



Universitetet
i Stavanger

FACULTY OF SCIENCE AND TECHNOLOGY

MASTER'S THESIS

Study programme/specialisation: Mathematics and Physics	Spring / Autumn semester, 2018/2019 Open/ Confidential
Author: Anders Holen	<i>Anders Holen</i> (signature of author)
Programme coordinator: Bjørn Henrik Auestad Supervisor(s): Tore Selland Kleppe	
Title of master's thesis: Estimating and Forecasting of Dynamic linear Gaussian State Space Models for Commodity Futures	
Credits: 60	
Keywords: Linear Gaussian State Space Models Nelson-Siegel Kalman Filter Forecasting	Number of pages:101..... + supplemental material/other:0..... Stavanger, 14.06.2019 date/year

Estimating and Forecasting of Dynamic linear
Gaussian State Space Models for Commodity
Futures

Anders Holen

Studentnr. 229678

Institutt for matematikk og naturvitenskap

Supervisor: Tore Selland Kleppe

14.06.2019

Abstract

The Kalman filter is used to estimate the parameters and forecast the observations in a dynamic Nelson-Siegel model a linear Gaussian state space representation for futures contracts on the commodities oil, natural gas, cotton, copper, gold and sugar. The three-factor Nelson-Siegel model is compared to three-factor Nelson-Siegel model with seasonality terms to check for seasonality in the different commodities. Using Wilks' Theorem we find that natural gas, cotton and sugar has an improved model fit by adding the seasonality term. The Kalman filter is shown to be a great model fit for most commodities except for natural gas and cotton, there needs to be further studies to find out why the parameter estimates for these two commodities are not as expected. For the forecasting of the observations, the Kalman filter performs very well with both three-factor model Nelson-Siegel and the three-factor Nelson-Siegel with a seasonality term. It was not possible to forecast the observations for sugar for the three-factor Nelson-Siegel model because the variance matrix of the prediction error is singular. Thus it does not have an inverse which is crucial for the Kalman filter. This should also be studied further to figure out why this happens for the sugar data.

Acknowledgment

I would like to extend my sincere gratitude to my supervisor, professor Tore Selland Kleppe, for the encouragement and advice he's given me in my years as a graduate student at the University of Stavanger. He has been a tremendous help throughout the process of writing this thesis, a great sparring partner whenever I encountered problems or obstacles in my analyses.

Contents

1	Introduction	1
2	Latent variable models	3
2.1	Dynamic factor models	3
3	Linear Gaussian state space models	6
3.1	Introduction	6
3.2	Local level model	7
3.3	Time series models	8
3.3.1	Trend Component	9
3.3.2	Seasonal Component	9
4	Nelson Siegel model	12
4.1	Three-factor model	12
5	Kalman filter	14
5.1	Introduction	14
5.2	Derivation of the Kalman filter	14
5.3	Missing observations	21
6	Forecasting	23
7	Maximum likelihood estimation	25
7.1	Parameter estimation	25
7.2	Goodness of fit	26
8	Delta method	27
9	Results	29
9.1	Crude oil	29
9.1.1	Three-factor model	29
9.1.2	Seasonal model	33
9.1.3	Hypothesis testing	38

9.2	Natural gas	40
9.2.1	Three-factor model	40
9.2.2	Seasonal model	44
9.2.3	Hypothesis testing	49
9.3	Copper	50
9.3.1	Three-factor model	50
9.3.2	Seasonal model	54
9.3.3	Hypothesis testing	59
9.4	Cotton	60
9.4.1	Three-factor model	60
9.4.2	Seasonal model	64
9.4.3	Hypothesis testing	68
9.5	Gold	70
9.5.1	Three-factor model	70
9.5.2	Seasonal model	72
9.5.3	Hypothesis testing	77
9.6	Sugar	79
9.6.1	Three-factor model	79
9.6.2	Seasonal model	82
9.6.3	Hypothesis testing	86
10	Conclusion	89

List of Figures

9.1	Factor loadings for three-factor model on natural gas data	30
9.2	Plot of β_{1t} from time $t = 0, \dots, 4865$	30
9.3	Plot of β_{2t} from time $t = 0, \dots, 4865$	31
9.4	Plot of β_{3t} from time $t = 0, \dots, 4865$	31
9.5	Forecasted observations for three-factor model and actual obser- vations of crude oil data	32
9.6	Factor loadings of seasonal model on crude oil data	33
9.7	Plot of β_{1t} from time $t = 0, \dots, 4865$	34
9.8	Plot of β_{2t} from time $t = 0, \dots, 4865$	34
9.9	Plot of β_{3t} from time $t = 0, \dots, 4865$	35
9.10	Seasonality on β_t with period $s = 253$	36
9.11	Forecasted observations for seasonal model and actual observa- tions of crude oil data	38
9.12	Factor loadings of three-factor model on natural gas data	41
9.13	Plot of β_{1t} from time $t = 0, \dots, 6059$	41
9.14	Plot of β_{2t} from time $t = 0, \dots, 6059$	42
9.15	Plot of β_{3t} from time $t = 0, \dots, 6059$	42
9.16	Plot of first column in Natural Gas dataset	43
9.17	Forecasted observations for three-factor model and actual obser- vations of natural gas data.	44
9.18	Factor loadings of the seasonal model on natural gas data	45
9.19	Plot of β_{1t} from time $t = 0, \dots, 6059$	46
9.20	Plot of β_{2t} from time $t = 0, \dots, 6059$	46
9.21	Plot of β_{3t} from time $t = 0, \dots, 6059$	47
9.22	Seasonality on β_t with period $s = 253$	48
9.23	Forecasted observations for seasonal model and actual observa- tions of natural gas data.	49
9.24	Factor loadings of three-factor model on the copper data	51
9.25	Plot of β_{1t} from time $t = 0, \dots, 8562$	51

9.27	Plot of β_{3t} from time $t = 0, \dots, 8562$	52
9.26	Plot of β_{2t} from time $t = 0, \dots, 8562$	52
9.28	Forecasted observations of three-factor model and actual observations of copper data	53
9.29	Factor loadings of seasonal model on the copper data	54
9.30	Plot of β_{1t} from time $t = 0, \dots, 8562$	55
9.31	Plot of β_{2t} from time $t = 0, \dots, 8562$	55
9.32	Plot of β_{3t} from time $t = 0, \dots, 8562$	56
9.33	Seasonality on β_t for period $s = 253$	57
9.34	Forecasted observations of seasonal model and actual observations of the copper data	58
9.35	Factor loadings of three-factor model on cotton data	60
9.36	Plot of β_{1t} from time $t = 0, \dots, 10380$	61
9.37	Plot of β_{2t} from time $t = 0, \dots, 10380$	61
9.38	Plot of β_{3t} from time $t = 0, \dots, 10380$	62
9.39	Plot of first column in Cotton dataset	63
9.40	Forecasted observations for three-factor model and actual observations of cotton data.	63
9.41	Factor loadings of the seasonal model on cotton data	64
9.42	Plot of β_{1t} from time $t = 0, \dots, 10380$	65
9.43	Plot of β_{2t} from time $t = 0, \dots, 10380$	65
9.44	Plot of β_{3t} from time $t = 0, \dots, 10380$	66
9.45	Seasonality on β_t with period $s = 253$	67
9.46	Forecasted observations for seasonal model and actual observations of cotton data.	68
9.47	Factor loadings of three-factor model on the gold data	70
9.48	Plot of β_{1t} from time $t = 0, \dots, 9205$	71
9.49	Plot of β_{2t} from time $t = 0, \dots, 9205$	71
9.50	Plot of β_{3t} from time $t = 0, \dots, 9205$	72
9.51	Forecasted observations of three-factor model and actual observations of gold data	73

9.52	Factor loadings of seasonal model on the gold data	74
9.53	Plot of β_{1t} from time $t = 0, \dots, 9205$	74
9.54	Plot of β_{2t} from time $t = 0, \dots, 9205$	75
9.55	Plot of β_{3t} from time $t = 0, \dots, 9205$	75
9.56	Seasonality on β_t for period $s = 253$	76
9.57	Forecasted observations of seasonal model and actual observations of the gold data	78
9.58	Factor loadings of three-factor model on the sugar data	80
9.59	Plot of β_{1t} from time $t = 0, \dots, 7407$	80
9.61	Plot of β_{3t} from time $t = 0, \dots, 7407$	81
9.60	Plot of β_{2t} from time $t = 0, \dots, 7407$	81
9.62	Factor loadings of seasonal model on the sugar data	82
9.63	Plot of β_{1t} from time $t = 0, \dots, 7407$	83
9.64	Plot of β_{2t} from time $t = 0, \dots, 7407$	83
9.65	Plot of β_{3t} from time $t = 0, \dots, 7407$	84
9.66	Seasonality on β_t for period $s = 253$	85
9.67	Forecasted observations of seasonal model and actual observations of the sugar data	87

1 Introduction

In this thesis I will look at estimating linear Gaussian state space models and forecasting prices of futures contracts on commodities using the Kalman filter. It is very interesting to look at predicting such prices and the results of these predictions are very valuable, especially for risk managers and producers within the different markets where prices are predicted. The data that I will work with in this thesis are future contracts for crude oil, natural gas, copper, cotton, gold and sugar, they are large data sets that consist of prices and maturities of future contracts on the different commodities. These are natural commodities to look at as they are some of the most traded commodities in the world. For example, the prices of crude oil are also an indicator of how the economy in the world is doing. For conveniency regarding normal approximations I have worked with log-prices instead of prices in my analyses.

To get an understanding of how and why the different analyses are done I will introduce the reader to some important theory to be able to understand the analyses in this thesis. The linear Gaussian state space model builds upon the latent variable model, hence to get a good understanding of what a linear Gaussian state space model is we need to start with establishing what a latent variable model is. In this thesis I am focusing on the Nelson-Siegel model, which is a dynamic factor model. To understand the Nelson-Siegel model it is therefore important to have a look at some of the foundation of dynamic factor models. Further on I will then look at what a linear Gaussian state space model is before I present the Nelson-Siegel model and how it looks in linear Gaussian state space form. The next step, after we have an understanding of what a linear Gaussian state space model with the Nelson-Siegel model as the observation equation is, we can go on and look at the method for estimating the parameters in the linear Gaussian state space model, namely the Kalman filter. To get an understanding of how the filter works I will derive the equations in the filter with regards to a linear Gaussian state space model. Further I will look at forecasting the observed data with the use of the Kalman filter. I will also

briefly look at maximum likelihood estimation for parameter estimation and goodness of fit of a model. And lastly I will present the Delta method, which will be used to present a confidence interval of the most important parameters in the data treatment.

All analyses have been done using the R programming language, there are several packages that utilizes a Kalman filter, but none of them handles the seasonal model used in this thesis to a satisfying degree. As such I have written my own Kalman filter in R and have used the package KFAS as a reference point for the three-factor model. The results presented in this thesis will solely be from my own Kalman filter code.

In this thesis I am following the notation used in James Durbin (2012), this book has also been the main source of information for the thesis.

2 Latent variable models

We have our data of futures contracts on commodities, the prices are different at different times. What is the cause of this? Can we observe some data that can explain why the prices are volatile? If not, how can we explain the volatility in our dataset? This is where latent variable models comes in to the picture.

Firstly, what is a latent variable? A variable that you cannot observe or measure directly and which is presumed to have an effect on a directly observed variable (manifest variable), is called a latent variable. Since one cannot measure a latent variable they need to be described by other variables that can be observed directly, for example quality-of-life is a latent variable in economics. One can't observe quality-of-life directly, it needs to be described by other manifest variables such as wealth, education, employment, physical and mental health etc. A latent variable is typically used in a latent variable model. A latent variable model is a statistical model which is used to describe the relations between a set of manifest variables and a set of latent variables. These types of models are used with a variety of goals in mind. Latent variable models can be used to explain the effect some unobserved covariates have on the observed variable. Another use can be to assume the latent variable is the "true" value of the observed variable and the directly observed variable is a "disturbed" version, thus one can account for measurement errors. There are a lot of distinct ways that a latent variable model can be applied to problems. A lot of well known models are latent variable models, such as General linear mixed models, Factor analysis models, Models for longitudinal/panel data based on a state-space formulation etc. An basic assumption that is made for latent variable models is local independence. This means that the observable items are conditionally independent of each other given an individual score on the latent variables.

2.1 Dynamic factor models

Factor models are a special form of latent-variable model. A factor model is a model where the observed vector is divided into two unobserved parts, which are

a *systematic* part and an *error* part. The different components in the error part is considered to be independent of each other. The systematic part is a linear combination of different factor variables. In a factor model both the manifest variables and the latent variables are continuous, and their relationship is linear. The basic form of a factor model is presented in equation 2.1:

$$y_i = \Lambda f_i + u_i \quad (2.1)$$

where the latent variable f_i , is a vector of m components, the components of f_i are sometimes referred to as *common factors*. The observation vector, y_i , is of dimension p . The vector y_i represents here a zero-mean vector that consists of observed or measured traits of subject number i , for $i = 1, \dots, n$. Λ is a $p \times m$ matrix of factor loadings and u_i is a noise vector of p components, with mean equal to 0 and a diagonal variance matrix Σ_u . u_i and f_i is assumed to be mutually and serially independent. We also assume that the components of u_i are uncorrelated. The variance of y_i , Σ_y is presented in 2.2:

$$Var(y_i) = \Sigma_y = \Lambda \Sigma_f \Lambda' + \Sigma_u \quad (2.2)$$

If one is to use a factor analysis in a time-series case, the observations in the vector y_i will then correspond to a time period t instead of a subject i . Thus we have y_t and not y_i . Now, the observations gets a time dependence to handle. This can be solved by substituting the assumption of serially independence for f_t with an assumption of serial dependence. One can, for example, make the assumption that f_t is modelled as a vector autoregressive process (VAR). There is also the case where one can let f_t depend linearly on *state vector* α_t (which we will look more at in the next section), that would be on the form:

$$f_t = U_t \alpha_t,$$

where U_t usually is a known matrix.

In this thesis we are looking at state space models and we can treat the dynamic factor model as one. The components of the state vector are the latent

factors in the model. Usually, the size, p , of the observation vector, y_t , is much larger than the size of the state vector, m , which is usually small, thus we have $p \gg m$. The dynamic factor model described in 2.3

$$y_t = \Lambda f_t + u_t \tag{2.3}$$

is a special case of a linear Gaussian state space model (which we will describe in detail in the next section), with $f_t = U_t \alpha_t$, $u_t = \epsilon_t$ and $\epsilon_t \sim N(0, H_t)$. Thus the factor model that will be described in the linear Gaussian state space model, 3.1, in the next section is:

$$y_t = Z_t \alpha_t + \epsilon_t \tag{2.4}$$

where $Z_t = \Lambda U_t$.

3 Linear Gaussian state space models

3.1 Introduction

In this section I will look at linear Gaussian state space models, how they are built up and how they can be used in time series analysis. A model where *state variables* are used to describe a system of first-order differential or difference equations is called a state space model. A state space model is built up of two equations, one which describes how a latent process changes over time, the other describes how the latent process is measured at each time-step. State variables are not themselves being measured or observed during data collection, but they can be recreated from the observed data. The general linear Gaussian state space model can be written in many different ways, in 3.1 the form that is used in this thesis is presented:

$$\begin{aligned}y_t &= Z_t \alpha_t + \epsilon_t, & \epsilon_t &\sim N(0, H_t) \\ \alpha_{t+1} &= T_t \alpha_t + R_t \eta_t, & \eta_t &\sim N(0, Q_t)\end{aligned}\tag{3.1}$$

where y_t is a $p \times 1$ vector of observations called the *observation vector*, and α_t is an unobserved $m \times 1$ vector called the state vector. How the system progresses over time is determined by α_t , specifically by the second equation in (3.1). Eventhough the system progression is determined by α_t , we can't observe α_t directly so the analysis must therefore be based on the observations y_t . The first and second equation in (3.1) are called the *observation equation* and the *state equation* respectively. Z_t , T_t , R_t , H_t and Q_t are assumed to be known matrices and the error terms ϵ_t and η_t are assumed to be independent of each other at all points in time and serially independent. The matrices Z_t and T_{t-1} can be allowed to be dependent on y_1, \dots, y_{t-1} . It is assumed that the initial state vector α_1 is $N(a_1, P_1)$ independently of $\epsilon_1, \dots, \epsilon_n$ and η_1, \dots, η_n , where a_1 and P_1 are assumed to be known. Some or all of the matrices Z_t , T_t , R_t , H_t and Q_t will in actuality usually depend on elements of a parameter vector ψ , which is unknown. This vector will be estimated using maximum likelihood estimation.

The observation equation in (3.1), has the structure of a linear regression model where α_t , the coefficient vector, differs over time. The state equation in (3.1) is representing a first order vector autoregressive (VAR) model. It's the Markovian nature of this model which accounts for many of the elegant properties of the state space model. The matrix R_t is the identity in many applications, in others it is possible to define $\eta_t^* = R_t\eta_t$ and $Q^* = R_tQ_tR_t'$, then one can continue without explicit inclusion of R_t and make the model look simpler. If R_t is $m \times r$ with $r < m$ and Q_t is nonsingular, then there is an advantage in working with nonsingular η_t instead of singular η_t^* . If we assume that R_t is a subset of the columns of I_m ; in that case R_t is called a selection matrix because it selects the rows of the state equation which have nonzero disturbance terms; however, much of the theory remains valid if R_t is a general $m \times r$ matrix. To keep it simple I will summarize the dimensions of the different parts in 3.1.

Vector		Matrix	
y_t	$p \times 1$	Z_t	$p \times m$
α_t	$m \times 1$	T_t	$m \times m$
ϵ_t	$p \times 1$	H_t	$p \times p$
η_t	$r \times 1$	R_t	$m \times r$
		Q_t	$r \times r$
a_1	$m \times 1$	P_1	$m \times m$

Table 1: Dimensions of state space model 3.1

3.2 Local level model

The local level model is a special case of the general linear Gaussian state space model. The local level model has the form:

$$\begin{aligned} y_t &= \alpha_t + \epsilon_t, & \epsilon_t &\sim N(0, \sigma_\epsilon^2) \\ \alpha_{t+1} &= \alpha_t + \eta_t, & \eta_t &\sim N(0, \sigma_\eta^2) \end{aligned} \tag{3.2}$$

for $t = 1, \dots, n$, where the ϵ_t 's and η_t 's are mutually independent and independent of α_1 .

The first line in 3.1 and 3.2 are the observation equation. An observation

equation is an equation that expresses measured values of some function of one (or more) unknown quantities. The unknown quantity that we are looking for here are α_t which is a random walk. The local level model may have a simple form, but it is not a special case made just for showing the principles of state space analysis. It provides the basis for the analysis of important real problems in practical time series analysis. The model shows the characteristic structure of a state space model in which there is a series of unobserved values $\alpha_1, \dots, \alpha_n$ called *states*.

3.3 Time series models

A *time series* is a set of observations y_1, \dots, y_n ordered in time. The basic model for representing a time series is the additive model:

$$y_t = \mu_t + \gamma_t + \epsilon_t, \quad t = 1, \dots, n \quad (3.3)$$

In this model μ_t is a slowly varying component called the *trend*, γ_t is a periodic component of fixed period called the *seasonal* and ϵ_t is an irregular component called the *error* or *disturbance*. In many applications, particularly in economics, the components combine multiplicatively, this gives the model:

$$y_t = \mu_t * \gamma_t * \epsilon_t \quad (3.4)$$

This model can be reduced to model (3.3) by working with logged values of each component.

A structural time series model is a model where the trend, seasonal and error terms in model (3.3), and other relevant terms, are modelled explicitly. In this section we will look at the cases where y_t is univariate. Later we will look at cases where y_t is multivariate.

3.3.1 Trend Component

Model (3.2) is a simple form of a structural time series model. If we add a slope term ν_t , which is generated by a random walk we have this model:

$$\begin{aligned} y_t &= \mu_t + \epsilon_t & \epsilon_t &\sim N(0, \sigma_\epsilon^2) \\ \mu_{t+1} &= \mu_t + \nu_t + \xi_t & \xi_t &\sim N(0, \sigma_\xi^2) \\ \nu_{t+1} &= \nu_t + \zeta_t & \zeta_t &\sim N(0, \sigma_\zeta^2) \end{aligned} \quad (3.5)$$

This model is called the *local linear trend* model. If $\xi_t = \zeta_t = 0$ then $\nu_{t+1} = \nu_t = \nu$, and then we have $\mu_{t+1} = \mu_t + \nu$. So the trend is exactly linear and model (3.5) reduces to the deterministic linear trend plus noise model. The form of model (3.5) with $\sigma_\xi^2 > 0$ and $\sigma_\zeta^2 > 0$ makes it so that the trend level and slope varies over time. An objection that is made by looking at the series of values of μ_t that is obtained by fitting model 3.5 is that the values doesn't look smooth enough to represent the idea of a trend. To meet this objection we can set $\sigma_\xi^2 = 0$ at the outset and then fitting the model under this restriction. The same effect can essentially be obtained if, in place of the second and third equation of 3.5, we use the model $\Delta^2 \mu_{t+1} = \zeta_t$, that is $\mu_{t+1} = 2\mu_t - \mu_{t-1} + \zeta_t$ where Δ is the first difference operator that is defined by $\Delta x_t = x_t - x_{t-1}$. Model 3.5 can be written in the form:

$$\begin{aligned} y_t &= \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} \mu_t \\ \nu_t \end{pmatrix} + \epsilon_t \\ \begin{pmatrix} \mu_{t+1} \\ \nu_{t+1} \end{pmatrix} &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} \mu_t \\ \nu_t \end{pmatrix} + \begin{pmatrix} \xi_t \\ \zeta_t \end{pmatrix} \end{aligned} \quad (3.6)$$

this is a special case of model 3.1.

