



BRNO UNIVERSITY OF TECHNOLOGY

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FACULTY OF MECHANICAL ENGINEERING

FAKULTA STROJNÍHO INŽENÝRSTVÍ

INSTITUTE OF MATHEMATICS

ÚSTAV MATEMATIKY

TIME SERIES DYNAMIC FACTOR ANALYSIS

DYNAMICKÁ FAKTOROVÁ ANALÝZA ČASOVÝCH ŘAD

MASTER'S THESIS

DIPLOMOVÁ PRÁCE

AUTHOR

AUTOR PRÁCE

Bc. Ľuboš Slávik

SUPERVISOR

VEDOUCÍ PRÁCE

doc. Mgr. Zuzana Hübnerová, Ph.D.

BRNO 2021

Assignment Master's Thesis

Institut: Institute of Mathematics
Student: **Bc. Ľuboš Slávik**
Degree program: Applied Sciences in Engineering
Branch: Mathematical Engineering
Supervisor: **doc. Mgr. Zuzana Hübnerová, Ph.D.**
Academic year: 2020/21

As provided for by the Act No. 111/98 Coll. on higher education institutions and the BUT Study and Examination Regulations, the director of the Institute hereby assigns the following topic of Master's Thesis:

Time series dynamic factor analysis

Brief Description:

Dynamic factor analysis is one of the dimension–reduction techniques. Unlike the principal component method, it aims at time series analysis. The general idea of the dynamic factor model is that each of a set of time series is assumed to be a linear combination of several latent time–varying factors. Dynamic factor analysis can be used to detect common patterns in a set of time series.

Master's Thesis goals:

Definition of the dynamic factor model, description of its properties, and its application in real data analysis.

Recommended bibliography:

MOLENAAR, Peter C.M. A dynamic factor model for the analysis of multivariate time series. *Psychometrika*. 1985, 50, 181–202. ISSN 0033-3123.

CARO, Ángela, PENA, Daniel. Estimation of the common component in Dynamic Factor Models. Madrid: Universidad Carlos III de Madrid. Departamento de Estadística, 2018. UC3M - Working Papers. *Statistics and Econometrics*, 18-3. ISSN 2387-0303.

Deadline for submission Master's Thesis is given by the Schedule of the Academic year 2020/21

In Brno,

L. S.

prof. RNDr. Josef Šlapal, CSc.
Director of the Institute

doc. Ing. Jaroslav Katolický, Ph.D.
FME dean

Abstrakt

Táto diplomová práca sa zaoberá novým prístupom k zhlukovaniu časových rád na základe dynamického faktorového modelu. Dynamický faktorový model je technika redukujúca dimenziu a rozširuje klasickú faktorovú analýzu o požiadavku autokorelačnej štruktúry latentných faktorov. Parametre modelu sa odhadujú pomocou EM algoritmu za použitia Kalmanovho filtra a vyhladzovača a taktiež sú aplikované nevyhnutné podmienky na model, aby sa stal identifikovateľným.

Po tom, ako je v práci predstavený teoretický koncept prístupu, dynamický faktorový model je aplikovaný na skutočné pozorované časové rady a práca skúma jeho správanie a vlastnosti na jednomesačných meteorologických dátach požiarneho indexu (Fire Weather Index) na 108 požiarnych staniciach umiestnených v Britskej Kolumbii. Postup výpočtu modelu odhadne záťažovú maticu (loadings matrix) spolu so zodpovedajúcim malým počtom latentných faktorov a kovariačnou maticou modelovaných časových rád. Diplomová práca aplikuje k-means zhlukovanie na výslednú záťažovú maticu a ponúka rozdelenie meteorologických staníc do zhlukov založené na redukovanej dimenzionalite pôvodných dát.

Vďaka odhadnutým priemerom zhlukov a odhadnutým latentným faktorom je možné získať aj priemerné trendy každého zhluku. Následne sú dosiahnuté výsledky porovnané s výsledkami získanými na dátach z rovnakých staníc avšak iného mesiaca, aby sa stanovila stabilita zhlukovania. Práca sa taktiež zaoberá efektom varimax rotácie záťažovej matice. Diplomová práca navyše navrhuje metódu detekovania odlahlých časových rád založenú na odhadnutej kovariačnej matici modelu a rozoberá dôsledky odlahlých hodnôt na odhanutý model.

Summary

This thesis studies a novel approach to time series clustering based on a dynamic factor model. Dynamic factor model is a dimension reduction technique enhancing classical factor analysis by a requirement of an autocorrelation structure of the latent factors. Parameters of the model are estimated via EM algorithm employing Kalman filtering and smoothing and necessary restrictions are placed on the model, so the model becomes identifiable.

After describing the theoretical concept of the approach, the dynamic factor model is applied to the real observed time series and the work discusses its behaviour and properties on one-month meteorological data of fire weather index at 108 fire stations located in British Columbia. The procedure of the model estimates a loadings matrix of the model with a corresponding small number of latent factors and a variance-covariance matrix of the modeled time series. The thesis applies k-means clustering to the resulted loadings matrix and provides a division of the stations into clusters based on the reduced dimensionality of the original data.

With the estimated cluster means and the latent factors, it is possible to obtain particular mean trends for each cluster. Then, the achieved clusters are compared with the results obtained for the same set of stations but within a different month to assess the stability of the clustering. The work discusses the effect of varimax rotation on the loadings matrix as well. Moreover, the thesis suggests a method for detecting possible time series outliers based on the estimated variance-covariance matrix of the model and discusses the effect of outliers on the estimated model.

Klíčová slova

dynamický faktorový model, zhlukovanie časových rád, Fire Weather Index, detekcia odľahlých časových rád, k-means

Keywords

dynamic factor model, time series clustering, Fire Weather Index, time series outlier detection, k-means

Rozšírený abstrakt

Diplomová práca skúma využitie dynamického faktorového modelu na modelovanie reálnych časových rád a využitie výstupu z tohto modelu na zhlukovanie časových rád. Práca je rozdelená do 3 hlavných častí. V 2. a 3. kapitole sa venujeme teoretickému výkladu dynamického faktorového modelu a metódam použitým pri analytickej práci s výsledkami modelovania. V 4. kapitole vysvetľujeme pôvod a charakter reálnych dát použitých v praktickej analýze. 5. kapitola obsahuje praktickú časť tejto diplomovej práce. Vysvetľujeme tu našu analýzu reálnych dát za použitia dynamického faktorového modelu.

Dynamický faktorový model je technika redukovajúca dimenziu vstupných dát, zatiaľ čo dynamicita ostáva zachovaná. Pracuje s časovými radami, ktoré rozloží na súčet súčiny záťažovej matice (loadings matrix) s malým počtom latentných faktorov a idiosynkratického komponentu (odhadovaná chyba). Metóda vychádza z klasickej faktorovej analýzy a je rozšírená o autokorelačnú požiadavku na latentných faktoroch. Každá časová rada je tým pádom reprezentovaná vektorom záťaží ako súčet záťaží na jednotlivých odhadnutých latentných faktoroch. V literatúre sa uvádza statická a dynamická forma dynamického faktorového modelu. Obidve formy sú ekvivalentné, avšak statická reprezentuje dynamicitu implicitne a je jednoduchšia na výpočet. Taktiež sa modely rozdeľujú na "exaktné" a "približné". Exaktný model nedovoľuje žiadnu kovariáciu zložiek idiosynkratického komponentu (odhadovaná chyba), zatiaľ čo viac realistický približný model dovoľuje istú koreláciu medzi časovými radami. V práci sme odvodili prevod medzi statickou a dynamickou formou, ktorý sa v žiadnej literatúre s ktorou sme pracovali nedá dohľadať a reprezentuje vlastný prínos autora. Nakoniec uvádzame zjednodušenú formu dynamického faktorového modelu ako autoregresný model s oneskorením o jeden časový index. Odhady modelu vychádzajú z predpokladu, že latentné faktory podliehajú viacrozmernému normálnemu rozdeleniu, rovnako ako idiosynkratický komponent a náhodná chyba na autoregresívnom procese latentných faktorov, u ktorých sa však predpokladá už nulová stredná hodnota. Dynamický faktorový model obsahuje veľa hyperparametrov, ktoré nepoznáme a potrebujeme odhadnúť. Tomuto účelu slúži EM algoritmus. Za použitia Kalmanovho filtra a vyhladzovača sa v prvom kroku odhaduje stredná hodnota logaritmu združenej vierohodnostnej funkcie (E krok) a následne sa upravujú hodnoty hyperparametrov tak, aby sa odhadnutá stredná hodnota z predchádzajúceho kroku maximalizovala (M krok). Takýmto spôsobom sa snažíme dosiahnuť optimálneho odhadu parametrov modelu. Následne taktiež vysvetľujeme nevyhnutné podmienky požadované v modeli tak, aby boli odhady parametrov identifikovateľné. Tieto podmienky spôsobujú, že je prvá časová rada modelovaná iba prvým latentným faktorom, druhá časová rada modelovaná prvými dvoma latentnými faktormi atď. Z tohto dôvodu vysvetľujeme varimax rotáciu záťažovej matice a latentných faktorov, ktorá rieši tento problém. V 3. kapitole vysvetľujeme koncept Akaikeho informačného kritéria (AIC), ktoré je často používané ako miera vhodnosti modelu na konkrétnych dátach. Toto kritérium používame v našej analýze na porovnávanie rôznych modelov a selekciu toho najlepšieho. Ďalej vysvetľujeme k-means zhlukovanie. Táto metóda sa snaží nájsť vhodné umiestnenie vopred určeného množstva centier a rozdelenie dátových hodnôt do zhlukov prislúchajúcich konkrétnemu centru tak, aby bola minimalizovaná suma vzdialeností dátových bodov od centra zhluku. Túto metódu aplikujeme v našej analýze na záťažovú maticu a vytvárame tak rozdelenie meteorologických staníc do zhlukov na základe výstupu z dynamického faktorového modelu. Tento prístup je novým prístupom v zhlukovaní časových rád.

Dáta, ktoré pri analýze spracovávame sú hodnoty tzv. Fire Weather Index (FWI). Ide o index, ktorý meria riziko intenzity požiaru v danej lokalite. Je výsledkom merania ďalších 5 indexov, ktoré sú získané na základe 4 základných meteorologických meraní teploty, množstva zrážok, vlhkosti a rýchlosti vetra. Ďalšie indikátory ako typ dreva v lesoch, vlhkosť zeme a podzemných zložiek či potenciál pre vznik požiaru sú tiež súčasťou týchto indexov. Fire Weather Index je jedna zložka Canadian Forest Fire Weather Index System, čo je systém na kontrolu požiarov v Britskej Kolumbii. Slúži predovšetkým na predpovedanie vzniku požiarov a ich intenzity za účelom adekvátnej prípravy a prerozdelenia dostupných a obmedzených ľudských či materiálnych zdrojov a prípravy obyvateľstva na potenciálne nebezpečenstvo.

V kapitole 5 sa venujeme praktickej časti tejto diplomovej práce, a to aplikovaniu dynamického faktorového modelu na FWI dáta. Pracujeme s dátami za mesiace júl a august z roku 2003 a programujeme v jazyku R. Máme hodnoty Fire Weather Index pre 108 meteorologických staníc počas 31 dní. Vybrali sme stanice, ktoré sa nachádzajú v južnej časti Britskej Kolumbie, kde je zväčša vnútrozemská klíma, vyššie teploty, a podľa istých meraní je to oblasť s prevažne najvyššími priemernými hodnotami FWI v rámci celej Kanady, a zároveň sme pre týchto 108 staníc mali plné dáta. Analýzu v práci demonštrujeme na mesiaci júl a neskôr výsledky porovnáваме s mesiacom august, ktorého výsledky analýzy uvádzame v prílohe. Pred odhadom samotného dynamického faktorového modelu sme dáta centralizovali, teda odpočítali od každej časovej rady ich priemernú hodnotu. Samotné hľadanie najlepšieho modelu prebiehalo nasledovne. Pri inicializácii modelu môžeme stanoviť dve premenné: počet faktorov, ktorý očakávame, a štruktúru kovariačnej matice. Dostupné štruktúry kovariačnej matice sú: "diagonálna rovnaká" - odhadujeme prvky na diagonále kovariačnej matice idiosynkratického komponentu, ktoré predpokladáme, že sú rovnaké, prvky mimo diagonály sú odhadnuté nulou; "diagonálna rozdielna" - predpokladáme, že prvky na diagonále sú rozdielne, mimo diagonály rovné nule; "zhodný rozptyl a kovariácia" - odhadujeme všetky prvky na diagonále aj mimo nej s tým, že predpokladáme, že prvky mimo diagonálu sú sebe rovné a prvky na diagonále sú tiež sebe rovné. Počet faktorov, ktoré sme sa pokúšali odhadnúť bol 2 až 5. Pre každú kombináciu kovariačnej štruktúry s daným počtom faktorov sme odhadli parametry zvoleného dynamického faktorového modelu. Výsledné modely sme porovnali na základe upraveného Akaikeho informačného kritéria a uprednostnili sme jednoduchší model (menej faktorov) pred komplexnejším. Ako najlepší model sme na základe týchto kritérií zvolili dynamický faktorový model s 4 latentnými faktormi a "diagonálnou rozdielnou" kovariačnou štruktúrou. V práci sme následne vykreslili histogramy vysvetľujúce konvergenciu parametrov (keďže ide o iteračný proces, máme podmienky konvergenencie pre odhad parametrov), rozdelenie záťaží na jednotlivých časových radách zo záťažovej matice či priebeh latentných faktorov. Záťažovú maticu spolu s latentnými faktormi sme uviedli a porovnali pred varimax rotáciou, ako aj po nej. Vďaka varimax rotácii sa nám podarilo prekonať spomenutú podmienku a prvé štyri časové rady sú riadne modelované všetkými štyrmi faktormi. Taktiež druhý sľubovaný efekt varimax rotácie, a to prehĺbenie rozdielu medzi záťažami jednotlivých faktorov, sa nám nepodarilo dostatočne preukázať. Odhad kovariačnej matice ponúka možnosť ako detekovať odľahlé časové rady. V práci sme použili prístup, v ktorom sme najprv pomocou Shapiro-Wilk testu dobrej zhody hľadali rozdelenie, ktorým by sa odhady rozptylov časových rád mohli riadiť, a potom sme stanovili 99 % percentil pre odhadnuté rozdelenie. Na základe zvoleného kritéria dobrej zhody sa ako najvhodnejšie javilo lognormálne rozdelenie. Hodnoty rozptylov,

ktoré boli mimo tohto 99% percentilu lognormálneho rozdelenia s odhadnutými parametrami sme prehlásili za odľahlé hodnoty. Takto bola identifikovaná jedna časová rada ako odľahlá. Ďalej sme v práci vytvorili zhluky na základe k-means zhlukovania. Ako vhodný počet zhlukov nám vyšiel 4. Stanice sme vykreslili na mape s farbami prislúchajúcimi daným zhlukom. Ďalej sme zobrazili trendy časových rád rozdelených do zhlukov spolu s priemerným trendom, čo je zaujímavý prínos tohto prístupu. Rovnako tak sme boli schopní vykresliť rozdelenie záťaží jednotlivých faktorov na časových radách rozdelených do zhlukov a sledovať zmeny v rámci samotných zhlukov. V poslednej časti tejto kapitoly sme porovnali dosiahnuté rozdelenie do zhlukov s najlepším modelom odhadnutým na augustových dátach. Z vizuálneho porovnania sme sledovali, že rozdelenie staníc do zhlukov sa trochu zmenilo, čo potvrdili aj číselné charakteristiky Rand Index a Adjusted Rand Index. Na augustových dátach už nepozorujeme také výrazné zhluky v rámci mapy, avšak hlavné centrum jednotlivých zhlukov spolu prislúchajúcimi stanicami zostalo pomerne zachované. V práci sme ukázali možnosti využitia nového prístupu k zhlukovaniu časových rád na základe dynamického faktorového modelu a k-means zhlukovaniu aplikovaného na záťažovú maticu výsledného dynamického faktorového modelu.

SLÁVIK, L. *Time series dynamic factor analysis*. Brno: Vysoké učení technické v Brně, Fakulta strojního inženýrství, 2021. 79 s. Vedoucí doc. Mgr. Zuzana Hübnerová, Ph.D.

I declare that I have written the Master's thesis *Time series dynamic factor analysis* on my own and under the guidance of my Master's thesis supervisor doc. Mgr. Zuzana Hübnerová, Ph.D. and using the sources listed in references.

Bc. Luboš Slávik

I would like to express huge gratitude to my supervisor doc. Mgr. Zuzana Hübnerová, Ph.D. for her incredibly human attitude and professional approach to me and to this thesis. Her guidance was a remarkable part of my studies and I will never forget your help. Thank you.

Bc. Luboš Slávik

Contents

1	Introduction	17
2	Dynamic factor model	18
2.1	Factor model	18
2.2	Lag operator	20
2.3	Dynamic factor model	21
2.3.1	Dynamic form	21
2.3.2	Static form	23
2.3.3	Simplified form	27
2.4	Joint density function	27
2.5	Kalman filter	28
2.6	Kalman smoother	29
2.7	Joint log-likelihood function	30
2.7.1	EM algorithm	31
2.8	Identification	32
2.9	Varimax rotation	33
3	Results assessment	35
3.1	Akaike information criterion	35
3.2	k-means clustering	36
4	Fire Weather Index	37
4.1	Canadian Forest Fire Danger Rating System	37
4.2	Canadian Forest Fire Weather Index System	38
5	Application of a dynamic factor model	40
5.1	Data summary	40
5.2	Data pre-processing	43
5.3	MARSS package	44
5.3.1	Variance-covariance matrix R	44
5.3.2	Convergence tests	45
5.3.3	Time consumption	46
5.4	Data analysis	46
5.4.1	Best model fit	51
5.4.2	Outliers detection	56
5.5	Clustering	57
5.5.1	Outliers	61
5.5.2	Model without outliers	62
5.6	Comparison with August data	62
6	Conclusion	65
	References	66
	Used symbols	69

CONTENTS

Attachments	71
MARSS initialization in R	71
Log-log slope test	73
Consequence of restriction on latent factors	74
Results for August data	75

1 Introduction

Time series are all around us. Every day we see changes in temperature, electricity usage, popularity charts, stock markets, or many other economical aspects. The one that is able to process this information in one's favor may benefit from its power and use it to his advantage. We want to predict, we want to systemize, we want to group, we want to find hidden patterns. The reason for this is very simple. There is much more information than we can see. And to get this information, we need to dig deeper. We need to create a model and apply it to real-world observations. With mathematical interpretation new doors open to data observation.

One of these models is also a Dynamic Factor Model. It is a dimension reduction technique applied to time series. In [Chapter 2](#) we will state its definition and discover its properties. We will find necessary restrictions that need to be placed on the model just to be able to estimate the parameters of the model. Then with restrictions placed on the model, we will discuss the EM algorithm together with Kalman filter and smoother, which are methods suitable for model estimation. Then we will explain Varimax rotation, which we need to rotate model output to overcome restrictions placed on the model.

In [Chapter 3](#) we will explain AIC, which we use for comparing the models and choosing the most suitable one for our observed data. In [\[1\]](#) many methods for time series clustering are explained. In this thesis, we introduce a new technique. By estimating a dynamic factor model we reduce the dimensionality of our original data and obtain an estimated loadings matrix. Then on this matrix, we apply k-means clustering. This way we are able to obtain time series division into clusters and also thanks to estimated latent factors and cluster means, we are able to plot individual cluster mean trends.

The data we will work with in this thesis are values of the "Fire Weather Index". In [Chapter 4](#) we will explain, that it is a component of the fire management system used to prevent uncontrolled wildfires and manage resources in British Columbia since the year 1970. The index measures potential fire intensity in case of fire ignition. It combines 5 different indices and is computed based on meteorological measurements of precipitation, wind speed, temperature, and moisture.

In [Chapter 5](#) we will provide our data analysis of Fire Weather Index measurements. We have measurements for months July and August in the year 2003. First, we will focus on the month July and break down each step of the analysis. We will examine the inputs and the outputs of an implementation of a dynamic factor model using the Multivariate Auto-Regressive State-Space (MARSS) package in R and provide graphs with possible visualizations gathered from the outputs. Then we will execute a k-means clustering on rotated loadings matrix of the most suitable dynamic factor model and divide stations into clusters. We will show a map of the southern location of British Columbia with stations divided into clusters and station trends within clusters with its cluster means. We will discuss also outliers detection as a possible result of the dynamic factor model. In the end, we will compare the obtained results from July data with next month's August data and see, how the clusters and trends change with next month's FWI measurements.

2 Dynamic factor model

Dynamic factor model is a model of dynamic factor analysis (DFA), which is a statistical method that allows us to reduce the dimensionality of data with preserving the information about the lags and their effect on our observations in time. When examining a big amount of time series a need for discovering underlying processes arises. Finding common patterns, latent factors as we call them, is an important part of data analysis, the basis for further work such as detection of similarities, finding common influences on the time series, classification, prediction. Dynamic factor model helps us understand our data, especially big amount of data, by describing the covariance relationship among many variables in terms of a few underlying factors [2] and simultaneously describing the effect of these factors on observations through their lags [3].

This chapter will focus on explaining the mathematical background of dynamic factor models and their basic description in a state-space form. The first section will introduce a classical factor model as a basis for understanding the origin of a dynamic factor model. Further, we'll explain the needed restrictions for solving the equation as the classical optimization problem is not identifiable. We will provide EM algorithm as a method for estimating the unknown parameters.

2.1 Factor model

According to [2] and [4] let \mathbf{X} be an observable p -dimensional random vector $\mathbf{X} = (X_1, X_2, \dots, X_p)'$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Then

$$\begin{aligned} X_1 &= l_{11}F_1 + l_{12}F_2 + \dots + l_{1m}F_m + \epsilon_1 \\ X_2 &= l_{21}F_1 + l_{22}F_2 + \dots + l_{2m}F_m + \epsilon_2 \\ &\vdots \qquad \qquad \qquad \vdots \\ X_p &= l_{p1}F_1 + l_{p2}F_2 + \dots + l_{pm}F_m + \epsilon_p \end{aligned} \tag{2.1}$$

is called *factor model*, where $\mathbf{F} = (F_1, F_2, \dots, F_m)'$ is a m -dimensional vector of *common factors*, coefficients l_{ij} are called *factor loadings* of j th factor on i th variable and $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \dots, \epsilon_p)$ is a random vector of additional sources of variation called *errors*. We normally center our data of observations and since l_{ij} creates a matrix, we denote

$$\mathbf{X} - \boldsymbol{\mu} = \mathbf{L}\mathbf{F} + \boldsymbol{\epsilon}, \tag{2.2}$$

where $\mathbf{X} - \boldsymbol{\mu}$ is a $p \times 1$ vector of centered random variables and $\mathbf{L} = (l_{ij})$ is a $p \times m$ matrix called *matrix of factor loadings*. In this model \mathbf{X} is observable, \mathbf{F} , $\boldsymbol{\epsilon}$ are unobservable and \mathbf{L} is to be estimated.

Since the model has so many unobservable quantities, it is convenient to have additional assumptions about \mathbf{F} and $\boldsymbol{\epsilon}$:

- (i) Mean values of both vectors are zero

$$E(\mathbf{F}) = \mathbf{0} \tag{2.3}$$

$$E(\boldsymbol{\epsilon}) = \mathbf{0}. \tag{2.4}$$

(ii) Variance matrix of factors is $m \times m$ identity matrix

$$\text{var}(\mathbf{F}) = \text{cov}(\mathbf{F}, \mathbf{F}) = E[\mathbf{F}\mathbf{F}'] = \mathbf{I}. \quad (2.5)$$

This means, that all of the factors are uncorrelated with one another.

(iii) Variance matrix of errors is a $p \times p$ diagonal matrix

$$\text{var}(\boldsymbol{\epsilon}) = E[\boldsymbol{\epsilon}\boldsymbol{\epsilon}'] = \boldsymbol{\Psi} = \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & 0 & \ddots & 0 \\ 0 & 0 & \cdots & \psi_p \end{bmatrix}, \quad (2.6)$$

which means there is no cross-sectional dependence.

(iv) \mathbf{F} and $\boldsymbol{\epsilon}$ are uncorrelated

$$\text{cov}(\boldsymbol{\epsilon}, \mathbf{F}) = \text{cov}(\mathbf{F}, \boldsymbol{\epsilon}) = E[\boldsymbol{\epsilon}\mathbf{F}'] = \mathbf{0}. \quad (2.7)$$

Altogether these assumptions constitute an *orthogonal factor model*.

Orthogonal factor model determines the covariance structure of \mathbf{X}

$$\begin{aligned} (\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})' &= (\mathbf{L}\mathbf{F} + \boldsymbol{\epsilon})(\mathbf{L}\mathbf{F} + \boldsymbol{\epsilon})' \\ &= (\mathbf{L}\mathbf{F} + \boldsymbol{\epsilon})((\mathbf{L}\mathbf{F})' + \boldsymbol{\epsilon}') \\ &= (\mathbf{L}\mathbf{F}(\mathbf{L}\mathbf{F})') + \boldsymbol{\epsilon}(\mathbf{L}\mathbf{F})' + \mathbf{L}\mathbf{F}\boldsymbol{\epsilon}' + \boldsymbol{\epsilon}\boldsymbol{\epsilon}' \end{aligned} \quad (2.8)$$

so

$$\begin{aligned} \boldsymbol{\Sigma} = \text{cov}(\mathbf{X}) &= E((\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})') \\ &= \mathbf{L}E(\mathbf{F}\mathbf{F}')\mathbf{L}' + E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}')\mathbf{L}' + \mathbf{L}E(\mathbf{F}\boldsymbol{\epsilon}') + E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') \\ &= \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}. \end{aligned} \quad (2.9)$$

Because \mathbf{F} and $\boldsymbol{\epsilon}$ are uncorrelated (2.7), we have

$$(\mathbf{X} - \boldsymbol{\mu})\mathbf{F}' = (\mathbf{L}\mathbf{F} + \boldsymbol{\epsilon})\mathbf{F}' = \mathbf{L}\mathbf{F}\mathbf{F}' + \boldsymbol{\epsilon}\mathbf{F}' \quad (2.10)$$

and

$$\text{cov}(\mathbf{X}, \mathbf{F}) = E(\mathbf{X} - \boldsymbol{\mu})\mathbf{F}' = \mathbf{L}E(\mathbf{F}\mathbf{F}') + E(\boldsymbol{\epsilon}\mathbf{F}') = \mathbf{L}. \quad (2.11)$$

This defines a covariance structure for the orthogonal factor model (2.9) as

$$\begin{aligned} \text{var}(X_i) &= l_{i1}^2 + \cdots + l_{im}^2 + \psi_i \\ \text{cov}(X_i, X_k) &= l_{i1}l_{k1} + \cdots + l_{im}l_{km} \end{aligned} \quad (2.12)$$

and (2.11) as

$$\text{cov}(X_i, F_j) = l_{ij}. \quad (2.13)$$

2.2. LAG OPERATOR

The defined orthogonal model is *linear* in the common factors. From (2.2) we can write

$$x_i = \sum_{j=1}^m l_{ij} f_j + \mu_i + \epsilon_i, \quad i = 1, 2, \dots, p, \quad (2.14)$$

and by specifying

$$\text{var}(X_i) = \sigma_{ii}, \quad (2.15)$$

together from (2.12) we obtain

$$\sigma_{ii} = \sum_{j=1}^m l_{ij}^2 + \psi_i. \quad (2.16)$$

Then by denoting

$$h_i^2 = \sum_{j=1}^m l_{ij}^2 \quad (2.17)$$

we have

$$\sigma_{ii} = h_i^2 + \psi_i, \quad i = 1, 2, \dots, p, \quad (2.18)$$

where the variance of i th variable contributed by the m common factors $l_{i1}^2 + l_{i2}^2 + \dots + l_{im}^2$ is called *i th communality* and ψ_i is called *specific variance*. We see that i th communality is the sum of squares of the loadings of the i th variable on m common factors.

