference between March and February and so on. Fitting and estimating of the distributions is done by functions of 'MASS' package [22] in R and the estimated parameters are presented in Table 6.1. The forecasts in Table 6.2 are obtained as mean of the random sample of 1000 observations from the estimated normal distributions for each month.

Months	i	μ_i	σ_i
Feb - Jan	1	22529985	14831919
Mar - Feb	2	14212082	17365164
Apr - Mar	3	7724419	21053482
May - Apr	4	8251935	21913596
Jun - May	5	85763564	25321350
Jul - Jun	6	-9950706	36261272
Aug - Jul	7	-43347780	32800860
Sep - Aug	8	267047	27424816
Oct - Sep	9	1990035	27395002
Nov - Oct	10	-16059808	27395002
Dec - Nov	11	53360559	25727154
Jan - Dec	12	-109117865	15266169

Table 6.1: Estimated normal distribution means and standard deviations for differences between TyEL payrolls with respect to the previous month.

Time	Real payroll	Pred. payroll	Difference EUR	Difference %
2018M01	412209073	438738383	26529309	6.44
2018M02	437740504	461560669	23820164	5.44
2018M03	477273043	476001221	-1271822	-0.27
2018M04	454212395	483949211	29736815	6.55
2018M05	475625854	493245934	17620080	3.70
2018M06	607400982	578863185	-28537797	-4.70
2018M07	534512864	569442627	34929763	6.53
2018M08	541101621	525117253	-15984367	-2.95
2018M09	495392123	524292280	28900157	5.83
2018M10	497039312	525287338	28248026	5.68
2018M11	537807242	509404947	-28402296	-5.28
2018M12	546454985	562208932	15753947	2.88

Table 6.2: Predicted monthly TyEL payrolls with estimated normal distributions $\mathcal{N}(\mu_i, \sigma_i)$.

The mean for absolute difference of monthly forecasts compared to the real values is 4.6875% with the normal approximation. On a yearly level the difference to the real value is only 2.182932%, making the yearly prediction slightly more accurate than obtained with most of the VAR models. Notable though, there are no large "outliers" in the predicted values, and the overall level of the predictions seems quite stable. Though, on a monthly level, the obtained forecasts differ from the real values on average more than with any of our estimated VAR models.

Another option to the prediction problem could be via Poisson distributions and processes. Compound Poisson variables are often used in the analysis of the total claim amounts of insurance companies. That is, one could estimate the monthly numbers of insured persons as a Poisson process and the distribution of the payrolls separately. This is still a fairly simple alternative, but we run to the problem with the economic trend. Poisson distribution is a discrete distribution with positive support, and thus the differenced data set cannot be used with it since it contains negative values. The effect of the trend should be cleared from the data set with different method or with use of some economic variable. At least without assumptions about the underlying economic trend these Poisson distribution-based alternatives don't seem very practical for this particular problem.

One could also consider a very simplified approach as an alternative. That could be by assuming that the payrolls depend from the level of payrolls of the last quarter, and predictions would be made by forecasting the values by assuming that the level of payrolls remains the same on the next quarter with only estimated seasonal variation added to forecasts. This, though, does not really seem like an interesting alternative since we have argued that the pension insurance companies can have substantial benefits by estimating the incoming TyEL cash flow efficiently. The assumption that the payroll levels would remain same than in the last quarter is clearly very prone to exogenous shocks for instance and generally a very poor one.

6.2 Conclusions

To briefly summarize the results of the VAR models and the alternative models, we will evaluate the forecast abilities with the mean squared prediction error (MSPE). The MSPE is defined as

$$\text{MSPE(M)} = \frac{1}{k}\sum_{h=1}^{k}\left[\hat{y}_{p,t+h|\Omega_M} - y_{p,t+h}\right]^2,$$

where $\hat{y}_{p,t+h|t}$ is the predicted value of the TyEL payroll at time point $t+h$ based on the estimated model M and the corresponding information set Ω_M, $y_{p,t+h}$ is the real observed value at time point $t+h$ and k is the forecast horizon. The MSPE's of the estimated VAR models and alternative models are given in Table 6.3.