3.3.2 Seasonal Component

The seasonal component of a time series is the part of the variations in the series that represents fluctuations occurring at specific intervals, examples of these intervals may be weekly, monthly or quarterly. The seasonal component is also known as the seasonality of a time series. If we are to model the seasonality

in 3.3, suppose we have s 'months' per 'year'. Then we have $s = 12$ for monthly data, for quarterly data we have $s = 4$ and, when modelling the weekly pattern we have $s = 7$ for daily data. We can model the seasonal values for months 1 to s by the constants $\gamma_1^*, \dots, \gamma_s^*$, where $\sum_{j=1}^s \gamma_j^* = 0$, if the seasonal pattern is constant over time. For the j th 'month' in 'year' i we have $\gamma_t = \gamma_j^*$ where $t = s(i-1) + j$ for $i = 1, 2, \dots$ and $j = 1, \dots, s$. It then follows that $\sum_{j=0}^{s-1} \gamma_{t+1-j} = 0$ so $\gamma_{t+1} = -\sum_{j=1}^{s-1} \gamma_{t+1-j}$ with $t = s-1, s, \dots$. In practice we often don't want the seasonal pattern to be constant, we want to allow it to change over time. One way of getting the seasonal pattern to change over time is to add an error term, ω_t , to this relation. Then we get the model:

$$\gamma_{t+1} = -\sum_{j=1}^{s-1} \gamma_{t+1-j} + \omega_t, \quad \omega_t \sim N(0, \sigma_\omega^2) \quad (3.7)$$

for $t = 1, \dots, n$. Harrison and Stevens (1976) suggested an alternative way of achieving the seasonality to change over time. They suggested to denote the effect of season j at time t by γ_{jt} and then let γ_{jt} be generated by the quasi-random walk

$$\gamma_{j,t+1} = \gamma_{jt} + \omega_{jt}, \quad t = (i-1)s + j, \quad i = 1, 2, \dots, \quad j = 1, \dots, s, \quad (3.8)$$

with an adjustment to make sure that every successive set of s seasonal components sums to zero. Often, it is preferable to express the seasonality in a trigonometric form, for a constant seasonality one such form is:

$$\gamma_t = \sum_{j=1}^{\lfloor s/2 \rfloor} (\tilde{\gamma}_j \cos \lambda_j t + \tilde{\gamma}_j^* \sin \lambda_j t), \quad \lambda_j = \frac{2\pi j}{s}, \quad j = 1, \dots, \lfloor s/2 \rfloor \quad (3.9)$$

where $\lfloor a \rfloor$ is the largest integer $\leq a$ and where the quantities $\tilde{\gamma}_j$ and $\tilde{\gamma}_j^*$ are given constants. For a seasonality that changes over time we can make this stochastic by replacing $\tilde{\gamma}_j$ and $\tilde{\gamma}_j^*$ by these random walks:

$$\tilde{\gamma}_{j,t+1} = \tilde{\gamma}_{jt} + \tilde{\omega}_{jt}, \quad \tilde{\gamma}_{j,t+1}^* = \tilde{\gamma}_{jt}^* + \tilde{\omega}_{jt}^*, \quad (3.10)$$

where $\tilde{\omega}_{jt}$ and $\tilde{\omega}_{jt}^*$ are independent $N(0, \sigma_\omega^2)$ variables, with $j = 1, \dots, [s/2]$ and $t = 1, \dots, n$. An alternative trigonometric form is the quasi-random walk model:

$$\gamma_t = \sum_{j=1}^{[s/2]} \gamma_{jt}, \quad (3.11)$$

where

$$\begin{aligned} \gamma_{j,t+1} &= \gamma_{jt} \cos \lambda_j + \gamma_{jt}^* \sin \lambda_j + \omega_{jt}, \\ \gamma_{j,t+1}^* &= -\gamma_{jt} \sin \lambda_j + \gamma_{jt}^* \cos \lambda_j + \omega_{jt}^*, \quad j = 1, \dots, [s/2] \end{aligned} \quad (3.12)$$

with the ω_{jt} and ω_{jt}^* terms are independent $N(0, \sigma_\omega^2)$ variables. When the stochastic term in 3.12 are zero, the values of γ_t defined by model 3.11 are periodic with the period being s . This can be shown by taking

$$\begin{aligned} \gamma_{jt} &= \tilde{\gamma}_j \cos \lambda_j t + \tilde{\gamma}_j^* \sin \lambda_j t, \\ \gamma_{jt}^* &= -\tilde{\gamma}_j \sin \lambda_j t + \tilde{\gamma}_j^* \cos \lambda_j t, \end{aligned} \quad (3.13)$$

which satisfy the deterministic part of 3.12. The required result follows considering γ_t that is defined by 3.9 is periodic with the period being s . Thus the deterministic part of 3.12 gives a recursion for 3.9. There is an advantage of choosing 3.11 over 3.10, the advantage is that in 3.11 the contributions of the errors ω_{jt} and ω_{jt}^* aren't amplified by the trigonometric functions $\cos \lambda_j t$ and $\sin \lambda_j t$. Model 3.7 is regarded as the main time domain model and model 3.11 is regarded as the main frequency domain model for the seasonality in structural time series analysis.

In our model the seasonality term will be presented in the state equation as a mean adjustment added to the model 3.1, this gives the model 3.14:

$$\begin{aligned} y_t &= Z_t \alpha_t + d_t + \epsilon_t & \epsilon_t &\sim N(0, H_t) \\ \alpha_{t+1} &= T_t \alpha_t + c_t + \eta_t & \eta_t &\sim N(0, Q_t) \\ & & \alpha_1 &\sim N(a_1, P_1) \end{aligned} \quad (3.14)$$

where d_t is a $p \times 1$ vector and c_t is $m \times 1$ vector. In the case of this thesis $d_t = 0$ and c_t is the seasonality term in our model.

4 Nelson Siegel model

Now that we have an understanding of what a dynamic factor model and a linear Gaussian state space model is we can move on with the model that is going to be used in this thesis. In this section I will talk about the Nelson-Siegel model. I will look at the three-factor Nelson-Siegel model and a three-factor Nelson-Siegel model with seasonality. There are more variants of the Nelson-Siegel model, such as a four- and five-factor model. These will not be considered due to them being computationally expensive. For a discussion on these models I refer you to Svensson (1994), Christensen et al. (2009) and de Rezende and Ferreira (2013).

4.1 Three-factor model

For the term structure of futures contracts on commodities, I will look at a dynamic version of the Nelson-Siegel model as was suggested by Grønborg and Lunde (2016). A three factor Nelson-Siegel model has the following form in terms of a linear regression model:

$$y_t(\tau_i) = \beta_{1t} + \beta_{2t} \left(\frac{1 - e^{-\lambda\tau_i}}{\lambda\tau_i} \right) + \beta_{3t} \left(\frac{1 - e^{-\lambda\tau_i}}{\lambda\tau_i} - e^{-\lambda\tau_i} \right) + \epsilon_{it}, \quad i = 1, \dots, N_t \quad (4.1)$$

where $y_t(\tau_i)$ denotes the price for future contract i with the maturity τ_i that is observed at time period t , ($t = 1, \dots, T$). ϵ_{it} denotes the error term and $\beta_t = (\beta_{1t}, \beta_{2t}, \beta_{3t})^T$ are interpreted as factors capturing the level (β_{1t}), the slope (β_{2t}) and the curvature (β_{3t}) of the yield curve. The parameter $\lambda > 0$ determines the exponential decay of the second and third component in (4.1). τ_i is the maturity of $y_t(\tau_i)$. We can write this model as a linear Gaussian state space model in the following way:

$$\begin{aligned} y_t &= Z_t \beta_t + \epsilon_t, & \epsilon_t &\sim N(0, \Sigma_y) \\ \beta_{t+1} &= \beta_t + \eta_t, & \eta_t &\sim N(0, \sigma_\eta^2) \end{aligned} \quad (4.2)$$

with $\Sigma_y = \sigma_y^2 I$.

For the case with the three-factor model with seasonality, the observation equation is the same. The seasonality term is added to the state equation. The three-factor model with seasonality is presented in 4.3, this model is referred to as the seasonal model in the rest of this thesis:

$$\begin{aligned} y_t &= Z_t \beta_t + \epsilon_t, & \epsilon_t &\sim N(0, \Sigma_y) \\ \beta_{t+1} &= \beta_t + c_t + \eta_t, & \eta_t &\sim N(0, \sigma_\eta^2) \end{aligned} \tag{4.3}$$

with $-\pi \leq \Omega \leq \pi$. The seasonality term is trigonometric, $c_t = \begin{pmatrix} \theta_1 \cos\left(\frac{2\pi t}{s} + \Omega\right) \\ \theta_2 \cos\left(\frac{2\pi t}{s} + \Omega\right) \\ \theta_3 \cos\left(\frac{2\pi t}{s} + \Omega\right) \end{pmatrix}$.

5 Kalman filter

5.1 Introduction

As the linear Gaussian state space and the dynamic factor models have been established, our next task is to estimate the β 's in the Nelson-Siegel model. This is where the Kalman filter comes in. A Kalman filter is a numerical method that uses different observations over time and produces estimates of unknown variables. The observations may include statistical noise and other inaccuracies and the Kalman filter uses a recursive method to compute a statistical optimal estimate. The filter is named after Rudolf E. Kálmán, one of the primary developers of this theory in the late 50s and early 60s. The Kalman filter has several applications in technology. Navigation, guidance and control of vehicles, especially aircraft and spacecraft is an area where the Kalman filter is used. Kalman filter is also a widely applied concept in time series analysis that are used in such fields as econometrics and signal processing.

In this section we put the Kalman filter under the microscope and derive the Kalman filter for a linear Gaussian state space model, (3.1).

5.2 Derivation of the Kalman filter

Before we can derive the Kalman filter recursions we need a little background theory. I will present some results that are well known in regression theory. If we let X be normally distributed with a nonsingular variance matrix, $X \sim N(\mu, \Sigma)$. From Anderson (2003) we have a theorem as such:

Theorem 2.5.1. *Let the components of X be divided into two groups composing the subvectors $X^{(1)}$ and $X^{(2)}$. Suppose the mean μ is similarly divided into $\mu^{(1)}$ and $\mu^{(2)}$, and suppose the covariance matrix Σ of X is divided into $\Sigma_{11}, \Sigma_{12}, \Sigma_{22}$, the covariance matrices of $X^{(1)}$, of $X^{(1)}$ and $X^{(2)}$, and of $X^{(2)}$, respectively. Then if the distribution of X is normal, the conditional distribution of $X^{(1)}$ given $X^{(2)} = x^{(2)}$ is normal with mean $\mu^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(x^{(2)} - \mu^{(2)})$*

and covariance matrix $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

To summarize this theorem, conditional distributions that is derived from joint normal distributions are normal. The means of such distributions does only depend linearly on the variates being held fixed. The variances and covariances does not depend on the values of the variates at all. Let x_1 and x_2 be joint normally distributed random vectors with:

$$E \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad Var \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma'_{12} & \Sigma_{22} \end{bmatrix}, \quad (5.1)$$

where Σ_{yy} a nonsingular matrix. Then the conditional distribution $x|y$ is a normal distribution with a mean vector and variance matrix:

$$E(x_1|x_2) = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \quad (5.2)$$

$$Var(x_1|x_2) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma'_{12}. \quad (5.3)$$

When Σ_{yy} is singular, 5.2 and 5.3 will still be valid if we interpret Σ_{yy}^{-1} as a generalised inverse.

Now we want to produce an estimator for the latent variable α_t and α_{t+1} given the data Y_t for $t = 1, \dots, n$. We have that the observations and the states are Markovian:

$$P(y_t|\alpha_1, \dots, \alpha_t, Y_{t-1}) = P(y_t|\alpha_t), \quad (5.4)$$

and

$$P(\alpha_{t+1}|\alpha_1, \dots, \alpha_t, Y_t) = P(\alpha_{t+1}|\alpha_t). \quad (5.5)$$

I will throughout this part use these definitions from James Durbin (2012):

$$a_{t|t} = E(\alpha_t|Y_t), \quad (5.6)$$

$$a_{t+1} = E(\alpha_{t+1}|Y_t), \quad (5.7)$$

$$P_{t|t} = Var(\alpha_t|Y_t), \quad (5.8)$$

$$P_{t+1} = Var(\alpha_{t+1}|Y_t). \quad (5.9)$$

Using the theorem from Anderson (2003) cited further up, we know that since all the distributions are normal then the conditional distributions of α_t and α_{t+1} given Y_t are also normal, given by $N(a_{t|t}, P_{t|t})$ and $N(a_{t+1}, P_{t+1})$ respectively. From this we can also conclude that the distribution of α_t given Y_{t-1} is $N(a_t, P_t)$, and we will use α_t and P_t to recursively calculate 5.6, 5.7, 5.8 and 5.9 for $t = 1, \dots, n$.

Now let's introduce v_t , the one-step ahead forecast error (or prediction error) of y_t given Y_{t-1} :

$$\begin{aligned}
v_t &= y_t - E(y_t|Y_{t-1}) \\
&= y_t - E(Z_t\alpha_t + \epsilon_t|Y_{t-1}) \\
&= y_t - Z_tE(\alpha_t|Y_{t-1}) - E(\epsilon_t|Y_{t-1}) \\
&= y_t - Z_t a_t
\end{aligned} \tag{5.10}$$

Now let:

$$\begin{aligned}
F_t &= Var(v_t|Y_{t-1}) \\
&= Var(y_t - Z_t a_t|Y_{t-1}) \\
&= Var(Z_t\alpha_t + \epsilon_t - Z_t a_t|Y_{t-1}) \\
&= Var(Z_t\alpha_t|Y_{t-1}) + Var(\epsilon_t|Y_{t-1}) \\
&= Z_t Var(\alpha_t|Y_{t-1}) Z_t' + H_t \\
&= Z_t P_t Z_t' + H_t
\end{aligned} \tag{5.11}$$

When v_t and Y_{t-1} are fixed, then Y_t is also fixed which then means:

$$\begin{aligned}
a_{t|t} &= E(\alpha_t|Y_t) = E(\alpha_t|Y_{t-1}, v_t), \\
a_{t+1} &= E(\alpha_{t+1}|Y_t) = E(\alpha_{t+1}|Y_{t-1}, v_t)
\end{aligned}$$

Now if we apply the Theorem from Anderson (2003), let x_1 and x_2 in the

theorem be α_t and v_t respectively, thus we have:

$$\begin{aligned} a_{t+1} &= E(\alpha_{t+1}|Y_{t-1}) + Cov(\alpha_{t+1}, v_t|Y_{t-1}) Var(v_t|Y_{t-1})^{-1} (v_t - E(v_t)) \\ &= E(\alpha_{t+1}|Y_{t-1}) + Cov(\alpha_{t+1}, v_t|Y_{t-1}) F_t^{-1} v_t, \end{aligned}$$

since $E(v_t) = E(E(v_t|Y_{t-1})) = E(0) = 0$. The covariance $Cov(\alpha_{t+1}, v_t|Y_{t-1})$ is:

$$\begin{aligned} Cov(\alpha_{t+1}, v_t|Y_{t-1}) &= Cov(T_t \alpha_t + R_t \eta_t, y_t - Z_t a_t | Y_{t-1}) \\ &= Cov(T_t \alpha_t + R_t \eta_t, Z_t \alpha_t + \epsilon_t - Z_t a_t | Y_{t-1}) \\ &= E((T_t \alpha_t + R_t \eta_t)(Z_t \alpha_t + \epsilon_t - Z_t a_t)' | Y_{t-1}) \\ &\quad - E(T_t \alpha_t + R_t \eta_t) E(Z_t \alpha_t + \epsilon_t - Z_t a_t | Y_{t-1}) \\ &= E(T_t \alpha_t (Z_t \alpha_t - Z_t a_t)' | Y_{t-1}) \\ &= T_t E(\alpha_t (\alpha_t - a_t) | Y_{t-1}) Z_t' \\ &= T_t Var(\alpha_t | Y_{t-1}) Z_t' \\ &= T_t P_t Z_t' \end{aligned}$$

Next we have:

$$\begin{aligned} E(\alpha_{t+1}|Y_{t-1}) &= E(T_t \alpha_t + R_t \eta_t | Y_{t-1}) \\ &= T_t E(\alpha_t | Y_{t-1}) \\ &= T_t a_t \end{aligned}$$

Thus we get a_{t+1} to be:

$$\begin{aligned} a_{t+1} &= T_t a_t + T_t P_t Z_t' F_t^{-1} v_t \\ &= T_t a_t + K_t v_t \end{aligned} \tag{5.12}$$

with $K_t = T_t P_t Z_t' F_t^{-1}$, K_t is called the Kalman gain.

Now we need to calculate P_{t+1} :

$$\begin{aligned}
P_{t+1} &= \text{Var}(\alpha_{t+1}|Y_t) \\
&= \text{Var}(\alpha_{t+1}|Y_{t-1}, v_t) \\
&= \text{Var}(\alpha_{t+1}|Y_{t-1}) - \text{Cov}(\alpha_{t+1}, v_t|Y_{t-1}) \text{Var}(v_t|Y_{t-1})^{-1} \text{Cov}(\alpha_{t+1}, v_t|Y_{t-1})' \\
&= \text{Var}(\alpha_{t+1}|Y_{t-1}) - T_t P_t Z_t' F_t^{-1} Z_t P_t' T_t' \\
&= \text{Var}(T_t \alpha_t + R_t \eta_t | Y_{t-1}) - T_t P_t Z_t' F_t^{-1} Z_t P_t' T_t' \\
&= T_t \text{Var}(\alpha_t | Y_{t-1}) T_t' + R_t Q_t R_t' - T_t P_t Z_t' F_t^{-1} Z_t P_t' T_t' \\
&= T_t P_t (T_t - K_t Z_t)' + R_t Q_t R_t' \tag{5.13}
\end{aligned}$$

Lastly we're going to calculate $a_{t|t}$ and $P_{t|t}$:

$$\begin{aligned}
a_{t|t} &= E(\alpha_t | Y_t) \\
&= E(\alpha_t | Y_{t-1}, v_t) \\
&= E(\alpha_t | Y_{t-1}) + \text{Cov}(\alpha_t, v_t | Y_{t-1}) \text{Var}(v_t | Y_{t-1})^{-1} v_t \\
\text{Cov}(\alpha_t, v_t | Y_{t-1}) &= E(\alpha_t (y_t - Z_t a_t)' | Y_{t-1}) \\
&= E(\alpha_t (Z_t \alpha_t + \epsilon_t - Z_t a_t)' | Y_{t-1}) \\
&= E(\alpha_t (\alpha_t - a_t) Z_t' | Y_{t-1}) \\
&= \text{Var}(\alpha_t | Y_{t-1}) Z_t' \\
&= P_t Z_t' \\
a_{t|t} &= a_t + P_t Z_t' F_t^{-1} v_t \tag{5.14}
\end{aligned}$$

$$\begin{aligned}
P_{t|t} &= \text{Var}(\alpha_t | Y_t) \\
&= \text{Var}(\alpha_t | Y_{t-1}, v_t) \\
&= \text{Var}(\alpha_t | Y_{t-1}) - \text{Cov}(\alpha_t, v_t | Y_{t-1}) \text{Var}(v_t | Y_{t-1})^{-1} \text{Cov}(\alpha_t, v_t | Y_{t-1})' \\
&= P_t - P_t Z_t' F_t^{-1} Z_t P_t' \tag{5.15}
\end{aligned}$$

The kalman filter equations are collected together in 5.16 for the conveniency

of the reader:

$$\begin{aligned}
v_t &= y_t - Z_t a_t, & F_t &= Z_t P_t Z_t' + H_t, \\
a_{t|t} &= a_t + P_t Z_t' F_t^{-1} v_t, & P_{t|t} &= P_t - P_t Z_t' F_t^{-1} Z_t P_t, \\
a_{t+1} &= T_t a_t + K_t v_t, & P_{t+1} &= T_t P_t (T_t - K_t Z_t)' + R_t Q_t R_t'
\end{aligned} \tag{5.16}$$

for $t = 1, \dots, n$, where $K_t = T_t P_t Z_t' F_t^{-1}$ with a_1 and P_1 as the mean vector and variance matrix of the initial state vector α_1 . Once $a_{t|t}$ and $P_{t|t}$ are computed, it is enough to use the relations in 5.17

$$a_{t+1} = T_t a_{t|t}, P_{t+1} = T_t P_{t|t} T_t' + R_t Q_t R_t', \tag{5.17}$$

for predicting α_{t+1} and its variance matrix at time t .

The Kalman filter for the seasonal model presented in 3.14 is not very different from the Kalman filter for 3.1. For the reader, the model 3.14 is presented again:

$$\begin{aligned}
y_t &= Z_t \alpha_t + d_t + \epsilon_t & \epsilon_t &\sim N(0, H_t) \\
\alpha_{t+1} &= T_t \alpha_t + c_t + \eta_t & \eta_t &\sim N(0, Q_t) \\
\alpha_1 &\sim N(a_1, P_1)
\end{aligned} \tag{5.18}$$

We use the same theorem and methods to calculate the Kalman filter for model 5.18.

$$\begin{aligned}
v_t &= y_t - E(y_t | Y_{t-1}) \\
&= y_t - E(Z_t \alpha_t + d_t + \epsilon_t | Y_{t-1}) \\
&= y_t - Z_t E(\alpha_t | Y_{t-1}) - E(d_t | Y_{t-1}) - E(\epsilon_t | Y_{t-1}) \\
&= y_t - Z_t a_t - d_t
\end{aligned} \tag{5.19}$$

$$\begin{aligned}
F_t &= \text{Var}(v_t|Y_{t-1}) \\
&= \text{Var}(y_t - Z_t a_t - d_t|Y_{t-1}) \\
&= \text{Var}(Z_t \alpha_t + d_t + \epsilon_t - Z_t a_t - d_t|Y_t) \\
&= \text{Var}(Z_t \alpha_t|Y_t) + \text{Var}(\epsilon_t|Y_t) \\
&= Z_t P_t Z_t' + H_t
\end{aligned} \tag{5.20}$$

$$\begin{aligned}
a_{t|t} &= E(\alpha_t|Y_t) \\
&= E(\alpha_t|Y_{t-1}, v_t) \\
&= E(\alpha_t|Y_{t-1}) + \text{Cov}(\alpha_t, v_t|Y_{t-1}) \text{Var}(v_t|Y_{t-1})^{-1} v_t
\end{aligned}$$

$$\begin{aligned}
\text{Cov}(\alpha_t, v_t|Y_{t-1}) &= E(\alpha_t (Z_t \alpha_t + d_t + \epsilon_t - Z_t a_t - d_t)' | Y_{t-1}) \\
&= E(\alpha_t (Z_t \alpha_t + \epsilon_t - Z_t a_t)' | Y_{t-1}) \\
&= P_t Z_t' \\
a_{t|t} &= a_t + P_t Z_t' F_t^{-1} v_t
\end{aligned} \tag{5.21}$$

$$\begin{aligned}
P_{t|t} &= \text{Var}(\alpha_t|Y_t) \\
&= \text{Var}(\alpha_t|Y_{t-1}, v_t) \\
&= \text{Var}(\alpha_t|Y_{t-1}) - \text{Cov}(\alpha_t, v_t|Y_{t-1}) \text{Var}(v_t|Y_{t-1}) \text{Cov}(\alpha_t, v_t|Y_{t-1})' \\
&= P_t - P_t Z_t' F_t^{-1} Z_t P_t'
\end{aligned} \tag{5.22}$$

$$\begin{aligned}
a_{t+1} &= E(\alpha_{t+1}|Y_t) \\
&= E(T_t \alpha_t + c_t + R_t \eta_t|Y_t) \\
&= T_t E(\alpha_t|Y_t) + E(c_t) \\
&= T_t a_{t|t} + c_t
\end{aligned} \tag{5.23}$$

$$\begin{aligned}
P_{t+1} &= \text{Var}(\alpha_{t+1}|Y_t) \\
&= \text{Var}(T_t\alpha_t + c_t + R_t\eta_t|Y_t) \\
&= T_t\text{Var}(\alpha_t|Y_t)T_t' + R_tQ_tR_t' \\
&= T_tP_{t|t}T_t' + R_tQ_tR_t'
\end{aligned} \tag{5.24}$$

For the convenience of the reader the filter equations for model 5.18 are collected together:

$$\begin{aligned}
v_t &= y_t - Z_t a_t - d_t, & F_t &= Z_t P_t Z_t' + H_t, \\
a_{t|t} &= a_t + P_t Z_t' F_t^{-1} v_t, & P_{t|t} &= P_t - P_t Z_t' F_t^{-1} Z_t P_t', \\
a_{t+1} &= T_t a_{t|t} + c_t, & P_{t+1} &= T_t P_{t|t} T_t' + R_t Q_t R_t',
\end{aligned} \tag{5.25}$$

In the table below the dimensions of the vectors and matrices of the Kalman filter are presented.

Vector		Matrix	
v_t	$p \times 1$	F_t	$p \times p$
		K_t	$m \times p$
a_t	$m \times 1$	P_t	$m \times m$
$a_{t t}$	$m \times 1$	$P_{t t}$	$m \times m$
d_t	$p \times 1$		
c_t	$m \times 1$		

Table 2: Dimensions of Kalman filter.

5.3 Missing observations

Lets assume that some observations y_j are missing for $j = \tau, \dots, \tau^*$ where $1 < \tau < \tau^* < n$.