When $m = p$, any covariance matrix Σ can be represented as in (2.9) with Ψ equal to zero matrix. However, factor analysis is most useful for us when $m < p$, so p observations are represented by much less factors and provides a more simple explanation of covariation in \mathbf{X} with fewer parameters than original $p(p+1)/2$ parameters in Σ . With $m > 1$ there is some ambiguity in the factor model because of possibility of two factor models leading to same properties. We will discuss this issue regarding time series in section 2.8.

2.2 Lag operator

Let $\{Y_t\}$ be a time series defined as in [5]. Then a *lag operator* L maps a sequence $\{Y_t\}$ into

$$LY_t = Y_{t-1}, \quad (2.19)$$

for all t which means, that lag operator produces the previous element of a particular time series. Lag operator may be applied repeatedly, so

$$L^k Y_t = Y_{t-k} \quad (2.20)$$

for all $k \in \mathbb{N}$. We may apply lag operator to a constant c or an inversion to a lag operator and obtain

$$\begin{aligned} Lc &= c \\ L^{-1}Y_t &= Y_{t+1}. \end{aligned} \quad (2.21)$$

Lag polynomials are commonly used to describe autoregressive (AR) processes. Let $\phi(L)$ be a one-sided lag polynomial of k th degree

$$\phi(L) = \sum_{i=0}^k \phi_i L^i = \phi_0 L^0 + \phi_1 L^1 + \dots + \phi_k L^k, \quad (2.22)$$

which means it is a polynomial of lag operators. Then a lag polynomial $\phi(L)$ applied to a process Y_t is

$$\begin{aligned} \phi(L)Y_t &= (\phi_0 L^0 + \phi_1 L^1 + \dots + \phi_k L^k)Y_t \\ &= \phi_0 L^0 Y_t + \phi_1 L^1 Y_t + \dots + \phi_k L^k Y_t \\ &= \phi_0 Y_t + \phi_1 Y_{t-1} + \dots + \phi_k Y_{t-k}. \end{aligned} \quad (2.23)$$

2.3 Dynamic factor model

As stated in [6], the dynamic factor model (DFM) represents the evolution of a vector of p observed time series \mathbf{X}_t , where our real data (observed time series) are modeled by a reduced number of unobserved common (latent) factors which evolve over time, plus uncorrelated disturbances which represent measurement error and/or idiosyncratic dynamics (noise) of the individual series. We distinguish between *dynamic* and *static* form of the dynamic factor model and between *exact* and *approximate* dynamic factor model. The dynamic form represents the dependence of observed time series \mathbf{X}_t on lags of the factors explicitly, it is more intuitive in terms of understanding the dynamics of the model and more realistic, but harder to estimate. The static form represents the dynamics implicitly. It can be equivalent to dynamic form and it is used by MARSS package [29] in R, which we will use further in real data application. The difference between exact and approximate model is in cross-sectional correlation of idiosyncratic component. Exact model does not allow any cross-sectional correlation, while approximate model, which is more realistic in applications, allows for existence of certain correlation across time series.

2.3.1 Dynamic form

Let \mathbf{X}_t be a $p \times 1$ vector of time series of interest depending on m unobserved latent factors \mathbf{f}_t and a mean-zero idiosyncratic component $\boldsymbol{\epsilon}_t$ as in (2.4), where latent factors and idiosyncratic terms are in general correlated, thus (2.7) does not hold. Dynamic form of the model can be written as

$$\mathbf{X}_t = \boldsymbol{\lambda}(L)\mathbf{f}_t + \boldsymbol{\epsilon}_t \quad (2.24)$$

$$\mathbf{f}_t = \boldsymbol{\Psi}(L)\mathbf{f}_{t-1} + \boldsymbol{\eta}_t, \quad (2.25)$$

where the lag polynomial matrix $\boldsymbol{\lambda}(L)$ is $p \times m$, $\boldsymbol{\Psi}(L)$ is $m \times m$ and $\boldsymbol{\eta}_t$ is $m \times 1$ vector of mean-zero error on factors. Errors are assumed to be uncorrelated at all leads and lags, so that

$$E(\boldsymbol{\epsilon}_t \boldsymbol{\eta}'_{t-k}) = 0 \quad \text{for all } k. \quad (2.26)$$

Idiosyncratic component $\boldsymbol{\epsilon}_t$ can be serially correlated [5]. Lag polynomial matrix $\boldsymbol{\lambda}(L)$ would look like

2.3. DYNAMIC FACTOR MODEL

$$\boldsymbol{\lambda}(L) = \begin{bmatrix} \lambda^{11}(L) & \lambda^{12}(L) & \cdots & \lambda^{1m}(L) \\ \lambda^{21}(L) & \lambda^{22}(L) & \cdots & \lambda^{2m}(L) \\ \vdots & \vdots & \ddots & \vdots \\ \lambda^{p1}(L) & \lambda^{p2}(L) & \cdots & \lambda^{pm}(L) \end{bmatrix}, \quad (2.27)$$

where $\lambda^{ij}(L)$ stands for a lag polynomial on the i th row and j th column.

For the i th time series X_{it} , the equation looks like

$$X_{it} = \lambda_i(L) \mathbf{f}_t + \epsilon_{it}, \quad (2.28)$$

where the i th row

$$\lambda_i(L) = (\lambda^{i1}(L), \lambda^{i2}(L), \dots, \lambda^{im}(L)) \quad (2.29)$$

of $\boldsymbol{\lambda}(L)$ is called the dynamic factor loading on the i th series and $\lambda_i(L) \mathbf{f}_t$ is called *common component* of the i th series. We treat the lag polynomial $\boldsymbol{\lambda}(L)$ as one sided.

Let us assume m latent factors $\mathbf{f}_t = (f_{1t}, f_{2t}, \dots, f_{mt})'$ and s th degree of a lag polynomial, which means

$$\lambda(L) = \lambda_0 L^0 + \lambda_1 L^1 + \dots + \lambda_s L^s. \quad (2.30)$$

Then for the particular time series X_{it} we could write

$$\begin{aligned} X_{it} &= \lambda_i(L) \mathbf{f}_t + \epsilon_{it} \\ &= [\lambda^{i1}(L) \quad \lambda^{i2}(L) \quad \cdots \quad \lambda^{im}(L)] \begin{bmatrix} f_{1t} \\ f_{2t} \\ \vdots \\ f_{mt} \end{bmatrix} + \epsilon_{it} \\ &= \lambda^{i1}(L) f_{1t} + \lambda^{i2}(L) f_{2t} + \cdots + \lambda^{im}(L) f_{mt} + \epsilon_{it} \\ &= \sum_{j=0}^s \lambda_j^{i1} f_{1t-j} + \sum_{j=0}^s \lambda_j^{i2} f_{2t-j} + \cdots + \sum_{j=0}^s \lambda_j^{im} f_{mt-j} + \epsilon_{it} \\ &= \lambda_0^{i1} f_{1t} + \lambda_1^{i1} f_{1t-1} + \cdots + \lambda_s^{i1} f_{1t-s} \\ &\quad + \lambda_0^{i2} f_{2t} + \lambda_1^{i2} f_{2t-1} + \cdots + \lambda_s^{i2} f_{2t-s} \\ &\quad + \cdots \\ &\quad + \lambda_0^{im} f_{mt} + \lambda_1^{im} f_{mt-1} + \cdots + \lambda_s^{im} f_{mt-s} + \epsilon_{it}, \end{aligned} \quad (2.31)$$

where

$$\begin{aligned} \lambda^{ij}(L) f_{jt} &= (\lambda_0^{ij} L^0 + \lambda_1^{ij} L^1 + \cdots + \lambda_s^{ij} L^s) f_{jt} \\ &= \lambda_0^{ij} f_{jt} + \lambda_1^{ij} f_{jt-1} + \cdots + \lambda_s^{ij} f_{jt-s}, \end{aligned} \quad (2.32)$$

where $i = 1, \dots, p$ denotes particular time series and $j = 1, \dots, m$ denotes latent factors. Then λ_0^{ij} is loading of 0th lag of j th latent factor on i th time series.

For the vector autoregression (2.25) of m factors \mathbf{f}_t that has r lags, we may write

$$\mathbf{f}_t = \begin{bmatrix} f_{1t} \\ f_{2t} \\ \vdots \\ f_{mt} \end{bmatrix} = \begin{bmatrix} \Psi_1(L) & 0 & \cdots & 0 \\ 0 & \Psi_2(L) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Psi_m(L) \end{bmatrix} \begin{bmatrix} f_{1t-1} \\ f_{2t-1} \\ \vdots \\ f_{mt-1} \end{bmatrix} + \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_m \end{bmatrix}, \quad (2.33)$$

where

$$\Psi_i(L) = \Psi_1^i L^0 + \Psi_2^i L^1 + \cdots + \Psi_r^i L^{r-1}. \quad (2.34)$$

For i th factor $i = 1, \dots, m$ (2.33) is written as

$$\begin{aligned} f_{it} &= \Psi_i(L) f_{it-1} + \eta_i = (\Psi_1^i L^0 + \Psi_2^i L^1 + \cdots + \Psi_r^i L^{r-1}) f_{it-1} + \eta_i \\ &= \Psi_1^i f_{it-1} + \Psi_2^i f_{it-2} + \cdots + \Psi_r^i f_{it-r} + \eta_i, \end{aligned} \quad (2.35)$$

where Ψ_j^i stands for autoregression coefficient of j th lag on i th factor.

2.3.2 Static form

Static form rewrites the dynamic form (2.24) and (2.25) to depend on n static factors \mathbf{F}_t . Dynamic form involves m dynamic factors \mathbf{f}_t , while it is assumed that $n \geq m$. According to [6], rewriting a dynamic form into a static form makes the model more amenable to principal component analysis.

Let the lag polynomial matrix $\boldsymbol{\lambda}(L)$ be of s th degree lag polynomials and let $\mathbf{F}_t = (\mathbf{f}'_t, \mathbf{f}'_{t-1}, \dots, \mathbf{f}'_{t-s})'$ be a $n \times 1$ vector of *static* factors in contrast to dynamic factors. For m dynamic factors we can write

$$\mathbf{F}_t = \begin{bmatrix} \mathbf{f}'_t \\ \mathbf{f}'_{t-1} \\ \vdots \\ \mathbf{f}'_{t-s} \end{bmatrix} = \begin{bmatrix} (f_{1t}, f_{2t}, \dots, f_{mt})' \\ (f_{1t-1}, f_{2t-1}, \dots, f_{mt-1})' \\ \vdots \\ (f_{1t-s}, f_{2t-s}, \dots, f_{mt-s})' \end{bmatrix} = \begin{bmatrix} f_{1t} \\ f_{2t} \\ \vdots \\ f_{mt} \\ f_{1t-1} \\ f_{2t-1} \\ \vdots \\ f_{mt-1} \\ \vdots \\ f_{1t-s} \\ f_{2t-s} \\ \vdots \\ f_{mt-s} \end{bmatrix} \quad (2.36)$$

Then let $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}_0, \boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_s)$ be a $p \times (m \cdot s)$ matrix of matrices $\boldsymbol{\lambda}_k$ ($k = 1, \dots, s$), where

$$\boldsymbol{\lambda}_k = \begin{bmatrix} \lambda_k^{11} & \lambda_k^{12} & \cdots & \lambda_k^{1m} \\ \lambda_k^{21} & \lambda_k^{22} & \cdots & \lambda_k^{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_k^{p1} & \lambda_k^{p2} & \cdots & \lambda_k^{pm} \end{bmatrix} \quad (2.37)$$

2.3. DYNAMIC FACTOR MODEL

is a $p \times m$ matrix of coefficients on the k th lag of $\boldsymbol{\lambda}(L)$ in (2.24). Also let $\Phi(L)$ be the matrix consisting of ones, zeros and elements of $\Psi(L)$ such that (2.25) is rewritten in terms of \mathbf{F}_t . Then we can rewrite dynamic form (2.24) and (2.25) into static form

$$\mathbf{X}_t = \Lambda \mathbf{F}_t + \boldsymbol{\epsilon}_t \quad (2.38)$$

$$\mathbf{F}_t = \Phi \mathbf{F}_{t-1} + \mathbf{G}\boldsymbol{\eta}_t. \quad (2.39)$$

If we join together (2.36), (2.37), (2.38) and break it down, we obtain

$$\begin{aligned} \mathbf{X}_t &= [\lambda_0 \quad \lambda_1 \quad \cdots \quad \lambda_s] \begin{bmatrix} \mathbf{f}'_t \\ \mathbf{f}'_{t-1} \\ \vdots \\ \mathbf{f}'_{t-s} \end{bmatrix} + \boldsymbol{\epsilon}_t = \lambda_0 \mathbf{f}'_t + \lambda_1 \mathbf{f}'_{t-1} + \cdots + \lambda_s \mathbf{f}'_{t-s} + \boldsymbol{\epsilon}_t \\ &= \begin{bmatrix} \lambda_0^{11} & \cdots & \lambda_0^{1m} \\ \vdots & \ddots & \vdots \\ \lambda_0^{p1} & \cdots & \lambda_0^{pm} \end{bmatrix} \begin{bmatrix} f_{1t} \\ f_{2t} \\ \vdots \\ f_{mt} \end{bmatrix} + \begin{bmatrix} \lambda_1^{11} & \cdots & \lambda_1^{1m} \\ \vdots & \ddots & \vdots \\ \lambda_1^{p1} & \cdots & \lambda_1^{pm} \end{bmatrix} \begin{bmatrix} f_{1t-1} \\ f_{2t-1} \\ \vdots \\ f_{mt-1} \end{bmatrix} + \cdots \\ &+ \begin{bmatrix} \lambda_s^{11} & \cdots & \lambda_s^{1m} \\ \vdots & \ddots & \vdots \\ \lambda_s^{p1} & \cdots & \lambda_s^{pm} \end{bmatrix} \begin{bmatrix} f_{1t-s} \\ f_{2t-s} \\ \vdots \\ f_{mt-s} \end{bmatrix} + \boldsymbol{\epsilon}_t. \end{aligned} \quad (2.40)$$

Then for particular time series X_{it} that depends on up to s th lagged values of \mathbf{f}_t , we get

$$\begin{aligned} X_{it} &= \lambda_0^{i1} f_{1t} + \lambda_0^{i2} f_{2t} + \cdots + \lambda_0^{im} f_{mt} \\ &+ \lambda_1^{i1} f_{1t-1} + \lambda_1^{i2} f_{2t-1} + \cdots + \lambda_1^{im} f_{mt-1} \\ &+ \cdots \\ &+ \lambda_s^{i1} f_{1t-s} + \lambda_s^{i2} f_{2t-s} + \cdots + \lambda_s^{im} f_{mt-s} + \epsilon_{it}, \end{aligned} \quad (2.41)$$

which is exactly the same as (2.31).

Now let us see, how dynamic form is equal with static form in vector autoregression function. Let us assume m dynamic factors \mathbf{f}_t and vector autoregression for \mathbf{f}_t has r lags, so $f_{it} = \Psi_1^i f_{it-1} + \Psi_2^i f_{it-2} + \cdots + \Psi_r^i f_{it-r}$. Then for (2.25) and (2.39) correspondence may be written as

$$\Phi = \begin{bmatrix} \Psi_1^1 & 0 & \cdots & 0 & \Psi_2^1 & 0 & \cdots & 0 & \cdots & \Psi_r^1 & 0 & \cdots & 0 \\ 0 & \Psi_1^2 & \cdots & 0 & 0 & \Psi_2^2 & \cdots & 0 & \cdots & 0 & \Psi_r^1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Psi_1^m & 0 & 0 & \cdots & \Psi_2^m & \cdots & 0 & 0 & \cdots & \Psi_r^m \\ 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix}, \quad (2.42)$$

$$\mathbf{F}_t = \Phi \mathbf{F}_{t-1} + \mathbf{G} \boldsymbol{\eta}_t$$

$$\begin{bmatrix} f_{1t} \\ f_{2t} \\ \vdots \\ f_{mt} \\ f_{1t-1} \\ f_{2t-1} \\ \vdots \\ f_{mt-1} \\ \vdots \\ f_{1t-r+1} \\ f_{2t-r+1} \\ \vdots \\ f_{mt-r+1} \end{bmatrix} = \Phi \begin{bmatrix} f_{1t-1} \\ f_{2t-1} \\ \vdots \\ f_{mt-1} \\ f_{1t-2} \\ f_{2t-2} \\ \vdots \\ f_{mt-2} \\ \vdots \\ f_{1t-r} \\ f_{2t-r} \\ \vdots \\ f_{mt-r} \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \boldsymbol{\eta}_t = \begin{bmatrix} \Psi_1^1 f_{1t-1} + \Psi_2^1 f_{1t-2} + \cdots + \Psi_r^1 f_{1t-r} \\ \Psi_1^2 f_{2t-1} + \Psi_2^2 f_{2t-2} + \cdots + \Psi_r^2 f_{2t-r} \\ \vdots \\ \Psi_1^m f_{mt-1} + \Psi_2^m f_{mt-2} + \cdots + \Psi_r^m f_{mt-r} \\ f_{1t-1} \\ f_{2t-1} \\ \vdots \\ f_{mt-1} \\ \vdots \\ f_{1t-r+1} \\ f_{2t-r+1} \\ \vdots \\ f_{mt-r+1} \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \boldsymbol{\eta}_t$$

$$= \begin{bmatrix} \Psi_1^1 L^0 f_{1t-1} + \Psi_2^1 L^1 f_{1t-1} + \cdots + \Psi_r^1 L^{r-1} f_{1t-1} \\ \Psi_1^2 L^0 f_{2t-1} + \Psi_2^2 L^1 f_{2t-1} + \cdots + \Psi_r^2 L^{r-1} f_{2t-1} \\ \vdots \\ \Psi_1^m L^0 f_{mt-1} + \Psi_2^m L^1 f_{mt-1} + \cdots + \Psi_r^m L^{r-1} f_{mt-1} \\ f_{1t-1} \\ f_{2t-1} \\ \vdots \\ f_{mt-1} \\ \vdots \\ f_{1t-r+1} \\ f_{2t-r+1} \\ \vdots \\ f_{mt-r+1} \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \boldsymbol{\eta}_t$$

2.3. DYNAMIC FACTOR MODEL

$$\begin{aligned}
&= \begin{bmatrix} (\Psi_1^1 L^0 + \Psi_2^1 L^1 + \dots + \Psi_r^1 L^{r-1}) f_{1t-1} \\ (\Psi_1^2 L^0 + \Psi_2^2 L^1 + \dots + \Psi_r^2 L^{r-1}) f_{2t-1} \\ \vdots \\ (\Psi_1^m L^0 + \Psi_2^m L^1 + \dots + \Psi_r^m L^{r-1}) f_{mt-1} \\ f_{1t-1} \\ f_{2t-1} \\ \vdots \\ f_{mt-1} \\ \vdots \\ f_{1t-r+1} \\ f_{2t-r+1} \\ \vdots \\ f_{mt-r+1} \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \boldsymbol{\eta}_t \\
&= \begin{bmatrix} \Psi_1(L) f_{1t-1} \\ \Psi_2(L) f_{2t-1} \\ \vdots \\ \Psi_m(L) f_{mt-1} \\ f_{1t-1} \\ f_{2t-1} \\ \vdots \\ f_{mt-1} \\ \vdots \\ f_{1t-r+1} \\ f_{2t-r+1} \\ \vdots \\ f_{mt-r+1} \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \boldsymbol{\eta}_t \rightarrow \begin{bmatrix} \Psi_1(L) & 0 & \dots & 0 \\ 0 & \Psi_2(L) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Psi_m(L) \end{bmatrix} \begin{bmatrix} f_{1t-1} \\ f_{2t-1} \\ \vdots \\ f_{mt-1} \end{bmatrix} + \boldsymbol{\eta}_t \\
&= \boldsymbol{\Psi}(L) \mathbf{f}_{t-1} + \boldsymbol{\eta}_t. \tag{2.43}
\end{aligned}$$

As we can see in (2.36), \mathbf{F}_t consists of stacked current and past values of \mathbf{f}_t , that is why we may call it also a *stacked* form. In general, the number of static factors m exceeds the number of dynamic factors n . When $n > m$, the static factors have a dynamic singularity, that means, $n - m$ linear combinations of \mathbf{F}_t are predictable from past \mathbf{F}_t . For a single dynamic factor $m = 1$, two static factors $n = 2$ and two lags on f_t we may write

$$\mathbf{F}_t = \begin{bmatrix} F_{1t} \\ F_{2t} \end{bmatrix} = \begin{bmatrix} f_t \\ f_{t-1} \end{bmatrix} = \begin{bmatrix} \Psi_1^1 & \Psi_2^1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} f_{t-1} \\ f_{t-2} \end{bmatrix} = \boldsymbol{\Phi} \mathbf{F}_{t-1} \tag{2.44}$$

where $n - m = 1$ linear combination may be written as

$$F_{2t} = f_{t-1} = F_{1t-1}. \tag{2.45}$$

According to [6], in macroeconomic data the difference between number of static and dynamic factors is often small, in some applications even equal.

In [3] it is stated, that in static form factors have only a contemporaneous effect on observations \mathbf{X}_t , while dynamic factors have effect on observations through their lags too.

Typically we need more static factors than dynamic factors. The dynamic form (2.24), (2.25) is more realistic but harder to estimate than static form (2.38), (2.39). However the estimation in dynamic form may be accomplished by going through the equivalent static model, which we explained above.

2.3.3 Simplified form

In the further work, we will work with a simplified version of static model according to [7]. This theoretical background is applied in derivation of EM algorithm [8], which is used in MARSS package [29], that we will use in our time series analysis. Now we join notation from (2.1) and works mentioned above, that will be used further.

Let us have a general formulation for static form of a dynamic factor model with m common trends (factors)

$$\begin{aligned}\mathbf{X}_t &= \mathbf{L}\mathbf{F}_t + \boldsymbol{\mu} + \boldsymbol{\epsilon}_t \\ \mathbf{F}_t &= \mathbf{F}_{t-1} + \boldsymbol{\eta}_t,\end{aligned}\tag{2.46}$$

where \mathbf{L} is a $p \times m$ matrix of unknown factor loadings, \mathbf{F}_t is a $m \times 1$ vector of factors at time t and $\boldsymbol{\mu}$ is called offset or level parameter. We assume, that

$$\mathbf{F}_0 \sim N_m(\boldsymbol{\pi}_0, \mathbf{H}_0)\tag{2.47}$$

$$\boldsymbol{\epsilon}_t \sim N_p(\mathbf{0}, \mathbf{R})\tag{2.48}$$

$$\boldsymbol{\eta}_t \sim N_m(\mathbf{0}, \mathbf{Q}).\tag{2.49}$$

The parameters $\mathbf{L}, \mathbf{R}, \mathbf{Q}, \boldsymbol{\mu}, \boldsymbol{\pi}_0$ and \mathbf{H}_0 are unknown and they are called hyperparameters. The variance of \mathbf{X}_t is given by

$$\text{var}(\mathbf{X}_t) = \mathbf{L} \text{var}(\mathbf{F}_t) \mathbf{L}' + \mathbf{R}\tag{2.50}$$

and mean by

$$E(\mathbf{X}_t) = \mathbf{L}E(\mathbf{F}_t) + \boldsymbol{\mu}.\tag{2.51}$$

As we can see, the covariance is very similar to (2.9).

Based on (2.46) and (2.47 - 2.49) the conditional distributions of involved random vectors are

$$\begin{aligned}(\mathbf{X}_t | \mathbf{F}_t = \mathbf{f}_t) &\sim N_p(\mathbf{L}\mathbf{f}_t + \boldsymbol{\mu}, \mathbf{R}) \\ (\mathbf{F}_t | \mathbf{F}_{t-1} = \mathbf{f}_{t-1}) &\sim N_m(\mathbf{f}_{t-1}, \mathbf{Q}).\end{aligned}\tag{2.52}$$

2.4 Joint density function

According to [2] let $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ be $p \times 1$ vectors of random sample of size n from a multivariate normal distribution. Vectors are mutually independent and each one has a distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Let \mathbf{x}_i be a specific realization of the random vector \mathbf{X}_i . The joint density function $f()$ of all the observations is the product of the marginal normal densities

$$\begin{aligned}f(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p) &= \prod_{j=1}^n \left\{ \frac{1}{(2\pi^{p/2}) |\boldsymbol{\Sigma}|^{1/2}} e^{-(\mathbf{x}_j - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_j - \boldsymbol{\mu}) / 2} \right\} \\ &= \frac{1}{(2\pi)^{np/2}} \frac{1}{|\boldsymbol{\Sigma}|^{n/2}} e^{-\sum_{j=1}^n (\mathbf{x}_j - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_j - \boldsymbol{\mu}) / 2},\end{aligned}\tag{2.53}$$

2.5. KALMAN FILTER

where $|\cdot|$ is a determinant of a covariance matrix.

In [8] this is discussed regarding stochastic processes. Let \mathbf{X}_t and \mathbf{F}_t be a random vectors with distributions given by dynamic factor model (2.46). Let $\mathbf{x} = \mathbf{x}_{1:T} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T)'$ and $\mathbf{f} = \mathbf{f}_{0:T} = (\mathbf{f}_0, \mathbf{f}_1, \dots, \mathbf{f}_T)'$ be a specific realizations of these random vectors. Then, the joint density function can be written as

$$f(\mathbf{x}, \mathbf{f}) = f(\mathbf{x}|\mathbf{F} = \mathbf{f})f(\mathbf{f}), \quad (2.54)$$

where

$$\begin{aligned} f(\mathbf{f}) &= f(\mathbf{f}_0) \prod_{t=1}^T f(\mathbf{f}_t | \mathbf{F}_{1:t-1} = \mathbf{f}_{1:t-1}) \\ f(\mathbf{x}|\mathbf{F} = \mathbf{f}) &= \prod_{t=1}^T f(\mathbf{x}_t | \mathbf{F} = \mathbf{f}) \end{aligned} \quad (2.55)$$

So joint density function of $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T, \mathbf{f}_0, \mathbf{f}_1, \dots, \mathbf{f}_T)$ is

$$\begin{aligned} f(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T, \mathbf{f}_0, \mathbf{f}_1, \dots, \mathbf{f}_T) &= \\ &= \prod_{t=1}^T f(\mathbf{x}_t | \mathbf{F} = \mathbf{f}) \times f(\mathbf{f}_0) \prod_{t=1}^T f(\mathbf{f}_t | \mathbf{F}_{1:t-1} = \mathbf{f}_{1:t-1}) \\ &= \prod_{t=1}^T f(\mathbf{x}_t | \mathbf{F}_t = \mathbf{f}_t) \times f(\mathbf{f}_0) \prod_{t=1}^T f(\mathbf{f}_t | \mathbf{F}_{t-1} = \mathbf{f}_{t-1}). \end{aligned} \quad (2.56)$$

In the equation $\mathbf{F}_{1:t-1}$ becomes \mathbf{F}_{t-1} because of Markov property of multivariate autoregressive process and \mathbf{F} becomes \mathbf{F}_t , because \mathbf{X}_t depends only on \mathbf{F}_t in (2.46).