Model M	Ω_M	MSPE
VAR(9)	Ω_{120}	$6.519 * 10^{14}$
VAR(3)	Ω_{120}	$8.161 * 10^{14}$
VAR(9)	Ω_{119}	$5.888 * 10^{14}$
VAR(2)	Ω_{119}	$5.322 * 10^{14}$
VAR(1)	Ω_{119}	$5.090 * 10^{14}$
$\mathcal{N}(\mu_i, \sigma_i)$	Ω_{119}	$6.25262 * 10^{14}$

Table 6.3: The mean squared prediction errors of the estimated models. The information set Ω_{119} corresponds the detrended data set, and the information set Ω_{120} the original data set and the estimated models correspond to the given information sets used in the estimation.

As can be seen from Table 6.3, measured in MSPE the VAR(1) and VAR(2) models estimated with the detrended data set have the best forecast abilities for our prediction problem. The alternative normal approximation-model performs the third best. When measured in the absolute mean difference of the monthly predictions, the VAR models generally outperform the normal approximation results. This supports the idea that the alternative normal approximation model performs decently on the yearly level, but loses on the monthly level accuracy for the VAR models.

Model M	Ω_M	Mean of the absolute differences $\lvert\hat{y}_{p,t+h\mid\Omega_M} - y_{p,t+h}\rvert$, % of the $y_{p,t+h}$
VAR(9)	Ω_{120}	4.111667
VAR(3)	Ω_{120}	3.915
VAR(9)	Ω_{119}	3.945833
VAR(2)	Ω_{119}	2.961667
VAR(1)	Ω_{119}	3.7075
$\mathcal{N}(\mu_i, \sigma_i)$	Ω_{119}	4.6875

The VAR(2) model has the lowest mean absolute difference in the monthly values, but its MSPE is 16% higher than VAR(1) models. Although, the predicted yearly payroll with VAR(2) differs only -2.54% of the observed one, which is generally quite a good result. Whether the monthly accuracy gained with the VAR(2) model is sufficient for the loss in the yearly prediction accuracy is left for debate.

Further investigation shows us that all the VAR models predicted worst on the same months - June, November and the higher order models on January also. On average the difference between observed values of TyEL payrolls on May and June is 85763564, when the difference between June 2018 and May 2018 is 131775128. This differs substantially lot from the mean difference. Similar phenomena can be obtained with the value of November 2018, as the mean difference between values observed on November and October is -16059808 with the difference in 2018 being 40767930,33. One could argue that these are outliers on the data caused by some exogenous shock.

The higher order models seem to predict the value of January worse than the lower order models due the higher number of lags in the model. The lower order models are able the predict the higher January values with the information of the last or last two months, when the VAR(9) models take into account a higher number of previous values and thus the affect of the values obtained on the couple of the last months is somewhat lesser in the model.

Without making any economical assumptions or using economic indicators VAR models seem superior approach to the prediction problem stated in Section 5.3. The normal approximation performs decently, but accuracy is clearly gained by fitting a VAR model. Models based on the Poisson distribution are somewhat out of the question without any additional assumptions made. The simplified approaches are as suggested quite prone, and could lead to very biased forecasts. With the use of the economical knowledge one could test the Poisson model, and probably make the simplified predictions more accurate, but in order to attain more accurate results than with the VAR models the Poisson approximation and distribution for the payrolls should intuitively be very well fitted to the underlying phenomena. Also, with the economical knowledge one could argue that the VAR models could be made even more accurate. That is, one could fit a model of the form

$$y_t = \nu_s + n_{1t}\nu_1 + \cdots + n_{11t}\nu_{11} + A_1 y_{t-1} + \cdots + A_p y_{t-p} + CD_t + u_t,$$

where C is a coefficient matrix of right dimension and D_t contains the values of exogenous economic variables. Whether this would increase the forecast accuracy should be justified. The use of the exogenous economical variable could also increase uncertainty, since it would probably be an estimate.