An easy way to handle missing values is as such, for the missing times $t = \tau, \dots, \tau^* - 1$ we set $Z_t = 0$ in the Kalman filter equations in 5.16 and 5.25. We then get these results:

$$\begin{aligned}
a_{t|t} &= E(\alpha_t|Y_t) \\
&= a_t
\end{aligned}$$

$$\begin{aligned}
P_{t|t} &= \text{Var}(\alpha_t | Y_t) \\
&= P_t
\end{aligned}$$

$$\begin{aligned}
a_{t+1} &= E(\alpha_{t+1} | Y_t) \\
&= T_t a_t
\end{aligned}$$

$$\begin{aligned}
P_{t+1} &= \text{Var}(\alpha_{t+1} | Y_t) \\
&= T_t P_t T_t' + R_t Q_t R_t'
\end{aligned}$$

Thus we can use the same Kalman filter equations as we derived above. This way of treating missing observations is very straightforward and is a reason for the appeal that state space methods has when it comes to state space analysis.

Now let us assume that not all observations in y_t at time t are missing, just some of the element of y_t are missing. A way to handle this case is to let the vector of values actually observed be set as y^* . We can then take y^* to be

$$y^* = W_t y_t$$

where W_t is a matrix whose rows are a subset of the rows in I . Therefore, at the times t where some of the elements in y_t are missing, the observation equation in 3.1 will then be replaced with the equation

$$y^* = Z_t^* \alpha_t + \epsilon_t^*, \epsilon_t^* \sim N(0, H_t^*),$$

where $Z_t^* = W_t Z_t$, $\epsilon_t^* = W_t \epsilon_t$ and $H_t^* = W_t H_t W_t'$. The filtering using the Kalman filter 5.16 or 5.25 can then continue as usual as long as y_t , Z_t and H_t are replaced by y_t^* , Z_t^* and H_t^* at the times t where y_t has some elements missing.

6 Forecasting

In this section we will look at forecasting. There are many ways to do forecasting and which method you choose depends on what your goal is. For example, if we want to forecast future observations of the state vector, that is the β_t 's in our model, then one can treat the future forecasted observed values, y_{n+j} for $j = 1, \dots, J$, as missing observations.

$$\begin{aligned}\bar{y}_{n+j} &= Z_{n+j} E(\alpha_{n+j} | Y_n) \\ &= Z_{n+j} \bar{a}_{n+j}\end{aligned}$$

$$\begin{aligned}\bar{F}_{n+j} &= E\left[(\bar{y}_{n+j} - y_{n+j})(\bar{y}_{n+j} - y_{n+j})' | Y_n\right] \\ &= Z_{n+j} \bar{P}_{n+j} Z_{n+j}' + H_{n+j}\end{aligned}$$

$$\begin{aligned}\bar{a}_{n+j+1} &= T_{n+j} E(\alpha_{n+j} | Y_n) \\ &= T_{n+j} \bar{a}_{n+j}\end{aligned}$$

where $\bar{a}_{n+1} = a_{n+1}$.

$$\begin{aligned}P_{n+j+1} &= E\left[(\bar{a}_{n+j+1} - \alpha_{n+j+1})(\bar{a}_{n+j+1} - \alpha_{n+j+1})' | Y_n\right] \\ &= T_{n+j} \bar{P}_{n+j} T_{n+j}' + R_{n+j} Q_{n+j} R_{n+j}'\end{aligned}$$

If one are to forecast the observations y_t then one way to do so is to do a one-day-ahead forecast. The premise of this forecast is to switch between doing a forecast one-day-ahead and updating the filter with observation that is done on the forecasted day. The first forecasted day would be:

$$\begin{aligned}\bar{y}_{n+1} &= Z_{n+1} E(\alpha_{n+1} | Y_n) \\ &= Z_{n+1} a_{n+1}\end{aligned}$$

where Z_{n+1} is the Z matrix for the first forecasted day and a_{n+1} is the predicted

value of the state vector, we get this vector from the last iteration of the Kalman Filter for the data we already have. Now we have forecasted one-day-ahead. The next thing we do is observe what y_t actually is on the forecasted day. Then we update the Kalman Filter with this observed y_t , this will give us an estimation of the state vector α_n and the predicted state vector, α_{n+1} , for the next day. This predicted state vector value will be used to forecast the next day again. In principal it will look like this. First we forecast day $n + 1$:

$$\bar{y}_{n+1} = Z_{n+1}a_{n+1}$$

then we observe the value of y_t when the forecasted day comes, $t = n + 1$. We update the filter equations with this vector:

$$\begin{aligned} v_{n+1} &= y_{n+1} - Z_{n+1}a_{n+1}, \\ F_{n+1} &= Z_{n+1}P_{n+1}Z'_{n+1} + H_{n+1}, \\ a_{n+1|n+1} &= a_{n+1} + P_{n+1}Z'_{n+1}F_{n+1}^{-1}v_{n+1}, \\ P_{n+1|n+1} &= P_{n+1} - P_{n+1}Z'_{n+1}F_{n+1}^{-1}Z_{n+1}P_{n+1}, \\ a_{n+2} &= T_{n+1}a_{n+1} + K_tv_{n+1}, \\ P_{n+2} &= T_{n+1}P_{n+1}(T_{n+1} - K_{n+1}Z_{n+1})' + R_{n+1}Q_{n+1}R'_{n+1} \end{aligned}$$

Now we can predict y_t for $t = n + 2$:

$$\bar{y}_{n+2} = Z_{n+2}a_{n+2}$$

and so we update the Kalman Filter again and so forth.

7 Maximum likelihood estimation

In this thesis I am using maximum likelihood estimation (MLE) for parameter estimation and to find the goodness of fit of a model. The reader is expected familiar with MLE, but in this section the application MLEs for this thesis will be expanded upon.

7.1 Parameter estimation

The matrices Z_t , R_t , T_t , H_t and Q_t may all or some of them depend on a vector of parameters ψ . We are interested in their values so that we get a better estimation from the Kalman filter. The different parameters of ψ are estimated using maximum likelihood estimation. The likelihood is:

$$\begin{aligned} L(Y_n) &= p(y_1, \dots, y_n) \\ &= p(y_1) \prod_{t=2}^n p(y_t|Y_{t-1}) \end{aligned} \quad (7.1)$$

under the assumption that the initial state vector has the density $N(a_1, P_1)$, with a_1 and P_1 being known. Further, it is the loglikelihood that we want to find:

$$\log(L(Y_n)) = \sum_{t=1}^n \log(p(y_t|Y_{t-1})) \quad (7.2)$$

where $p(y_1|Y_0) = p(y_1)$. For the linear Gaussian state space model, 3.1, we have that $E(y_t|Y_{t-1}) = Z_t a_t$, $v_t = y_t - Z_t a_t$ and $F_t = Var(y_t|Y_{t-1}) = Z_t P_t Z_t' + H_t$. If we now substitute the density $N(Z_t a_t, F_t)$ with $p(y_t|Y_{t-1})$ in the loglikelihood in 7.2 and with a dependence on the parameter vector ψ , we get:

$$\log(L_{Y_n}(\psi)) = -\frac{np}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^n \left(\log(|F_t(\psi)|) + v(\psi)_t' F(\psi)_t^{-1} v(\psi)_t \right) \quad (7.3)$$

with $L(Y_n) = L_{Y_n}$. v_t and F_t are computed using the Kalman filter. As such it is easy to update the loglikelihood with the output of each iteration of the Kalman filter. If the reader is interested, further discussions on the subject of parameter estimation and MLE can be found here Shumway and Stoffer (2017)

and here James Durbin (2012).

7.2 Goodness of fit

Suppose you have a model that fits your observed data, now how good does the model fit? What if someone comes with a different model and says that it is better than yours? This is where a model goodness of fit comes in to play. For time series models the goodness of fit is often associated with the errors in the forecasted observations. Usually that would be handled by mean squared errors (MSE). If we denote the forecasted observations as y_t^* , the mean squared error would be:

$$MSE = \sum_{t=1}^n (y_t^* - y_t)^2$$

If you have two or more competing models and you want to find out which is the best one, it is common to compare the loglikelihood, $\log(L_{Y_n}(\psi))$, of the different models. The initial problem here is that the greater the number of parameters in ψ the greater the loglikelihood value will be. Thus, if we are to compare the models fairly we use the Akaike information criterion (AIC) or the Bayesian information criterion (BIC). These are given as:

$$AIC = -2 \log(L_{Y_n}(\psi)) + 2\omega$$

$$BIC = -2 \log(L_{Y_n}(\psi)) + \omega \log(n)$$

with ω being the number of parameters in ψ and n is the number of observations you have.

8 Delta method

When you're doing an analysis, there will always be uncertainties in your results. Thus it is normal to calculate confidence intervals so that we can say with more certainty that our results are within the area found. Thus the Delta method is here to calculate the variance or in other words, the length of the confidence interval.

The delta method is, in its essence, a method to determine the variance of a function of random variables that are asymptotically normal with known variances. I will first show the univariate delta method. Let X_n be a sequence of random variables that satisfies

$$\sqrt{n} [X_n - \theta] \xrightarrow{D} N(0, \sigma^2), \quad (8.1)$$

where θ and σ^2 are finite constants, then

$$\sqrt{n} [g(X_n) - g(\theta)] \xrightarrow{D} N\left(0, \sigma^2 [g'(\theta)]^2\right) \quad (8.2)$$

for any function g that satisfies the property of $g'(\theta)$ existing and being non-zero.

The delta method is easily generalized to the multivariate case. Let B be an consistent estimator, which means as the number of observations increases (meaning as n grows) the estimator converges toward the true value it is supposed to estimate. If we have an consistent estimator B we often can use the central limit theorem to achieve asymptotic normality:

$$\sqrt{n} (B - \beta) \xrightarrow{D} N(0, \Sigma), \quad (8.3)$$

here Σ is a symmetric positive semi-definite (non-negative) covariance matrix. Next we need to estimate the variance of a function h of the estimator B , $h(B)$. We can estimate $h(B)$ by taking the first two terms of a Taylor Series as such:

$$h(B) \approx h(\beta) + \nabla h(\beta)^T (B - \beta), \quad (8.4)$$

and that implies that the variance of $h(B)$ can be approximated by:

$$\begin{aligned} \text{Var}(h(B)) &\approx \text{Var}\left(h(\beta) + \nabla h(\beta)^T (B - \beta)\right) \\ &= \nabla h(\beta)^T \left(\frac{\Sigma}{n}\right) \nabla h(\beta) \end{aligned} \quad (8.5)$$

It is thus indicated by the delta method:

$$\sqrt{n}(h(B) - h(\beta)) \xrightarrow{D} N\left(0, \nabla h(\beta)^T \Sigma \nabla h(\beta)\right) \quad (8.6)$$

9 Results

In this paper I have been working with several types of futures contracts of commodities. I have used a Kalman filter and MLE to estimate parameters β_{1t} , β_{2t} , β_{3t} , λ and σ_y . These have been estimated for both the three-factor model and the seasonal model. In addition, the seasonality terms c_{1t} , c_{2t} and c_{3t} have been estimated for the seasonal model. The results from these estimations and the hypothesis testing following these estimates are presented below, with regards to the different data sets. For the hypothesis testing our null hypothesis for all data sets is that there is no seasonality, meaning that we don't get a better fit by adding a seasonality term to the three-factor model.

9.1 Crude oil

9.1.1 Three-factor model

Using the Kalman filter for a linear Gaussian state space model and optimizing the likelihood function we get an estimate for the variables β_{1t} , β_{2t} , β_{3t} and λ . To get a 95% confidence interval of λ and of σ_y we use the *Delta method* to obtain this. The 95% confidence interval of λ is:

$$\lambda = (5.419 \times 10^{-3}, 5.479 \times 10^{-3})$$

Using the optimized λ value we can plug it in the Z -matrix and get the factor loadings for the three-factor model. The factor loadings for the three-factor model with its optimized λ are showed in the following figure:

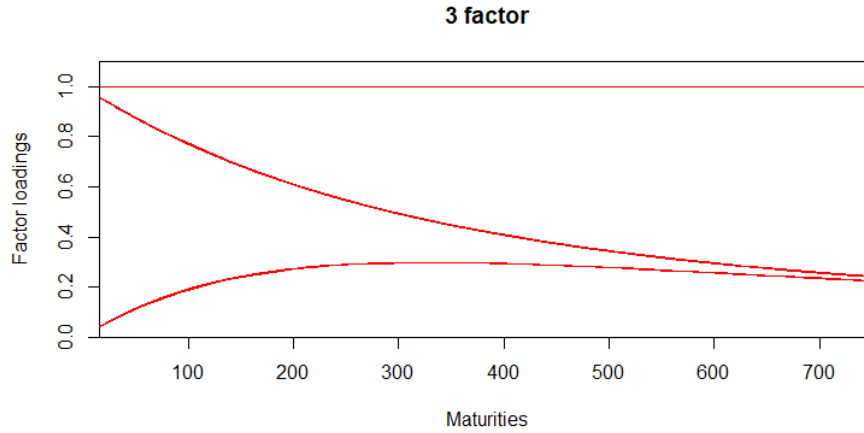


Figure 9.1: Factor loadings for three-factor model on natural gas data

From the optimizing we also get a 95% confidence interval of σ_y that is used in the variance in the error term in equation 4.2.

$$\sigma_y = (2.633 \times 10^{-3}, 2.656 \times 10^{-3})$$

The time series estimates for the β_{1t} 's are:

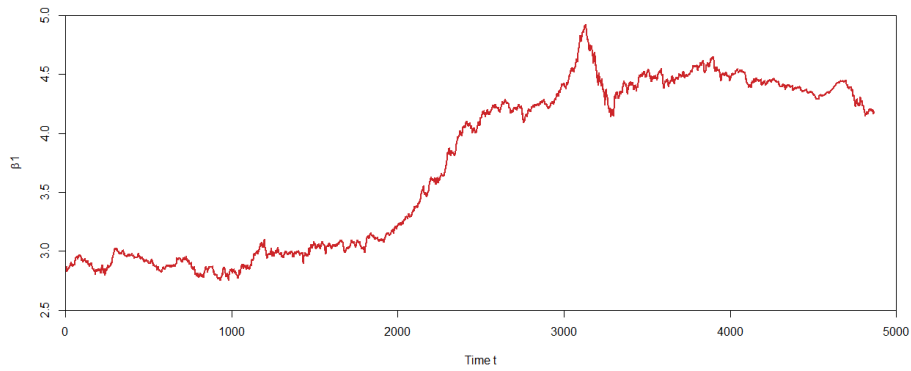


Figure 9.2: Plot of β_{1t} from time $t = 0, \dots, 4865$

The plot for the β_{2t} 's are:

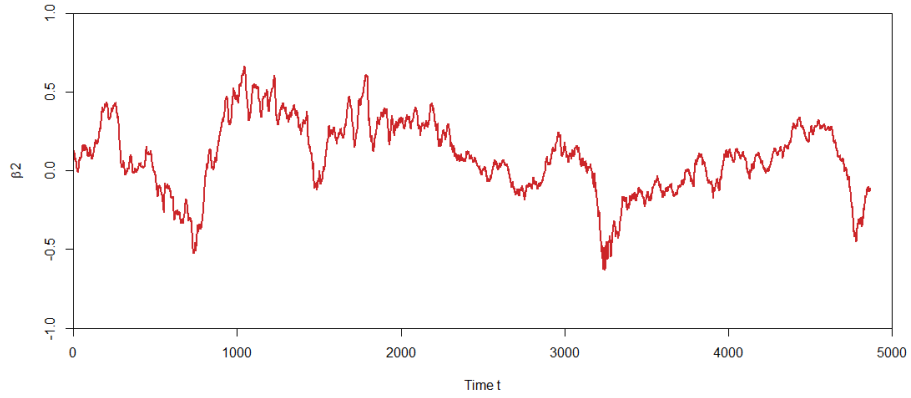


Figure 9.3: Plot of β_{2t} from time $t = 0, \dots, 4865$

The estimates of β_{3t} 's are:

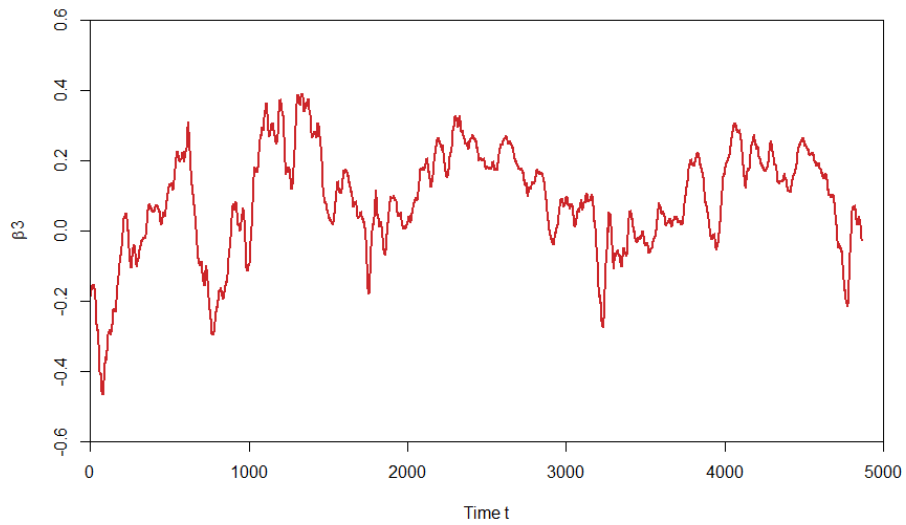


Figure 9.4: Plot of β_{3t} from time $t = 0, \dots, 4865$

The values for the β_{1t} 's are increasing over the crude oil sample size, except for a little decrease after hitting the highpoint, this is an indicator that in general

the prices has increased over time. The β_{2t} 's start out as mainly positive for a while before becoming mainly negative in latter part of the sample size. The β_{3t} 's changes from negative to positive and back several times throughout the sample size.

Forecasting of future observations was also done for both the three-factor model and the seasonal model. I use mean squared error, MSE, to see how good the forecasting is. By doing a one-day-ahead forecast of the future observations, they look like this:

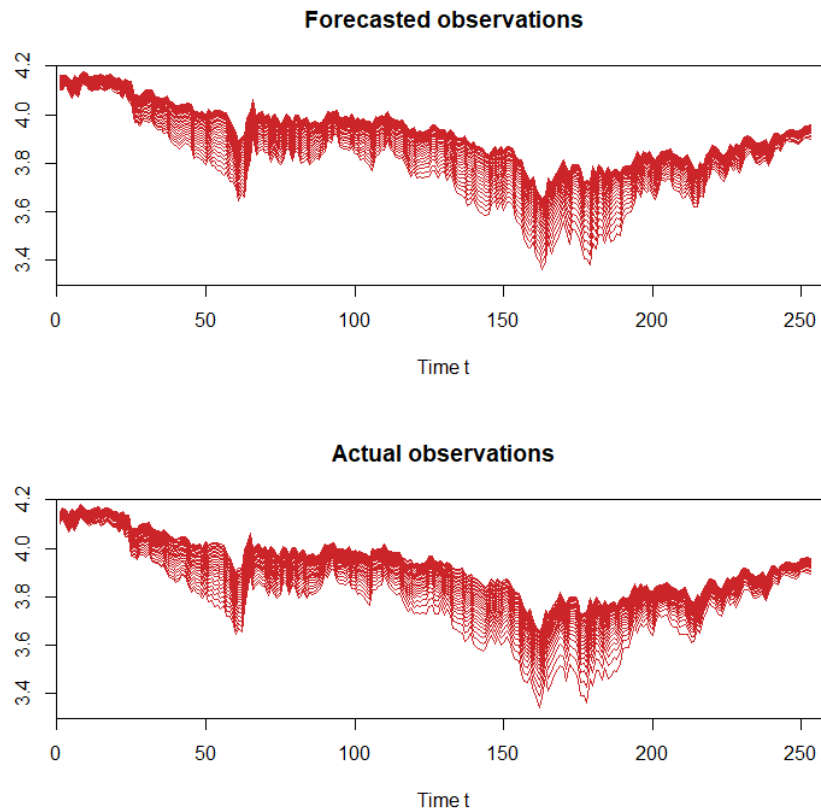


Figure 9.5: Forecasted observations for three-factor model and actual observations of crude oil data

As we can see from the two plots, the forecasted observations look very good.

And with a $MSE = 5.41 \times 10^{-4}$ it is indeed a very good forecast.