2.5 Kalman filter

Kalman filter, and further Kalman smoother, is a technique to obtain estimators for the underlying unobserved signal \mathbf{x}_t given the data $\mathbf{y}_{1:s} = \{\mathbf{y}_1, \dots, \mathbf{y}_s\}$ to time s . This perfectly fits our problem with observed time series \mathbf{X}_t , unobserved latent factors \mathbf{F}_t and unknown hyperparameters in our dynamic factor model. For $s < t$ we call the problem *forecasting* or *prediction*. For $s = t$ the problem is called *filtering* and finally for $s > t$ we call the problem *smoothing*. We will focus on just two cases, filtering and smoothing. This whole section is gathered from [9].

We will use the following definitions for the terms $\mathbf{F}_{t|T}$ and $\mathbf{H}_{t|T}$:

$$\begin{aligned} \mathbf{F}_{t|T} &:= E(\mathbf{F}_t | \mathbf{x}_{1:T}) \\ \mathbf{H}_{t|T} &:= E\{(\mathbf{F}_t - \mathbf{F}_{t|T})(\mathbf{F}_t - \mathbf{F}_{t|T})'\} \\ &= E\{(\mathbf{F}_t - E(\mathbf{F}_t | \mathbf{x}_{1:T}))(\mathbf{F}_t - E(\mathbf{F}_t | \mathbf{x}_{1:T}))'\}. \end{aligned} \quad (2.57)$$

Also

$$\begin{aligned} \mathbf{H}_{t,t-1|T} &= E\{(\mathbf{F}_t - \mathbf{F}_{t|T})(\mathbf{F}_{t-1} - \mathbf{F}_{t-1|T})' | \mathbf{x}_{1:T}\} \\ &= E\{(\mathbf{F}_t - E(\mathbf{F}_t | \mathbf{x}_{1:T}))(\mathbf{F}_{t-1} - E(\mathbf{F}_{t-1} | \mathbf{x}_{1:T}))' | \mathbf{x}_{1:T}\} \end{aligned} \quad (2.58)$$

is conditioned covariance matrix of \mathbf{F}_t and \mathbf{F}_{t-1} .

Let us have a dynamic factor model (2.46) with initial conditions $\mathbf{F}_{0|0} = \boldsymbol{\pi}_0$ and $\mathbf{H}_{0|0} = \mathbf{H}_0$ for $t = 1, \dots, T$ defined in (2.58). Then

$$\mathbf{F}_{t|t-1} = \mathbf{F}_{t-1|t-1} \quad (2.59)$$

$$\mathbf{H}_{t|t-1} = \mathbf{H}_{t-1|t-1} + \mathbf{Q} \quad (2.60)$$

and

$$\mathbf{F}_{t|t} = \mathbf{F}_{t|t-1} + \mathbf{K}_t(\mathbf{X}_t - \mathbf{L}\mathbf{F}_{t|t-1} - \boldsymbol{\mu}) \quad (2.61)$$

$$\mathbf{H}_{t|t} = (\mathbf{I} - \mathbf{K}_t\mathbf{L})\mathbf{H}_{t|t-1}, \quad (2.62)$$

where

$$\mathbf{K}_t = \mathbf{H}_{t|t-1}\mathbf{L}'(\mathbf{L}\mathbf{H}_{t|t-1}\mathbf{L}' + \mathbf{R})^{-1} \quad (2.63)$$

is called *Kalman gain*. Important byproducts of this filter are the innovations

$$\mathbf{e}_t = \mathbf{X}_t - E(\mathbf{X}_t|\mathbf{X}_{1:t-1}) = \mathbf{X}_t - \mathbf{L}\mathbf{F}_{t|t-1} - \boldsymbol{\mu} \quad (2.64)$$

and the corresponding variance-covariance matrices

$$\mathbf{H}_t := \text{var}(\mathbf{e}_t) = \text{var}(\mathbf{L}(\mathbf{F}_t - \mathbf{F}_{t|t-1}) + \boldsymbol{\epsilon}_t) = \mathbf{L}\mathbf{H}_{t|t-1}\mathbf{L}' + \mathbf{R} \quad (2.65)$$

for $t = 1, \dots, T$. Thus we may rewrite (2.61) and (2.63) into a form

$$\mathbf{F}_{t|t} = \mathbf{F}_{t|t-1} + \mathbf{K}_t\mathbf{e}_t, \quad (2.66)$$

$$\mathbf{K}_t = \mathbf{H}_{t|t-1}\mathbf{L}'(\mathbf{H}_t + \mathbf{R})^{-1}. \quad (2.67)$$

The scheme in Figure 2.1 shows, how the prediction step and the correction step follows each other. We start with initial conditions, then we create a prediction conditioned on the initial conditions. After the prediction is made, we compute the Kalman gain which we substitute together with the prediction into a correction of the estimate and the covariance matrix. After the correction, we move again to prediction and the whole iteration repeats.

2.6 Kalman smoother

Kalman smoother is the next step after obtaining linear estimates from Kalman filter. As mentioned in section 2.5, it is the case when $s > t$. Let us have initial conditions $\mathbf{F}_{T|T}$ and $\mathbf{H}_{T|T}$ for $t = T, T-1, \dots, 1$ obtained from Kalman filter (2.61) and (2.62). Then Kalman smoother is specified as

$$\mathbf{F}_{t-1|T} = \mathbf{F}_{t-1|t-1} + \mathbf{J}_{t-1}(\mathbf{F}_{t|T} - \mathbf{F}_{t|t-1}), \quad (2.68)$$

$$\mathbf{H}_{t-1|T} = \mathbf{H}_{t-1|t-1} + \mathbf{J}_{t-1}(\mathbf{H}_{t|T} - \mathbf{H}_{t|t-1})\mathbf{J}'_{t-1}, \quad (2.69)$$

where

$$\mathbf{J}_{t-1} = \mathbf{H}_{t-1|t-1}(\mathbf{H}_{t|t-1})^{-1}. \quad (2.70)$$

2.7. JOINT LOG-LIKELIHOOD FUNCTION

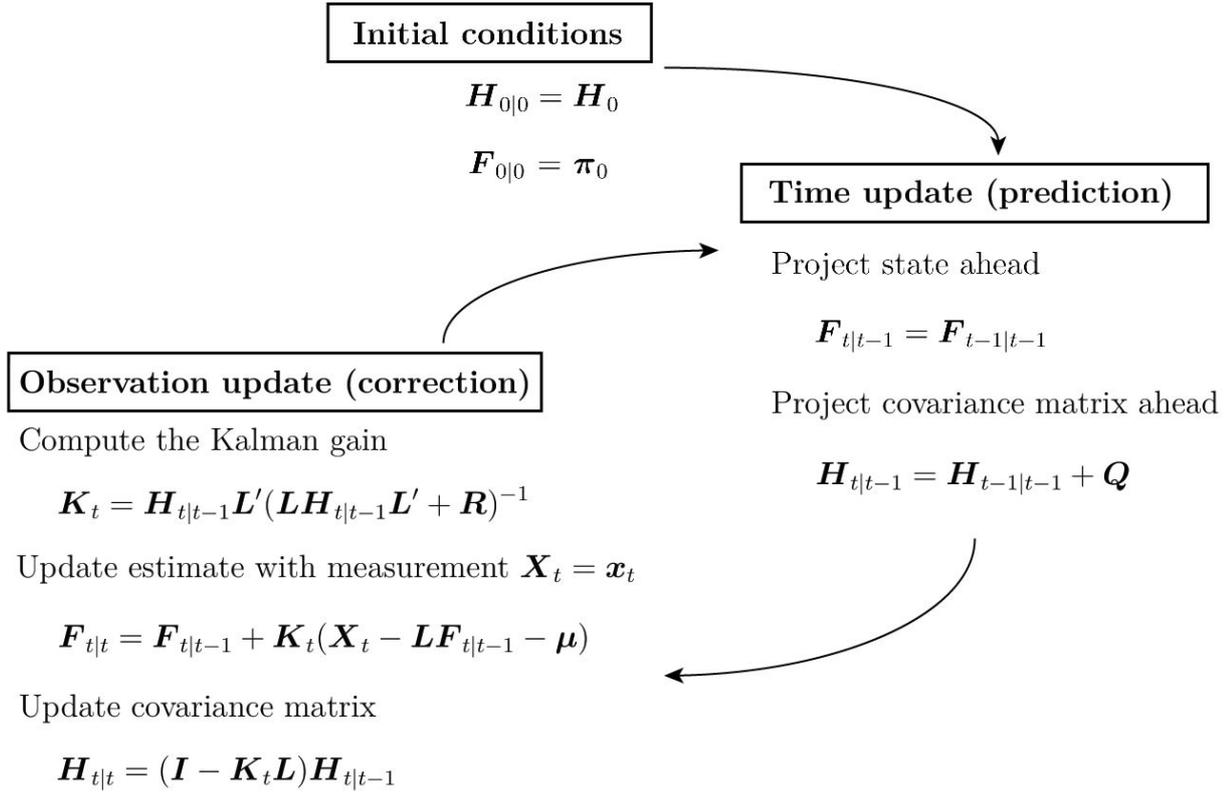


Figure 2.1: Scheme of Kalman filtering.

2.7 Joint log-likelihood function

As [2] states, we wish to have the "best" explanation of the data. In terms of our hyperparameters, the "best" may mean to select the parameter values that *maximize* the joint density function from (2.56). This is called *maximum likelihood estimation* and the maximizing parameter values are called *maximum likelihood estimates*. In this approach we handle the joint density function as a function of parameters under given observations $\mathbf{x}_1, \dots, \mathbf{x}_T$, which is named *likelihood function*. Similarly to [7] and [8] the joint log likelihood function of the observations $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T$ and the factors (if we could observe them) $\mathbf{f}_0, \mathbf{f}_1, \dots, \mathbf{f}_T$ is

$$\begin{aligned}
 \log L(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T, \mathbf{f}_0, \mathbf{f}_1, \dots, \mathbf{f}_T) = & \\
 & - \frac{1}{2} \sum_{t=1}^T (\mathbf{x}_t - \mathbf{L} \mathbf{f}_t - \boldsymbol{\mu})' \mathbf{R}^{-1} (\mathbf{x}_t - \mathbf{L} \mathbf{f}_t - \boldsymbol{\mu}) - \frac{T}{2} \log |\mathbf{R}| \\
 & - \frac{1}{2} \sum_{t=1}^T (\mathbf{f}_t - \mathbf{f}_{t-1})' \mathbf{Q}^{-1} (\mathbf{f}_t - \mathbf{f}_{t-1}) - \frac{T}{2} \log |\mathbf{Q}| \\
 & - \frac{1}{2} (\mathbf{f}_0 - \boldsymbol{\pi}_0)' \mathbf{H}_0^{-1} (\mathbf{f}_0 - \boldsymbol{\pi}_0) - \frac{T}{2} \log |\mathbf{H}_0| \\
 & + \text{constant},
 \end{aligned} \tag{2.71}$$

Same equation is derived also in [9].

This log likelihood function of observations $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T, \mathbf{f}_0, \mathbf{f}_1, \dots, \mathbf{f}_T)$ is called a *complete data* likelihood. Since the factors are unknown, we possess only *incomplete*

data and the function cannot be optimized directly. *EM algorithm* serves exactly this purpose. Thanks to it, it is possible to obtain a maximum likelihood estimates of the hyperparameters based on the incomplete data $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T$.

2.7.1 EM algorithm

Expectation-maximization algorithm is a iterative method used in statistics to find a maximum likelihood estimates of the parameters, when the observations may be viewed as the incomplete data [10]. The iteration alternates between performing an *expectation* (E) step, which creates a function for the expectation of the log-likelihood function evaluated using the current estimate for the parameters and a *maximization* (M) step, which computes parameter-estimates by maximizing the expected log-likelihood function found in the E step. These parameter-estimates are then used to determine the distribution of the latent variables in the next E step. First formulation of this algorithm was provided in [10]. The EM algorithm explained here is from [7], which is derived from [9] and program implementation used further in the code is from [8].

In the algorithm, first we maximize the conditional expectation of the complete data likelihood function

$$E[\log L(\mathbf{x}_1, \dots, \mathbf{x}_T, \mathbf{F}_0, \mathbf{F}_1, \dots, \mathbf{F}_T) | \mathbf{x}_1, \dots, \mathbf{x}_T; \Theta^{j-1}], \quad (2.72)$$

where Θ^{j-1} contains all the hyperparameters ($\Theta = \mathbf{L}, \mathbf{R}, \mathbf{Q}, \boldsymbol{\mu}, \boldsymbol{\pi}_0, \mathbf{H}_0$) estimated in the $(j-1)$ th iteration. Then in the next step, the conditional expectation is maximized with respect to the hyperparameters.

E-step: Conditional expectation of the complete data likelihood function looks like

$$\begin{aligned} E[\log L(\mathbf{x}_1, \dots, \mathbf{x}_T, \mathbf{F}_0, \mathbf{F}_1, \dots, \mathbf{F}_T) | \mathbf{x}_1, \dots, \mathbf{x}_T; \Theta^{j-1}] = \\ - \text{tr}\{\mathbf{R}^{-1} \frac{1}{2} \sum_{t=1}^T ((\mathbf{x}_t - \mathbf{L}\mathbf{F}_t - \boldsymbol{\mu})(\mathbf{x}_t - \mathbf{L}\mathbf{F}_t - \boldsymbol{\mu})' + \mathbf{L}\mathbf{H}_{t|T}\mathbf{L}')\} - \frac{1}{2} \log |\mathbf{R}| \\ - \frac{1}{2} \text{tr}\{\mathbf{Q}^{-1}(\mathbf{C} - 2\mathbf{B} + \mathbf{A})\} - \frac{1}{2} \log |\mathbf{Q}| \\ - \frac{1}{2} \text{tr}\{\mathbf{H}_0^{-1}(\mathbf{H}_{0|T} + (\mathbf{F}_0 - \boldsymbol{\pi}_0)(\mathbf{F}_0 - \boldsymbol{\pi}_0)')\} - \frac{1}{2} \log |\mathbf{H}_0| \\ + \text{constant}, \end{aligned} \quad (2.73)$$

where $\text{tr}\{\cdot\}$ is trace of a matrix and

$$\begin{aligned} \mathbf{A} &= \sum_{t=1}^T (\mathbf{F}_{t-1|T} \mathbf{F}'_{t-1|T} + \mathbf{H}_{t-1|T}) \\ \mathbf{B} &= \sum_{t=1}^T (\mathbf{F}_{t|T} \mathbf{F}'_{t-1|T} + \mathbf{H}_{t,t-1|T}) \\ \mathbf{C} &= \sum_{t=1}^T (\mathbf{F}_{t|T} \mathbf{F}'_{t|T} + \mathbf{H}_{t|T}). \end{aligned} \quad (2.74)$$

2.8. IDENTIFICATION

The terms $\mathbf{F}_{t|T}$ and $\mathbf{H}_{t|T}$ are defined in (2.57) and represent the best linear estimator for \mathbf{F}_t and corresponding variance matrix \mathbf{H}_t given all observations. These we obtain from Kalman filter and smoother explained in sections 2.5 and 2.6.

M-step: We obtain updated equations for hyperparameters ($\Theta = \mathbf{L}, \mathbf{R}, \mathbf{Q}, \boldsymbol{\mu}, \boldsymbol{\pi}_0, \mathbf{H}_0$) by maximizing the conditional expectation of the complete data likelihood function (2.72) with respect to the hyperparameters. In [7] it is shown, that arguments

$$\mathbf{Q} = \frac{1}{T}(\mathbf{C} - 2\mathbf{B} + \mathbf{A}) \quad (2.75)$$

$$\mathbf{R} = \frac{1}{T} \sum_{t=1}^T ((\mathbf{x}_t - \mathbf{L}\mathbf{F}_{t|T} - \boldsymbol{\mu})(\mathbf{x}_t - \mathbf{L}\mathbf{F}_{t|T} - \boldsymbol{\mu})' + \mathbf{L}\mathbf{H}_{t|T}\mathbf{L}') \quad (2.76)$$

$$\boldsymbol{\pi}_0 = \mathbf{F}_{0|T} \quad (2.77)$$

$$\mathbf{H}_0 = \mathbf{H}_{0|T} \quad (2.78)$$

$$\boldsymbol{\mu} = \frac{1}{T} \sum_{t=1}^T (\mathbf{x}_t - \mathbf{L}\mathbf{F}_{t|T}) \quad (2.79)$$

$$\mathbf{L} = \mathbf{E}_2 \mathbf{E}_1^{-1}, \quad (2.80)$$

where

$$\mathbf{E}_1 = \sum_{t=1}^T (\mathbf{F}_{t|T} \mathbf{F}_{t|T}' + \mathbf{H}_{t|T}) \quad (2.81)$$

$$\mathbf{E}_2 = \sum_{t=1}^T (\mathbf{x}_t \mathbf{F}_{t|T}' - \boldsymbol{\mu} \mathbf{F}_{t|T}') \quad (2.82)$$

maximize the conditional expectation of the complete data likelihood function.

2.8 Identification

According to [11] the dynamic factor model is not identifiable. This occurs because if \mathbf{T} is a $m \times m$ non-singular matrix, matrix of factor loadings $\mathbf{L}\mathbf{T}^{-1}$ and common trends $\mathbf{T}\mathbf{F}_t$ would lead to equivalent model fits

$$\begin{aligned} \mathbf{X}_t &= \mathbf{L}\mathbf{F}_t + \boldsymbol{\mu} + \boldsymbol{\epsilon}_t \\ \mathbf{F}_t &= \mathbf{F}_{t-1} + \boldsymbol{\eta}_t \end{aligned} \quad (2.83)$$

and

$$\begin{aligned} \mathbf{X}_t &= \mathbf{L}\mathbf{T}^{-1}\mathbf{T}\mathbf{F}_t + \boldsymbol{\mu} + \boldsymbol{\epsilon}_t \\ \mathbf{T}\mathbf{F}_t &= \mathbf{T}\mathbf{F}_{t-1} + \mathbf{T}\boldsymbol{\eta}_t, \end{aligned} \quad (2.84)$$

according to [7] and [24]. There is infinite number of parameter sets which would lead to same density function for the observations [11]. To solve this problem, additional restrictions are addressed in dynamic factor model, similar to restrictions in factor analysis. In [11], these three restrictions are suggested:

- (i) The covariance matrix of latent factors is the identity matrix

$$\mathbf{Q} = \mathbf{I}. \quad (2.85)$$

This is not a sufficient restriction to make the model identifiable, since if \mathbf{T} is an orthogonal matrix, (2.84) still satisfies all of the restrictions of the original model because $\text{var}(\mathbf{T}\boldsymbol{\eta}_t) = \mathbf{T}\mathbf{Q}\mathbf{T}' = \mathbf{T}\mathbf{I}\mathbf{T}' = \mathbf{T}\mathbf{T}' = \mathbf{I}$. This restriction implies that the latent factors will be uncorrelated with each other.

- (ii) The ij th element of \mathbf{L} is set to zero for $j > i$, $i = 1, \dots, m - 1$. For example, if we assume three latent factors ($m = 3$), then loadings matrix would look like

$$\mathbf{L} = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \\ \vdots & \vdots & \vdots \\ l_{p1} & l_{p2} & l_{p3} \end{bmatrix}. \quad (2.86)$$

Alternatively we could set \mathbf{Q} equal to a diagonal matrix while $l_{ij} = 0$ for $j > i$ and $l_{ii} = 1$ for $i = 1, \dots, m$.

- (iii) The first m elements of the level parameter $\boldsymbol{\mu}$ are set to zero. For example, if we assume three latent factors ($m = 3$), then the level parameter would look like

$$\boldsymbol{\mu} = [0 \ 0 \ 0 \ \mu_4 \ \cdots \ \mu_p]'. \quad (2.87)$$

This restriction is already assumed in the model in [11].

With these restrictions set on the model, solution of the factor loadings \mathbf{L} exists and is unique. The problem with these restrictions is that the first observed time series X_{1t} is determined only by the first latent factor F_{1t} , second time series is determined by the first two latent factors and so on up to m th time series. Therefore a factor rotation may be applied to the factor loadings and common factors after parameter estimation of the model to avoid this problem. Many of the methods of factor rotation exist in factor analysis literature. We will explain *varimax* rotation, which will be used in our data analysis.

2.9 Varimax rotation

Varimax rotation, originally proposed in [12] is commonly used in factor analysis mentioned in section 2.1. It is a method to obtain principal components with high correlation for some variables and no correlation with other variables. It minimizes the number of principal components and is suited for locating clusters that lie at right angles to each other [13]. This is also useful in time series clustering. Advantage of this rotation of the factor loadings and latent factors in time series applications is to overcome the problem of first m time series being not properly modeled by the factors due to the restrictions placed on the model because of identifiability of the model mentioned in section 2.8. Second advantage is, that varimax rotation changes the loadings matrix in a way, that creates the largest difference between the loadings [29], thus each observed time series

2.9. VARIMAX ROTATION

will be modeled by the least number of the latent factors, which suits potential clustering analysis in the future. For small number of the observed time series it is important to overcome the problem with restrictions and the rotation is needed.

Let us have estimated dynamic factor model (2.46). Then

$$\begin{aligned}\mathbf{X}_t &= \mathbf{L}^* \mathbf{F}_t^* + \boldsymbol{\mu} + \boldsymbol{\epsilon}_t \\ \mathbf{F}_t^* &= \mathbf{F}_{t-1}^* + \boldsymbol{\eta}_t^*\end{aligned}\tag{2.88}$$

is equivalent dynamic factor model with rotated loadings matrix $\mathbf{L}^* = (l_{ij}^*)_{p \times m}$ and rotated common trends, where

$$\begin{aligned}\mathbf{L}^* &= \mathbf{L}\mathbf{T}^{-1} \\ \mathbf{F}_t^* &= \mathbf{T}\mathbf{F}_t \\ \boldsymbol{\eta}_t^* &= \mathbf{T}\boldsymbol{\eta}_t\end{aligned}\tag{2.89}$$

and

$$\text{var}(\boldsymbol{\eta}_t^*) = \mathbf{T}\mathbf{Q}\mathbf{T}'.\tag{2.90}$$

Let

$$\tilde{l}_{ij}^* = \frac{l_{ij}^*}{h_i}\tag{2.91}$$

be the rotated coefficients scaled by the square root of the communalities defined in (2.17). Then the *normal varimax criterion* procedure selects the orthogonal transformation \mathbf{T} for (2.89) that maximize

$$V = \frac{1}{p} \sum_{j=1}^m \left[\sum_{i=1}^p \tilde{l}_{ij}^{*4} - \frac{1}{p} \left(\sum_{i=1}^p \tilde{l}_{ij}^{*2} \right)^2 \right].\tag{2.92}$$

Scaling the rotated coefficients by communalities give time series with smaller communalities (less reliable variables) more weight in the determination of the structure. Without the scaling less reliable variables are given less weight than those, whose variance is well explained by the m latent factors. After the transformation \mathbf{T} is determined, the rotated loadings \tilde{l}_{ij}^* may be multiplied by h_i and the original communalities are preserved. Equation (2.92) may be interpreted as a sum of m variances of squares of scaled loadings for each latent factor [2].

3 Results assessment

From a dynamic factor model we obtain parameter estimates and maximized likelihood value. We will use the maximum likelihood to compute AIC, which will serve as a criterion for model selection. k-means method, which will be used for clustering of our observed time series, will be executed on the loadings matrix as an output of a dynamic factor model.

3.1 Akaike information criterion

The Akaike information criterion (AIC) is a common method for selecting the most appropriate model. It is an approximately unbiased estimator of the expected Kullback-Leiber information of a fitted model [14]. We suppose, that for our observable time series \mathbf{X}_t and unobservable latent factors \mathbf{F}_t with joint density function (2.56), there exists a generating parameter structure Θ_0 with an unknown parametric likelihood function $L(\mathbf{x}, \mathbf{f}; \Theta)$. To preserve the notation from [15], we denote this parametric likelihood function $L(\Theta|\mathbf{Y})$.

Now $L(\Theta_0|\mathbf{Y})$ refers to a true or generating model (2.72), which we are trying to get to as close as possible. We estimate models with different hyperparameters $\Theta_k (k \in K)$ (in our model this means different number of expected latent factors or different variance-covariance structure) and obtain hyperparameter estimates $\hat{\Theta}_k$. Let $L(\Theta|\mathbf{Y})$ be an approximating model. Then we consider Kullback-Leiber information [15] estimated as

$$\delta(\Theta_0, \Theta) = E_0\{-2 \log L(\Theta|\mathbf{Y})\}, \quad (3.1)$$

where E_0 denotes the expectation under the generating model and $L(\Theta|\mathbf{Y})$ represent the likelihood corresponding to the approximated model obtained from EM algorithm (2.72). We are able to estimate hyperparameters from approximating model $\Theta = \hat{\Theta}$, but we are not able to possess the estimations of the true parameters Θ_0 , which are needed for the computation of (3.1) [15]. However, $-2 \log L(\Theta|\mathbf{Y})$ serves as a biased estimator of (3.1). Therefore a bias adjustment (or penalty term) is introduced [16] and is approximated by $2k$, where k stands for a dimension of $\hat{\Theta}$ (number of hyperparameters being estimated). Then

$$AIC = -2 \log L(\hat{\Theta}|\mathbf{Y}) + 2k \quad (3.2)$$

is *Akaike Information Criterion*. Under appropriate conditions [16], AIC should be asymptotically close to the expected value of (3.1).

AIC tends to favor higher dimensional models [16] (may be our case, since we will be comparing models with different dimensions), but especially for a small sample settings (which is not our case). Therefore a corrected AIC was introduced to overcome this problem. It contains adjusted penalty term, and so

$$AICc = -2 \log L(\hat{\Theta}|\mathbf{Y}) + 2 \frac{kn}{n - k - 1}, \quad (3.3)$$

where $L(\hat{\Theta}|\mathbf{Y})$ is estimated likelihood of our model, k stands for a number of estimated parameters¹ and n is dimension of the sample. In our analysis with $p = 108$ time series and $T = 31$ days, we have a sample size $n = pT = 3348$.

¹In our dynamic factor models k will be obtained as $(pm - \frac{m(m-1)}{2}) + \sum_i \text{unique}(\sigma_i)$, where $\text{unique}(\sigma_i)$ are unique parameters estimated in variance-covariance matrix \hat{R} (section 5.3.1).