Another option of further model development could be by adding an another variable to the VAR model. The assumption of correlation between the number of insured persons and the payrolls seems intuitive, but a third, maybe an economical variable could also be considered. As mentioned in the introduction of the thesis, VAR models offer a possibility of analysing the dynamic structures between its variables (see e.g [14]). If it is an interest to fit a higher order model in order to capture the real underlying data generation process more accurately, the method presented in Section 4.4.1 could be considered, specifically when a couple of more years of data is available.

Altogether, the VAR models seem highly appropriate approach for estimation of the incoming TyEL cash flow of a pension insurance company. Even without any economical assumptions the results are very satisfactory with the test data. Estimated VAR models are able to capture the seasonal variations, and they seem not to have any significant deficiencies. Whether to include some economic variables could be decided, but it really don't seem necessary if not especially wanted. Estimation of the VAR models seems not too complicated or computationally inefficient. All in all, no reason is seen why VAR models should not be used and in the light of the results attained we'd generally recommend the use of the VAR models for this particular prediction problem of a pension insurance company.

Appendix A

The data set

Time t	$y_{t,p}$	$y_{t,i}$	Time t	$y_{t,p}$	$y_{t,i}$
2008M01	282697827.3	8833	2010M02	298641175.4	9129
2008M02	342345808.7	10696	2010M03	314292341.8	9608
2008M03	319662773.5	9988	2010M04	328251490.3	10034
2008M04	325543560.4	10171	2010M05	317253373.3	9698
2008M05	366289012.5	11444	2010M06	397624227.9	12155
2008M06	428037274.9	13374	2010M07	429349565.2	13125
2008M07	396533059.4	12389	2010M08	353631760.1	10810
2008M08	382251148.4	11943	2010M09	366744899.6	11211
2008M09	359988169.4	11248	2010M10	383665079.5	11728
2008M10	395692947	12363	2010M11	351516737.6	10745
2008M11	347386483.2	10854	2010M12	421735484.3	12892
2008M12	388131935.3	12127	2011M01	296735425.2	8888
2009M01	289407475.9	8949	2011M02	317301246.7	9504
2009M02	299108285.2	9249	2011M03	329472855.4	9868
2009M03	298704084.8	9237	2011M04	350038677	10484
2009M04	316084701.3	9774	2011M05	343743017.3	10296
2009M05	307596493.3	9512	2011M06	423068329	12672
2009M06	381160963.4	11786	2011M07	450769231.5	13501
2009M07	400966782.3	12399	2011M08	379838132.7	11377
2009M08	324168709.1	10024	2011M09	415933248	12458
2009M09	331040115.6	10236	2011M10	379418422	11364
2009M10	349229132.9	10799	2011M11	375641026.2	11251
2009M11	312851098.3	9674	2011M12	449090388.9	13451
2009M12	375502158	11611	2012M01	326243727.1	9468
2010M01	269453865.1	8237	2012M02	343765768.7	9977

2012M03	375889511.6	10909	2015M02	348594225.5	9478
2012M04	347937683.4	10098	2015M03	358751399.9	9754
2012M05	361705001.8	10497	2015M04	391660644.9	10649
2012M06	495206271.2	14372	2015M05	381097183.6	10361
2012M07	427621253.5	12410	2015M06	464792300.6	12637
2012M08	426369679.1	12374	2015M07	481856353.6	13101
2012M09	397166276.4	11526	2015M08	403849254.2	10980
2012M10	392159978.8	11381	2015M09	405474402.1	11024
2012M11	417608658.3	12120	2015M10	431070481.6	11720
2012M12	420946190.1	12217	2015M11	396536088.7	10781
2013M01	323429616.1	9200	2015M12	470886605.3	12803
2013M02	349105206.6	9931	2016M01	341228086	9072
2013M03	376437286.9	10708	2016M02	367476400.3	9770
2013M04	355731165.5	10119	2016M03	389213285.6	10348
2013M05	400042265.3	11380	2016M04	424894587.8	11297
2013M06	460090017.3	13088	2016M05	404798222.2	10763
2013M07	439383895.9	12499	2016M06	496257192.3	13194
2013M08	437313283.8	12440	2016M07	508971219.6	13532
2013M09	400870510.1	11403	2016M08	433507316	11526
2013M10	402526999.8	11450	2016M09	471649397.7	12540
2013M11	417021284.8	11863	2016M10	436378225.3	11602
2013M12	427788467.9	12169	2016M11	429816146.8	11428
2014M01	344691318	9600	2016M12	504869920.5	13423
2014M02	349237606.4	9727	2017M01	382929970.8	9934
2014M03	353370595.8	9842	2017M02	407822493.2	10580
2014M04	369075955.6	10279	2017M03	449724906.1	11667
2014M05	396766984.8	11051	2017M04	433544766.5	11247
2014M06	459175125.1	12789	2017M05	445991027.7	11570
2014M07	443056466.3	12340	2017M06	577506521.5	14982
2014M08	431897394.9	12029	2017M07	504903330.9	13098
2014M09	398833479.5	11108	2017M08	477106680.8	12377
2014M10	432723992.8	12052	2017M09	504903330.9	13098
2014M11	389740902.8	10855	2017M10	469638924.1	12183
2014M12	438510178	12213	2017M11	473787677.8	12291
2015M01	341281059.9	9279	2017M12	548050369.9	14217