9.1.2 Seasonal model

The results for the seasonal model regarding λ and σ_y are quite similar to the three factor model. Down to the 4th decimal, the results are the same for the λ 's and for the σ_y 's the results are the same down to the 5th decimal. The 95% confidence interval, using the Delta method, for λ is:

$$\lambda = (5.420 \times 10^{-3}, 5.480 \times 10^{-3})$$

The factor loadings for the seasonal model:

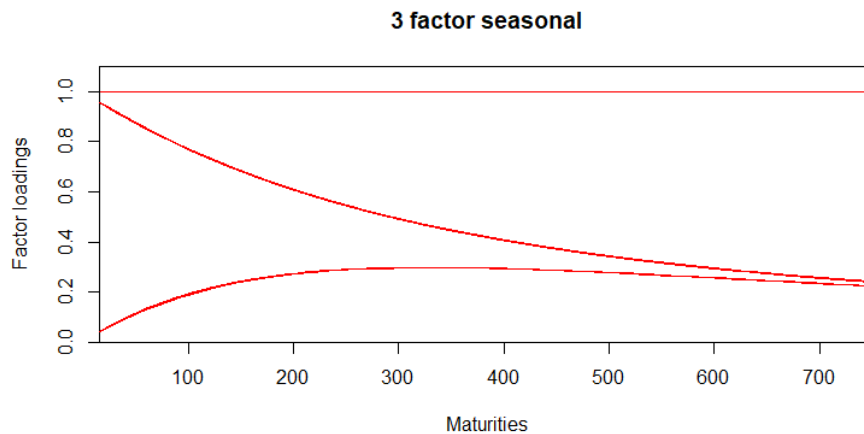


Figure 9.6: Factor loadings of seasonal model on crude oil data

For σ_y , the 95% confidence interval is:

$$\sigma_y = (2.631 \times 10^{-3}, 2.655 \times 10^{-3})$$

The estimates for the β_t 's are also quite similar to the ones from the three factor model. The estimates for β_{1t} are:

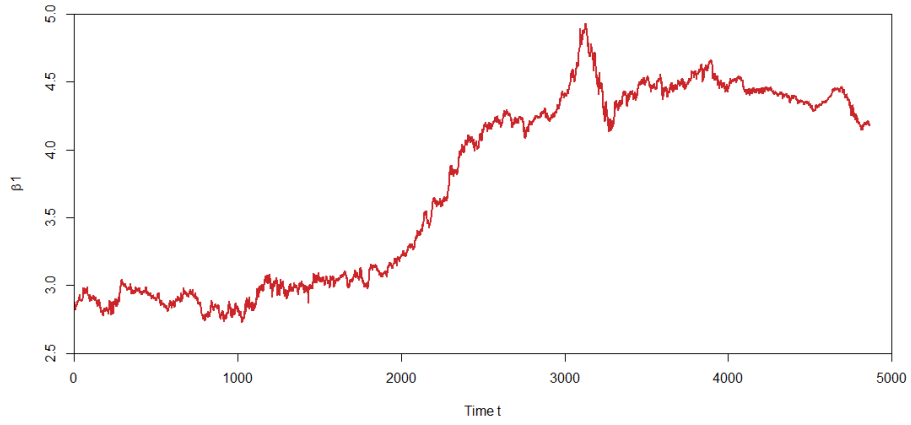


Figure 9.7: Plot of β_{1t} from time $t = 0, \dots, 4865$

The estimates for β_{2t} are:

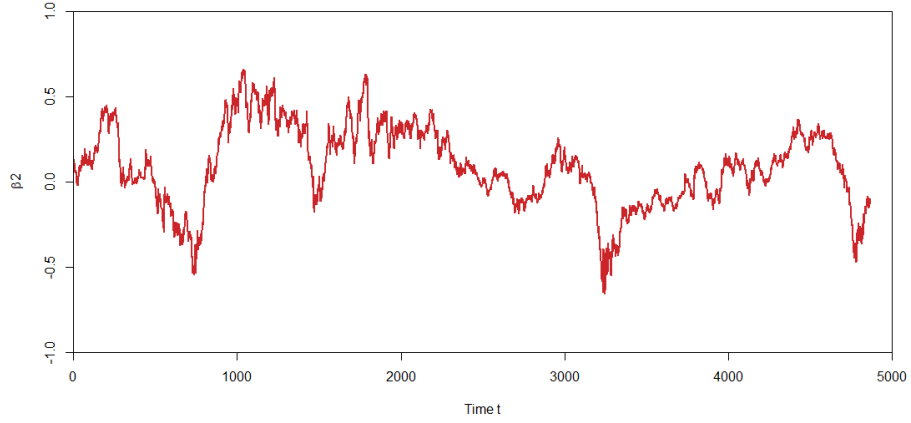


Figure 9.8: Plot of β_{2t} from time $t = 0, \dots, 4865$

And the estimates for β_{3t} are:

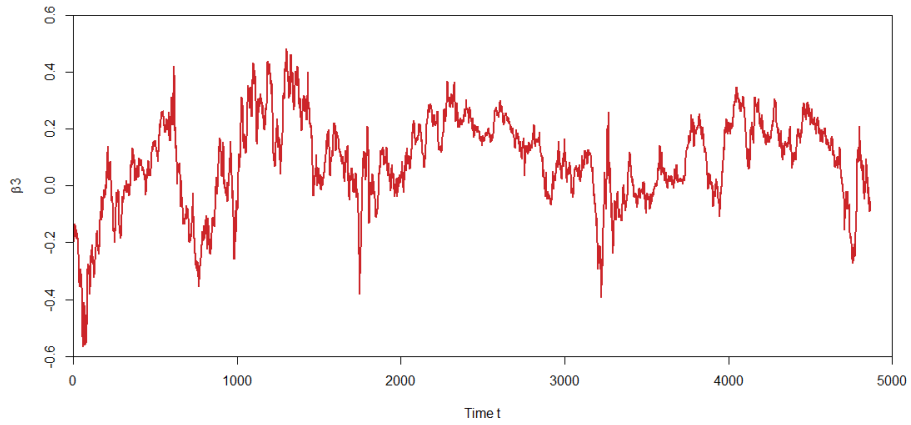


Figure 9.9: Plot of β_{3t} from time $t = 0, \dots, 4865$

The seasonality related to the β_t 's is:

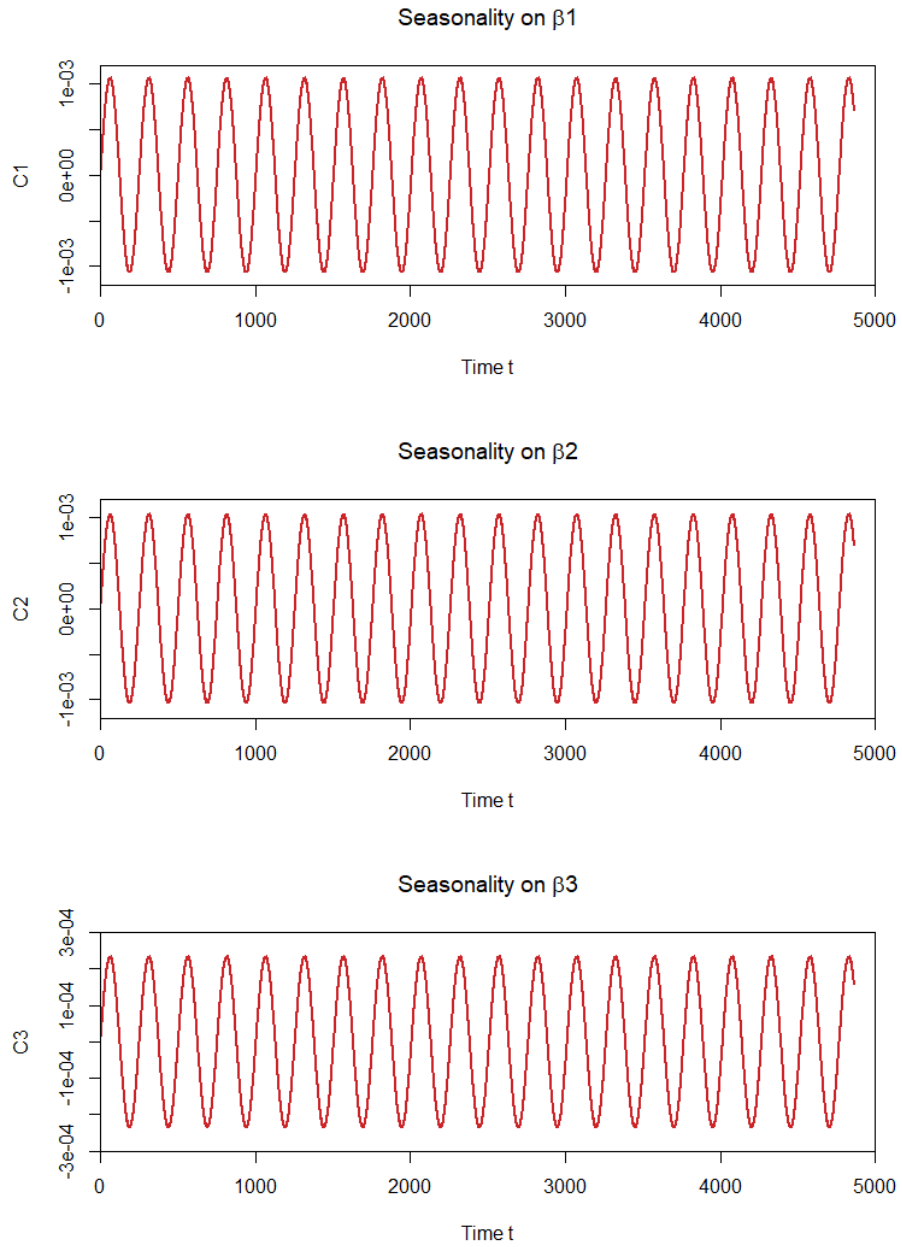


Figure 9.10: Seasonality on β_t with period $s = 253$

The β_t 's for the seasonal model behaves much the same way as for the

three-factor model, where the β_{1t} 's increase throughout, except for the same decrease when hitting the highpoint. The β_{2t} 's are for the most part positive to begin with and the mainly negative at the latter part of the sample, while the β_{3t} 's changes sign several times throughout the sample size. The dataset consists of data collected each day for 253 days a year, thus excluding weekends etc. Therefore the period of the seasonality is $s = 253$. As we can see, the seasonality terms are very small, with the seasonality on the β_{1t} 's and β_{2t} 's having the biggest amplitude.

As mentioned, I also forecasted future observations using the seasonal model. The forecasting method is the same, one-day-ahead forecasting, and to check how the forecast is I use MSE. The next figure shows the forecasted data for the seasonal model and the actual observed data:

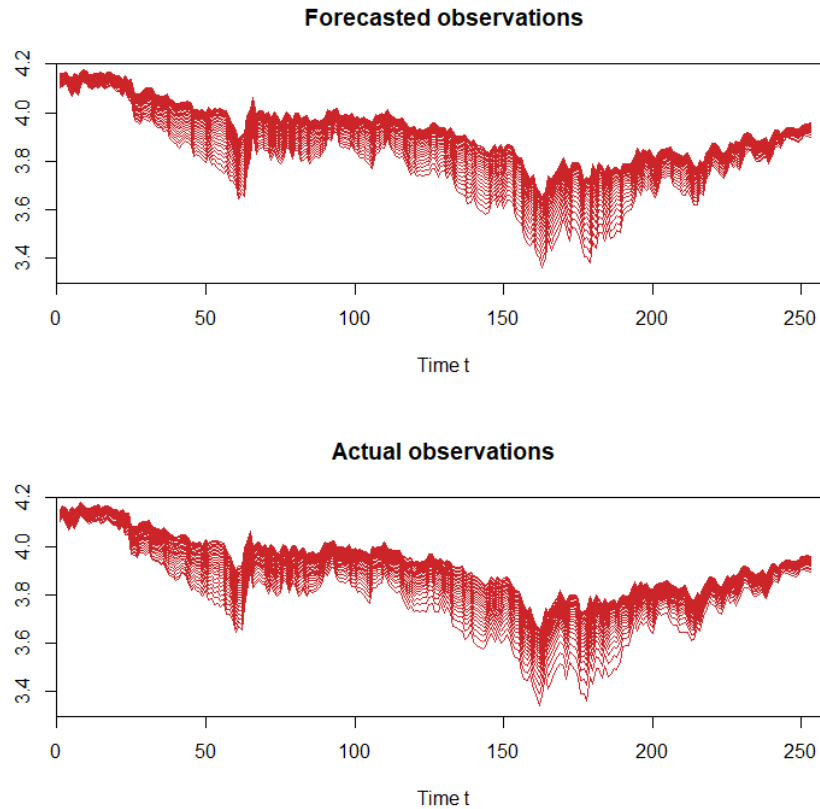


Figure 9.11: Forecasted observations for seasonal model and actual observations of crude oil data

As with the three-factor model, the forecasting looks very good in the plot. Now to check how good it is, $MSE = 5.37 \times 10^{-4}$. A very good result. In fact it is slightly better than the mean squared error value that I got from the forecasting of the three-factor model.

9.1.3 Hypothesis testing

I've been comparing the two models to check if the data for crude oil has seasonality to be aware of. To compare the models I did a likelihood ratio test. Specifically I used Wilks' theorem to perform the test. In short, the theory,

developed by Samuel S. Wilks, says that when $n \rightarrow \infty$ then the test statistic, $-2\log(\Lambda)$, of a nested model, when H_0 is true, will be asymptotically chi-squared distributed with degrees of freedom equal to the difference in dimensionality of the two models, Θ and Θ_0 , being tested, (χ_{p-q}^2) . This means that we only need to calculate the likelihood ratio of the three-factor model and the seasonal model, Λ , and then compare the value of $-2\log(\Lambda)$ with the value of the χ^2 distribution with the corresponding degrees of freedom.

For the three-factor model we have a log-likelihood value of $\log(L(Y_n|\psi_0)) = 499436.7$, and for the seasonal model we have $\log(L(Y_n|\psi)) = 499438.1$. The test statistic is thus:

$$\begin{aligned} D &= -2\log(\Lambda) \\ &= 2 \times (\log(L(Y_n|\psi)) - \log(L(Y_n|\psi_0))) \\ &= 2 \times (499438.1 - 499436.7) \\ &= 2.8 \end{aligned}$$

The parameter vector for the three-factor model, ψ_0 , has a dimensionality of $q = 8$ and the parameter vector for the seasonal model, ψ , has a dimensionality of $p = 12$. Thus the degrees of freedom for the test statistic is $p - q = 12 - 8 = 4$. Now to see if the seasonal model has a significantly better fit than the three-factor model we need to check if $D = 2.8$ is a significant value in a χ_4^2 -distribution. For a chi-squared distribution with degrees of freedom equal to 4 and a significance level of $\alpha = 0.05$, then the test statistic needs to be higher than 9.49. $D = 2.8 < 9.49$, thus we can conclude that there are no seasonality of significance in the dataset.

Furthermore, the AIC and the BIC have been calculated for both models. AIC and BIC are goodness of fit numbers and if they are used to compare models, then one would like to choose the model with the lowest AIC and BIC

value. For the three-factor model:

$$AIC = -998857$$

$$BIC = -998780$$

and for the seasonal model:

$$AIC = -998852$$

$$BIC = -998736$$

which suggests that there is nothing to gain by choosing the seasonal model over the three-factor model.

9.2 Natural gas

9.2.1 Three-factor model

The results from estimating β_{1t} , β_{2t} , β_{3t} , λ and σ_y of the natural gas futures contracts are as follows, the 95% confidence interval for λ :

$$\lambda = (1.412 \times 10^{-3}, 2.217 \times 10^{-3})$$

for σ_y :

$$\sigma_y = (3.419 \times 10^{-2}, 3.464 \times 10^{-2})$$

and the factor loadings for the three-factor model are:

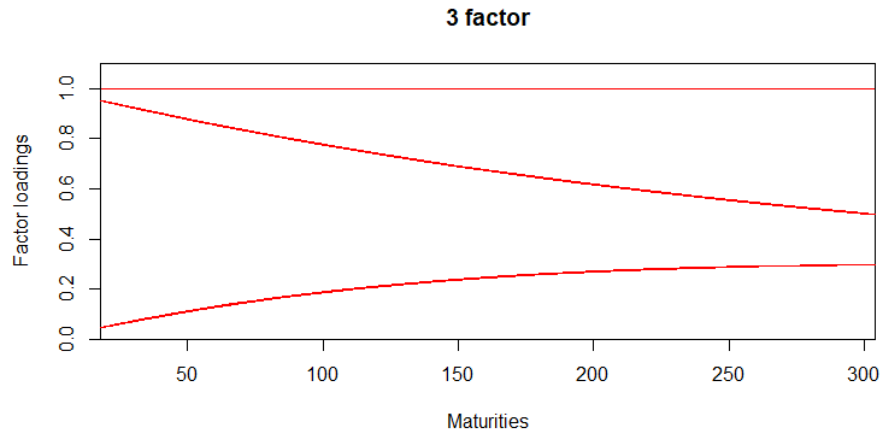


Figure 9.12: Factor loadings of three-factor model on natural gas data

The estimates for the β_{1t} 's are:

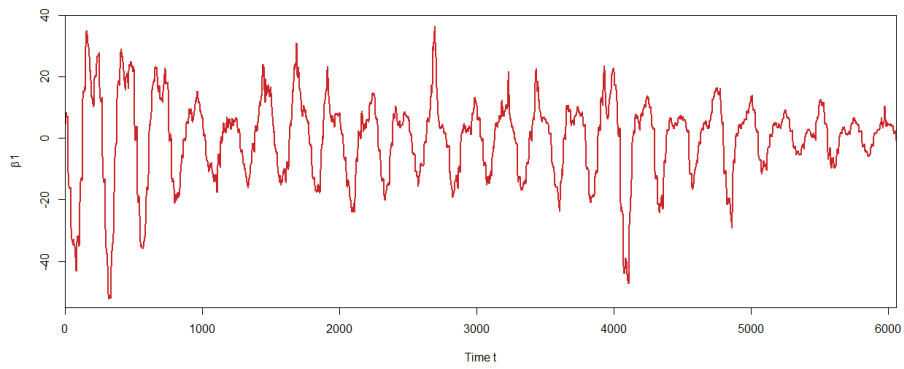


Figure 9.13: Plot of β_{1t} from time $t = 0, \dots, 6059$

For the β_{2t} 's:

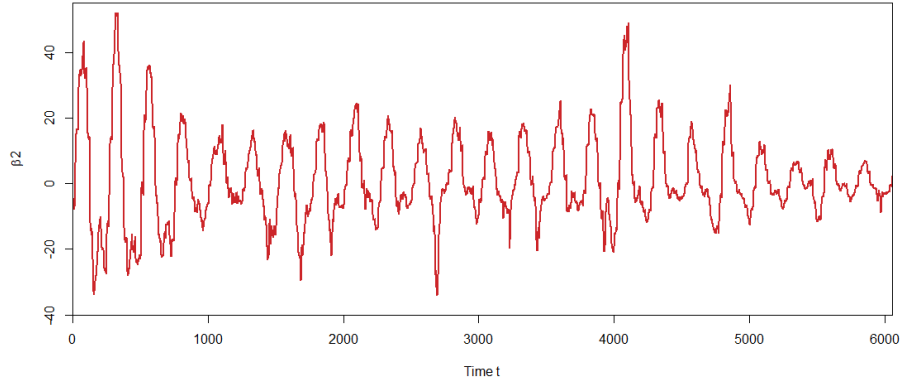


Figure 9.14: Plot of β_{2t} from time $t = 0, \dots, 6059$

And for the β_{3t} 's:

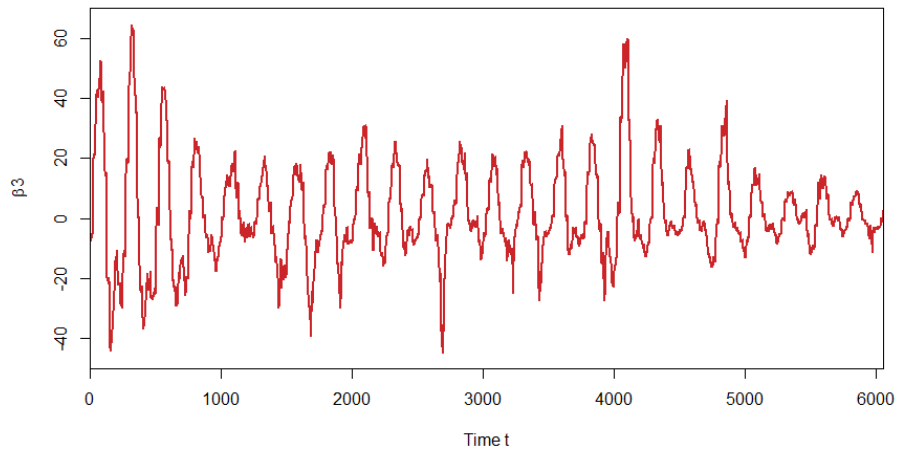


Figure 9.15: Plot of β_{3t} from time $t = 0, \dots, 6059$

The plot for the β_{1t} 's are not how the plot for the β_{1t} 's should look like, hence it is difficult to draw any conclusions from the plot of either of the β_t 's. The estimates for β_{1t} should look like the first column in our dataset due to the β_{1t} 's governing the level of the yield curve. It is an indicator of how the

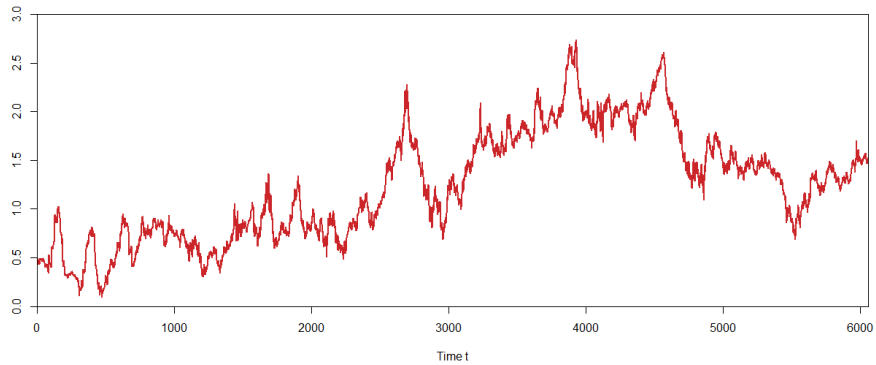


Figure 9.16: Plot of first column in Natural Gas dataset

prices varies in time, if they go up or down. From the estimates of the β_{1t} 's it is impossible to tell if the yield curve is going up or down. Also worth noting is that the estimates for the β_{1t} 's and the β_{2t} 's are mirror opposites of each other, and the β_{3t} 's have the same shape as the β_{2t} 's with higher maximum and minimum values, this should be studied further to find the cause of this problem. The plot of β_{1t} should look like this:

The fact that it does not have this shape at all could be an indicator that the model is not good for this purpose. As the other β_t 's are probably compensating for the β_{1t} 's not having the shape it should have. Although the β_t 's appear to not be the way they should, the one-day-ahead forecasting of the observed data does look very good, with a mean squared error, $MSE = 1.49 \times 10^{-3}$. As can be seen from the plots below, the forecasted observations follow the actual observations very closely.

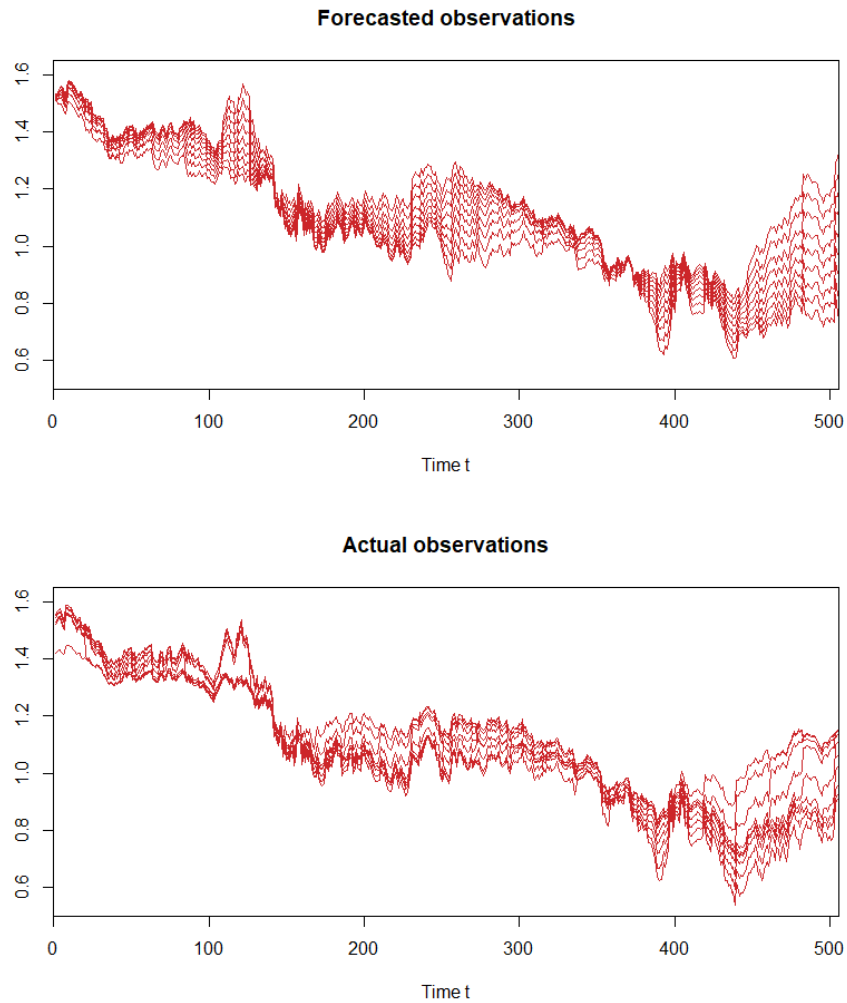


Figure 9.17: Forecasted observations for three-factor model and actual observations of natural gas data.