3.2 k-means clustering

It is a clustering technique, which aims to group values into k number of clusters based on their distance to the "mean" of the cluster (also named centroid) and the total distance from all values within the cluster [2]. The process starts with the initialization of k clusters, which we decide in advance based on other indices. Each cluster contains its mean (centroid). For each cluster a within cluster sum of squares (WSS) is computed [17]

$$W(C_k) = \sum_{i=1}^{|C_k|} (\mathbf{x}_i - \boldsymbol{\mu}_k)^2, \quad (3.4)$$

where C_k stands for k th cluster (set of points belonging to k th cluster), $|C_k|$ is cardinality, $\boldsymbol{\mu}_k = (\mu_{k1}, \dots, \mu_{km})'$ is the m -dimensional mean (centroid) of k th cluster and $\mathbf{x}_i = (x_{i1}, \dots, x_{im})'$ are m -dimensional points belonging to k th cluster. Then the goal of k-means is to change the clusters and points belonging to a cluster in a way, that the total within sum of squares of all the clusters is minimized

$$\sum_k W(C_k) \longrightarrow \min. \quad (3.5)$$

The algorithm may be found in [18]. To determine correct number of clusters is up to discussion. We will use mainly two methods explained in section 5.5. Other indices possible to use are mentioned in [37].

4 Fire Weather Index

Forest fires endanger humans for ages. Although occasional wildfires are beneficial for the overall health of the forest, at the same time they may be extremely destructive and harmful for the forest resources and species, public safety, and their property. In September and August 2003, a one-in-a-hundred-year wildfire in The Okanagan Mountain Park burned down around $250km^2$ of forest and parkland, forced to evacuate 27 000 people, and destroyed around 239 homes. Ignited by a lightning strike on August 16th, over 1 000 forestry firefighters, contractors, and loggers, as well as 1 400 members of the Canadian Armed Forces, tried to take control of the fire for over a month [19]. Around \$1 402 000 were donated to those impacted by the fires, even though claims submitted by the individuals created a total of \$67 134 880 [20]. The Canadian Forest Fire Danger Rating System (CFFDRS) is a fire management system trying to provide a complex analysis of the fire danger situation based on the integration of the scientific, technological and human elements [21]. Fire Weather Index (FWI) is the sixth and final component of the fire weather index system. It combines altogether 5 other numerical values and creates a subsystem of CFFDRS [22]. Our data analysis in section 5 processes Fire Weather Index values.

4.1 Canadian Forest Fire Danger Rating System

CFFDRS is the principal source of fire intelligence for all forest fire management agencies in Canada. The system supports fire management decision making regarding strategies, fire prevention or firefighter safety. It is a important element of effective fire organization. Fire danger refers to an assessment of both fixed and variable factors of the fire environment (fuels, weather, topography) that determine the ease of ignition, rate of spread, difficulty of control and impact of wildfires. Principally the system serves as a economic activity that helps to reduce costly and limited resources and efficiently allocate an appropriate level of resources across a region or a country from day to day or place to place, on the basis of existing and forecasted fire danger levels [21]. In Figure 4.1 we may see the fire management actions included in the CFFDRS.

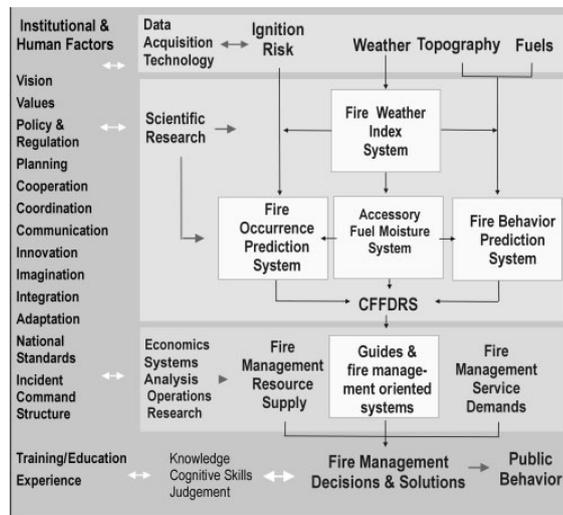


Figure 4.1: The structure of the Canadian Forest Fire Danger Rating System [21].

4.2 Canadian Forest Fire Weather Index System

The purpose of the FWI system is to evaluate how flammable the forest is at any time. Flammability of the forest and therefore fire danger changes continuously because of the continuous weather changes. The system was completed in 1970 and since then, it provides a uniform method of evaluating fire weather severity across Canada, relates the effects of past and current weather on fuel flammability and provides numerical ratings of relative fire potential. As seen in Figure 4.2, system joins weather variables, fuel moisture characteristics and fire behaviour characteristics [22].

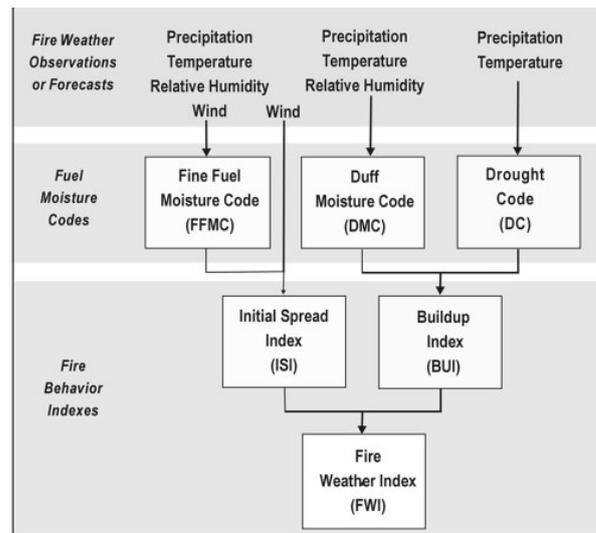


Figure 4.2: The structure of the Canadian Forest Fire Weather Index System consisting of 6 components [21].

Information about the system is gathered from [22]. The system is composed of 3 moisture codes:

- *Fine Fuel Moisture Code (FFMC)* - a numerical rating of the moisture content of the surface litter layers 1 or 2 centimeters in depth. It is related to ease of ignition because it represents the moisture content of fine fuels, where most fires ignite. A fire will not start in most fine fuels until the moisture content is below about 25-30%.
- *Duff Moisture Code (DMC)* - a numerical rating of the dryness of forest floor layers 5 to 10 centimeters deep. Duff consumption is the main source of the energy produced by the moving flame. In many fuel types, the duff starts to be involved in combustion when the code is about 20.
- *Drought Code (DC)* - a long-term drought indicator. It represents loss of water from the soil by evaporation from the surface and by transpiration from plant in a 200mm moisture reservoir with a full reservoir equivalent of 800%. This code is well correlated with the moisture content of deep compact organic layers 10-20 centimeters thick. It is an indicator of the effect of seasonal drought on forest fuels, total fuel consumption and smoldering in deep organic layers and large fallen logs.

and 3 fire behaviour indices:

4. FIRE WEATHER INDEX

- *Initial Spread Index (ISI)* - a numerical rating of relative fire spread without the effect of slope or fuel consumption. It combines wind speed with fine fuel moisture and increases exponentially with wind speed.
- *Buildup Index (BUI)* - a numerical rating of the amount of fuel available for combustion. It is a weighted combination of DMC and DC without any other weather elements.
- *Fire Weather Index (FWI)* - combines ISI (representing rate of speed) and BUI (representing fuel consumption) into one numerical rating of fire intensity. It depends only on weather inputs in order to give uniform results regardless of fuel type. The calculation looks as follows:

$$\begin{aligned}
 f(D) &= \begin{cases} 0.626U^{0.809} + 2 & U \leq 80 \\ \frac{1000}{25 + 108.64e^{-0.023U}} & U > 80 \end{cases} \\
 B &= 0.1Rf(D) \\
 S &= \begin{cases} e^{(2.72(0.434 \ln B)^{0.647})} & B > 1 \\ B & B \leq 1, \end{cases}
 \end{aligned} \tag{4.1}$$

where U is the buildup index (BUI), R is the initial spread index (ISI) and S is the resulted today's fire weather index values [23].

5 Application of a dynamic factor model

In our research, we will apply a dynamic factor model to a Fire Weather Index (FWI) time series measured and computed on the meteorological stations in British Columbia. The first goal is to determine the most suitable dynamic factor model that explains the data. From the model, we obtain a loadings matrix and a number of latent factors. After we compute the varimax rotation of the loadings and the factors, we apply k-means clustering to the loadings matrix and try to determine time series division into the clusters. Clusters join together stations with similar behaviour. This is a significant part of fire management. By properly selecting similar FWI trends we are able to distinguish between more dangerous areas, provide a better controlling system, effectively distribute available and limited resources and prepare for possible fire ignition. Therefore, being able to observe the collective behaviour of the stations and not just individual stations helps us prepare better for the possible threats and save resources.

In the analysis, we follow steps and adjust codes from [24] and [25].

5.1 Data summary

In the analysis we use a FWI measurements (mentioned in section 4.2) from 108 meteorological stations. We work with the data from July and August in 2003. Time period of the measurements is from 1.7.2003 to 31.8.2003, which is 31 days per month. We may summarize the data specifications in Table 5.1.

number of stations	$p = 108$
number of time points (per month)	$T = 31$
date start	1.7.2003
date ends	31.8.2003
complete data	yes

Table 5.1: Summary of the data used in the analysis.

We will execute the analysis for months individually. In whole section we will provide specific analysis with visualizations of the data for month July. The results and visualizations for month August are shown in [Attachments](#). In the end of this section we compare the results for both months.

In the original data we had following information about the stations:

- *station_code* - number assigned to a particular meteorological station (1-933)
- *station_name* - name of the meteorological station
- *station_category* - if the station is still active or no longer in use (archived)
- *latitude* and *longitude* - specification of the position on the map
- *install_date* - when the station was installed (from 1970 to 2003)
- *elevation* - height above the sea level

5. APPLICATION OF A DYNAMIC FACTOR MODEL

- *danger_region* - British Columbia is divided into 3 danger regions of the different fire climates [21], see Figure 5.1
- *weather_zone* - the number of zone in which the station is located (we use southern zones)

Fire Weather Zones



Figure 5.1: Map of British Columbia with weather zones [26]. Grey area are the weather zones used in our data analysis.

In the original dataset we had data from around 933 stations with FWI values from the years 1970 to 2014. Not all of them are active at the same time. In the year 1970 there were 23 active stations, in 2014 it was around 191 and right now in 2021 it is 260 weather stations. Stations change in the course of time and some are being created, meanwhile some get deactivated and the data is archived. Therefore, at first we selected stations that

5.1. DATA SUMMARY

had measurements for the year 2003. In the data there are 5 types of FWI measurements - actual, estimate, substitution - partial, forecast, startup. To ensure reliability of the model, we selected only stations with the complete measurements, i.e. that have values *actual* for every time point in the time series. Due to the long computation time with all the stations (186 so far), we decided to focus only on specific weather zones to decrease the number of stations in the model. As we can see in the map of British Columbia in Figure 5.1, whole province is divided into 30 weather zones. In our analysis we covered the stations in the weather zones considered to be more dangerous with respect to the possibility of fires (southern location, prevailing higher temperatures and inland climate). In Figure 5.2 we may see, that the region covered the most extreme area (in years 1957-1966) in terms of mean FWI measurements for the summer months within whole Canada.

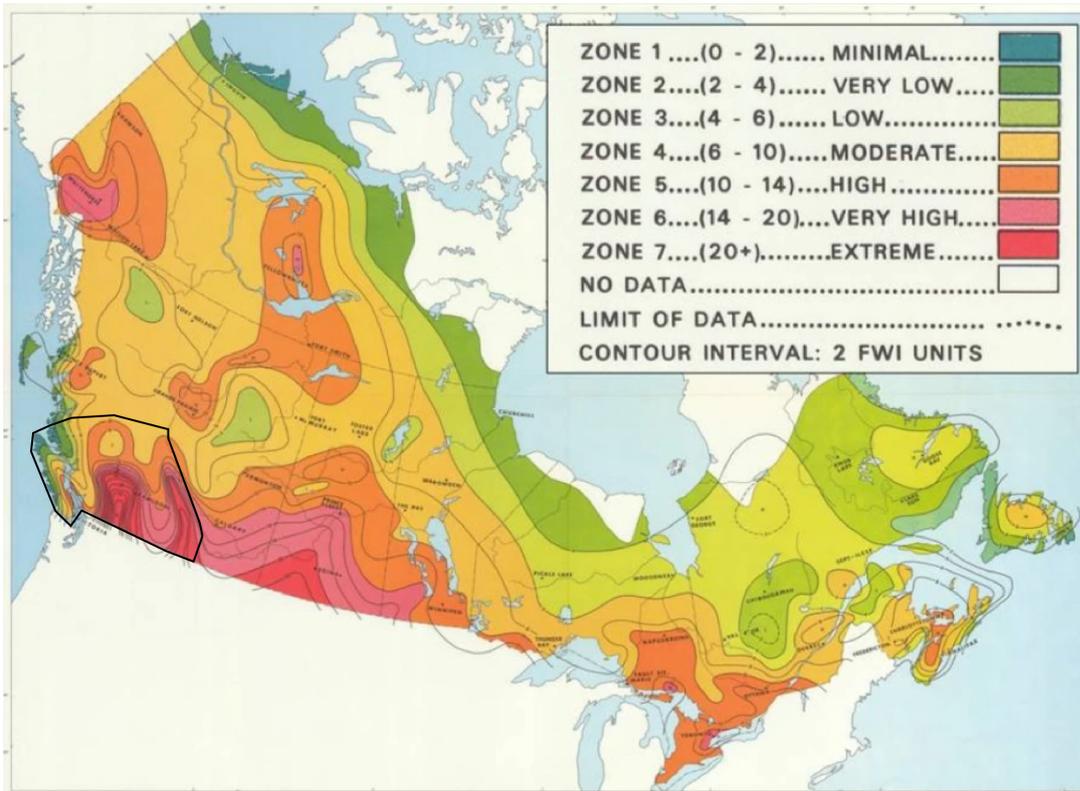


Figure 5.2: Fire weather zones of Canada calculated from mean fire weather index values for the months June, July, and August for 364 weather stations for the period 1957 to 1966 [27].

The used weather zones marked in Figure 5.1 are 1, 2, 3, 4, 5, 7, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27. Later we repeat

We also had to select stations that have complete measurements ("actual" FWI measurements) for both months July and August. Altogether this selection makes 108 stations in 20 southern weather zones covering almost half of the province. In Figure 5.3 we may see the actual map of the selected area with the location of the stations.

5. APPLICATION OF A DYNAMIC FACTOR MODEL

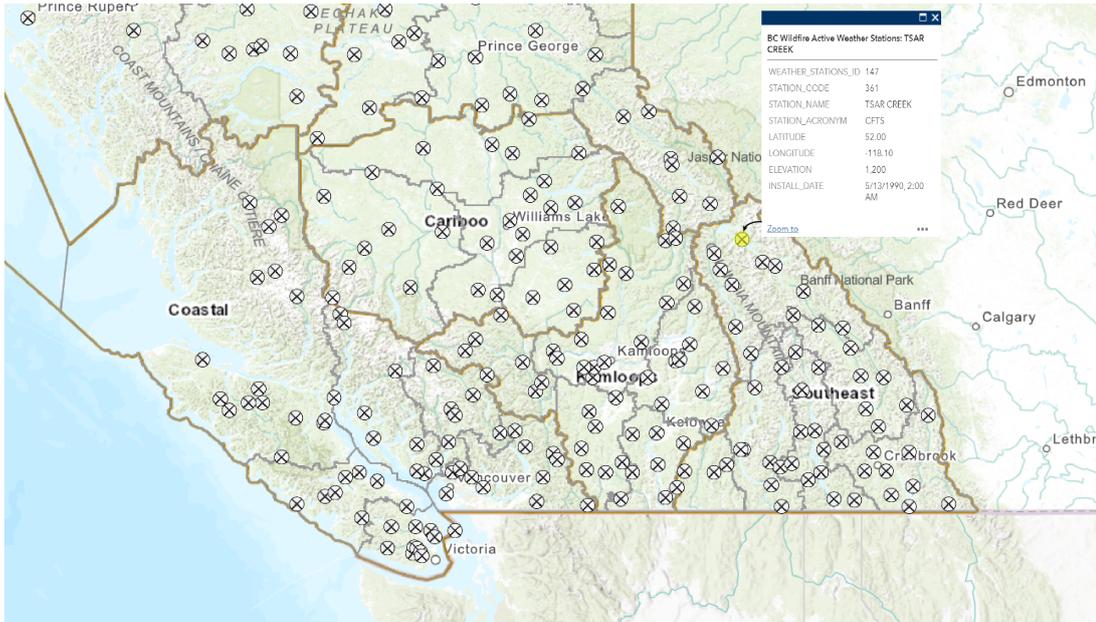


Figure 5.3: Map of the stations located in the selected area with the information about the particular station [28]. Same data was obtained from the original dataset. In the figure there are stations active to date 29.3.2021.

The measurements from all 108 stations may be seen plotted in Figure 5.4.

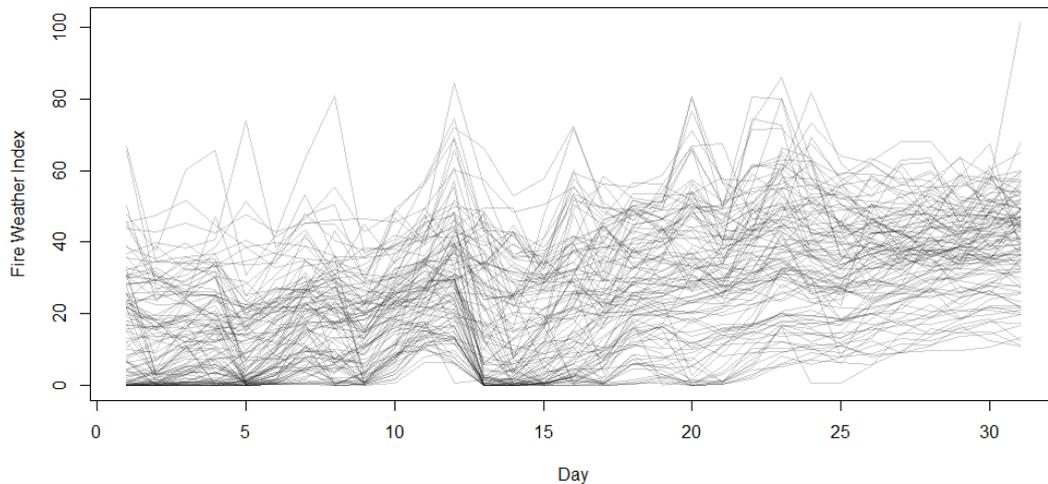


Figure 5.4: Plot of 108 observed time series used in the data analysis.

5.2 Data pre-processing

Before computation of the model, we had to do two adjustments of the time series. In the first place, we rounded the data to two decimal places, since some stations contained FWI measurements not rounded. In the second place, we had to center (de-mean) the data. Previous analyses of the data were performed using centered data, thus we also wanted to keep the convention for the further comparison of the results. Also according to [7], EM

5.3. MARSS PACKAGE

algorithm from section 2.7.1 for the model with the third constraint (2.87) is not robust and it takes long time to converge [25]. Therefore EM estimates are more stable if after each Kalman smoothing iteration the average of each latent factor $\overline{\mathbf{F}}$ is calculated and subtracted from the estimated latent factor $\mathbf{F}_{t|T}$, thus we constrain \mathbf{F}_t to have a zero mean across time $t = 1, \dots, T$. With this approach, all elements of $\boldsymbol{\mu}$ are estimated and they represent the average level of \mathbf{X}_t relative to $\mathbf{L}(\mathbf{F}_t - \overline{\mathbf{F}}_t)$. For this reason we center observed time series

$$\mathbf{X}_t^\dagger = \mathbf{X}_t - \overline{\mathbf{X}} \quad (5.1)$$

and set all elements of $\boldsymbol{\mu}$ to zero.

MARSS package provides an option to standardize the data (z-score) mentioned in [7], which speeds up the computation time up to 2-3 times. It is a possibility to do, when the higher number of iterations is demanded in the EM algorithm within practical computation time or for the model comparisons. We do not standardize the data, because we want to preserve the information of the variance structure of the observed time series. Dynamic factor model of standardized sets may be seen in [7], [25], [34] or in [35].

5.3 MARSS package

Multivariate Auto-Regressive State-Space (MARSS) package is a R package used for estimating the parameters of linear state-space models with Gaussian errors for time series [29]. We mostly use the information from [30] and the codes from MARSS User Guide [25]. Dynamic factor model is a special case of a state-space model [9]. Package provides computation of Kalman filter, Kalman Smoother and EM algorithm for estimating parameters of our model, which is exactly what we need. Basic initialization of a MARSS model for dynamic factor analysis in R with possible inputs may be seen in [Attachments](#).

In our analysis we work with the setting from Table 6.1. To find the best model fit we will be estimating models with $m= 2, 3, 4, 5$ latent factors and variance-covariance structures "diagonal and equal", "diagonal and unequal", "equal variance covariance" and "unconstrained" (see section 5.3.1).

5.3.1 Variance-covariance matrix R

The most important parameter we set in the model is the expected variance-covariance structure of our modeled time series. The expected form of R directly influence the model structure and computation time.

Let us assume 5 observed time series ($p = 5$). Then

- $R =$ "diagonal and equal"

$$R = \begin{bmatrix} \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 \end{bmatrix} \quad (5.2)$$

means we expect time series to be modeled with the same variance and no covariances between the different time series. Only one parameter σ^2 is being estimated, which means shorter computation time (which allows higher number of

5. APPLICATION OF A DYNAMIC FACTOR MODEL

EM iterations) is demanded. This would represent conditional independence and homoscedasticity of the measurements at different stations.

- $R =$ "diagonal and unequal"

$$R = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma_4^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma_5^2 \end{bmatrix} \quad (5.3)$$

means we expect different variances of the time series and no covariance between them. p parameters (equal to number of time series) have to be estimated. This may represent conditional independence and heteroscedasticity of the measurements at different stations.

- $R =$ "equal variance covariance"

$$R = \begin{bmatrix} \sigma^2 & c & c & c & c \\ c & \sigma^2 & c & c & c \\ c & c & \sigma^2 & c & c \\ c & c & c & \sigma^2 & c \\ c & c & c & c & \sigma^2 \end{bmatrix} \quad (5.4)$$

means we expect the same variance for each time series and the same non-zero covariance for every two time series. Here just the 2 parameters are estimated. This allows for conditional dependence between the stations, which might be far from true in the spatial case.

- $R =$ "unconstrained"

$$R = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \sigma_{24} & \sigma_{25} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{34} & \sigma_{35} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 & \sigma_{45} \\ \sigma_{51} & \sigma_{52} & \sigma_{53} & \sigma_{54} & \sigma_5^2 \end{bmatrix} \quad (5.5)$$

means we do not restrict the variance-covariance matrix to have any pre-defined structure. This is basically the best option, since we do not lose any information possible to get from the data. The problem is, it is computationally demanding to estimate such number of parameters for a larger set of time series. We compute $\frac{p(p+1)}{2}$ parameters just for the variance-covariance matrix. This structure would be able to reflect the conditional spatial dependence between the stations, which could lead to a much more complex results.

5.3.2 Convergence tests

There are two ways how the EM algorithm ends. Either maximum number of iterations had been reached or convergence tests pass. The MARSS package works with two convergence tests *abstol* and *con.test.slope.tol*.

5.4. DATA ANALYSIS

- *abstol* - Absolute tolerance of the difference between the last and the previous iteration of the log-likelihood function. If the absolute value is less than 0.0001, we consider log-likelihood function to be converged.

$$|\log L_{(j)} - \log L_{(j-1)}| < 0.0001 \quad (5.6)$$

This test applies only to the log-likelihood function.

- *conv.test.slope.tol* - Tests slope of the log parameter against the log iterations (log-log slope convergence test). If the absolute value of the slope is less than 0.5, we consider the parameter to be converged. For elaborate explanation see [Attachments](#). This test applies to the parameters of the loadings matrix \mathbf{L} , variance-covariance parameters \mathbf{R} and the log-likelihood function $\log L$.
- *conv.test.deltaT* - Number of iterations to use in the computation of the log-log slope test.

If both tests are passed for the log-likelihood function and the log-log slope test is passed for the parameter estimations of \mathbf{L} and \mathbf{R} , then EM iteration process is ended. This is demanded situation.

5.3.3 Time consumption

Unfortunately, EM algorithm is quite time consuming and therefore MARSS package provides an option to set a maximum number of iterations to be passed through EM algorithm. If the maximum number of iterations is computed and not every parameter had converged, the computation ends and the output states, which of the parameters had converged and which not. Creating an appropriate model means to deal with a problem of exchanging complexity of a model for a computational time. More complex model, such as having more latent factors, more time series and more complex variance-covariance structure, means we are able to do less number of iterations in the same amount of time, thus possibility of not reaching the MLEs.

Because of this issue, in our analysis we decided to do as much iterations per model as possible, but not to exceed more than around 12 hours of computational time per model. If the model is less complex, we computed more iterations, if the model is difficult to estimate, we did fewer iterations. Though the best would be to compute the same number of iterations for each model, thus the comparison of the models would be more relevant, we are not able to do that and we consider the issue as an disadvantage of the technique.

5.4 Data analysis

Main part of this thesis is to find the most suitable model for the observed data. For July data we tried fitting different number of factors and different variance-covariance structures data according to setting in Table 6.1. In Table 5.2 we may see the resulted models.

5. APPLICATION OF A DYNAMIC FACTOR MODEL

R	m	logLik	K	AICc	time	convergence	warnings	maxit	p
diagonal and equal	2	-10623.286	216	21708.51	2010.81	0	0	3732	108
diagonal and equal	3	-10225.161	322	21163.09	17178.24	10	81	10000	108
diagonal and equal	4	-9990.506	427	20960.19	43630.98	10	79	10000	108
diagonal and equal	5	-9787.76	531	20838.15	39213.58	10	47	5000	108
diagonal and unequal	2	-10088.3	323	20891.81	1115.13	0	0	1457	108
diagonal and unequal	3	-9650.341	429	20285.12	19588.65	10	1	10000	108
diagonal and unequal	4	-9319.98	534	19911.08	35710.92	10	11	9000	108
diagonal and unequal	5	-9168.171	638	19913.32	40031.75	10	76	5000	108
equalvarcov	2	-10604.421	217	21673.07	5680.3	10	19	10000	108
equalvarcov	3	-10210.927	323	21137.07	23800.17	10	54	10000	108
equalvarcov	4	-9981.426	428	20944.66	37093.29	10	114	8000	108
equalvarcov	5	-9775.758	532	20816.98	49754.04	10	62	5000	108

Table 5.2: Resulted models. K is the number of parameters being estimated. Convergence number has a specific meaning within MARSS [33] (10 - abstol convergence of the log-likelihood has been reached but some parameters did not converge in the log-log slope test, 0 - both tests passed). Warnings column states the number of parameters that did not converge. Maxit states number of computed EM iterations.

We were not able to compute a model with "unconstrained" variance-covariance structure (5.5). Due to complexity of the structure, with more than 30 time series in the model we obtained information that the computation became unstable because of R update matrix being not positive definite.