Table A.1: Information set used for simulation.

Time t	$y_{t,p}$	$y_{t,i}$	Time t	$y_{t,p}$	$y_{t,i}$
2018M01	412209073.3	10367	2018M07	534512864.3	13444
2018M02	437740504.4	11010	2018M08	541101620.7	13609
2018M03	477273042.9	12004	2018M09	495392123.1	12460
2018M04	454212395.5	11424	2018M10	497039312.2	12501
2018M05	475625853.8	11962	2018M11	537807242.5	13526
2018M06	607400982.1	15277	2018M12	546454985.3	13744

Table A.2: Observed values of $y_{t,p}$ and $y_{t,i}$ of the year 2018.

Appendix B

R codes

Here 'data' refers to the data set obtained by combining the data presented in Tables A.1 and A.2. Time points 1-132 refer respectively to the time points $1 = 2008M01, 2 = 2008M02, \ldots, 120 = 2017M12, \ldots, 2018M12 = 132$. Essential parts of the codes used estimating the VAR models are provided. Codes for plotting the figures, comparing the predictions to the observed values etc. are omitted.

B.1 Estimation of the VAR models

```
#The data set consists of observed values of monthly
#TyEL payrolls and numbers of insured persons under TyEL.
#Payrolls are referred as 'Wage' and the number of insured persons as 'Insured'.

#NOTE! Data is not detrended.

train_data <- data.frame(Wage = data$Wage[1:120],
                         Insured = data$Insured[1:120])

#Information criteria suggestions for VAR order with maximums 20,13 and 6
vars::VARselect(train_data, lag.max = 20, season = 12, type = "trend")
vars::VARselect(train_data, lag.max = 13, season = 12, type = "trend")
vars::VARselect(train_data, lag.max = 6, season = 12, type = "trend")

#Choosing the model order candidates
p1 <- 15
p2 <- 13
p3 <- 3
p4 <- 6

#Estimation of the VAR models
fit1 <- vars::VAR(train_data, p = p1, season = 12, type = "trend")
fit2 <- vars::VAR(train_data, p = p2, season = 12, type = "trend")
fit3 <- vars::VAR(train_data, p = p3, season = 12, type = "trend")
fit4 <- vars::VAR(train_data, p = p4, season = 12, type = "trend")
```

```
#Checking the stability of the estimated models
vars::roots(fit1)
vars::roots(fit2)
vars::roots(fit3)
vars::roots(fit4)

#fit1 and fit2 are not stable. Proceeding with fit3 and fit4.

#Residual diagnostics
residuals3 <- resid(fit3)
residuals4 <- resid(fit4)


acf(residuals3[,1])
acf(residuals3[,2])
ccf(residuals3[,1])


acf(residuals4[,1])
acf(residuals4[,2])
ccf(residuals4[,1])

#Fitting a higher order model.

p5 <- 9
fit5 <- vars::VAR(train_data, p = p5, season = 12, type = "trend")
vars::roots(fit5)
#Stable.

#Residual diagnostics.
residuals5 <- resid(fit5)
acf(residuals5[,1])
acf(residuals5[,2])
ccf(residuals5[,1])