9.2.2 Seasonal model

The estimates of λ for the seasonal model are quite different from the three-factor model. The 95% confidence interval for the λ in the seasonal model is:

$$\lambda = (3.132 \times 10^{-3}, 3.721 \times 10^{-3})$$

while the confidence interval for σ_y is similar to the three-factor model, down to the 1st significant number:

$$\sigma_y = (3.469 \times 10^{-2}, 3.515 \times 10^{-2}).$$

The factor loadings for the seasonal model are a little different from the three-factor model:

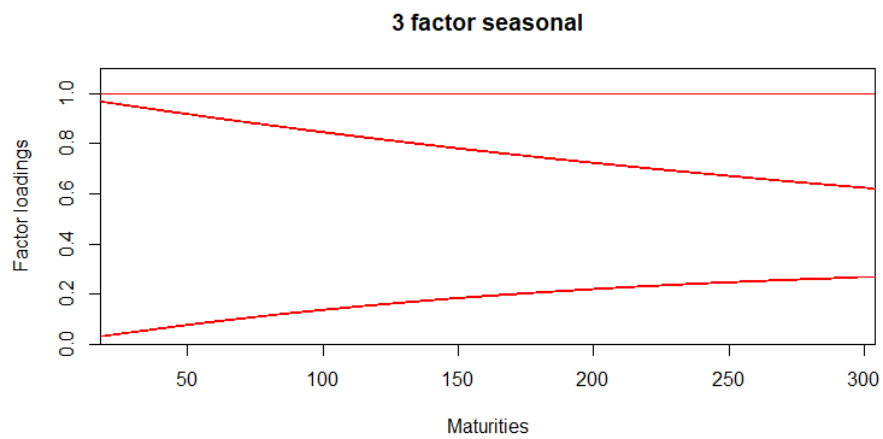


Figure 9.18: Factor loadings of the seasonal model on natural gas data

For the seasonal model we get a lot smaller values for the β_t 's. The estimates for the β_{1t} 's are:

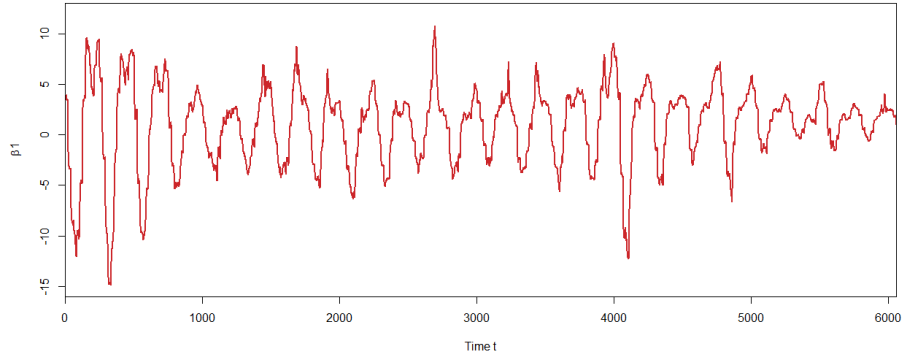


Figure 9.19: Plot of β_{1t} from time $t = 0, \dots, 6059$

For the β_{2t} 's:

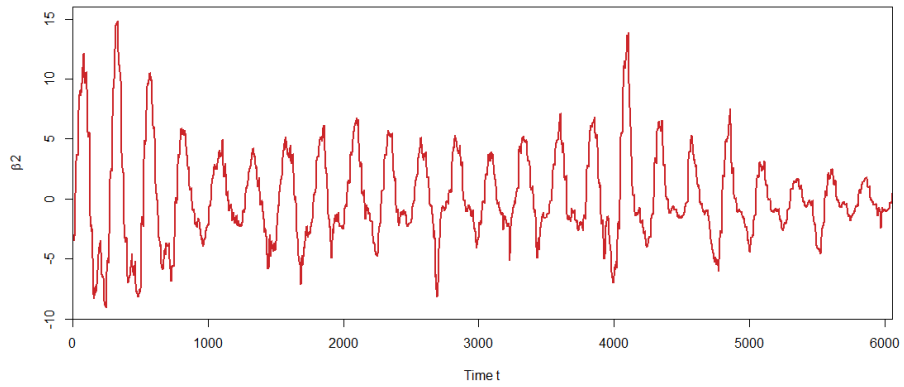


Figure 9.20: Plot of β_{2t} from time $t = 0, \dots, 6059$

And for the β_{3t} 's we have:

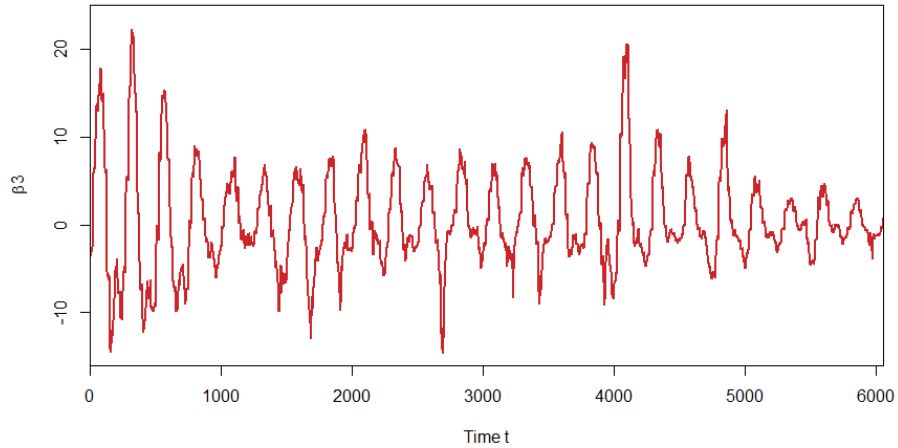


Figure 9.21: Plot of β_{3t} from time $t = 0, \dots, 6059$

Here we also have the seasonality term for the β_t 's:

Again, these estimates for β_t 's does not look to have the shape they should have, even though the end maximum and minimum values of these β_t 's are a lot smaller than for the seasonal model, thus it is difficult to draw any conclusions for these β_t 's aswell. As for the three-factor model, the β_{1t} 's and β_{2t} 's are mirror opposites of each other and the β_{3t} 's have the same shape as the β_{2t} 's, but with higher maximum and minimum values. The seasonality terms are quite small except for the seasonality on the β_{3t} 's which has the highest amplitude of the seasonality terms.

For the one-day-ahead forecasting of the observations, these are even better for the seasonal model, with an mean squared error, $MSE = 8.85 \times 10^{-4}$. Indeed a very good forecast:

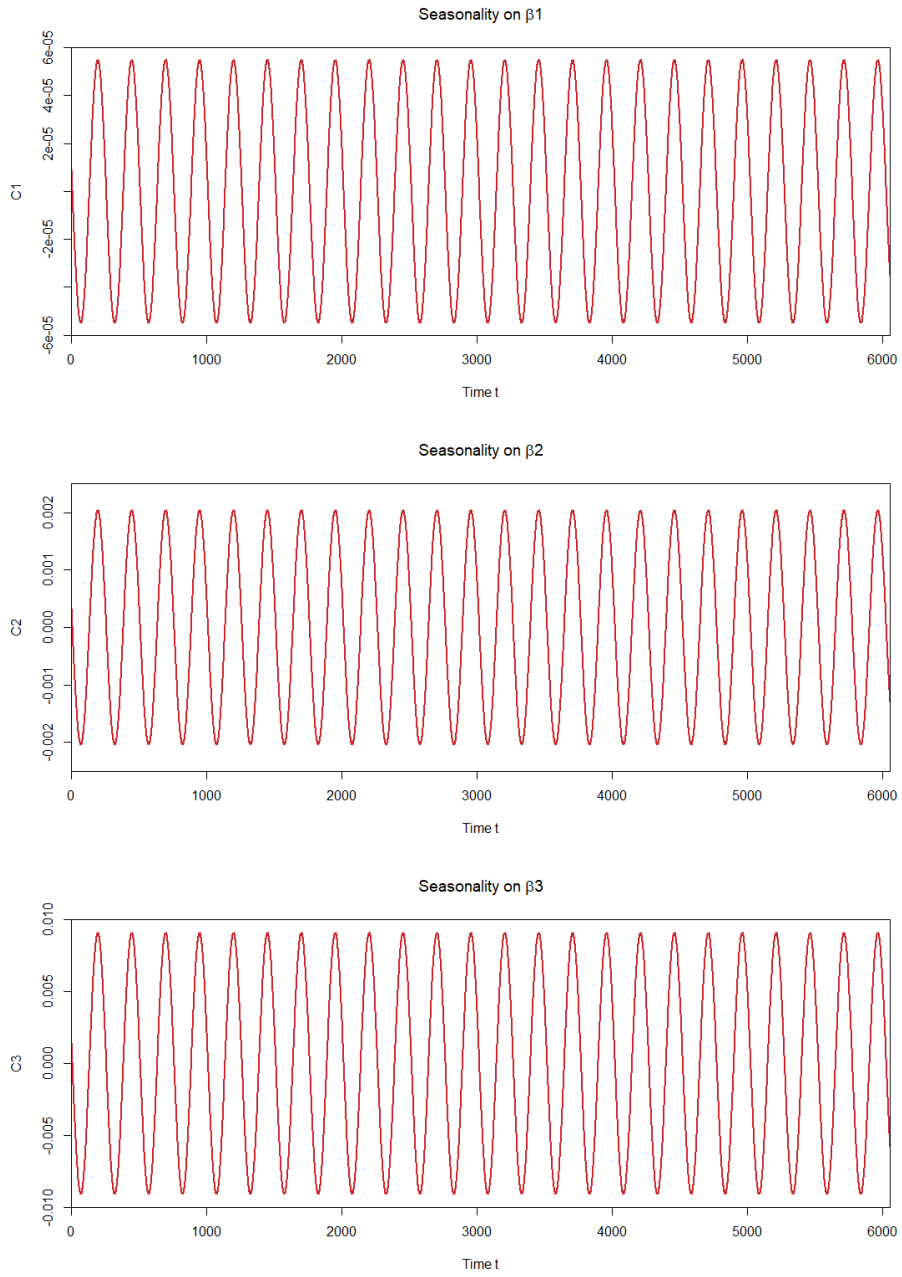


Figure 9.22: Seasonality on β_t with period $s = 253$

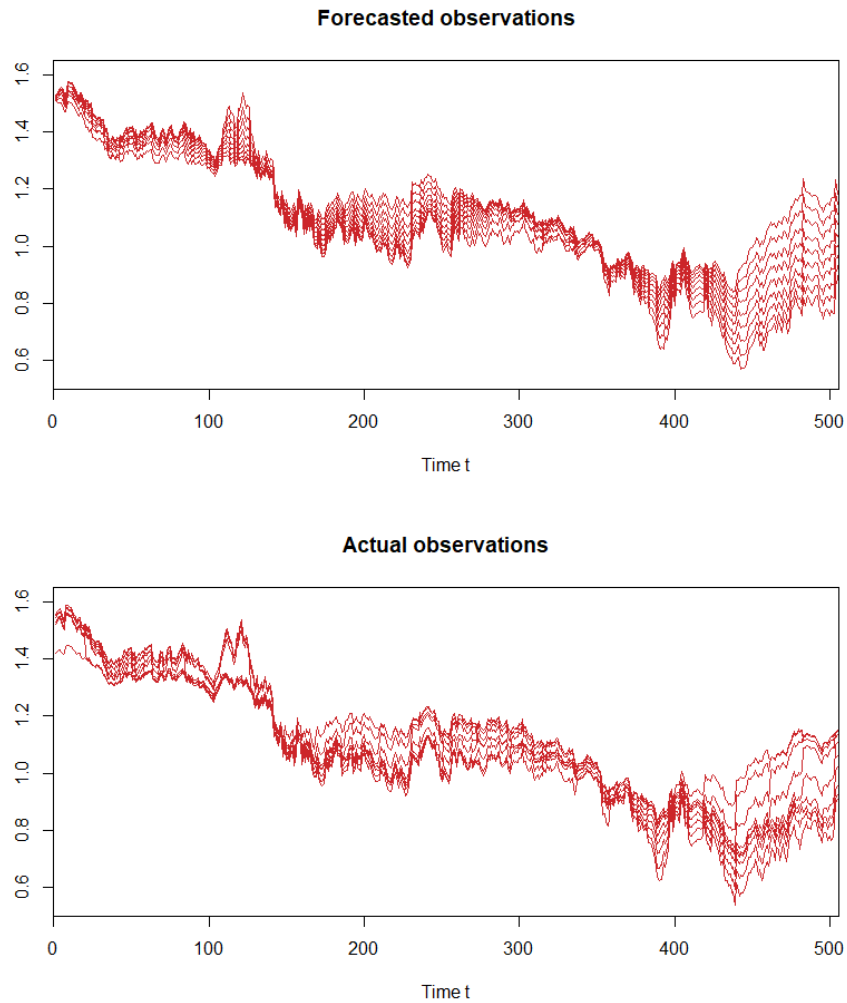


Figure 9.23: Forecasted observations for seasonal model and actual observations of natural gas data.

9.2.3 Hypothesis testing

Now we have estimated parameters and forecasted observations for the natural gas data. The next thing to do is to compare the three-factor model and the seasonal model to check if there is a significant improvement to the model if we use the seasonality term. For the three-factor model we have a

log-likelihood value of $\log(L(Y_n|\psi_0)) = 100238$, and for the seasonal model we have $\log(L(Y_n|\psi)) = 100276.1$. The test statistic is thus:

$$\begin{aligned} D &= -2\log(\Lambda) \\ &= 2 \times (\log(L(Y_n|\psi)) - \log(L(Y_n|\psi_0))) \\ &= 2 \times (100276.1 - 100238) \\ &= 76.2 \end{aligned}$$

The degrees of freedom for the test statistic is $p - q = 12 - 8 = 4$ as it was further up for the crude oil data. We have the test statistic $D = 76.2$, to see if the seasonal model has a better fit to the data we need to check if the value $D = 76.2$ is significant in a χ_4^2 -distribution. As mentioned in the hypothesis testing for the crude oil data, with a significance level of $\alpha = 0.05$ and 4 degrees of freedom, we need $D > 9.49$. We have $D = 76.2 > 9.49$, thus we can conclude that there are seasonality of significance in the dataset.

The AIC and BIC calculated for the three-factor model:

$$\begin{aligned} AIC &= -200460 \\ BIC &= -200388 \end{aligned}$$

and for the seasonal model:

$$\begin{aligned} AIC &= -200538 \\ BIC &= -200421 \end{aligned}$$

and this indicates that the seasonal model has a better fit.

9.3 Copper

9.3.1 Three-factor model

For the three-factor model, the 95% confidence interval for λ and σ_y are:

$$\lambda = (6.32 \times 10^{-3}, 6.49 \times 10^{-3})$$

$$\sigma_y = (2.16 \times 10^{-3}, 2.19 \times 10^{-3})$$

The factor loadings for the three-factor model are:

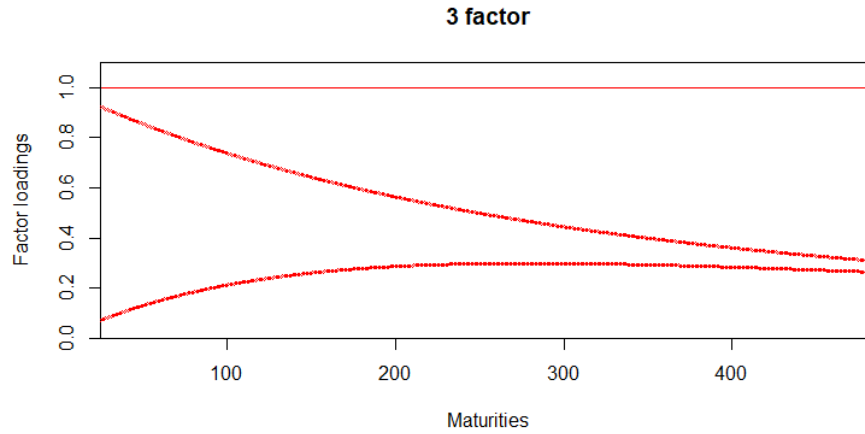


Figure 9.24: Factor loadings of three-factor model on the copper data

The results from estimating the β_{1t} 's, β_{2t} 's and β_{3t} 's for the three-factor model are as follows, for the β_{1t} 's:

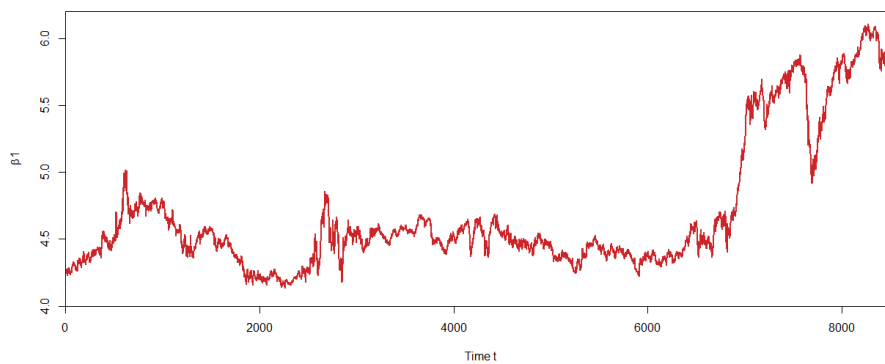


Figure 9.25: Plot of β_{1t} from time $t = 0, \dots, 8562$

For the β_{2t} 's:

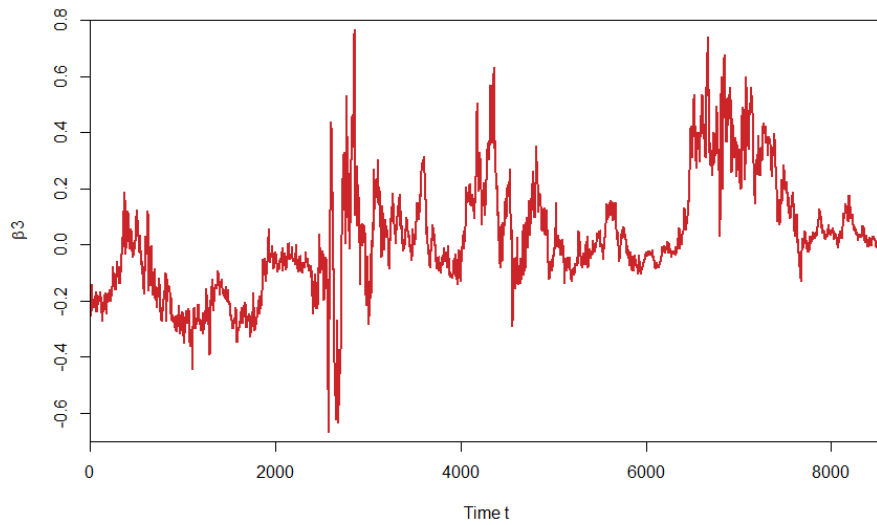


Figure 9.27: Plot of β_{3t} from time $t = 0, \dots, 8562$

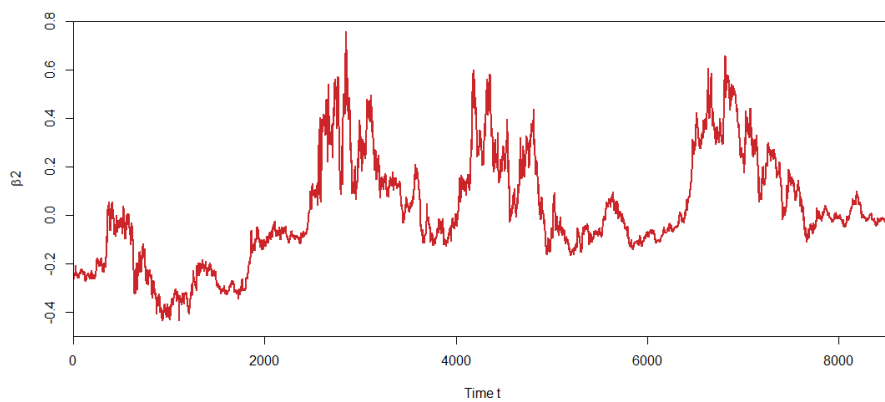


Figure 9.26: Plot of β_{2t} from time $t = 0, \dots, 8562$

And for the β_{3t} 's:

The values for the β_{1t} 's are predominantly varying around 4.5 over the sample size, but at end of the sample the β_{1t} 's increase, this is an indicator that

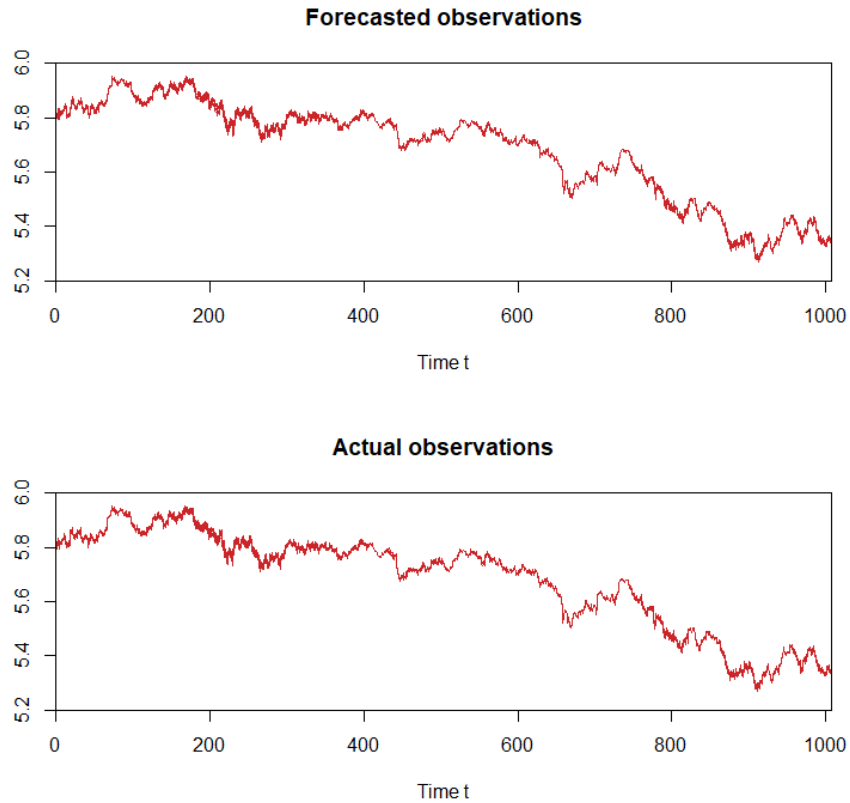


Figure 9.28: Forecasted observations of three-factor model and actual observations of copper data

the prices in general have been steady over time before going up in the end of the sample size. The β_{2t} 's start out as mainly negative for a while before becoming predominantly positive in the latter 3/4th's of the sample size. The β_{3t} 's changes from negative to positive and back several times in the beginning before becoming mainly positive in the latter half of the sample.