In the literature, the main criteria used for the comparison of the models and choice of the best model fit is AICc (3.3). The model with the smaller value of AICc is considered to be the better model. In Table 5.3 we may see the order of the models from the best to the worst regarding their difference in the AICc from the lowest (best) value. As we

R	m	logLik	K	Δ AICc	convergence	warnings
diagonal and unequal	4	-9319.98	534	-	10	11
diagonal and unequal	5	-9168.171	638	2.243111	10	76
diagonal and unequal	3	-9650.341	429	374.03731	10	1
equalvarcov	5	-9775.758	532	905.895129	10	62
diagonal and equal	5	-9787.76	531	927.072189	10	47
diagonal and unequal	2	-10088.3	323	980.732351	0	0
equalvarcov	4	-9981.426	428	1033.574577	10	114
diagonal and equal	4	-9990.506	427	1049.106205	10	79
equalvarcov	3	-10210.927	323	1225.987313	10	54
diagonal and equal	3	-10225.161	322	1252.00439	10	81
equalvarcov	2	-10604.421	217	1761.987351	10	19
diagonal and equal	2	-10623.286	216	1797.431816	0	0

Table 5.3: Order of the model fits from the best to the worst with respect to the difference of AICc.

can see from Table, the best model fit to our data seems to be a "diagonal and unequal" variance-covariance structure (5.4). Also it seems that more factors provide a better fit to our data in general. For the structures "equalvarcov" and "diagonal and equal" higher number of factors are always a better fit to the data. For "diagonal and unequal"

5.4. DATA ANALYSIS

structure 4 and 5 factors seem to model the data quite similarly. We were able to compute more iterations for 4 factor model and we obtained more converged values. Therefore we conclude that models with 4 and 5 latent factors for "diagonal and unequal" structure are the best model fit to our data. In Figure 5.5 we may see the visual comparison of the models based on the AICc.

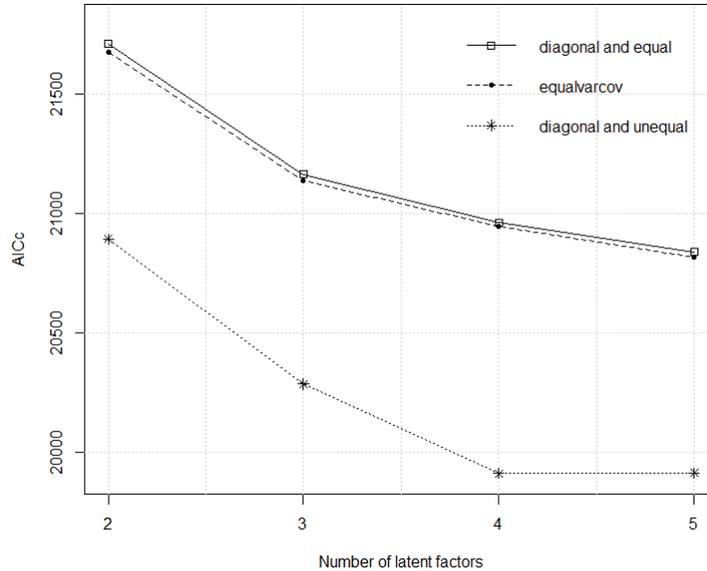


Figure 5.5: Comparison of the models based on the resulted value of AICc. Smaller AICc means a better fit to the data (compared to other models).

Let us see, how the fits of different models for particular time series look like. We have chosen two time series based on their estimated variance from a dynamic factor model, which for station 316 is low and for station 297 is high. In Figure 5.6 we may see their centered FWI measurements with a fit of model with "diagonal and unequal" structure and 4 latent factors.

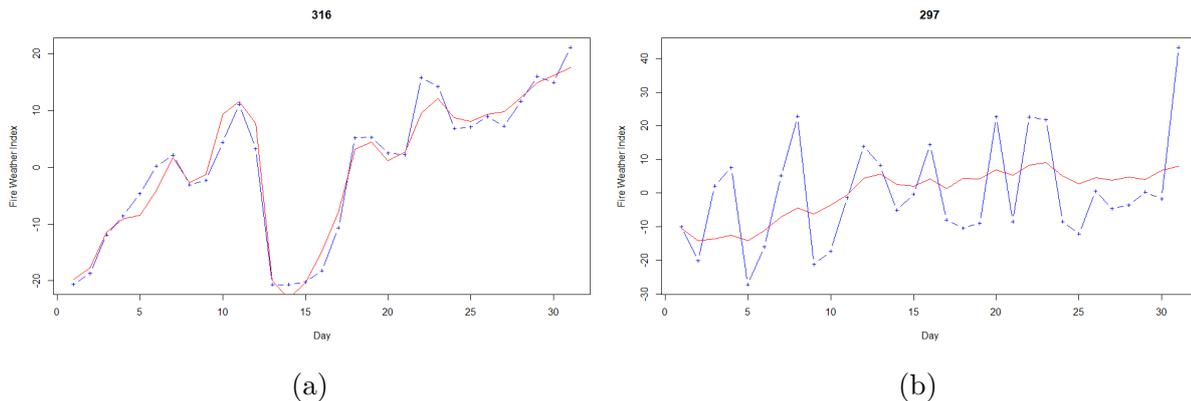


Figure 5.6: Particular time series (blue lines) and their estimation by a "diagonal and unequal" model with 4 latent factors (red lines). (a) station 316 with estimated variance $\hat{\sigma}_{316}^2 = 6.7829$, (b) station 297 with estimated variance $\hat{\sigma}_{297}^2 = 182.3466$.

We may observe that station 316 has a significant trend and the model is able to capture it. On the other hand, station 297 has greater volatility and no distinct trend.

Let us now see, how different models fit these two stations. First we will demonstrate

5. APPLICATION OF A DYNAMIC FACTOR MODEL

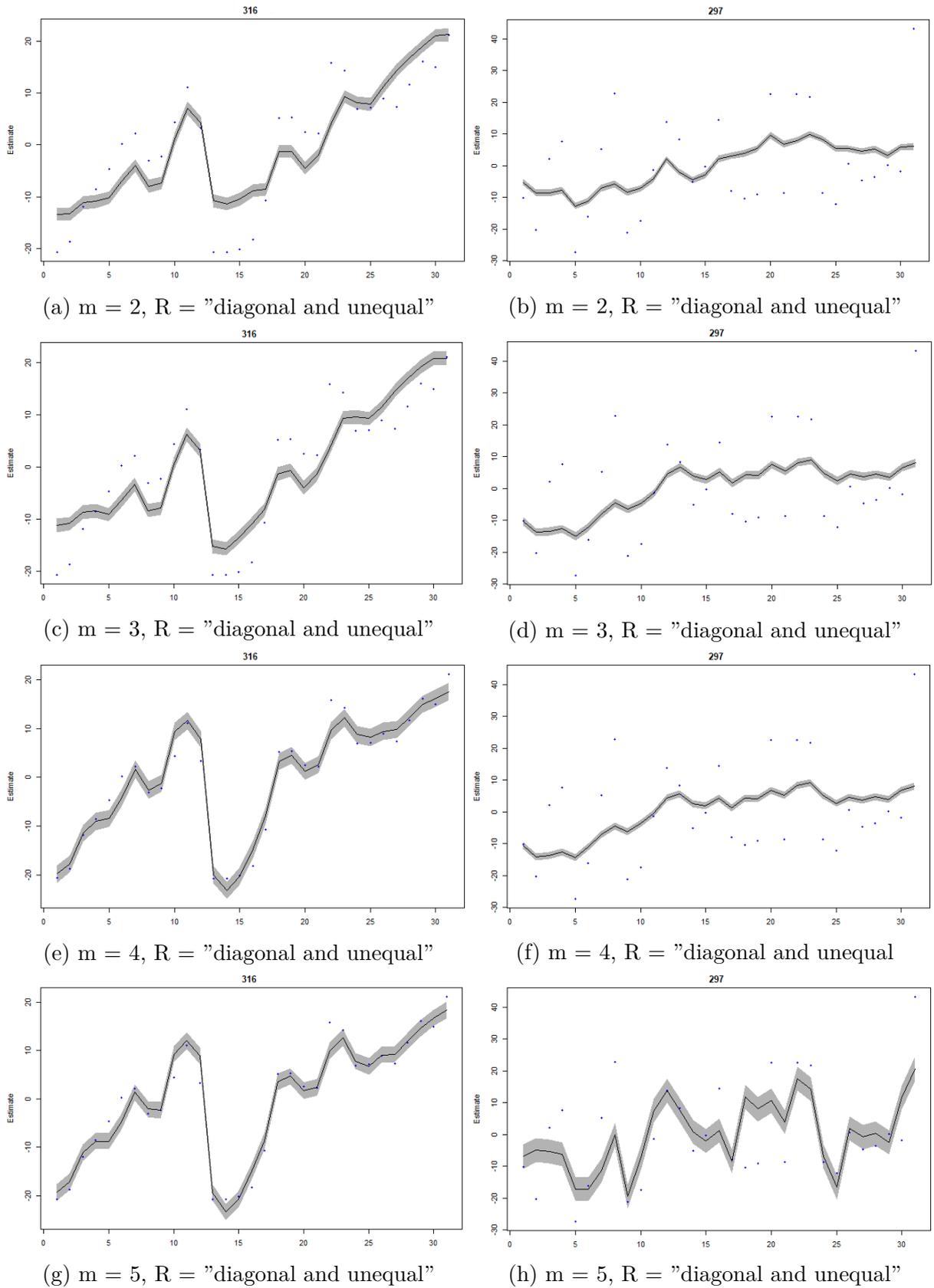


Figure 5.7: Plots of the model fits with same variance-covariance structure but different number of factors.

5.4. DATA ANALYSIS

the difference on models with the same variance-covariance structure (diagonal and unequal) and different number of latent factors. As we can see in Figure 5.7, for station 316 with growing number of latent factors the fit gets closer to the real time series values and most of the real values are caught in the confidence interval of the model fit. The fit changes mostly for 2-4 latent factors and the difference between fit of 4 and 5 factors is negligible. For station 297, there is little difference in the fit of the models with 2-4 latent factors and the real change happens with 5 latent factors. Fit gets closer to the real time series values and the estimated variance gets smaller.

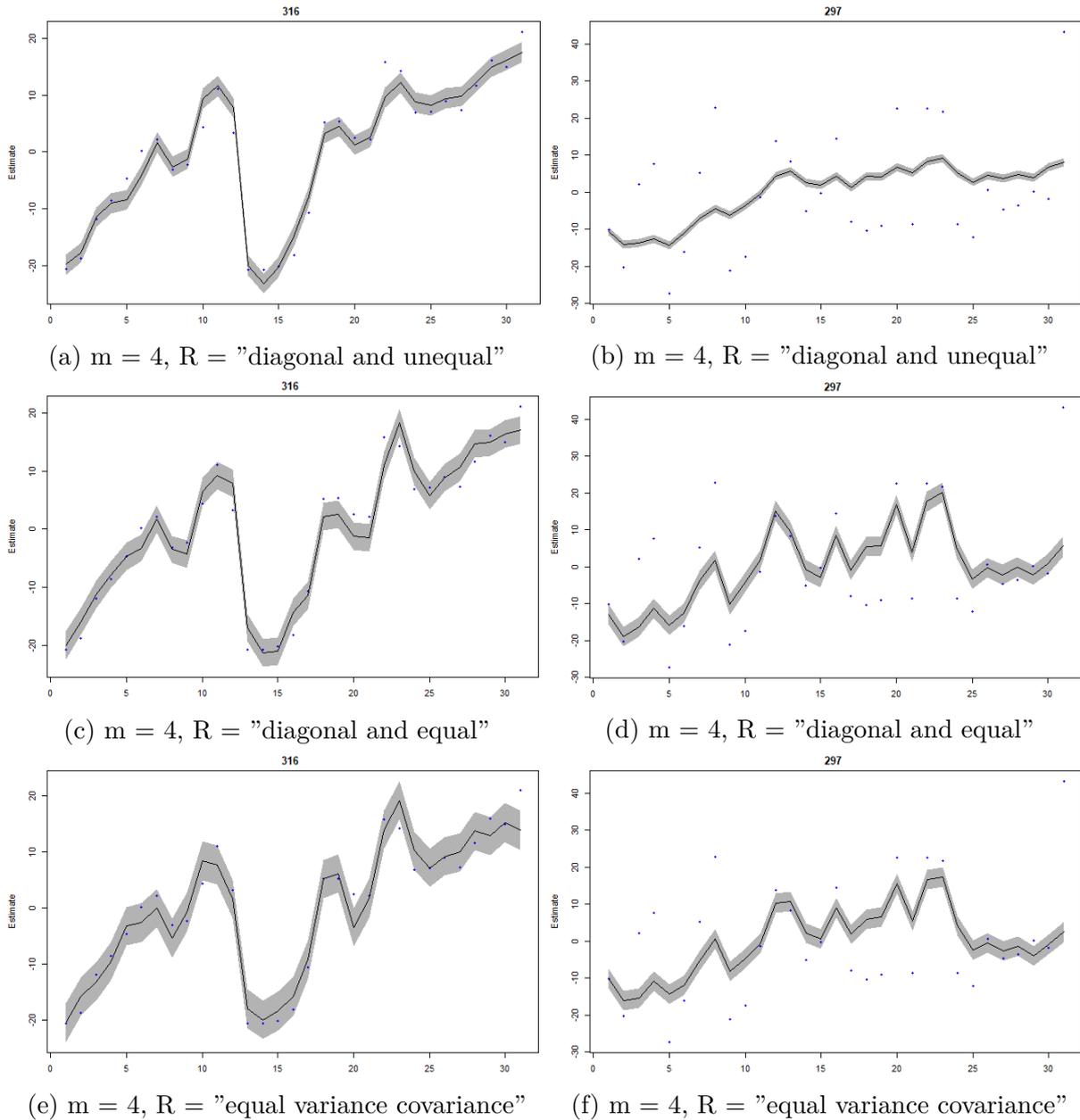


Figure 5.8: Plots of the model fits with the same number of factors but different variance-covariance structure.

When comparing different variance-covariance structures with same number of factors in Figure 5.8, model fits behave similarly for station 316. Each structure is able to catch the trend and for "equal variance covariance" structure we observe bigger confidence

interval. Confidence interval for i th time series is computed as a fitted value plus-minus $1 - \alpha/2$ quantile of a normal distribution times standard error $\mathbf{l}_i \mathbf{F}_t \pm z_{0.975}(\mathbf{l}_i \mathbf{H}_{t|T} \mathbf{l}_i')$, where \mathbf{l}_i is the i th row of the loadings matrix [30]. For station 297 structures "diagonal and equal" and "equal variance covariance" behave very similarly, we also do not observe such a grow in the confidence interval. Neither of the structures is able to model real time series values for this station.

5.4.1 Best model fit

Based on the AICc from Table 5.3, we should consider two models as the best fits for the data, diagonal and unequal variance-covariance structure with 4 and 5 latent factors. Since the 5 factors model is more complex, we were able to compute less number of iterations than for 4 factors model. Therefore the difference between converged estimates is 65 parameters. We do not want to over-fit the data, therefore if it is possible, we choose simpler model over more complex one. As we can see in Figure 5.7, more factors provide better fit for the high variance stations (297) but for the low variance, it does not change much. We do not want to adjust model selection to fit the potential outliers better, but we want to find the equilibrium between the good data fit and the model complexity.

Taking all of this into consideration we conclude, that the best model fit for our data is a dynamic factor model with diagonal and unequal variance-covariance structure (5.4) and 4 latent factors.

R	m	logLik	K	AICc	time (s)	convergence	warnings	maxit	p
diagonal and unequal	4	-9319.98	534	19911.08	35710.92	10	11	9000	108

Table 5.4: Best model fit.

It is an exact dynamic factor model, as mentioned in section 2.3, and in the following analysis we will work with this model.

As we can see, with this model we reached the abstol convergence of the log-likelihood function and 11 parameters did not pass log-log slope test. Histogram of parameter slopes with and without converged parameters and histogram of differences between the last two iterations may be seen in Figure 5.9.

5.4. DATA ANALYSIS

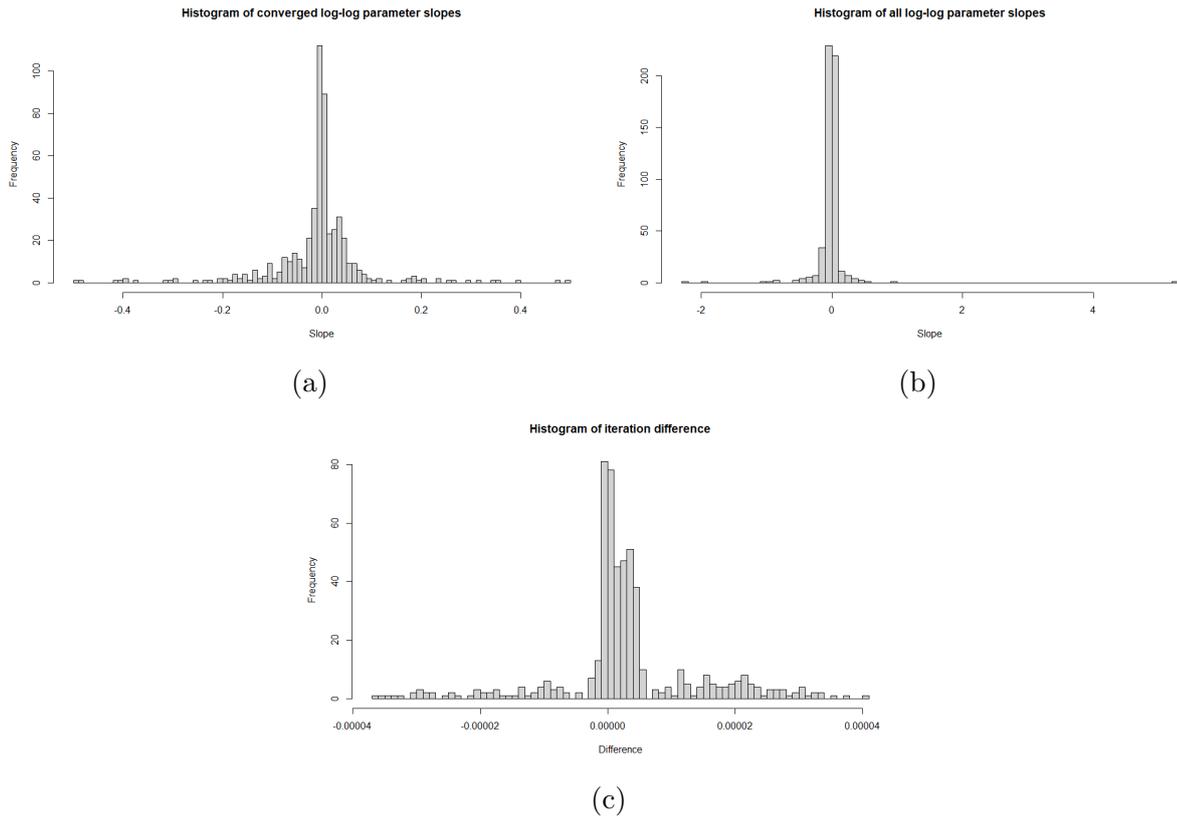
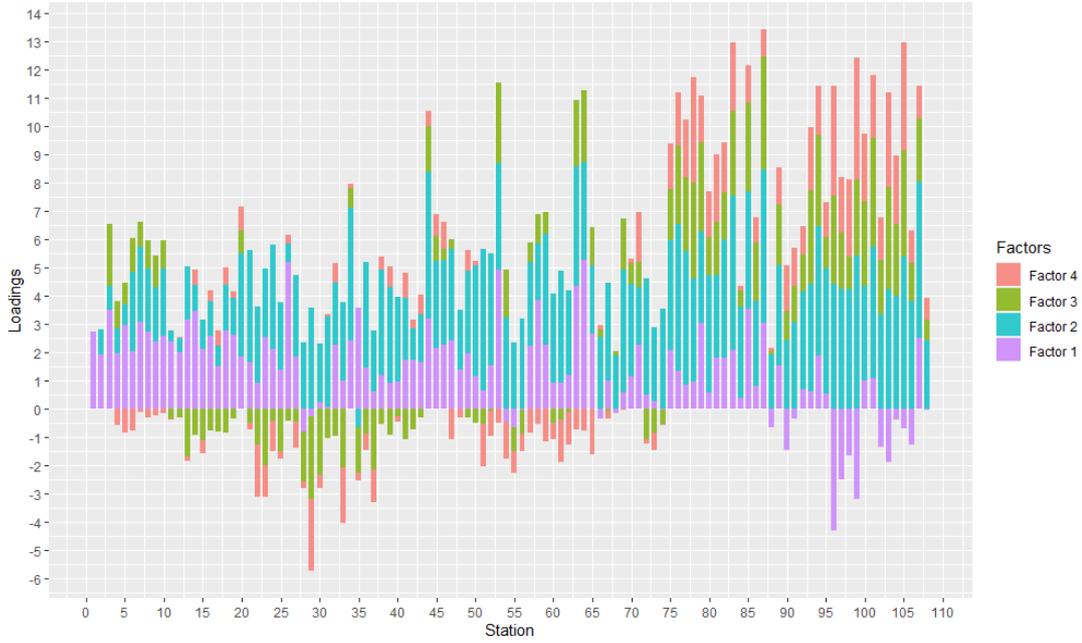


Figure 5.9: (a) Histogram of slopes of last 10 iterations for each parameter that passed log-log slope test. (b) Histogram of slopes with parameters that did not pass the log-log slope test. (c) Histogram of the parameter differences between the 8999th and the 9000th iteration.

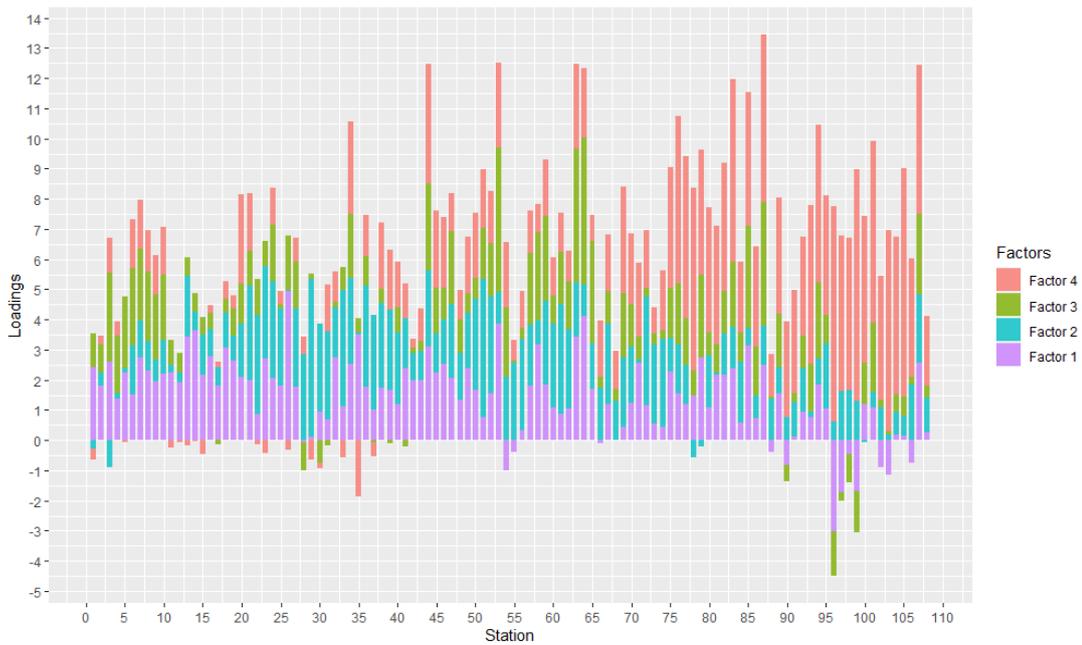
We can see that the differences are small even for not converged values. The model seems to be reasonable regarding convergence tests.

Let us now see how the loadings matrix \mathbf{L} and the latent factors \mathbf{F}_t change after varimax rotation. In Figure 5.10 we may see the stacked histogram with the loadings of the particular time series.

5. APPLICATION OF A DYNAMIC FACTOR MODEL



(a) before varimax rotation



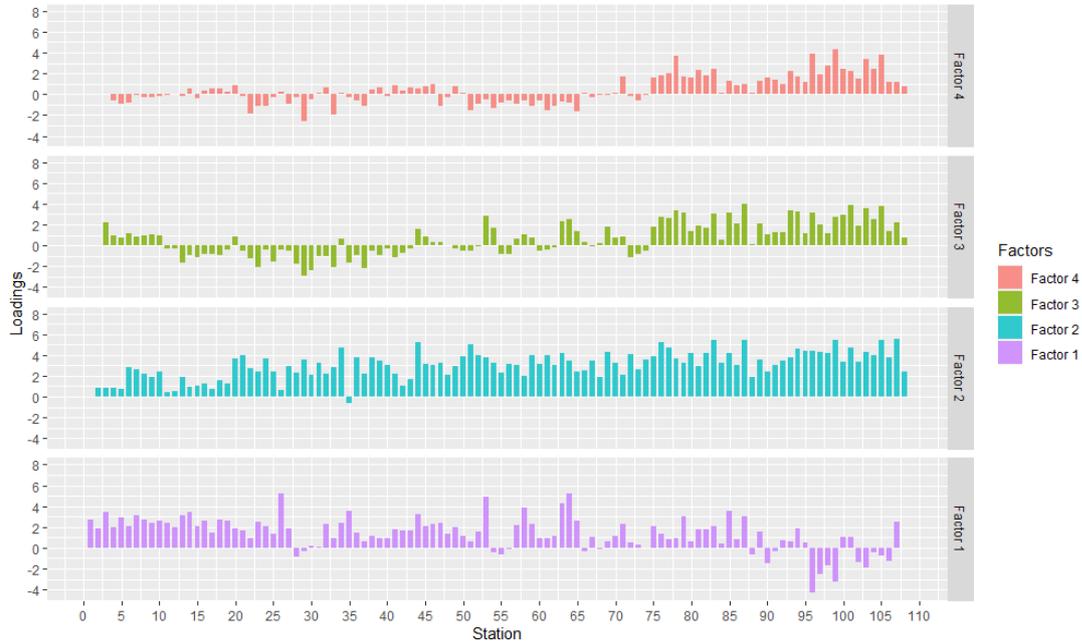
(b) after varimax rotation

Figure 5.10: Stacked histogram of loadings values of latent factors for each modeled time series (in other words estimated l_{ij} values). A column in the graph is the sum of the factor loadings for a particular time series. To obtain a loading of the particular factor on the particular time series, we have to search for that series on the x-axis, then find the proper colour representing the factor we want to find loading for and then subtract the lower bound from the higher bound of that particular color in the column.