#Portmanteau tests for estimated VAR(3) (fit3) and VAR(9) (fit5) models.
h1 <- 24
h2 <- 36
h3 <- 48

vars::serial.test(fit3, lags.pt = h1, type = "PT.adjusted")
vars::serial.test(fit3, lags.pt = h2, type = "PT.adjusted")
vars::serial.test(fit3, lags.pt = h3, type = "PT.adjusted")
vars::serial.test(fit5, lags.pt = h1, type = "PT.adjusted")
vars::serial.test(fit5, lags.pt = h2, type = "PT.adjusted")
vars::serial.test(fit5, lags.pt = h3, type = "PT.adjusted")

#Predictions
predval3 <- predict(fit3, n.ahead = 12, ci = 0.95)
predval5 <- predict(fit5, n.ahead = 12, ci = 0.95)

#Estimation with detrended data set.

#Detrending the data.
detrend_wage <- diff(data$Wage, differences = 1)
detrend_insured <- diff(data$Insured, differences = 1)

train_data <- data.frame(Wage = detrend_wage[1:119],
                         Insured = detrend_insured[1:119])
```

```
#Information criteria suggestions for VAR order with maximums 20,13 and 6.
#NOTE! As the data is detrended, we don't set
#the argument 'type' as 'trend'.
vars::VARselect(train_data, lag.max = 20, season = 12)
vars::VARselect(train_data, lag.max = 13, season = 12)
vars::VARselect(train_data, lag.max = 6, season = 12)

#Choosing the model order candidates
p1_d <- 12
p2_d <- 6
p3_d <- 2

#Estimation of the VAR models
fit1_d <- vars::VAR(train_data, p = p1_d, season = 12)
fit2_d <- vars::VAR(train_data, p = p2_d, season = 12)
fit3_d <- vars::VAR(train_data, p = p3_d, season = 12)

#Checking the stability conditions, residual diagnostics and Portmanteau tests
#are done similarly as before.

#Fitting models of order 9 and 1.
p4_d <- 9
p5_d <- 1
fit1_d <- vars::VAR(train_data, p = p4_d, season = 12)
fit2_d <- vars::VAR(train_data, p = p5_d, season = 12)

#Predictions in order VAR(1), VAR(2), VAR(9)
predval5_d <- predict(fit1 , n.ahead = 12 , ci = 0.95)
predval3_d <- predict(fit1 , n.ahead = 12 , ci = 0.95)
predval4_d <- predict(fit1 , n.ahead = 12 , ci = 0.95)

#Converting the predicted values to the
#estimated payrolls and # of insured pers. with VAR(1) model.
start_wage <- data$Wage[120]
start_ins <- data$Insured[120]

predwage <- start_wage + predval5_d$fcst$Wage[1,1]
predval_real_wage <- data.frame("Estim" = c(0,0,0,0,0,0,0,0,0,0,0,0))
predval_real_wage[1,1] <- predwage

predins <- start_ins + predval5_d$fcst$Insured[1,1]
predval_real_ins <- data.frame("Estim" = c(0,0,0,0,0,0,0,0,0,0,0,0))
predval_real_ins[1,1] <- predins

for (i in 2:12){
  predval_real_wage[i,1] <- predval_real_wage[i-1,1] + predval5_d$fcst$Wage[i,1]
  predval_real_ins[i,1] <- predval_real_ins[i-1,1] + predval5_d$fcst$Insured[i,1]
}
```