The forecasted observations looks very good with an $MSE = 1.47 \times 10^{-4}$ the one-day-ahead forecasted results are:

9.3.2 Seasonal model

The 95% confidence interval for λ and σ_y for the seasonal model are:

$$\lambda = (6.33 \times 10^{-3}, 6.49 \times 10^{-3})$$

$$\sigma_y = (2.15 \times 10^{-3}, 2.19 \times 10^{-3})$$

The factor loadings for the seasonal model are:

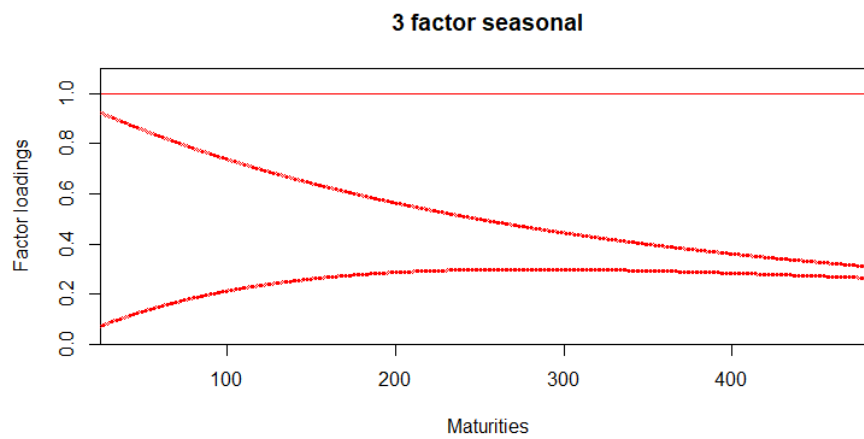


Figure 9.29: Factor loadings of seasonal model on the copper data

The estimates of β_{1t} 's are:

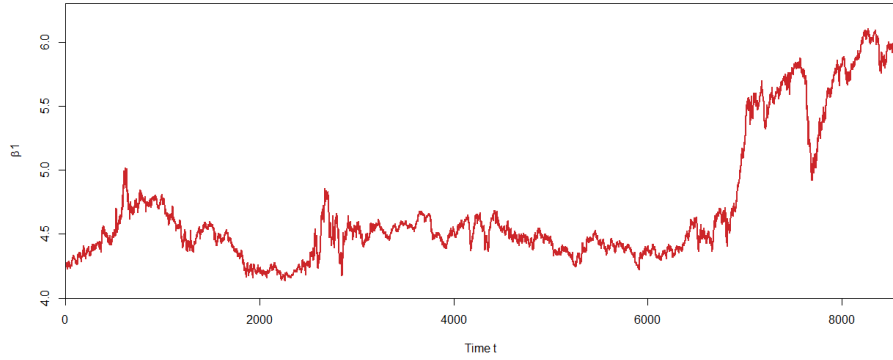


Figure 9.30: Plot of β_{1t} from time $t = 0, \dots, 8562$

Estimates of β_{2t} 's:

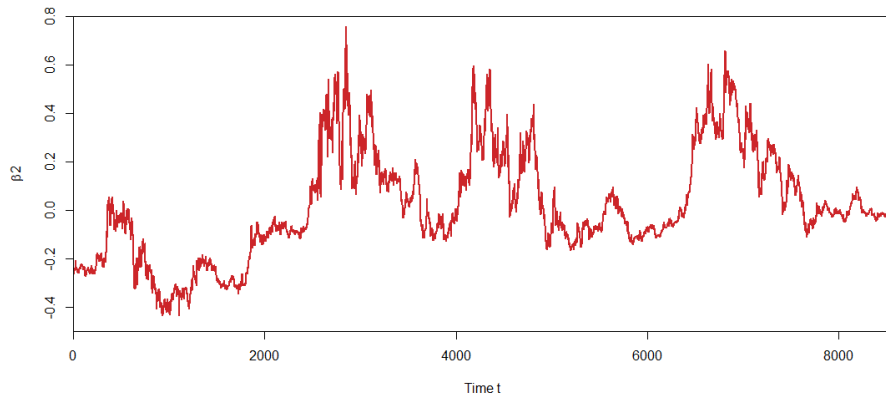


Figure 9.31: Plot of β_{2t} from time $t = 0, \dots, 8562$

And of β_{3t} 's:

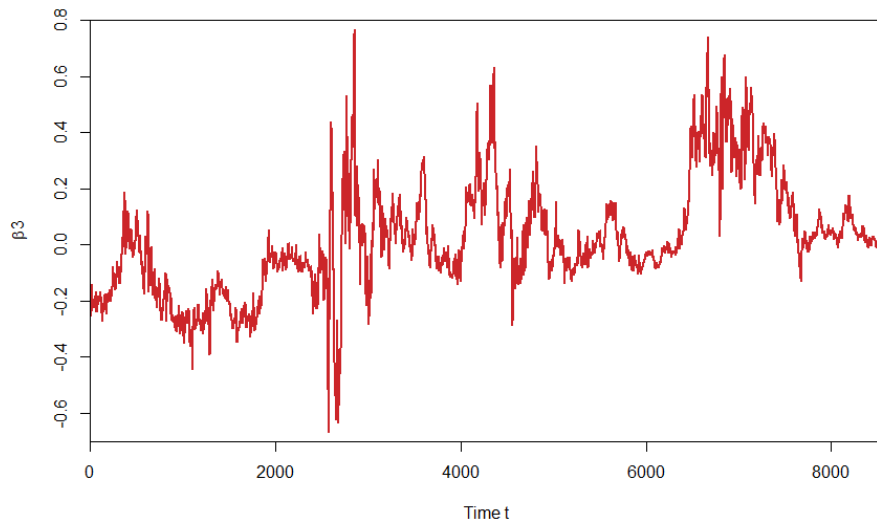


Figure 9.32: Plot of β_{3t} from time $t = 0, \dots, 8562$

The seasonality on the β_t 's are also estimated:

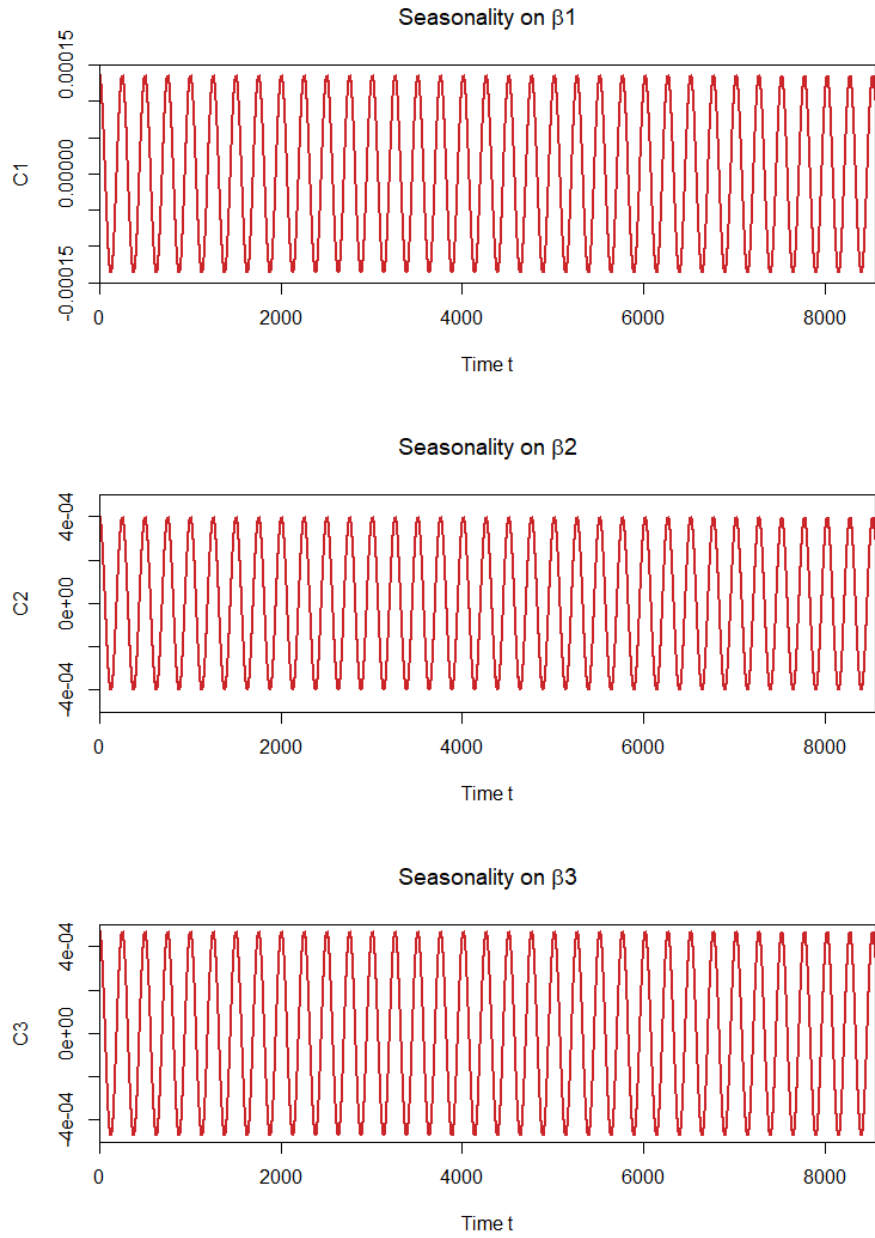


Figure 9.33: Seasonality on β_t for period $s = 253$

The β_t 's for the seasonal model behaves much the same way as for the three-

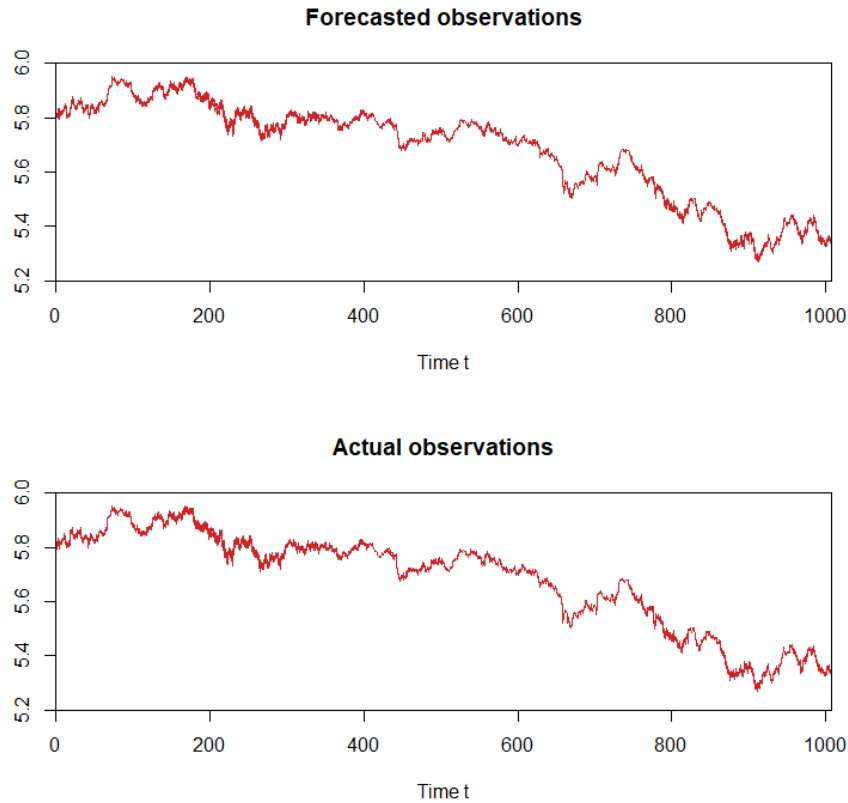


Figure 9.34: Forecasted observations of seasonal model and actual observations of the copper data

factor model, where the β_{1t} 's vary around 4.5 before increasing at the very end. The β_{2t} 's are for the most part negative to begin with and then mainly positive at the latter 3/4th's of the sample, while the β_{3t} 's changes sign several times in the beginning of the sample size and getting mainly positive in the latter half. As we can see, the seasonality terms are very small, with the seasonality on the β_{3t} 's slightly having the biggest amplitude.

The one-day-ahead forecasting is very good for the seasonal model as well as for the three-factor model. The $MSE = 1.47 \times 10^{-4}$ of the seasonal model is the same as for the three-factor model. The forecasted observations are:

9.3.3 Hypothesis testing

For the three-factor model we have a log-likelihood value of $\log(L(Y_n|\psi_0)) = 207409.9$ and the seasonal model have a log-likelihood value of $\log(L(Y_n|\psi)) = 207412.2$. Comparing the three-factor model and the seasonal model to check if there is significant improvement by adding the seasonality term, we get the test statistic:

$$\begin{aligned} D &= -2 \log(\Lambda) \\ &= 2 \times (\log(L(Y_n|\psi)) - \log(L(Y_n|\psi_0))) \\ &= 2 \times (207412.2 - 207409.9) \\ &= 4.6 \end{aligned}$$

The degrees of freedom for the test statistic is $p - q = 12 - 8 = 4$. We have the test statistic $D = 4.6$, to see if the seasonal model has a better fit to the data we need to check if the value $D = 4.6$ is significant in a χ_4^2 -distribution. As mentioned in the hypothesis testing for the crude oil data, with a significance level of $\alpha = 0.05$ and 4 degrees of freedom, we need $D > 9.49$. We have $D = 4.6 < 9.49$, thus we cannot reject the null hypothesis.

The AIC and BIC calculated for the three-factor model:

$$AIC = -414804$$

$$BIC = -414733$$

and for the seasonal model:

$$AIC = -414800$$

$$BIC = -414694$$

and this seems to indicate that there is no improvement by adding the seasonality term to the model.

9.4 Cotton

9.4.1 Three-factor model

The results from estimating λ and σ_y of the cotton futures contracts are as follows, the 95% confidence interval for λ :

$$\lambda = (6.897 \times 10^{-3}, 7.403 \times 10^{-3})$$

for σ_y :

$$\sigma_y = (1.460 \times 10^{-2}, 1.478 \times 10^{-2})$$

and the factor loadings for the three-factor model are:

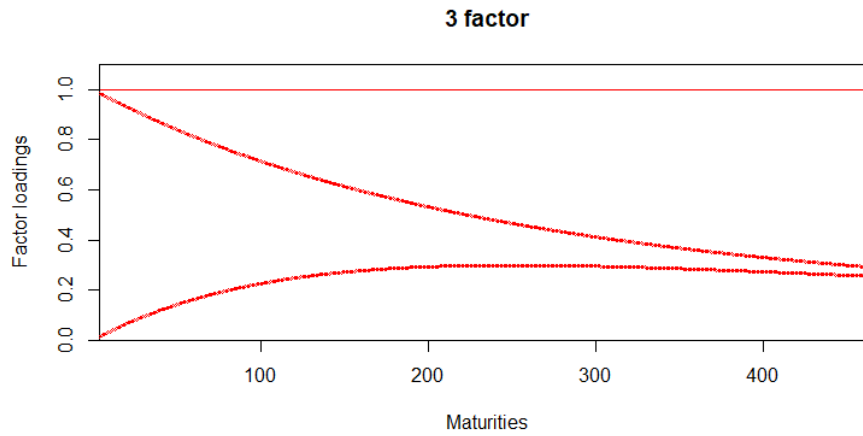


Figure 9.35: Factor loadings of three-factor model on cotton data

The estimates for the β_{1t} 's are:

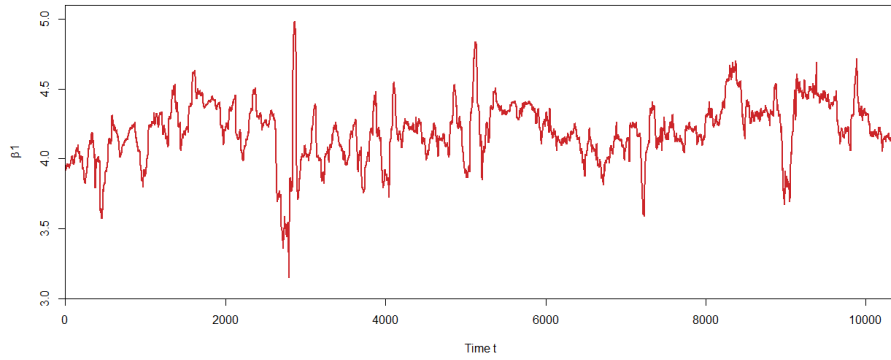


Figure 9.36: Plot of β_{1t} from time $t = 0, \dots, 10380$

For the β_{2t} 's:

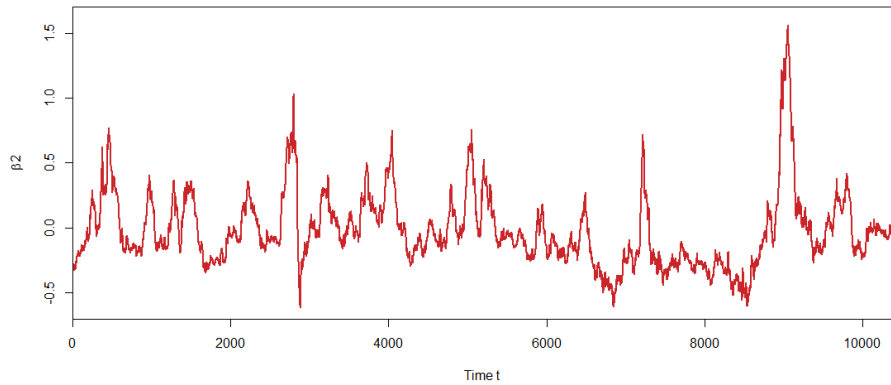


Figure 9.37: Plot of β_{2t} from time $t = 0, \dots, 10380$

And for the β_{3t} 's:

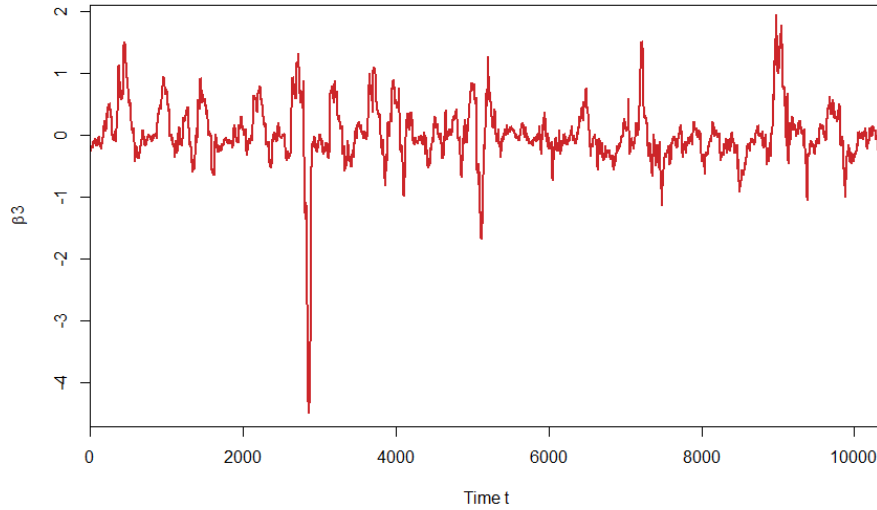


Figure 9.38: Plot of β_{3t} from time $t = 0, \dots, 10380$

As with the β_{1t} 's for the natural gas data the plot for the β_{1t} 's are not how the plot for the β_{1t} 's should look like, hence it is difficult to draw any conclusions from the plot of either of the β_t 's. The estimates for β_{1t} should look like the first column in our dataset due to the β_{1t} 's governing the level of the yield curve. It is an indicator of how the prices varies in time, if they go up or down. From the estimates of the β_{1t} 's it is impossible to tell if the yield curve is going up or down. As stated earlier, the estimates for β_{1t} should look like the first column in our dataset. The plot of β_{1t} should look like this:

The fact that it does not have this shape at all could be an indicator that the model is not very good for this purpose. As the other β_t 's are probably compensating for the β_{1t} 's not having the shape it should have. Although the β_t 's appear to be not the way they should, the one-day-ahead forecasting of the observed data does look very good, with a mean squared error, $MSE = 2.14 \times 10^{-4}$. As can be seen from the plots below, the forecasted observations follow the actual observations very closely.

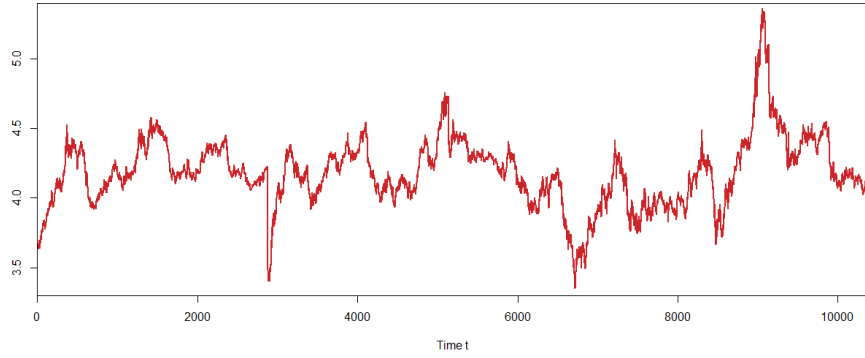


Figure 9.39: Plot of first column in Cotton dataset

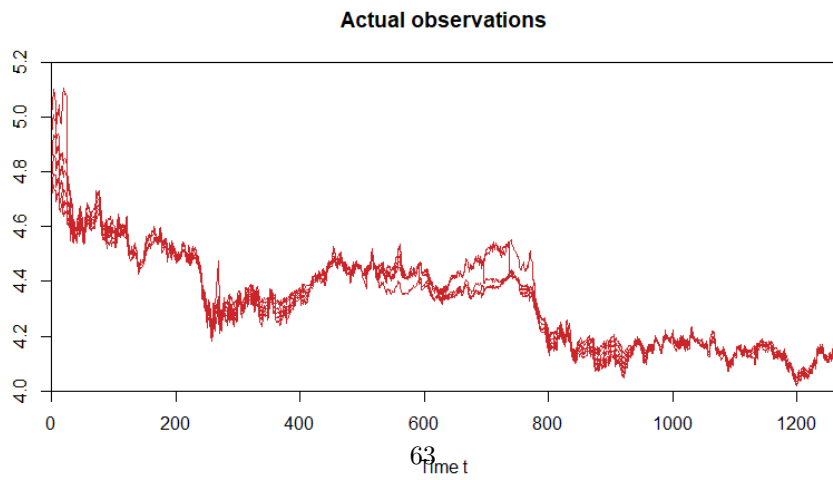
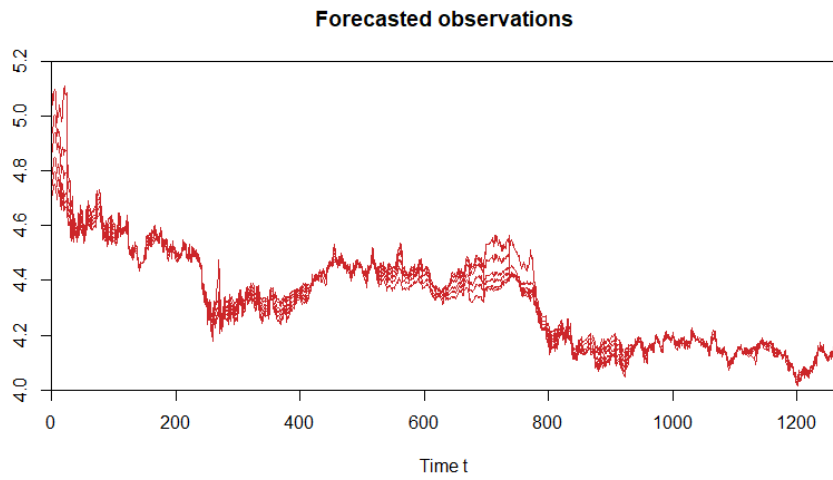


Figure 9.40: Forecasted observations for three-factor model and actual observations of cotton data.