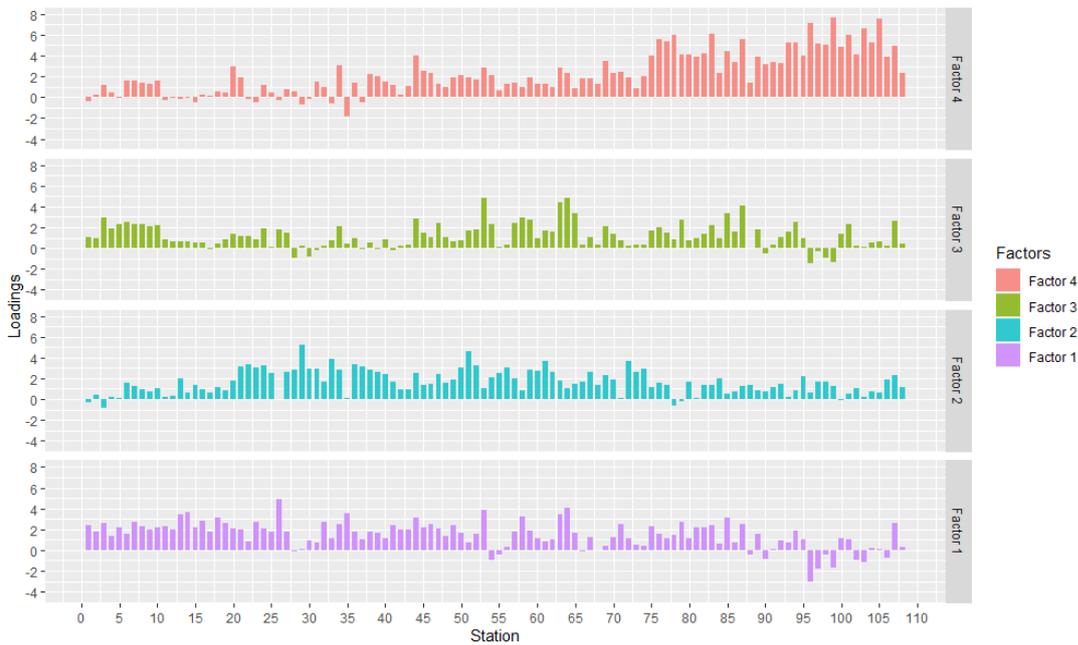
We can observe some small changes. The most significant seems to be the change on last 30 time series, where factor 4 became quite dominant after varimax rotation. The problem of first time series being modeled just by the first latent factor, following up to

5.4. DATA ANALYSIS

4th time series, is solved. Besides values changing sign, we do not observe the promised result of the varimax rotation, that is to create bigger difference between the loadings and promote just a subset of the factors. We were expecting to obtain one or two dominant factors for each time series. If we compare loadings of the individual factors in Figure 5.11, we observe that factor 4 got multiplied within last 30 stations. On the other hand,



(a) before varimax rotation



(b) after varimax rotation

Figure 5.11: Histogram of the loadings separated by factors.

factor 2 got decreased. Factor 3 has multiple changes, first 10 stations increased and last 30 stations decreased. Factor 1 has almost no significant changes.

5. APPLICATION OF A DYNAMIC FACTOR MODEL

Varimax rotation changes the loadings matrix and also the factors. Let us compare the factors before and after rotation in Figure 5.12. We may observe that the factors have

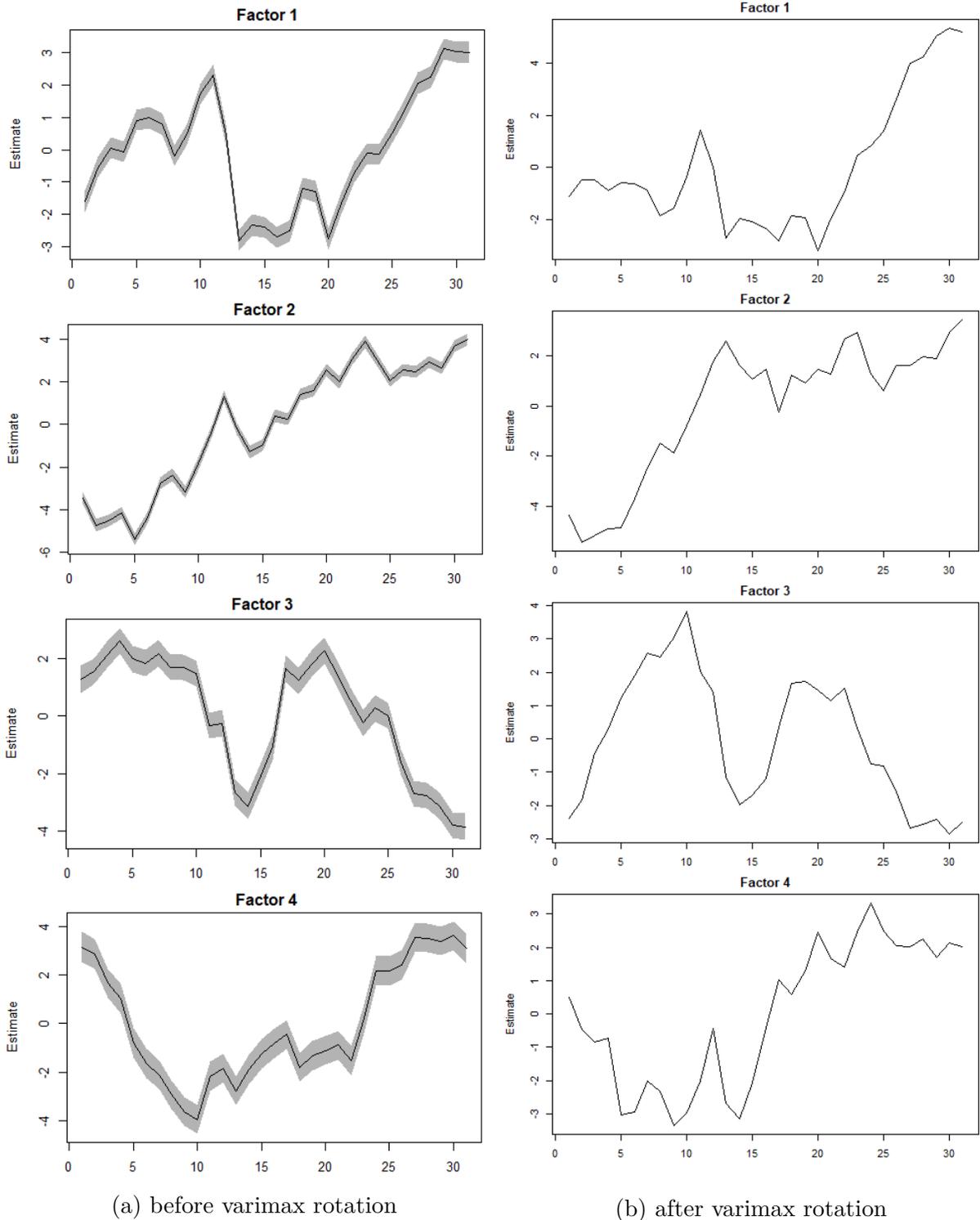


Figure 5.12: Latent factors F_t before and after varimax rotation.

very similar trends and they differ mostly in the amplitude of the change between days. The most significant change is in the beginning of the factor 3, which changed the sign and is rising instead of being more or less constant.

5.4.2 Outliers detection

As we could see in Figures 5.7 and 5.8, the station 297 is modeled with high variance. Therefore, the station may be considered as the *outlier* of the dataset. If we assume the best model fit for the data is as in Table 5.4, then if the variance estimated for a particular time series is too high, time series may be considered as an outlier. Let us see this in the histogram and the boxplot of the model variances in Figure 5.13. The boxplot detected

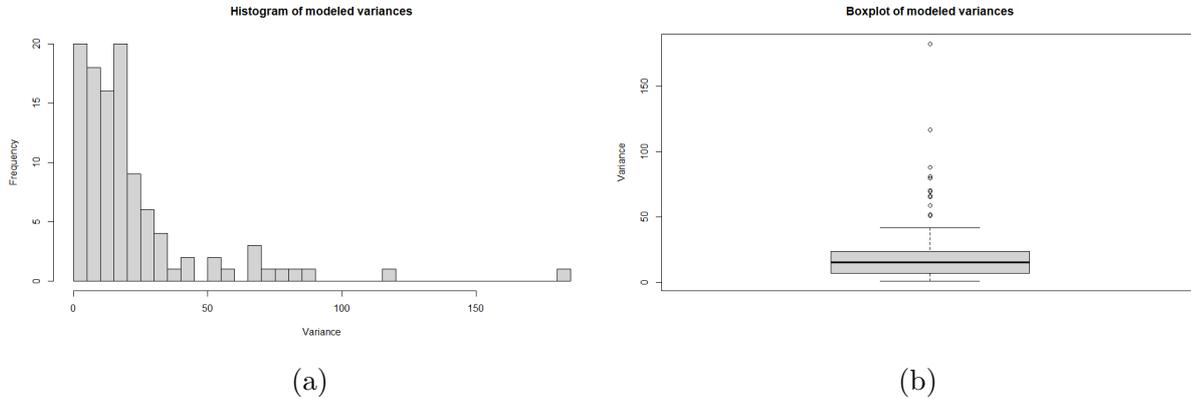


Figure 5.13: (a) Histogram of the estimated variances. (b) Boxplot of the estimated variances.

12 high variance time series with the values from 50.9 - 182.35.

We determine the outliers using 99% quantile of a data distribution fit [32]. Based on the Shapiro-Wilk goodness of fit test [31] we found out, that we do not reject hypothesis, that variances follow a lognormal distribution.

Shapiro-Wilk Goodness of Fit test

```
data: Variances
W = 0.9884743, p-value = 0.4868429
alternative hypothesis: True cfd does not equal
Lognormal distribution
```

Table 5.5: Output of the Shapiro-Wilk Goodness of Fit test for a lognormal distribution [31].

We also tested for Weibull and Gamma distribution and both tests rejected the null hypothesis.

So if we fit a lognormal distribution in Figure 5.14, we obtain a lognormal distribution with estimated parameters $\mu = 2.602$, $\sigma^2 = 0.989$. Now using a 99% quantile we will select the distribution outliers. For $LN(2.602, 0.989)$ a 99% quantile is $q_{0.99} = 133.1376$. Therefore we will consider higher values as the outliers. That makes 1 outlier with variance: station 297 - 182.35.

How to set the right threshold of variance for claiming a time series to be an outlier is up to discussion. Dynamic factor model provides us with the information only about estimated variances. We decided to consider a time series as an outlier, if it does not belong to a 99% quantile of a fitted distribution following [32]. Further, we will observe the influence of the selected outlier on the model estimation and the conclusion.

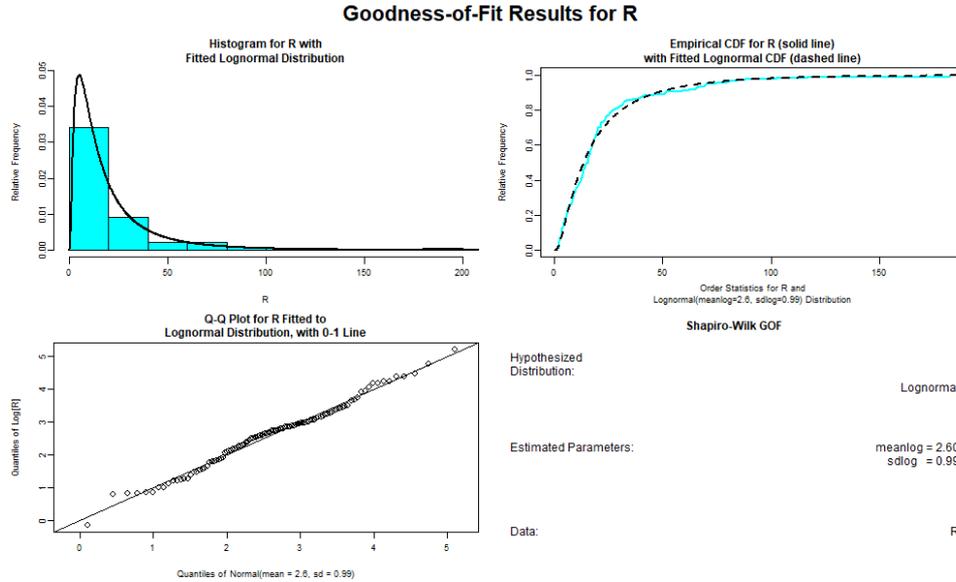


Figure 5.14: Lognormal distribution fit of the estimated variances (diagonal of estimated variance-covariance matrix R). Estimated distribution parameters are: $LN(\mu, \sigma^2) = LN(2.602, 0.989)$.

Let us see the relation of estimated variances with sample means and sample variance estimates. In Figure 5.15 we may observe, that with rising sample means estimated

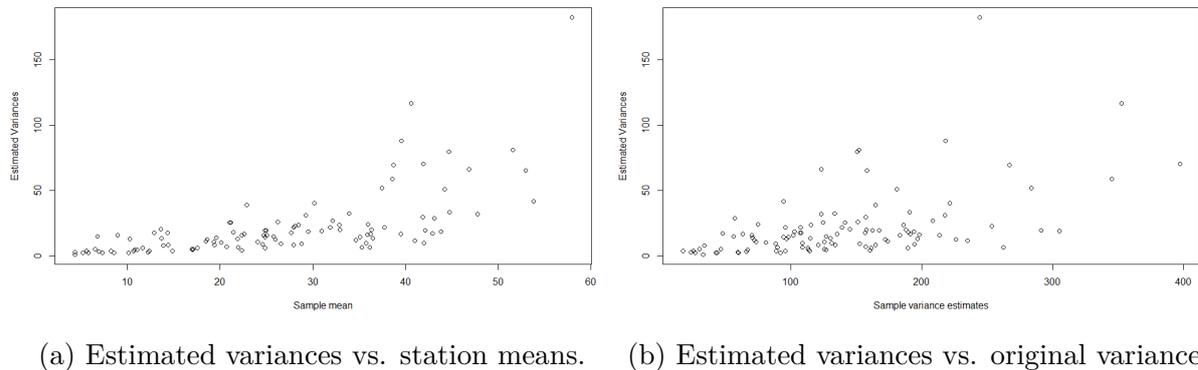


Figure 5.15: Scatter plot of estimated variances (diagonal of R) vs sample means (used when centering data in section 5.2) and sample variance estimates.

variances stay low, while with sample means over 38 some variances get much higher. Especially, our detected outlier has the highest sample mean. On the other hand, when compared to sample variance estimates, we observe that the outlier has still high variance but there are 9 values bigger than our outlier.

5.5 Clustering

To create clusters out of our 108 stations is the main purpose of this thesis. Dynamic factor model provides us with a loadings matrix, which may be a good base for creating clusters. We will be using a common clustering method *k-means*. For the computation of *k-means* we will use the R package *stats*. For determining number of clusters and

5.5. CLUSTERING

visualizations of the techniques we will use the package *NbClust*. Following analysis of determining the number of clusters is based on [36] and [37].

There are many methods how to determine number of clusters for k-means. We will demonstrate two of them.

- *Elbow method* - Method based on computing total within sum of squares (WSS) for different number of expected clusters. We search for a number of clusters for which WSS is the smallest and at the same time, if we add one more cluster, WSS does not change significantly. In Figure 5.16 we can see that the elbow is at number 2. Up to

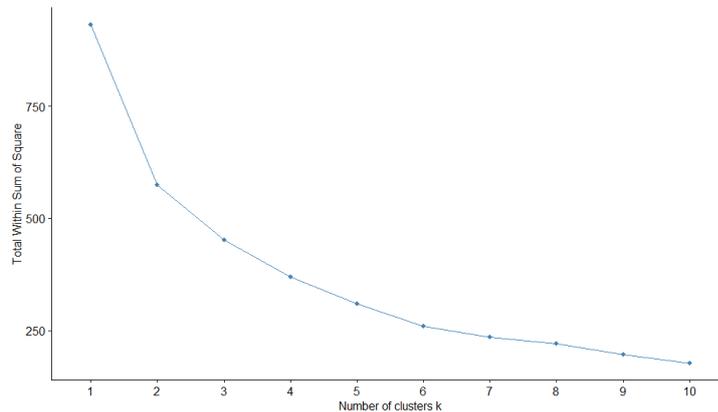


Figure 5.16: Elbow method for determining the number of clusters.

number 4, WSS still changes quite a lot, then after 4 and 5 clusters, WSS changes just a little bit. Therefore from this method we may conclude, that choosing 2-4 clusters might be relevant.

- *Gap statistic* - Explained in [37], gap statistic measures total within intra-cluster variation for different values of k with their expected values under null reference distribution of the data. The number of clusters that maximizes the gap statistics is considered to be the most appropriate. We observe that the highest values of

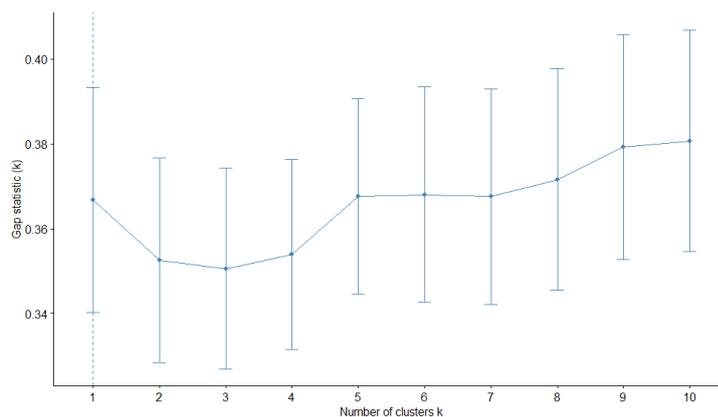


Figure 5.17: Gap statistic for determining the number of clusters.

gap statistic are for 5 and more clusters and mostly the same for 2-4 clusters. This suggests that 5 clusters should be appropriate. Result contradict the conclusion from elbow method.

5. APPLICATION OF A DYNAMIC FACTOR MODEL

We decided to choose 4 clusters, as it seems to be appropriate for our 4 latent factor model.

After performing k-means on the rotated loadings matrix with 4 assumed clusters, we obtained a map with stations divided into clusters in Figure 5.18.

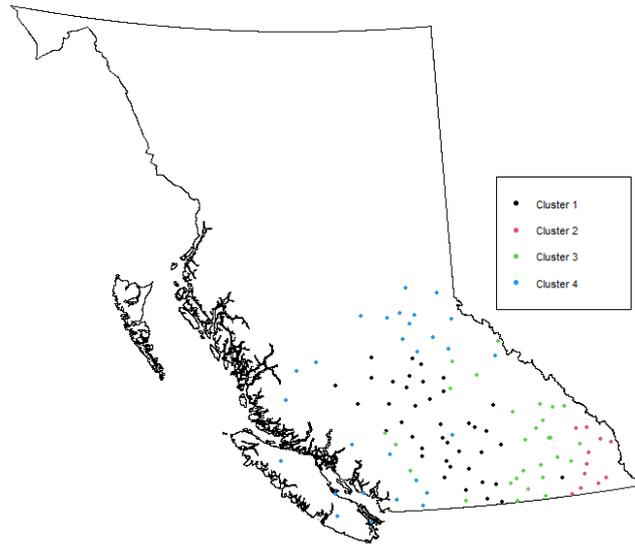


Figure 5.18: Map of the stations divided into clusters based on k-means clustering of the rotated loadings matrix obtained from a dynamic factor model.

For each cluster we obtained also a cluster mean (represented as a point in a 4-dimensional space), therefore multiplied with the rotated latent factors we may plot a corresponding cluster mean as a time series. In Figure 5.19 we may see clusters with their cluster means.

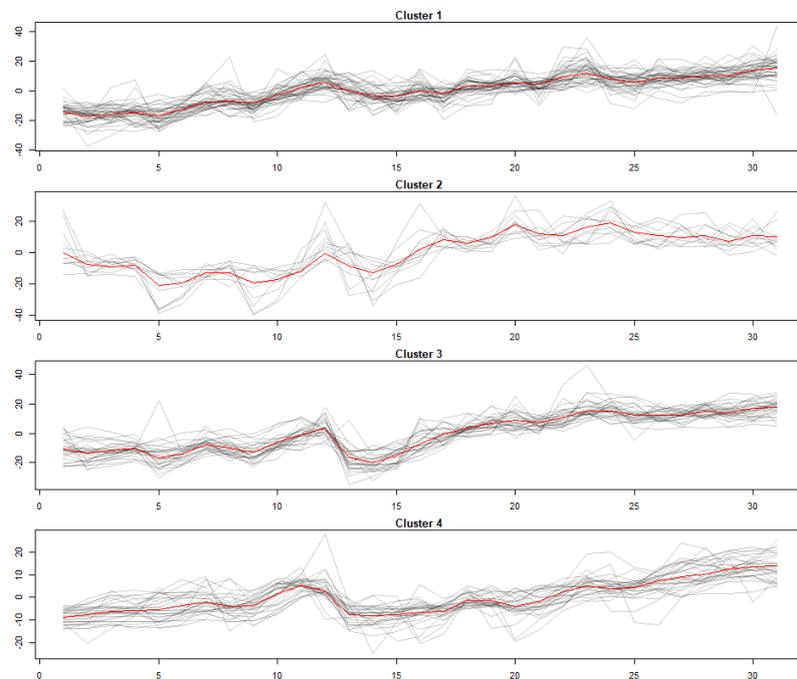
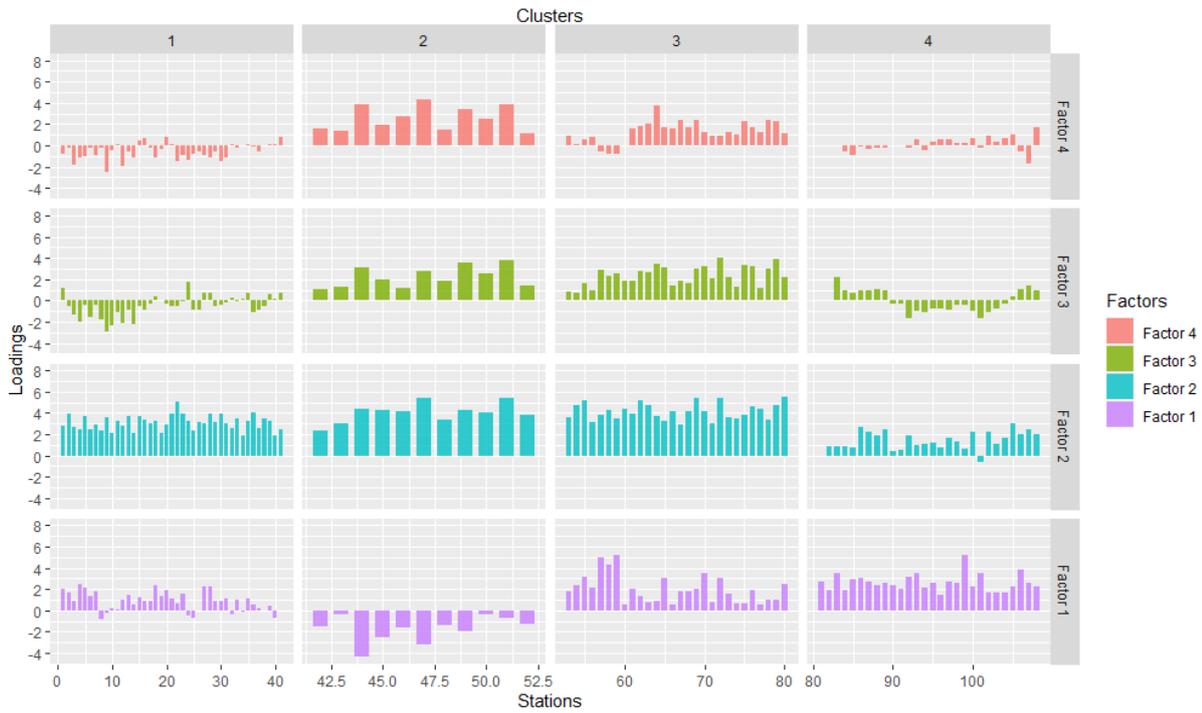


Figure 5.19: Plot of the observed centered time series divided into k-means clusters computed from the rotated loadings matrix of a dynamic factor model with cluster means (red lines).

5.5. CLUSTERING

Using the cluster division we may now plot in Figure 5.20 how the loadings change after varimax rotation within the clusters. We are able to see some significant changes. In



(a) before varimax rotation



(b) after varimax rotation

Figure 5.20: Factor loadings of the stations divided into clusters based on k-means clustering.

Figure 5.11 we observed a significant change in the factor 2 and 4. From Figure 5.20 we

may see that the change occurred mostly in the cluster 2 and 3 and the change within these clusters at the factors 2 and 4 is very significant. Besides this, we may see some little change in the cluster 1 at factors 3 and 4. Cluster 4 seems to change the least.

5.5.1 Outliers

Let us see, where the outlier and the time series with high variances, determined in section 5.4.2, are within the clusters and how do they behave. In Figure 5.21 we may observe, that 5 out of 11 stations in cluster 2 have high variances. Cluster 4 contains only one

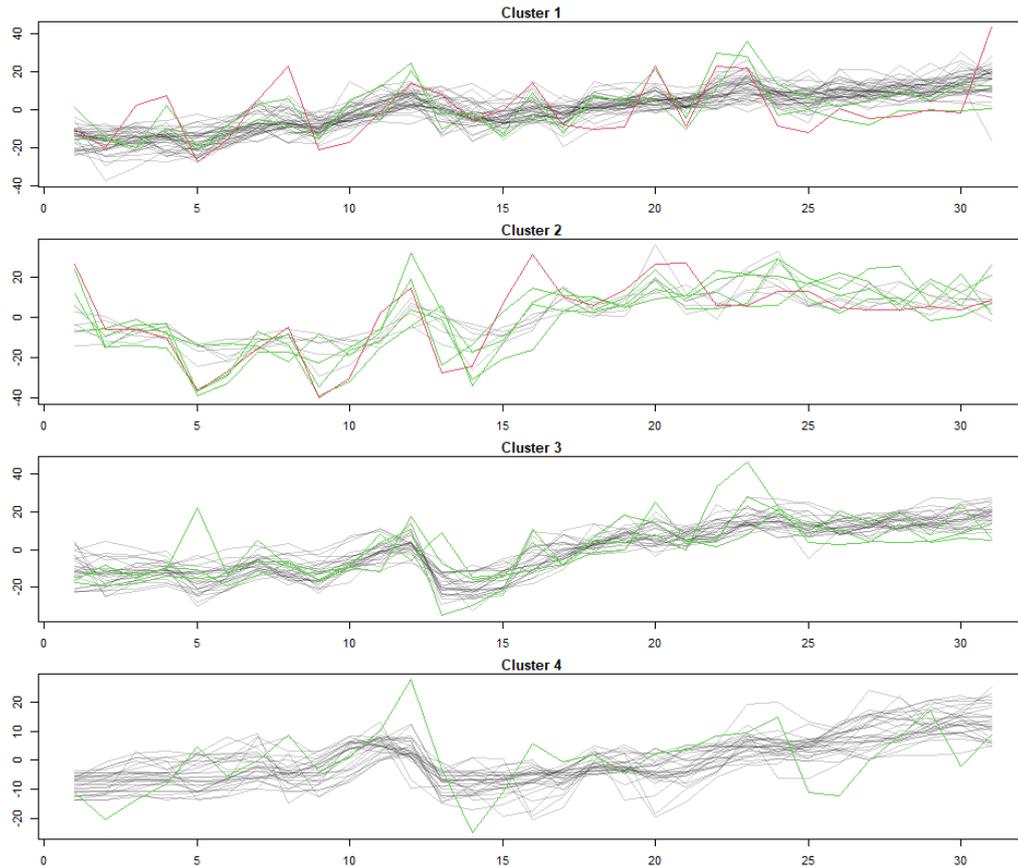


Figure 5.21: Plot of the outlier (red line) and the time series with high variance (green lines) within the clusters.

high variance station and the outlier belongs to cluster 1.