B.2 Normal approximation

The R code for normal approximation in Section 6.1.

```
#Detrending the data.
detrend_wage <- diff(data$Wage, differences = 1)
detrend_insured <- diff(data$Insured, differences = 1)
```

```
#Monthly values.
data_1 <- c(0,0,0,0,0,0,0,0,0,0)
data_2 <- c(0,0,0,0,0,0,0,0,0,0)
data_3 <- c(0,0,0,0,0,0,0,0,0,0)
data_4 <- c(0,0,0,0,0,0,0,0,0,0)
data_5 <- c(0,0,0,0,0,0,0,0,0,0)
data_6 <- c(0,0,0,0,0,0,0,0,0,0)
data_7 <- c(0,0,0,0,0,0,0,0,0,0)
data_8 <- c(0,0,0,0,0,0,0,0,0,0)
data_9 <- c(0,0,0,0,0,0,0,0,0,0)
data_10 <- c(0,0,0,0,0,0,0,0,0,0)
data_11 <- c(0,0,0,0,0,0,0,0,0,0)
data_12 <- c(0,0,0,0,0,0,0,0,0)

for (i in 1:10) {
  data_1[i] <- detrend_wage[1+(i-1)*12]
}

for (i in 1:10) {
  data_2[i] <- detrend_wage[2+(i-1)*12]
}

for (i in 1:10) {
  data_3[i] <- detrend_wage[3+(i-1)*12]
}

for (i in 1:10) {
  data_4[i] <- detrend_wage[4+(i-1)*12]
}

for (i in 1:10) {
  data_5[i] <- detrend_wage[5+(i-1)*12]
}

for (i in 1:10) {
  data_6[i] <- detrend_wage[6+(i-1)*12]
}

for (i in 1:10) {
  data_7[i] <- detrend_wage[7+(i-1)*12]
}

for (i in 1:10) {
  data_8[i] <- detrend_wage[8+(i-1)*12]
}

for (i in 1:10) {
  data_9[i] <- detrend_wage[9+(i-1)*12]
}

for (i in 1:10) {
  data_10[i] <- detrend_wage[10+(i-1)*12]
}

for (i in 1:10) {
  data_11[i] <- detrend_wage[11+(i-1)*12]
}
```

```
for (i in 1:9) {
  data_12[i] <- detrend_wage[12+(i-1)*12]
}

#Fit the normal distributions for each month.
fit1 <- MASS::fitdistr(data_1,"normal")
fit2 <- MASS::fitdistr(data_2,"normal")
fit3 <- MASS::fitdistr(data_3,"normal")
fit4 <- MASS::fitdistr(data_4,"normal")
fit5 <- MASS::fitdistr(data_5,"normal")
fit6 <- MASS::fitdistr(data_6,"normal")
fit7 <- MASS::fitdistr(data_7,"normal")
fit8 <- MASS::fitdistr(data_8,"normal")
fit9 <- MASS::fitdistr(data_9,"normal")
fit10 <- MASS::fitdistr(data_10,"normal")
fit11 <- MASS::fitdistr(data_11,"normal")
fit12 <- MASS::fitdistr(data_12,"normal")

#Estimated parameters.
para1 <- fit1$estimate
para2 <- fit2$estimate
para3 <- fit3$estimate
para4 <- fit4$estimate
para5 <- fit5$estimate
para6 <- fit6$estimate
para7 <- fit7$estimate
para8 <- fit8$estimate
para9 <- fit9$estimate
para10 <- fit10$estimate
para11 <- fit11$estimate
para12 <- fit12$estimate

#Predictions
set.seed(33)

bs1 <- rnorm(1000,para1[1],para1[2])
bs2 <- rnorm(1000,para2[1],para2[2])
bs3 <- rnorm(1000,para3[1],para3[2])
bs4 <- rnorm(1000,para4[1],para4[2])
bs5 <- rnorm(1000,para5[1],para5[2])
bs6 <- rnorm(1000,para6[1],para6[2])
bs7 <- rnorm(1000,para7[1],para7[2])
bs8 <- rnorm(1000,para8[1],para8[2])
bs9 <- rnorm(1000,para9[1],para9[2])
bs10 <- rnorm(1000,para10[1],para10[2])
bs11 <- rnorm(1000,para11[1],para11[2])
bs12 <- rnorm(1000,para12[1],para12[2])

predval <- c(0,0,0,0,0,0,0,0,0,0,0,0)

predval[1] <- data$Wage[120] + mean(bs12)
predval[2] <- predval[1] + mean(bs1)
predval[3] <- predval[2] + mean(bs2)
predval[4] <- predval[3] + mean(bs3)
predval[5] <- predval[4] + mean(bs4)
predval[6] <- predval[5] + mean(bs5)
predval[7] <- predval[6] + mean(bs6)
predval[8] <- predval[7] + mean(bs7)
predval[9] <- predval[8] + mean(bs8)
```

```
predval[10] <- predval[9] + mean(bs9)
predval[11] <- predval[10] + mean(bs10)
predval[12] <- predval[11] + mean(bs11)
```