This suggests certain specificity. k-means clustering based on rotated loadings matrix selected cluster 2 as a cluster with few stations and simultaneously many of them being high variance stations. This may mean, that the stations in cluster 2 are somehow different from the others, that they belong to special climate or that it is difficult to predict their behaviour. On the other hand, cluster 4 contains only one high variance station. Model of these stations seem to be quite stable.

Let us see where the outlier and high variance time series are placed on the map in Figure 5.22. In the map we can see two areas at the east border and in the center, where the presence of the stations with high variances is quite significant. This can suggest places with specific climate or specific technical details with the stations. In terms of FWI measurements this may mean different things and should be discussed with experts.

5.6. COMPARISON WITH AUGUST DATA

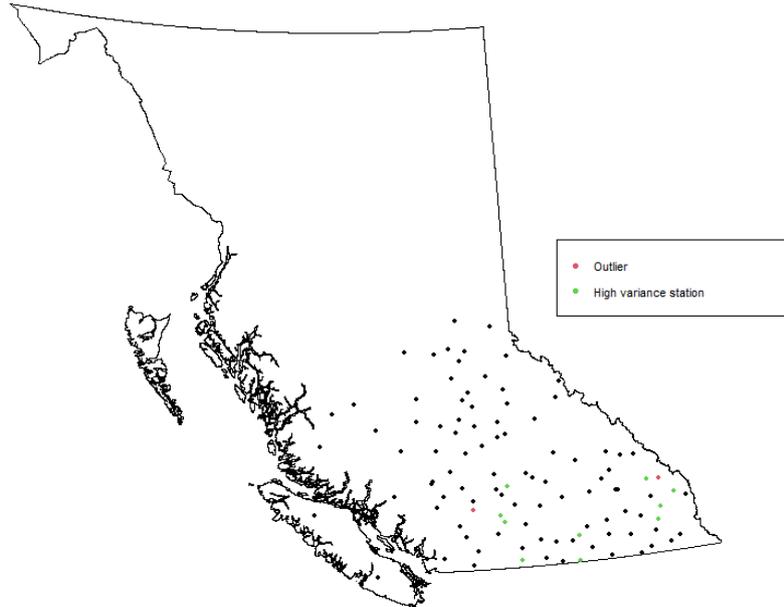


Figure 5.22: Map of the outlier and the stations with high variances.

5.5.2 Model without outliers

Outliers are considered to be time series that may significantly influence the model estimation. Therefore we tried to remove the outlier (stations 297) detected in section 5.4.2 and estimate the model again to see, if there would be any significant changes in the model and following cluster analysis.

We found out that removing an outlier from a dataset did not change the resulted clustering and mean trends within clusters. We observe same map and same cluster trends, therefore we do not provide the figures.

5.6 Comparison with August data

As mentioned in section 5.1, we chose 108 stations on purpose. All those stations have complete data also for the next month in year 2003, August. We repeated whole data analysis for the FWI measurements in August and focused on comparing the results obtained from clustering. All the relevant graphs created for August may be found in [Attachments](#).

For August FWI data the most suitable model is in Table 5.6. Same as for July, we

R	m	logLik	K	AICc	time (s)	convergence	warnings	maxit	p
diagonal and unequal	5	-10584	638	22745	30287	1	159	3000	108

Table 5.6: Best model fit for August data.

rotated loadings matrix together with latent factors and created 4 clusters with k-means. Resulted August map compared to July clusters may be seen in Figure 5.23. As we can see clusters changed. Cluster 2 contains more stations and reaches more to the center of the observed area in August. Cluster 4 seems to change the least. We may observe some specific groups of stations within clusters that did not change, i.e. cluster 1 in the

5. APPLICATION OF A DYNAMIC FACTOR MODEL

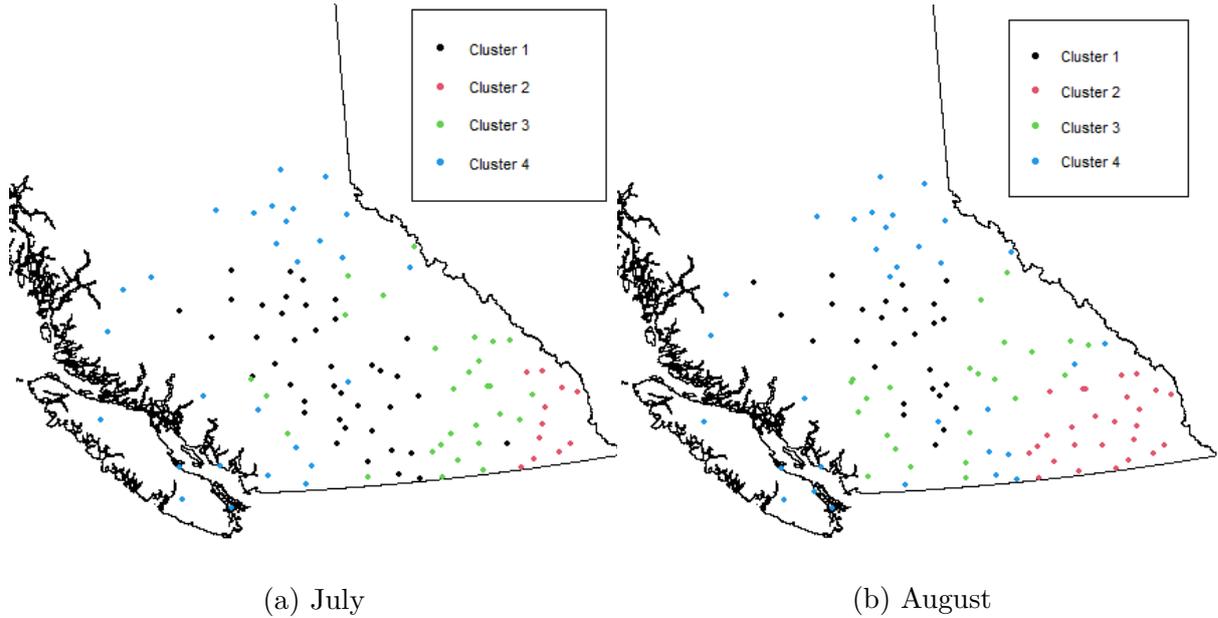


Figure 5.23: Map of stations divided into cluster for month July and August.

center, cluster 4 in the surrounding, cluster 2 in the South East corner. There are some interchanges between the clusters, mainly from cluster 3.

Clusters with mean trends may be seen in Figure 5.24.

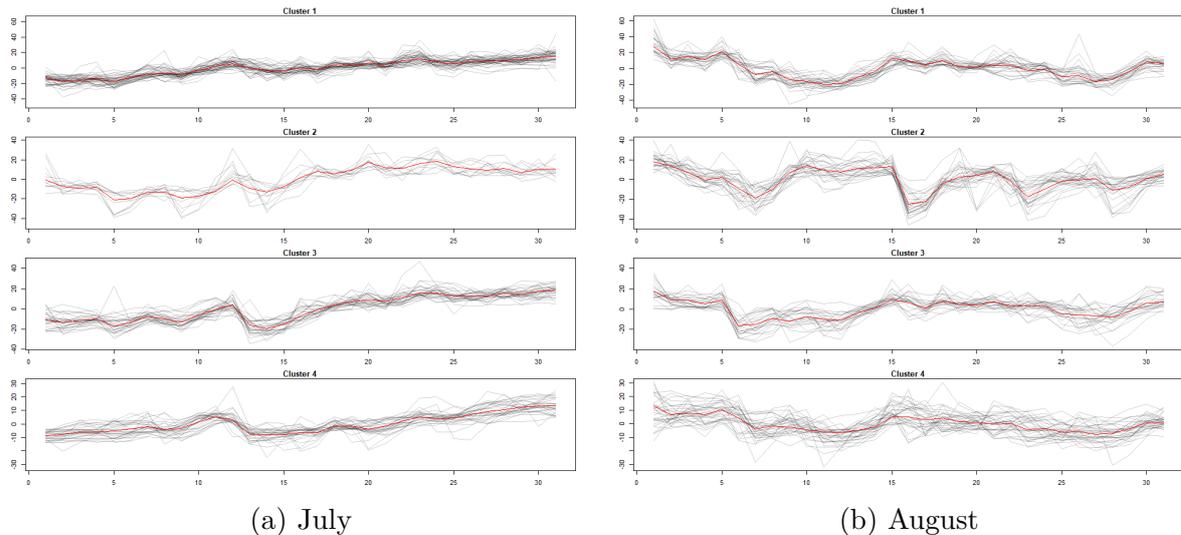


Figure 5.24: Division of stations into clusters with mean trends for months July and August.

In August we may observe two big drops in the FWI. In cluster 2 between 15th and 16th day and in cluster 3 between 5th and 6th day. Clusters were able to capture this anomaly and select stations with this behaviour.

Also we can observe that character of the clusters changed between two months. In August, stations seem to have greater variance than in July, especially in cluster 4. In cluster 2 we observe again, that there are periods in which some time series differ from mean trend significantly. Based on the [38] we calculated the Rand Index (RI), which measures the difference between two cluster partitioning and the Adjusted Rand Index

5.6. COMPARISON WITH AUGUST DATA

(ARI), which rescales the index and takes into account that by random chance some stations would belong to the same cluster. Result is a value between 0 and 1, where 1 means two partitioning are identical. We obtained $RI = 0.721703$ and $ARI = 0.2858123$. We see that Rand Index is quite high and suggests that two clusterings are similar, although taking random chance into consideration with ARI, the partitioning does not seem to be that similar.

Nevertheless, clustering of different month's data preserved some underlying structure of the station division. The results should be discussed with the experts to analyze, what may be the causes of such changes.

6 Conclusion

The goal of this thesis was to explore the possible use of the dynamic factor model in time series clustering. We explained the difference between the dynamic and the static form and stated a simplified form used in our data analysis. We discussed how the computation of a dynamic factor model is executed, what restrictions need to be placed on the model to overcome the problem of unidentifiability of the model, and how to use varimax rotation to overcome issues connected to the restrictions.

Then using FWI measurements from July 2003, we estimated dynamic factor model parameters and found out that a dynamic factor model with diagonal and unequal variance-covariance structure and 4 latent factors fits our observed data the best regarding AICc. We discussed how the loadings matrix changed with varimax rotation and how the model can determine outliers based on their estimated variance. The loadings matrix of a dynamic factor model proved to be a suitable dimension reduction tool for clustering. Thanks to it we were able to put stations into clusters and also see the cluster mean trends thanks to the multiplication of cluster means and estimated latent factors. This is a useful property of a dynamic factor model. In the end, we compared the results obtained for the month of July with results obtained for the month of August and we observed how the cluster division and cluster mean trends changed.

References

- [1] MAHARAJ, Elizabeth A., Pierpaolo D'URSO and Jorge CAIADO. *Time Series Clustering and Classification*. 1st edition. Chapman and Hall/CRC Press, 2019. DOI: 10.1201/9780429058264. ISBN 978-1-4987-7321-8.
- [2] JOHNSON, Richard A. and Dean W. WICHERN. *Applied Multivariate Statistical Analysis*. 6th edition. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1992. ISBN 978-0-13-187715-3.
- [3] BARIGOZZI, Matteo. *Dynamic Factor Models*. United Kingdom: London School of Economics and Political Science, Statistics Department, 2020. Lecture notes. Available from: http://www.barigozzi.eu/MB_DF_lecture_notes_online.pdf
- [4] MARDIA, Kanti V., James T. KENT and J. M. BIBBY. *Multivariate analysis*. 10th printing. London: Academic Press, 1995. ISBN 0-12-471250-9.
- [5] BROCKWELL, Peter J. and Richard A. DAVIS. *Introduction to Time Series and Forecasting*. 2nd edition. New York: Springer, 2002. ISBN 0-387-95351-5.
- [6] STOCK, James H. and Mark W. WATSON. Chapter 8 - Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics. *Handbook of Macroeconomics* [online]. 2016, vol. **2**, p. 415-525 [cit. 2021-01-28]. DOI: 10.1016/bs.hesmac.2016.04.002.
- [7] ZUUR, Alain F., R. J. FRYER, I. T. JOLLIFFE, R. DEKKER and J. J. BEUKEMA. Estimating common trends in multivariate time series using dynamic factor analysis. *Environmetrics*. 2003, **14**(7), p. 665-685. ISSN 1180-4009. Available from: doi:10.1002/env.611
- [8] HOLMES, Elizabeth E. *Derivation of an EM algorithm for constrained and unconstrained multivariate autoregressive state-space (MARSS) models*. Seattle, 2013. arXiv: 1302.3919 [stat.ME]. Technical report. Available from: <https://cran.r-project.org/web/packages/MARSS/vignettes/EMDerivation.pdf>
- [9] SHUMWAY, Robert H. and David STOFFER. *Time series analysis and its applications: with R examples*. 4th ed. Cham: Springer, 2017. Springer texts in statistics. ISBN 978-3-319-52451-1.
- [10] DEMPSTER, A.P., N.M. LAIRD and D.B. RUBIN. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1977, **39**(1), 1-38 [cit. 2021-03-15]. JSTOR: 2984875. MR: 0501537. Available from: <https://web.mit.edu/6.435/www/Dempster77.pdf>
- [11] HARVEY, Andrew C. *Forecasting, structural time series models and the Kalman Filter*. New York: Cambridge University Press, 1989. ISBN 978-0-521-40573-7.
- [12] KAISER, Henry F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika* [online]. 1958, **23**(3), 187-200 [cit. 2021-03-02]. ISSN 0033-3123. DOI: 10.1007/BF02289233.

- [13] BASILEVSKY, Alexander. *Statistical Factor Analysis and Related Methods: Theory and Applications*. Hoboken, NJ, USA: John Wiley and Sons, Inc., 1994. Wiley Series in Probability and Statistics. ISBN 9780470316894.
- [14] HURVICH, Clifford and Chih-Ling TSAI. A corrected Akaike Information Criterion for vector autoregressive model selection. *Journal of Time Series Analysis* [online]. 2008, **14**, 271-279. DOI: 10.1111/j.1467-9892.1993.tb00144.x
- [15] CAVANAUGH, Joseph E. and Robert H. SHUMWAY. A bootstrap variant of AIC for state-space model selection. *Statistica Sinica*. 1997, **7**(2), 473-496. Available from: www.jstor.org/stable/24306089
- [16] BENGTTSSON, Thomas and Joseph E. CAVANAUGH. An improved Akaike information criterion for state-space model selection. *Computational Statistics and Data Analysis* [online]. 2006, **50**(10), 2635-2654. ISSN 0167-9473. DOI: 10.1016/j.csda.2005.05.003
- [17] MACQUEEN, J. Some Methods for Classification and Analysis of Multivariate Observations. In: *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*. 1967, **1**, 281-297. Available from: <https://sci2s.ugr.es/keel/pdf/algorithm/congreso/1967-MacQueen-MSP.pdf>
- [18] HARTIGAN, J. A. and M. A. WONG. Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. 1979, **28**(1), 100-108. DOI: 10.2307/2346830. Available from: <https://www.jstor.org/stable/2346830>
- [19] Okanagan Mountain Park fire. In: *City of Kelowna* [online]. British Columbia, 2021 [cit. 2021-04-13]. Available from: <https://www.kelowna.ca/city-services/safety-emergency-services/fire-department/okanagan-mountain-park-fire>
- [20] RUSH, David M. *Final report of the Kelowna and area Okanagan fire recovery society*. Kelowna, 2004. Available from: https://www.kelowna.ca/sites/files/1/docs/city-services/ok_mountain_park_fire_-_final_report.pdf
- [21] TAYLOR, Stephen W. and Martin E. ALEXANDER. Science, technology, and human factors in fire danger rating: the Canadian experience. *International Journal of Wildland Fire* [online]. 2006, **15**(1) [cit. 2021-03-24]. ISSN 1049-8001. DOI: 10.1071/WF05021
- [22] TAYLOR, Stephen W. and B.D. LAWSON. *The Canadian Forest Fire Weather Index (FWI) System: the foundation of fire danger rating in Canada*. Canadian Forest Service. Victoria, B.C. 2007.
- [23] VAN WAGNER, C. E. and T. L. PICKETT. *Equations and FORTRAN program for the Canadian Forest Fire Weather Index System*. 1985. Canadian Forestry Service, Petawawa National Forestry Institute, Chalk River, Ontario. Forestry Technical Report 33. 18 p. Available from: <https://cfs.nrcan.gc.ca/publications?id=19973>

REFERENCES

- [24] HOLMES, Elizabeth E., Mark D. SCHEUERELL and E. J. WARD. *Applied time series analysis for fisheries and environmental data*. NOAA Fisheries, Northwest Fisheries Science Center, 2725 Montlake Blvd E., Seattle, WA 98112. Available from: <https://nwfsc-timeseries.github.io/atsa-labs/sec-dfa.html>
- [25] HOLMES, Elizabeth E., Eric J. WARD and Mark SCHEUERELL. *Analysis of multivariate time-series using the MARSS package*. Version 3.11.3. Seattle, 2020. Available from: <https://cran.r-project.org/web/packages/MARSS/vignettes/UserGuide.pdf>
- [26] *Guide to the Ministry of Forests & Range* [CD-ROM]. Protection Program. Historic fire weather data.
- [27] SIMARD, A.J. *Fire weather zones of Canada*. 1973. Environment Canada Forestry Service, Ottawa, Ontario, Canada.
- [28] *BCWS Weather Station Map* [online]. British Columbia. Available from: <https://governmentofbc.maps.arcgis.com/apps/webappviewer/index.html?id=c36baf74b74a46978cf517579a9ee332>
- [29] HOLMES, Elizabeth E., Eric J. WARD and Kellie WILLS. MARSS: Multivariate Autoregressive State-space Models for Analyzing Time-series Data. *The R Journal* [online]. 2012, **4**(1), 11-19. DOI: 10.32614/RJ-2012-002
- [30] *RDocumentation* [online]. 2021, version 3.11.3. , [cit. 2021-04-20]. Available from: <https://www.rdocumentation.org/packages/MARSS/versions/3.11.3>
- [31] ROYSTON, P. Approximating the Shapiro-Wilk W-test for non-normality. *Statistics and Computing*. 1992, **2**, 117–119. DOI: 10.1007/BF01891203
- [32] BARNETT, Vic and Toby LEWIS. *Outliers in Statistical Data*. John Wiley & Sons Ltd, Great Britain: The Pitman Press, 1978. ISBN 0-471-99599-1.
- [33] *RDocumentation, MARSSkem* [online]. 2021, version 3.11.3. , [cit. 2021-04-25]. Available from: <https://www.rdocumentation.org/packages/MARSS/versions/3.11.3/topics/MARSSkem>
- [34] MOLENAAR, Peter C.M. A dynamic factor model for the analysis of multivariate time series. *Psychometrika*. 1985, **50**, 181–202. ISSN 0033-3123.
- [35] CARO, Ángela and Daniel PENA. *Estimation of the common component in Dynamic Factor Models*. Madrid: Universidad Carlos III de Madrid. Departamento de Estadística, 2018. DES - Working Papers. Statistics and Econometrics.
- [36] THORNDIKE, Robert L. Who belongs in the family? *Psychometrika* [online]. 1953, **18**(4), 267-276 [cit. 2021-4-28]. ISSN 0033-3123. DOI: 10.1007/BF02289263
- [37] CHARRAD, Malika, Nadia GHAZZALI, Véronique BOITEAU and Azam NIKNAFS. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software* [online]. 2014, **61**(6) [cit. 2021-04-10]. ISSN 1548-7660. DOI: 10.18637/jss.v061.i06
- [38] HUBERT, L. and P. ARABIE. Comparing Partitions. *Journal of Classification*, 1985, **2**, 193-218. DOI: doi.org/10.1007/BF01908075

Used symbols

L	lag operator
\mathbf{X}_t	random vector of measurements at time t
\mathbf{f}_t	vector of dynamic factors or realizations of latent factors at time t
$\lambda(L), \Psi(L)$	lag polynomial matrices
ϵ_t	idiosyncratic component
$\boldsymbol{\eta}_t$	vector of mean-zero error on factors
\mathbf{F}_t	vector of static factors
Λ, \mathbf{L}	loadings matrix
$N_m(\boldsymbol{\mu}, \Sigma)$	m -dimensional multivariate normal distribution with expected value $\boldsymbol{\mu}$ and variance-covariance matrix Σ
$(\mathbf{X}_t \mathbf{F}_t = \mathbf{f}_t)$	conditional distribution
$f(\cdot)$	joint density function
\mathbf{R}	variance-covariance matrix of modeled time series in dynamic factor model
\mathbf{Q}	variance-covariance matrix of latent factors
$\mathbf{F}_{t T}$	conditional expected value of latent factors in Kalman filter
$\mathbf{H}_{t T}$	conditional variance-covariance matrix of latent factors in Kalman filter
\mathbf{K}_t	Kalman gain
\mathbf{e}_t	innovations
$\log L$	log-likelihood function
Θ	vector of hyperparameters
\mathbf{T}	transformation matrix
\mathbf{I}	identity matrix
$\boldsymbol{\mu}$	level parameter of a dynamic factor model
\mathbf{L}^*	rotated loadings matrix after varimax rotation
AIC	Akaike information criterion
$W(C_k)$	within cluster sum of squares
$\boldsymbol{\mu}_k$	centroid of k th cluster

USED SYMBOLS

<i>FWI</i>	Fire Weather Index
<i>p</i>	number of stations, observed time series
<i>T</i>	number of time points
$LN(\mu, \sigma^2)$	lognormal distribution
<i>RI</i>	Rand Index
<i>ARI</i>	Adjusted Rand Index

Attachments

MARSS initialization in R

```

MARSS(data,
       model = list(  m,
                      R,
                      tinitx),
       method = "kem",
       form = "dfa",
       inits = list(  x0
                     Q
                     V0),
       control = list( maxit,
                       allow.degen,
                       abstol,
                       conv.test.slope.tol,
                       conv.test.deltaT),
       demean,
       zscore),

```

where the parameters are:

- *data* - $p \times T$ complete matrix of FWI index, p time series over $1, \dots, T$ time points.
- *m* - Number of latent factors to be estimated.
- *R* - Form of variance-covariance structure for time series ("diagonal and equal", "diagonal and unequal", "equal variance covariance", "unconstrained").
- *tinitx* - Whether to consider initial states \mathbf{F}_0 and \mathbf{H}_0 at time $t = 0$ or $t = 1$.
- *method* - "kem" performs estimation using EM algorithm and calls Kalman Filtering and Smoothing. Another option "BFGS" performs a quasi-Newton estimation [9].
- *form* - "dfa" is a helper function to put a state-space model into a simplified form of a dynamic factor model from section 2.3.3.
- *inits* - Set of initial values for the estimation.
- *x0* - $m \times 1$ vector of the mean $\boldsymbol{\pi}_0$. It is the initial condition $\mathbf{F}_{0|0} = E(\mathbf{F}_0|\mathbf{x}_0)$ from the Kalman filter 2.5.
- *Q* - Variance-covariance matrix of the latent factors. It is set to $m \times m$ identity matrix according to the restriction (2.85).

ATTACHMENTS

- $V0$ - Initial variance-covariance matrix $\mathbf{H}_{0|0} = E\{(\mathbf{F}_0 - \mathbf{F}_{0|0})(\mathbf{F}_0 - \mathbf{F}_{0|0})'\}$.
- *control* - Control parameters within the computation of the EM algorithm.
- *maxit* - If the conditions of the convergence for the parameters and the log-likelihood function are not met, how many EM iterations to do before the computation is ended.
- *allow.degen* - If the values of \mathbf{Q} or \mathbf{R} appear to be going to zero, algorithm tries to set the values to zero.
- *abstol* - Absolute tolerance convergence test for the log-likelihood function. If the difference in the absolute value between the last and the previous iteration of the log-likelihood function is less than *abstol* value, the log-likelihood function is considered to be maximized.
- *conv.test.slope.tol* - Convergence test for the parameters. If the slope of the log parameter versus log iteration is less than this value, the parameter passed log-log convergence test.
- *conv.test.deltaT* - How many of the last EM iterations to include in the log-log convergence test for the parameters.
- *demean* or *zscore* - If to center or standardize the input data.

m	2, 3, 4, 5
R	"unconstrained" "diagonal and unequal" "diagonal and equal" "equalvarcov"
<i>tinitx</i>	1
<i>method</i>	"kem"
<i>form</i>	"dfa"
$x0$	$\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}_{m \times 1}$
Q	$I_{m \times m}$
$V0$	$\begin{bmatrix} 5 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 5 \end{bmatrix}_{m \times m}$
<i>maxit</i>	5000 – 10000
<i>allow.degen</i>	TRUE
<i>abstol</i>	0.0001
<i>conv.test.slope.tol</i>	0.5
<i>conv.test.deltaT</i>	9
<i>demean</i>	TRUE
<i>zscore</i>	FALSE
μ	0

Table 6.1: Setting of the MARSS() function in R used for our data analysis.

Log-log slope test

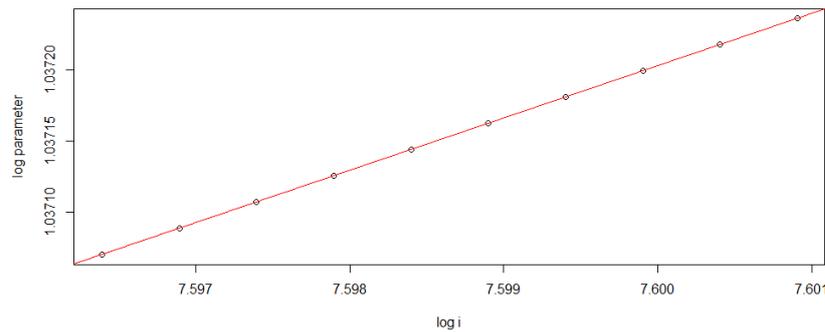
Let us have a regression line

$$\log Y = \beta_0 + \beta_1 \log x, \quad (6.1)$$

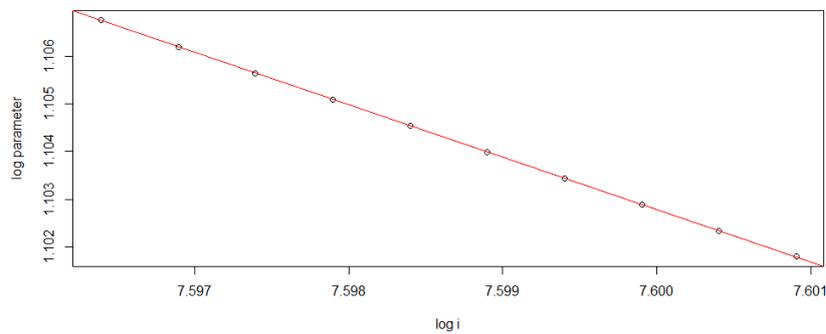
where Y is the parameter estimate and x is the iteration number. Parameter estimates are based on last 10 values. Then if

$$|\hat{\beta}_1| < 0.5, \quad (6.2)$$

we consider parameter estimate to be converged. In Figure 6.1 we may see both cases.



(a)



(b)

Figure 6.1: The regression lines of the log-log convergence test. We plot last 10 log of the parameter estimations against log of the last 10 iterations (1991 - 2000). (a) parameter converged, the slope is $\hat{\beta}_1 = 0.02609$ (b) parameter did not converge, the slope is $\hat{\beta}_1 = -1.098$.

Consequence of restriction on latent factors

As mentioned in section 2.9, after the model fit is computed we have to do a rotation of the factor loadings and the latent factors. The reason for the rotation is, that with the restrictions from section 2.8 placed on the model, first latent factor is determined only by the data of the first time series, second latent factor is determined by first two observed time series and so on.

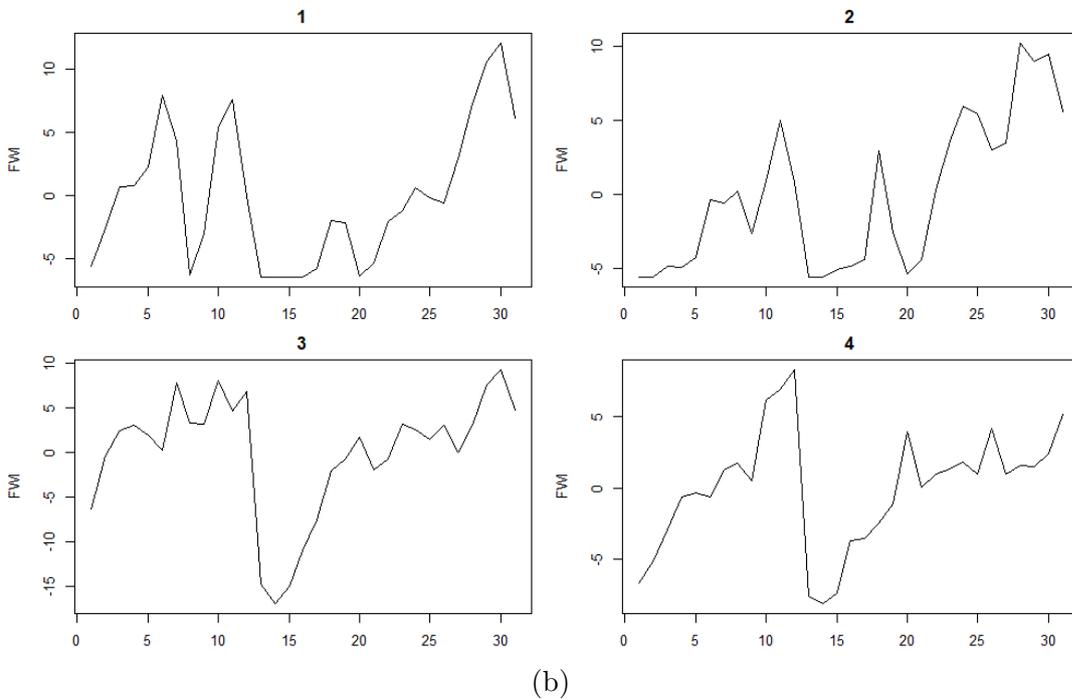
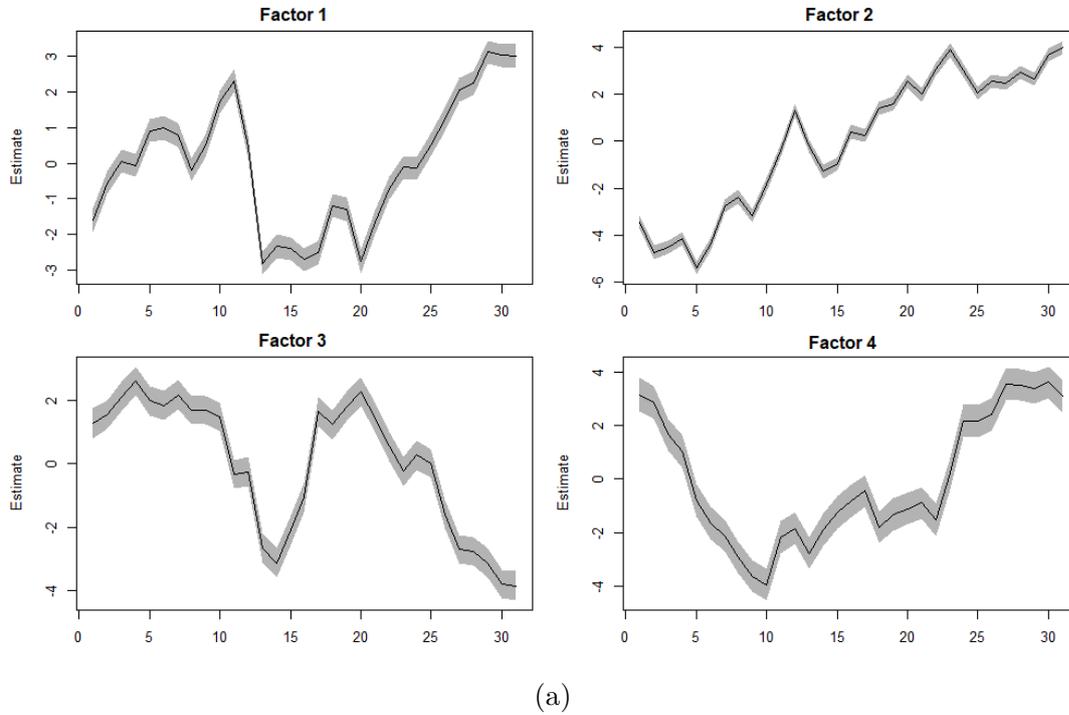


Figure 6.2: (a) Plot of the estimated latent factors of the best model fit. (b) Plot of the first four (1-4) centered observed time series in our dataset.

Results for August data

R	m	logLik	K	AICc	time	convergence	maxit	N_ts
diagonal and equal	3	-11299.81	322	23312.38	23524.79	10	10000	108
diagonal and equal	4	-11091.93	427	23163.03	39246.37	10	8000	108
diagonal and equal	5	-10881.43	531	23025.48	38493.13	10	4000	108
diagonal and unequal	3	-10972.53	429	22929.49	25416.25	10	10000	108
diagonal and unequal	4	-10773.45	534	22818.01	32062.69	10	6000	108
diagonal and unequal	5	-10584.03	638	22745.04	30286.86	1	3000	108
equalvarcov	3	-11285.69	323	23286.59	20382.59	10	10000	108
equalvarcov	4	-11075.86	428	23133.52	32618.75	10	7000	108
equalvarcov	5	-10860.09	532	22985.64	46230.16	10	5000	108

Table 6.2: Resulted models for August data

R	m	logLik	K	Δ AICc	convergence	maxit
diagonal and unequal	5	-10584.03	638	0	1	3000
diagonal and unequal	4	-10773.45	534	72.97331	10	6000
diagonal and unequal	3	-10972.53	429	184.45048	10	10000
equalvarcov	5	-10860.09	532	240.59719	10	5000
diagonal and equal	5	-10881.43	531	280.4455	10	4000
equalvarcov	4	-11075.86	428	388.48471	10	7000
diagonal and equal	4	-11091.93	427	417.99195	10	8000
equalvarcov	3	-11285.69	323	541.5561	10	10000
diagonal and equal	3	-11299.81	322	567.34406	10	10000

Table 6.3: Order of the model fits from the best to the worst with respect to the difference of AICc on August data.

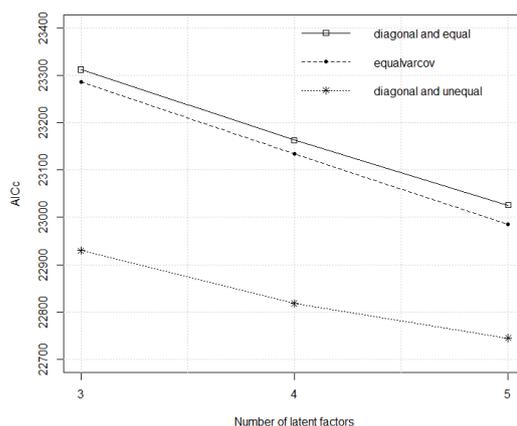


Figure 6.3: Comparison of the models based on the resulted value of AICc. Smaller AICc means a better fit to the data (compared to other models).

ATTACHMENTS

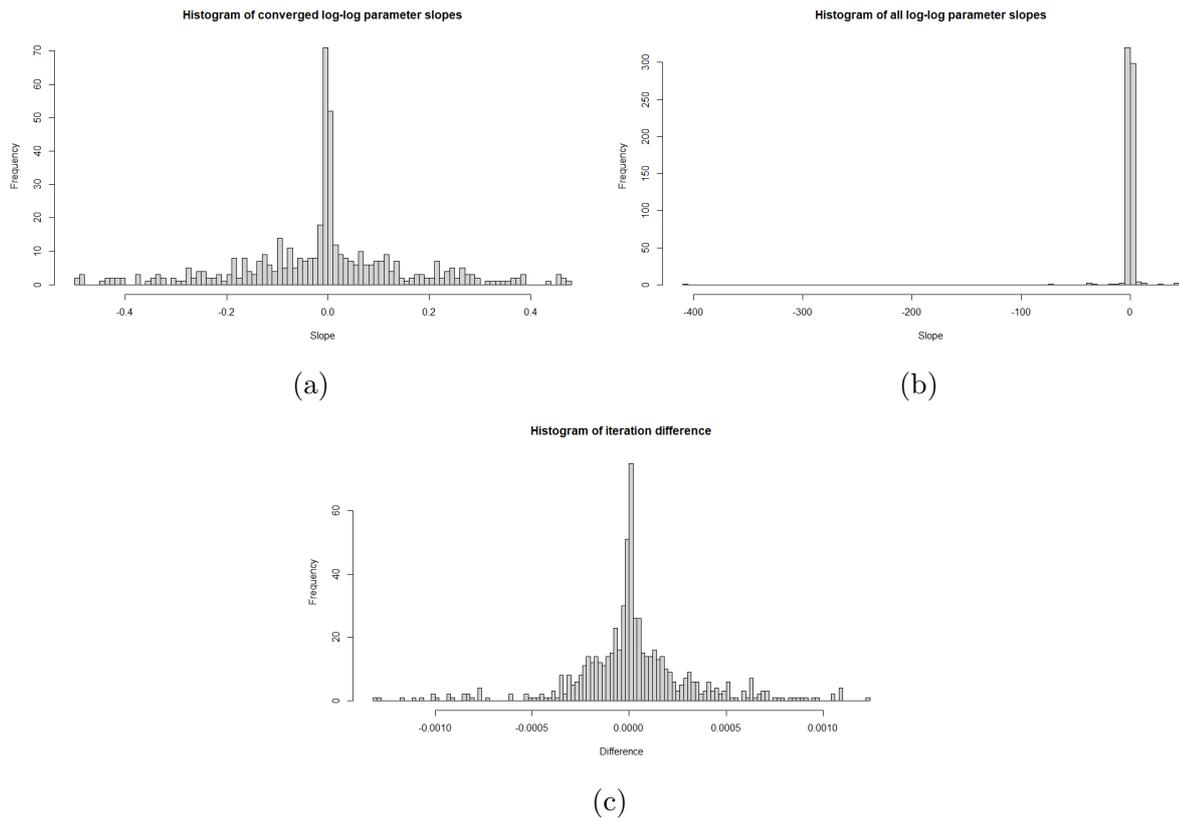


Figure 6.4: (a) Histogram of slopes of last 10 iterations for each parameter that passed log-log slope test. (b) Histogram of slopes with parameters that did not pass the log-log slope test. (c) Histogram of the parameter differences between the 2999th and the 3000th iteration.

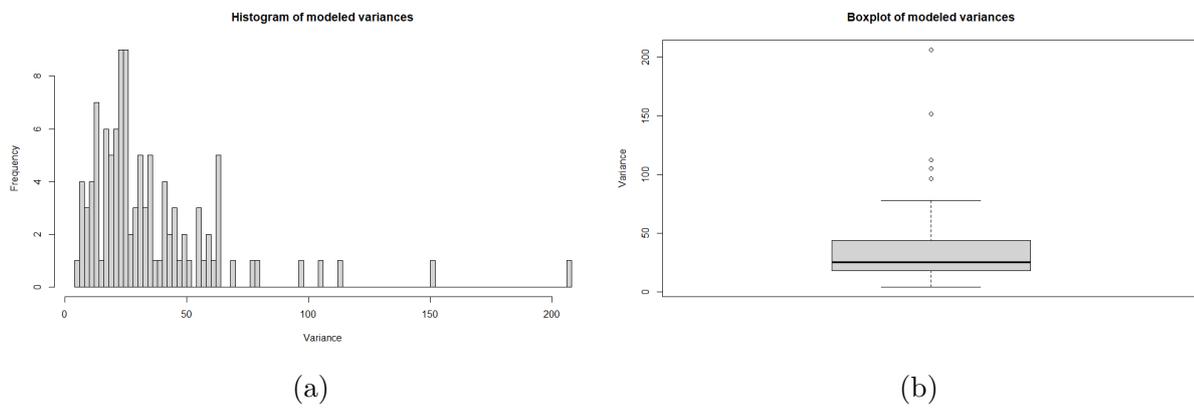
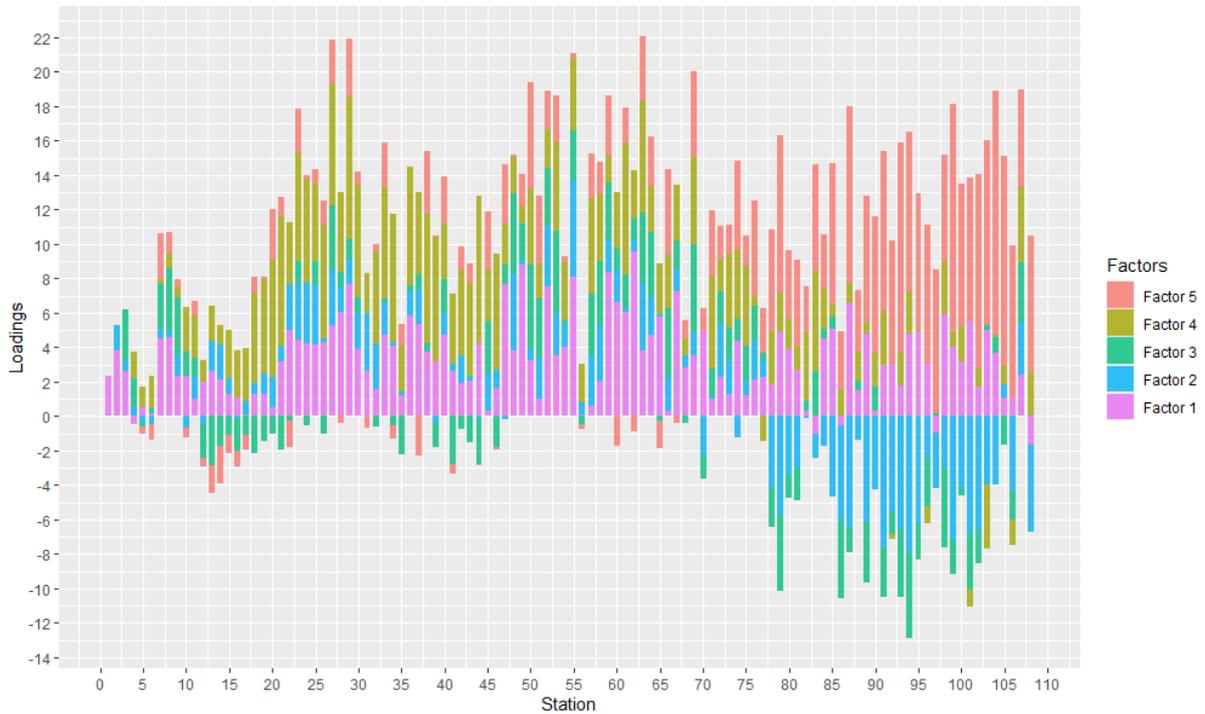
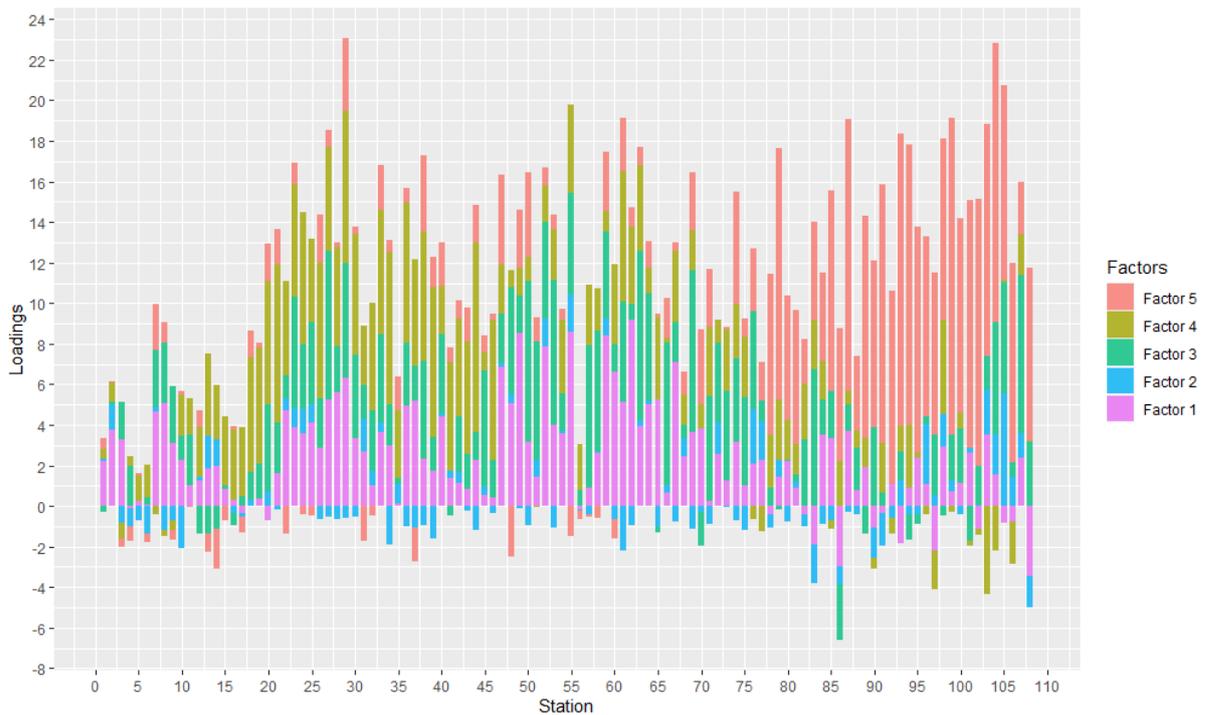


Figure 6.5: (a) Histogram of the estimated variances. (b) Boxplot of the estimated variances.

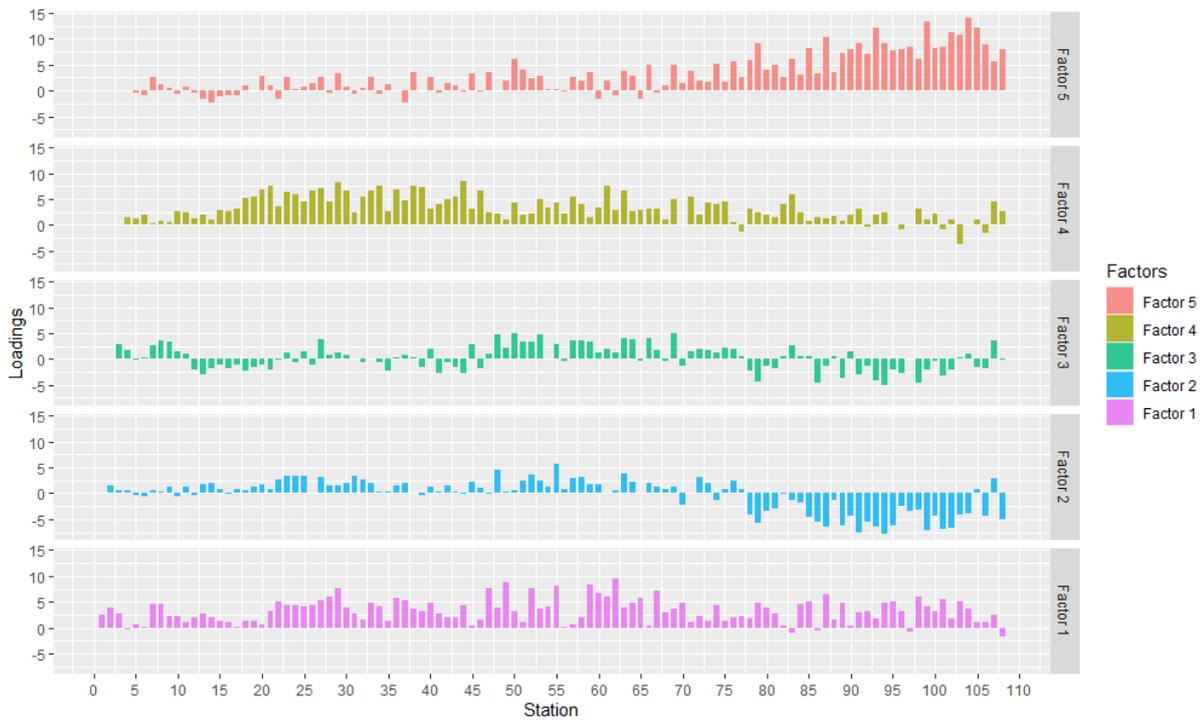


(a) before varimax rotation

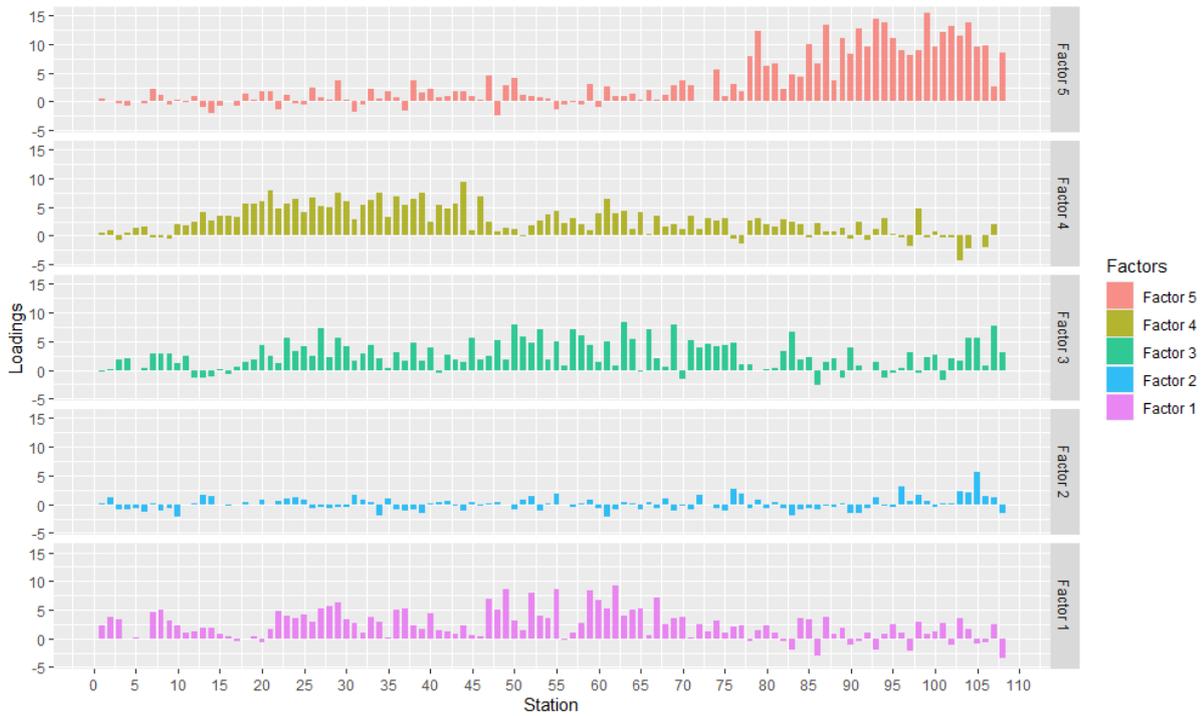


(b) after varimax rotation

Figure 6.6: Stacked histogram of loadings values of latent factors for each modeled time series (in other words estimated l_{ij} values). A column in the graph is the sum of the factor loadings for a particular time series. To obtain a loading of the particular factor on the particular time series, we have to search for that series on the x-axis, then find the proper colour representing the factor we want to find loading for and then subtract the lower bound from the higher bound of that particular color in the column.



(a) before varimax rotation

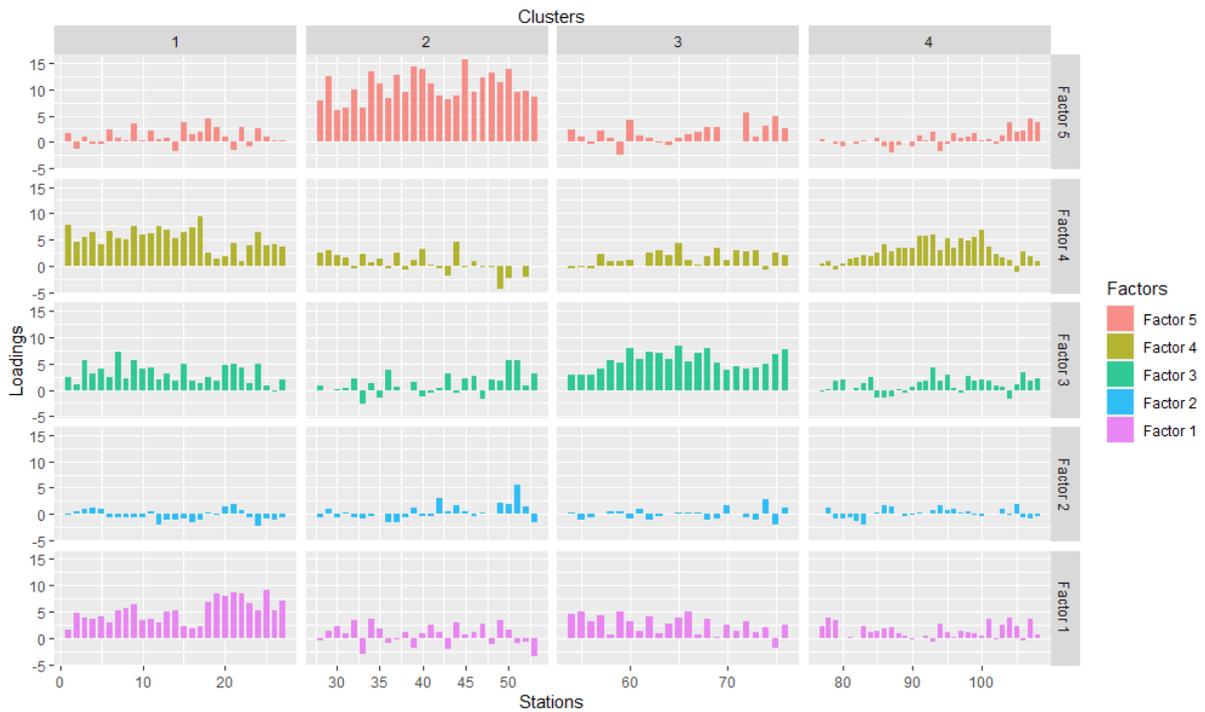


(b) after varimax rotation

Figure 6.7: Histogram of the loadings separated by factors.



(a) before varimax rotation



(b) after varimax rotation

Figure 6.8: Factor loadings of the stations divided into clusters based on k-means clustering.