Bibliography

[1] Sung K. Ahn. Distribution for residual autocovariances in multivariate autoregressive models with structured parameterization. *Biometrika*, 75(3):590–593, 1988.

[2] Nelville Davies, C. M. Triggs, and Paul Newbold. Significance levels of the Box-Pierce portmanteau statistic in finite samples. *Biometrika*, 64(3):517–522, 12 1977.

[3] Jean-Marie Dufour. Unbiasedness of predictions from etimated vector autoregressions. *Econometric Theory*, 1(3):387–402, 1985.

[4] David Edgerton and Ghazi Shukur. Testing autocorrelation in a system perspective. *Econometric Reviews*, 18(4):343–386, 1999.

[5] Statistics Finland. *111m – Wage and salary indices by industry monthly (2015=100), 1995M01-2020M03*, Accessed March 18, 2020. `https://pxnet2.stat.fi/PXWeb/pxweb/en/StatFin/StatFin__pal__ktps/statfin_ktps_pxt_111m.px/`.

[6] Finnish Centre for Pensions. Työeläkelaitosten tilinpäätöstiedot 2018. *Työeläkelehti*, (3), 2019. Available in Finnish and Swedish only.

[7] Finnish Centre for Pensions. *Contributions under TyEL of employers with an insurance contract*, Accessed March 18, 2020. `https://tilastot.etk.fi/pxweb/en/ETK/ETK__180tyoelakkeiden_rahoitus__50yritysten_tyoelakemaksut/rahti02.px/`.

[8] Wayne A. Fuller. *Introduction to statistical time series.* Wiley series in probability and mathematical statistics. Wiley, New York, 1976.

[9] Clive W. J. Granger. Prediction with a generalized cost of error function. *Operations Research Quarterly*, 20(2):199–207, 1969.

[10] Clive W. J. Granger and Paul Newbold. *Forecasting Economic Time Series.* Elsevier, 2nd edition, 1986.

[11] Jonathan R. M. Hosking. The multivariate portmanteau statistic. *Journal of the American Statistical Association*, 75(371):602–608, 1980.

[12] Richard A. Johnson. *Applied multivariate statistical analysis*. Pearson Prentice Hall, Upper Saddle River, 6th ed edition, 2007.

[13] Greta M. Ljung and George E.P. Box. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303, 1978.

[14] Helmut Lütkepohl. *New introduction to multiple time series analysis*. Springer, Berlin, 2006.

[15] Bernhard Pfaff. *Analysis of Integrated and Cointegrated Time Series with R*. Springer, New York, second edition, 2008. ISBN 0-387-27960-1.

[16] Bernhard Pfaff. Var, svar and svec models: Implementation within R, package vars. *Journal of Statistical Software*, 27(4), 2008.

[17] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020.

[18] Rao C. Radhakrishna and Rao M. Bhaskara. *Matrix algebra and its applications to statistics and econometrics*. World Scientific Pub. Co, Singapore ; Hackensack, N.J., 1998.

[19] James H. Stock and Mark W. Watson. Vector autoregressions. *Journal of Economic Perspectives*, 15(4):101–115, December 2001.

[20] Ruey S. Tsay. *Analysis of financial time series*. Wiley series in probability and statistics. Wiley, Hoboken, New Jersey, 3rd edition, 2010.

[21] Ruey S. Tsay. *Multivariate time series analysis : with R and financial applications*. Wiley series in probability and statistics. Wiley, Hoboken, New Jersey, 2014.

[22] William N. Venables and Brian D. Ripley. *Modern Applied Statistics with S*. Springer, New York, 4th edition, 2002.

[23] Arnold Zellner. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, 57(298):348–368, 1962.