9.4.2 Seasonal model

The 95% confidence interval for the λ in the seasonal model is:

$$\lambda = (7.116 \times 10^{-3}, 7.622 \times 10^{-3})$$

the confidence interval for σ_y is the same as for the three-factor model:

$$\sigma_y = (1.460 \times 10^{-2}, 1.478 \times 10^{-2}).$$

The factor loadings for the seasonal model:

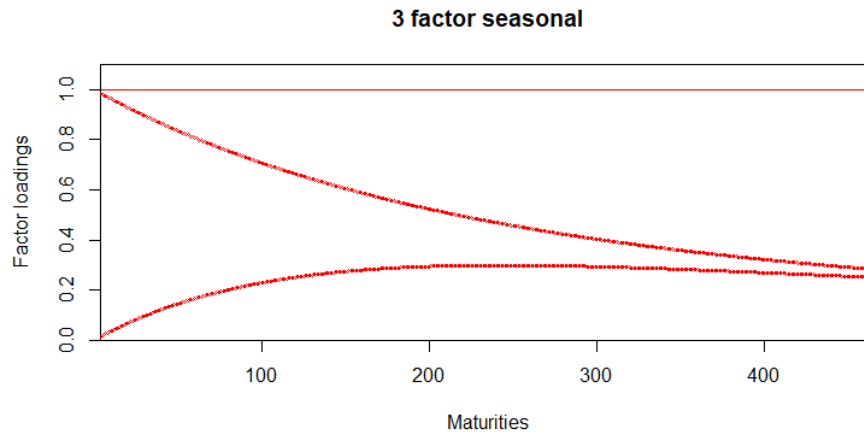


Figure 9.41: Factor loadings of the seasonal model on cotton data

For the seasonal model the estimates for the β_{1t} 's are:

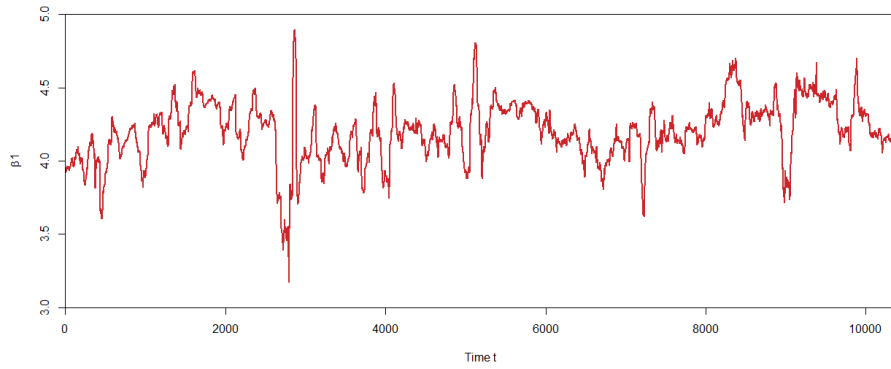


Figure 9.42: Plot of β_{1t} from time $t = 0, \dots, 10380$

For the β_{2t} 's:

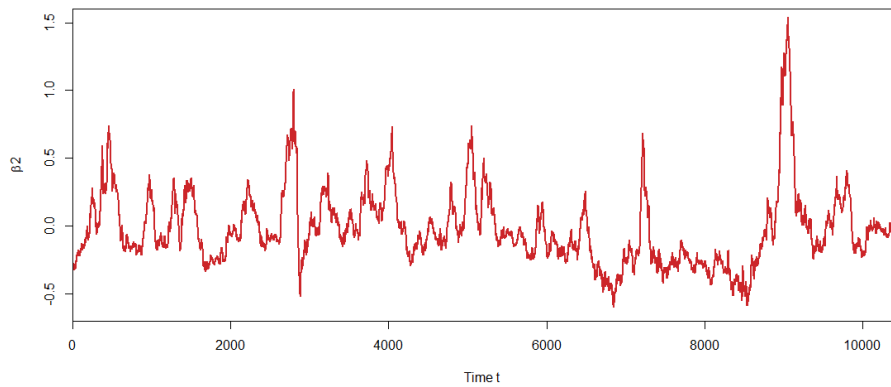


Figure 9.43: Plot of β_{2t} from time $t = 0, \dots, 10380$

And for the β_{3t} 's we have:

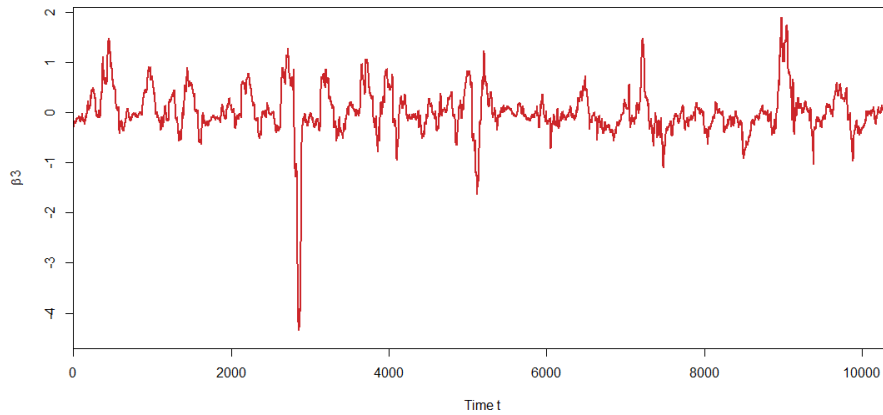


Figure 9.44: Plot of β_{3t} from time $t = 0, \dots, 10380$

Here we also have the seasonality term for the β_t 's:

Again, these estimates for β_t 's does not look to have the shape they should have, thus it is difficult to draw any conclusions for these β_t 's as well. The seasonality on the β_{1t} 's has the highest amplitude of the seasonality terms.

For the one-day-ahead forecasting of the observations, these are good, with an mean squared error, $MSE = 2.14 \times 10^{-4}$, the same as for the three-factor model. Indeed a good forecast:

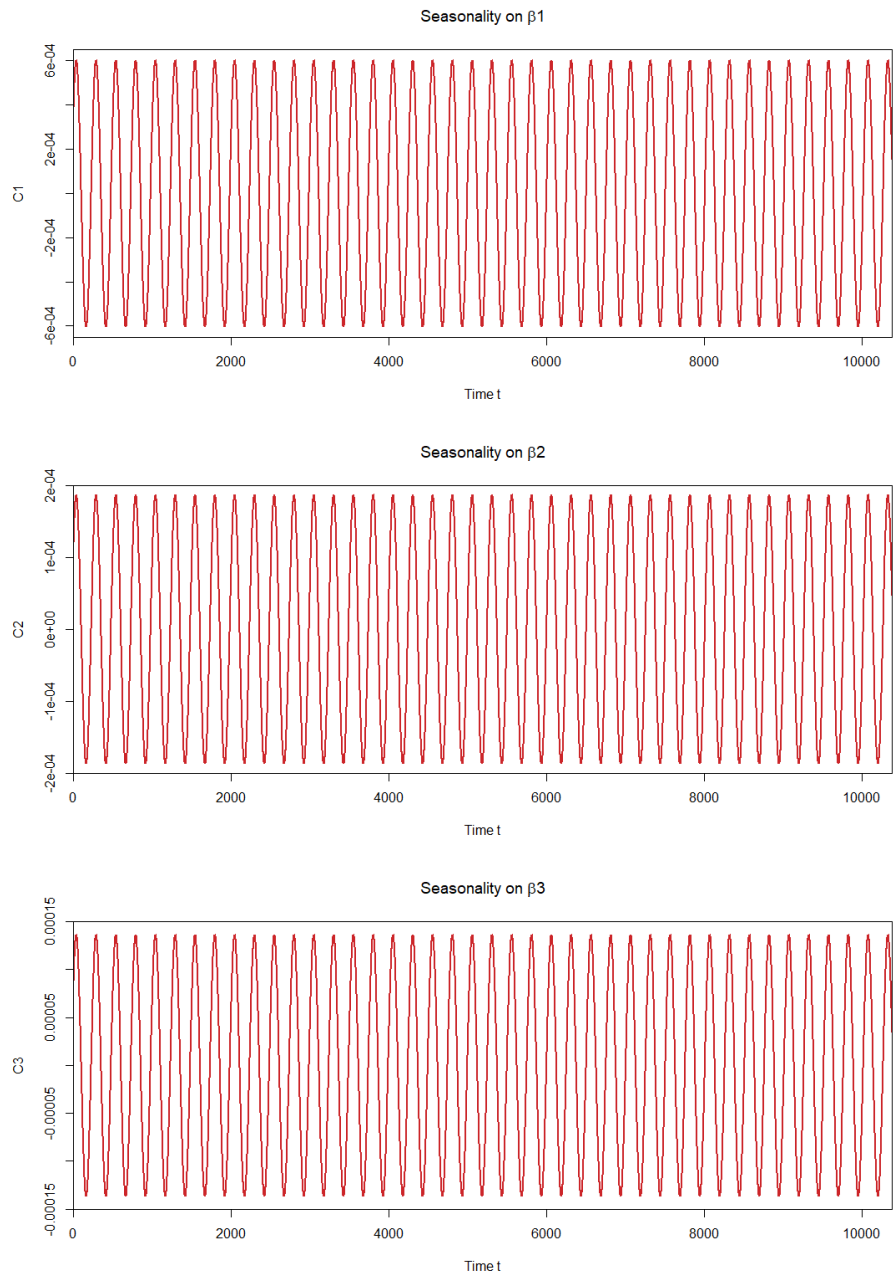


Figure 9.45: Seasonality on β_t with period $s = 253$

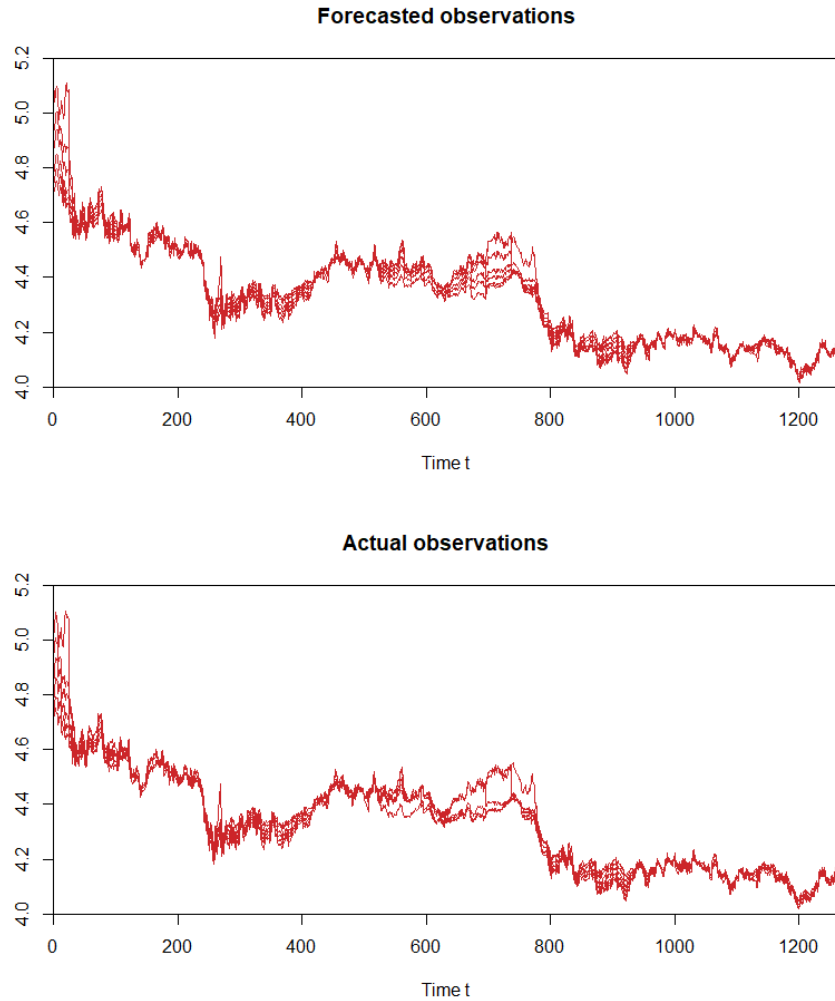


Figure 9.46: Forecasted observations for seasonal model and actual observations of cotton data.

9.4.3 Hypothesis testing

For the three-factor model we have a log-likelihood value of $\log(L(Y_n|\psi_0)) = 162514.3$ and the seasonal model have a log-likelihood value of $\log(L(Y_n|\psi)) = 162520.2$. Comparing the three-factor model and the seasonal model to check if there is significant improvement by adding the seasonality term, we get the test

statistic:

$$\begin{aligned} D &= -2 \log(\Lambda) \\ &= 2 \times (\log(L(Y_n|\psi)) - \log(L(Y_n|\psi_0))) \\ &= 2 \times (162520.2 - 162514.3) \\ &= 11.8 \end{aligned}$$

The degrees of freedom for the test statistic is $p - q = 12 - 8 = 4$. We have the test statistic $D = 11.8$, to see if the seasonal model has a better fit to the data we need to check if the value $D = 11.8$ is significant in a χ_4^2 -distribution. As mentioned in the hypothesis testing for the crude oil data, with a significance level of $\alpha = 0.05$ and 4 degrees of freedom, we need $D > 9.49$. We have $D = 11.8 > 9.49$, thus we can reject the null hypothesis and conclude that there is seasonality in the cotton data.

The AIC and BIC calculated for the three-factor model:

$$\begin{aligned} AIC &= -325012 \\ BIC &= -324940 \end{aligned}$$

and for the seasonal model:

$$\begin{aligned} AIC &= -325016 \\ BIC &= -324907 \end{aligned}$$

Here we have a bit of a disagreement, the AIC seems to favour the seasonal model while the BIC seems to favour the three-factor model. This is probably due to the BIC giving a higher penalty to more complex models hence the BIC is more likely to choose a simpler model than the AIC. On the grounds of $D = 11.8$ which gives a $p - value = 1.89 \times 10^{-2}$ we choose to agree with the AIC and choose the seasonal model.

9.5 Gold

9.5.1 Three-factor model

For the three-factor model, the 95% confidence interval for λ and σ_y are:

$$\lambda = (1.55 \times 10^{-3}, 1.73 \times 10^{-3})$$

$$\sigma_y = (4.79 \times 10^{-4}, 4.86 \times 10^{-4})$$

The factor loadings for the three-factor model are:

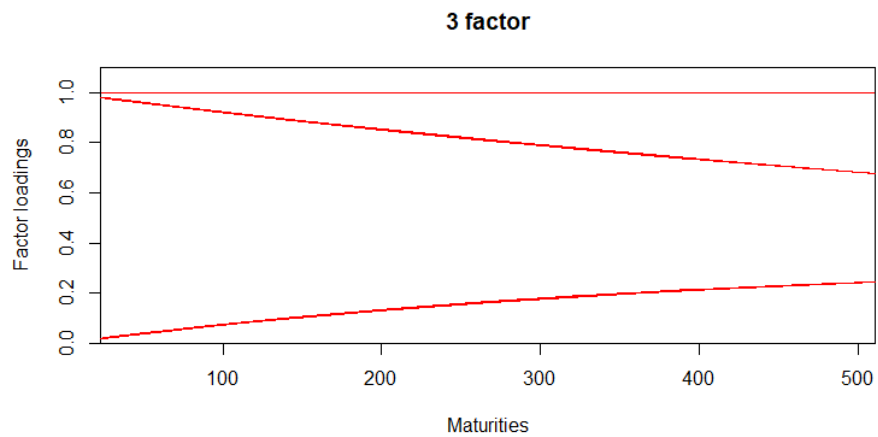


Figure 9.47: Factor loadings of three-factor model on the gold data

The results from estimating the β_{1t} 's, β_{2t} 's and β_{3t} 's for the three-factor model are as follows, for the β_{1t} 's:

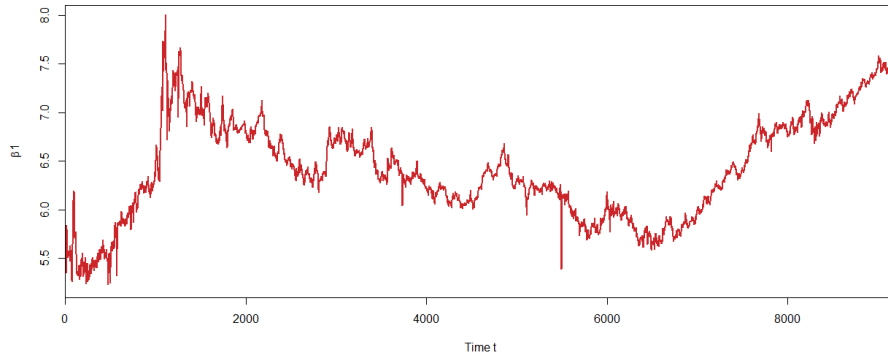


Figure 9.48: Plot of β_{1t} from time $t = 0, \dots, 9205$

For the β_{2t} 's:

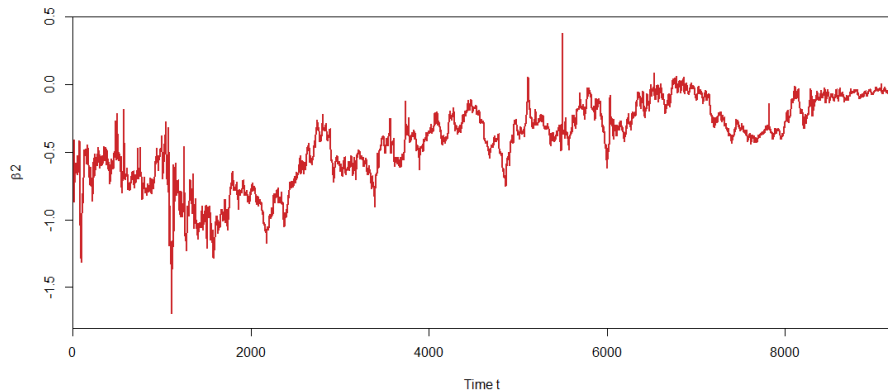


Figure 9.49: Plot of β_{2t} from time $t = 0, \dots, 9205$

And for the β_{3t} 's:

The values for the β_{1t} 's are increasing in the beginning, decreasing in the middle part of the sample before it increases again in the end, this is an indicator that the prices in general have varied over time, but overall the prices has increased over time. The β_{2t} 's are predominantly negative throughout the sample size. The β_{3t} 's are also mainly negative, but with several times where

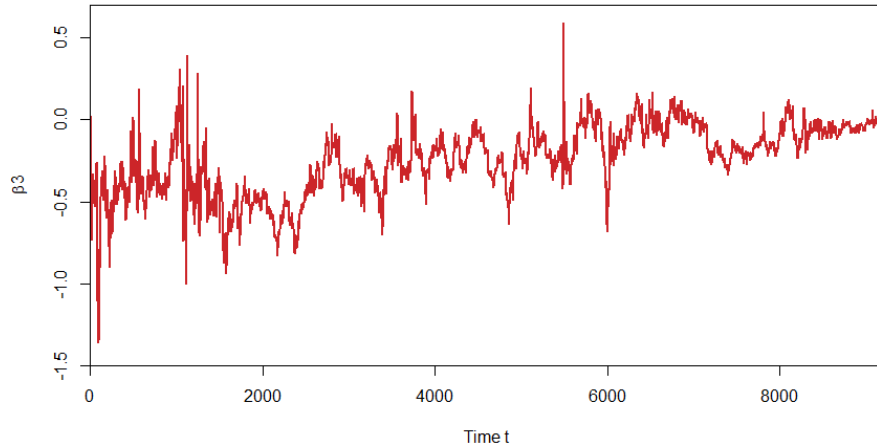


Figure 9.50: Plot of β_{3t} from time $t = 0, \dots, 9205$

the β_{3t} 's become positive.

The forecasted observations looks very good with an $MSE = 1.19 \times 10^{-4}$
the one-day-ahead forecasted results are:

9.5.2 Seasonal model

The 95% confidence interval for λ and σ_y for the seasonal model are:

$$\lambda = (1.41 \times 10^{-3}, 1.46 \times 10^{-3})$$

$$\sigma_y = (4.81 \times 10^{-4}, 4.87 \times 10^{-4})$$

The factor loadings for the seasonal model are:

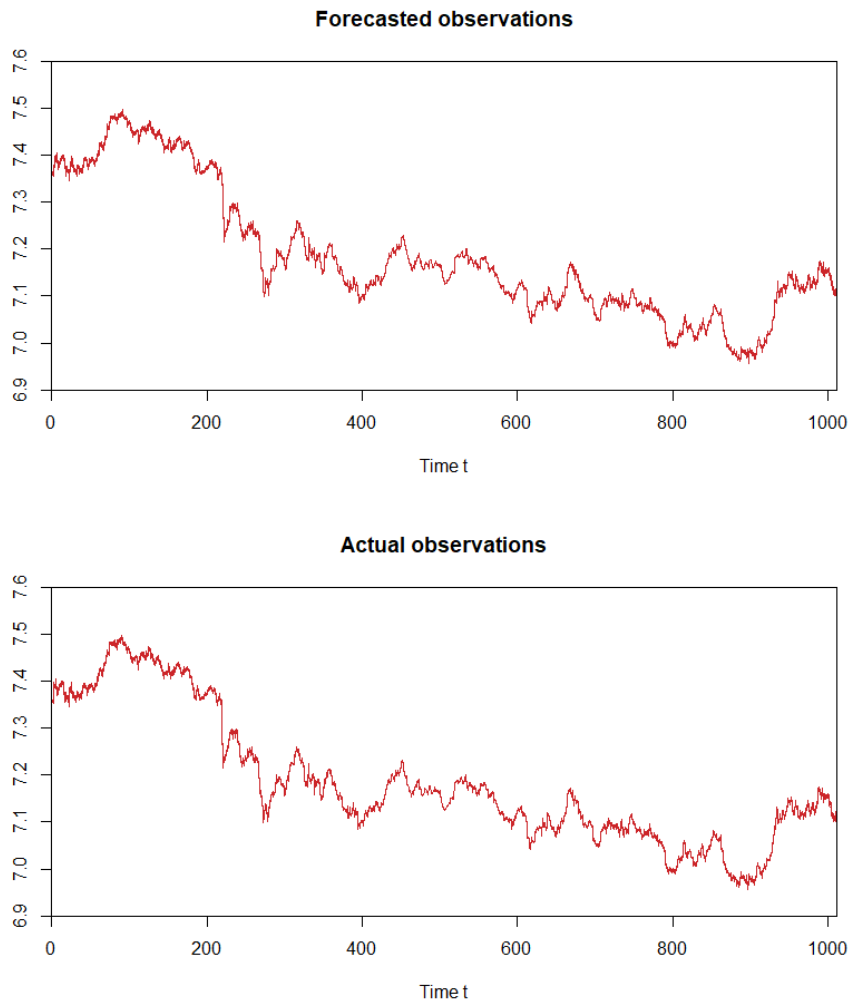


Figure 9.51: Forecasted observations of three-factor model and actual observations of gold data

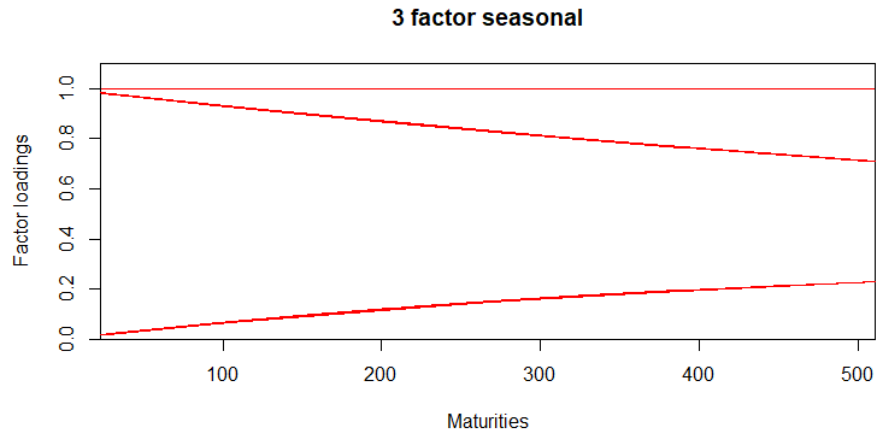


Figure 9.52: Factor loadings of seasonal model on the gold data

The estimates of β_{1t} 's are:

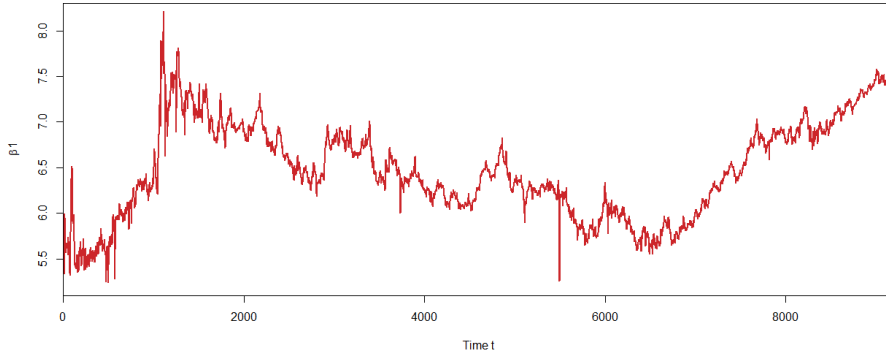


Figure 9.53: Plot of β_{1t} from time $t = 0, \dots, 9205$

Estimates of β_{2t} 's:

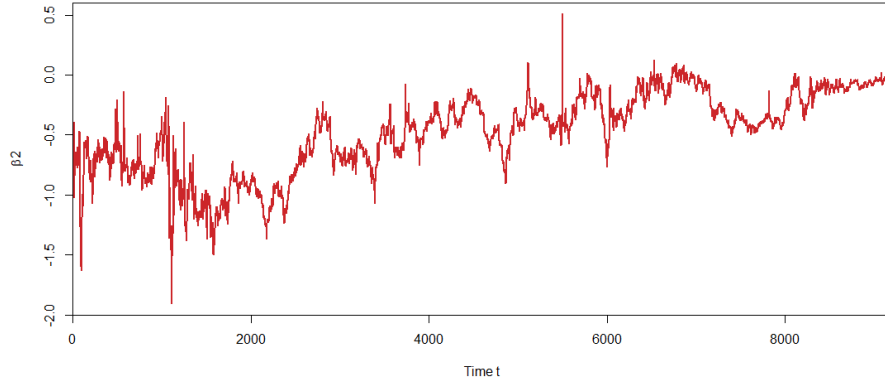


Figure 9.54: Plot of β_{2t} from time $t = 0, \dots, 9205$

And of β_{3t} 's:

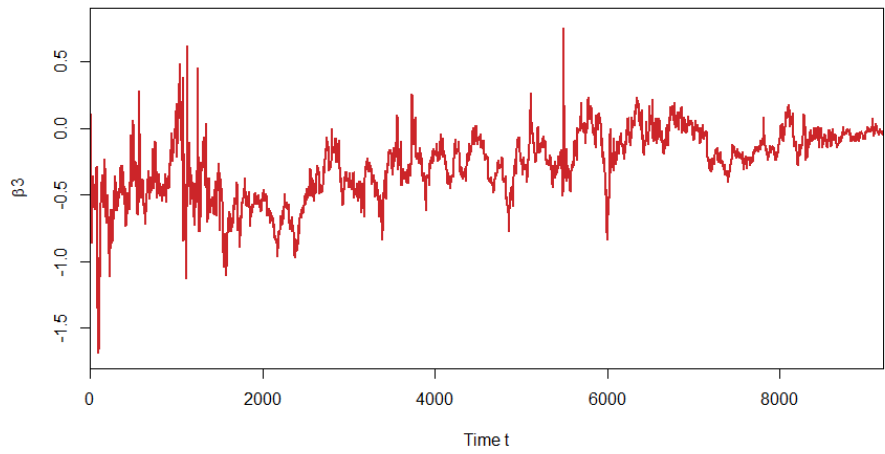


Figure 9.55: Plot of β_{3t} from time $t = 0, \dots, 9205$

The seasonality on the β_t 's are also estimated:

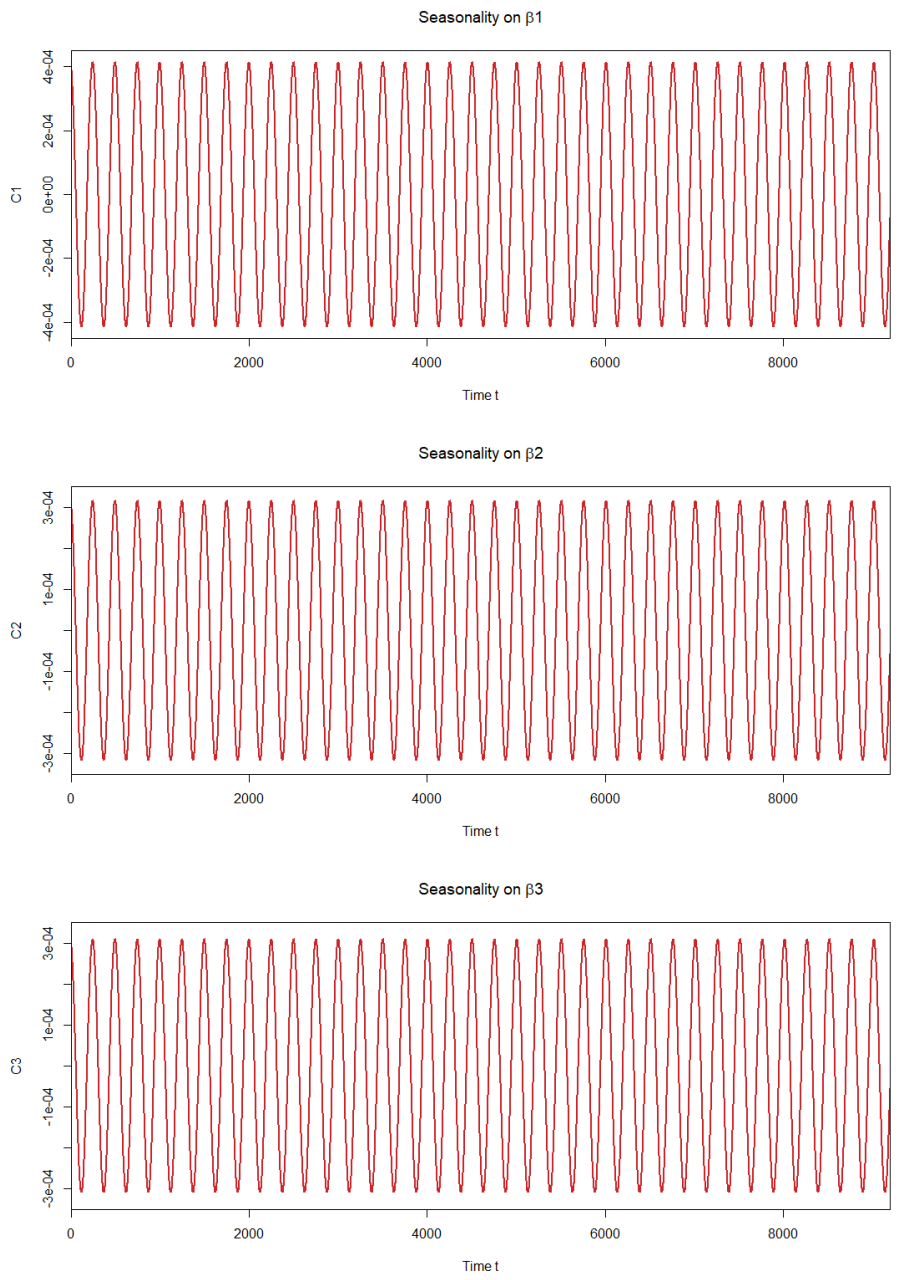


Figure 9.56: Seasonality on β_t for period $s = 253$

The β_t 's for the seasonal model behaves much the same way as for the three-

factor model, where the β_{1t} 's increases a lot in the beginning, then it decreases slowly over the middle part of the sample before it increases again in the end part. The β_{2t} 's are predominantly negative throughout the sample size, while the β_{3t} 's are also mainly negative, but with several times where the β_{3t} 's become positive. As we can see, the seasonality terms are small, with the seasonality on the β_{1t} 's having the biggest amplitude.

The one-day-ahead forecasting is very good for the seasonal model as well as for the three-factor model. The $MSE = 1.19 \times 10^{-4}$ of the seasonal model is the same as for the three-factor model. The forecasted observations are:

9.5.3 Hypothesis testing

For the three-factor model we have a log-likelihood value of $\log(L(Y_n|\psi_0)) = 293513.2$ and the seasonal model have a log-likelihood value of $\log(L(Y_n|\psi)) = 293513.6$. Comparing the three-factor model and the seasonal model to check if there is significant improvement by adding the seasonality term, we get the test statistic:

$$\begin{aligned} D &= -2 \log(\Lambda) \\ &= 2 \times (\log(L(Y_n|\psi)) - \log(L(Y_n|\psi_0))) \\ &= 2 \times (293513.6 - 293513.2) \\ &= 0.8 \end{aligned}$$

The degrees of freedom for the test statistic is $p - q = 12 - 8 = 4$. We have the test statistic $D = 0.8$, to see if the seasonal model has a better fit to the data we need to check if the value $D = 0.8$ is significant in a χ_4^2 -distribution. As mentioned in the hypothesis testing for the crude oil data, with a significance level of $\alpha = 0.05$ and 4 degrees of freedom, we need $D > 9.49$. We have $D = 0.8 < 9.49$, thus we cannot reject the null hypothesis.

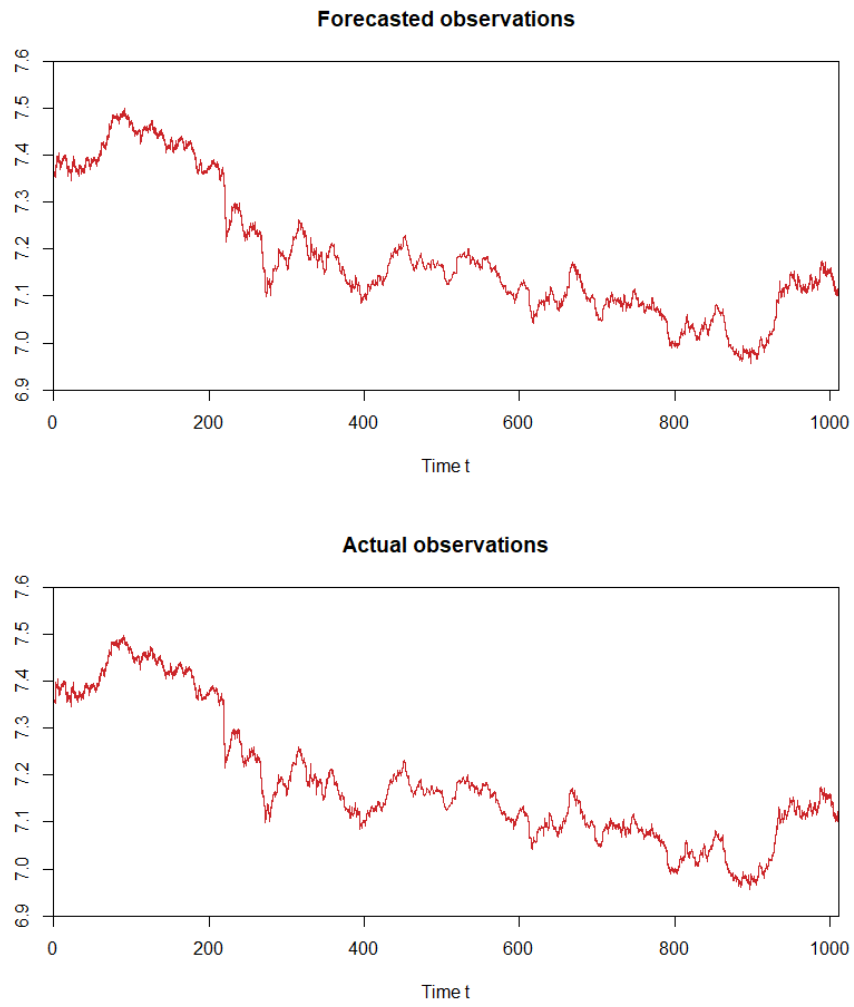


Figure 9.57: Forecasted observations of seasonal model and actual observations of the gold data

The AIC and BIC calculated for the three-factor model:

$$AIC = -587010$$

$$BIC = -586939$$

and for the seasonal model:

$$AIC = -587003$$

$$BIC = -586896$$

this seems to indicate that there is no improvement by adding the seasonality term to the model.

9.6 Sugar

9.6.1 Three-factor model

For the three-factor model, the 95% confidence interval for λ and σ_y are:

$$\lambda = (9.53 \times 10^{-3}, 9.87 \times 10^{-3})$$

$$\sigma_y = (9.838 \times 10^{-3}, 9.995 \times 10^{-4})$$

The factor loadings for the three-factor model are:

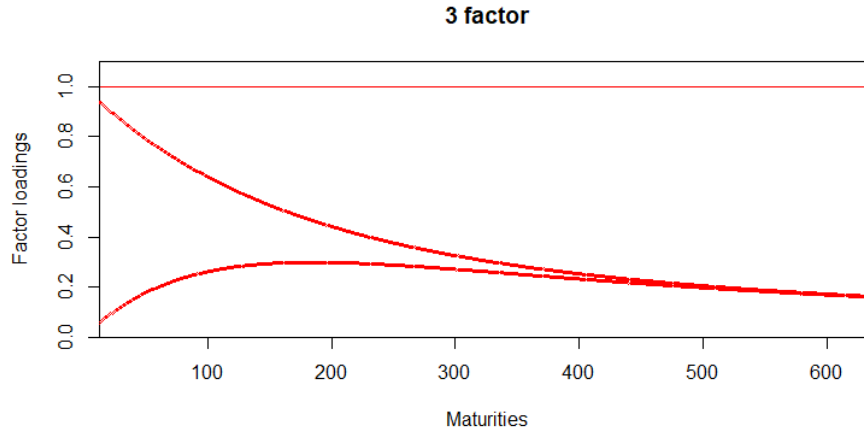


Figure 9.58: Factor loadings of three-factor model on the sugar data

The results from estimating the β_{1t} 's, β_{2t} 's and β_{3t} 's for the three-factor model are as follows, for the β_{1t} 's:

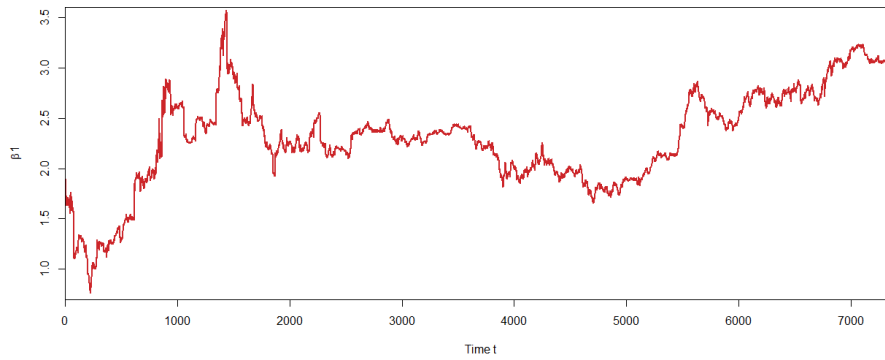


Figure 9.59: Plot of β_{1t} from time $t = 0, \dots, 7407$

For the β_{2t} 's:

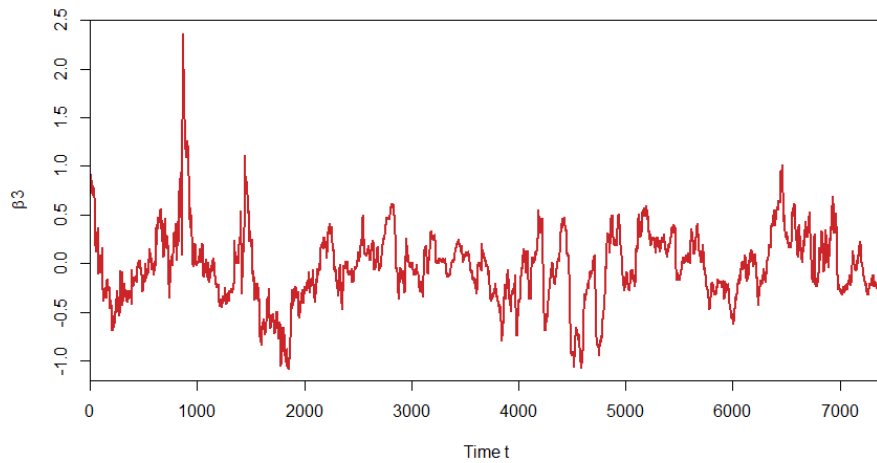


Figure 9.61: Plot of β_{3t} from time $t = 0, \dots, 7407$

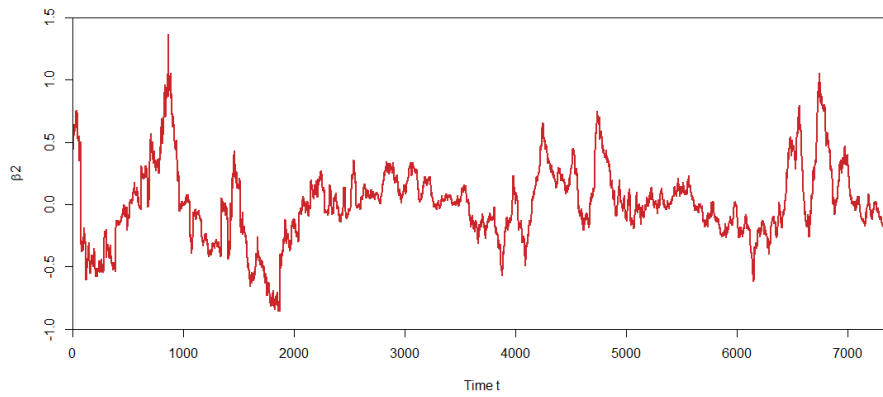


Figure 9.60: Plot of β_{2t} from time $t = 0, \dots, 7407$

And for the β_{3t} 's:

The values for the β_{1t} 's are for the most part increasing in the beginning, decreasing a little and being stable in the middle part of the sample before it increases again in the latter part of the sample, this is an indicator that the prices increased for the first few years, then had a several years where the price

decreased or was stable before it had several years where the price increased again. The β_{2t} 's are changing sign several times over time, and the same goes for the β_{3t} 's.

The variance matrix, F_t , for the prediction error, v_t , is singular, meaning the determinant of the matrix is equal to zero, $\det(F_t) = 0$. This means that the matrix F_t does not have an inverse, hence it is not possible, using the kalman filter on the three-factor model, to forecast the observations of the sugar data. Thus we move on to the seasonal model.

9.6.2 Seasonal model

The 95% confidence interval for λ and σ_y for the seasonal model are:

$$\lambda = (9.49 \times 10^{-3}, 9.84 \times 10^{-3})$$

$$\sigma_y = (9.835 \times 10^{-3}, 9.992 \times 10^{-3})$$

The factor loadings for the seasonal model are:

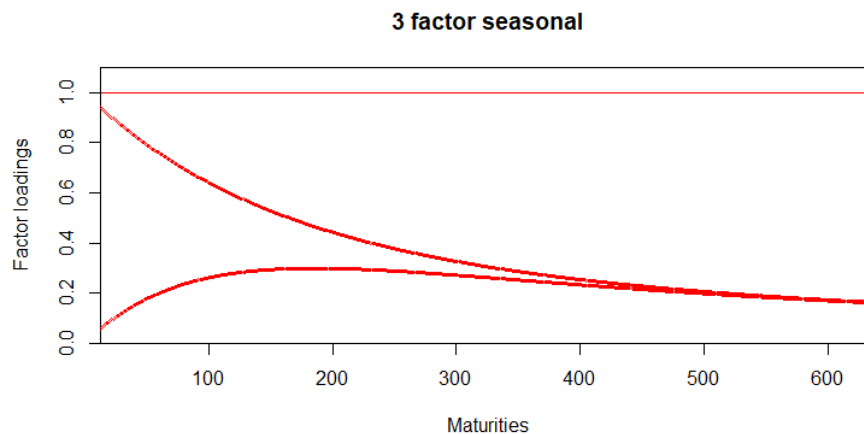


Figure 9.62: Factor loadings of seasonal model on the sugar data

The estimates of β_{1t} 's are:

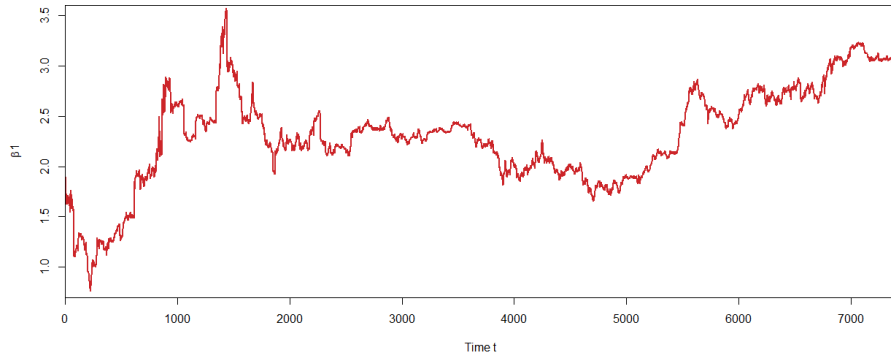


Figure 9.63: Plot of β_{1t} from time $t = 0, \dots, 7407$

Estimates of β_{2t} 's:

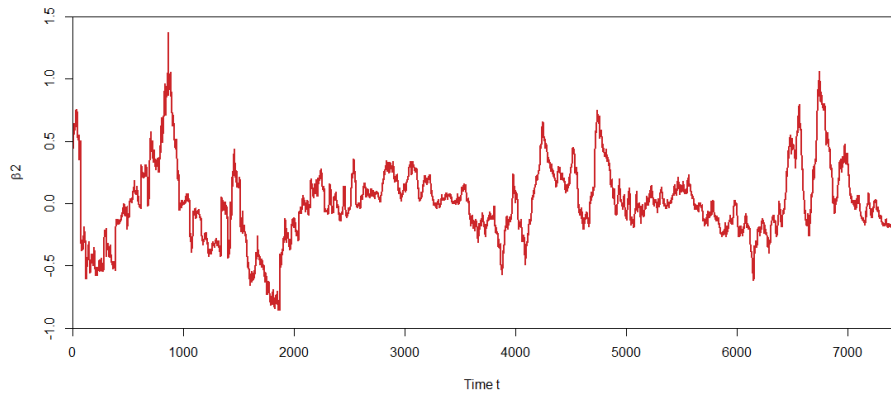


Figure 9.64: Plot of β_{2t} from time $t = 0, \dots, 7407$

And of β_{3t} 's:

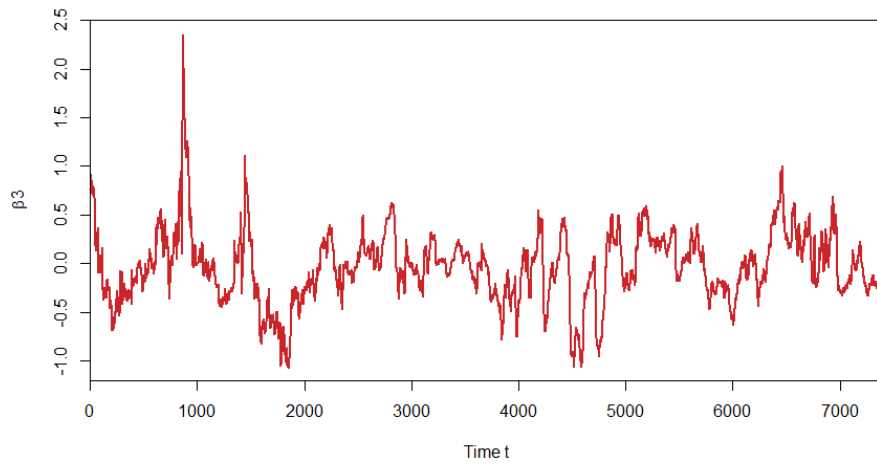


Figure 9.65: Plot of β_{3t} from time $t = 0, \dots, 7407$

The seasonality on the β_t 's are also estimated:

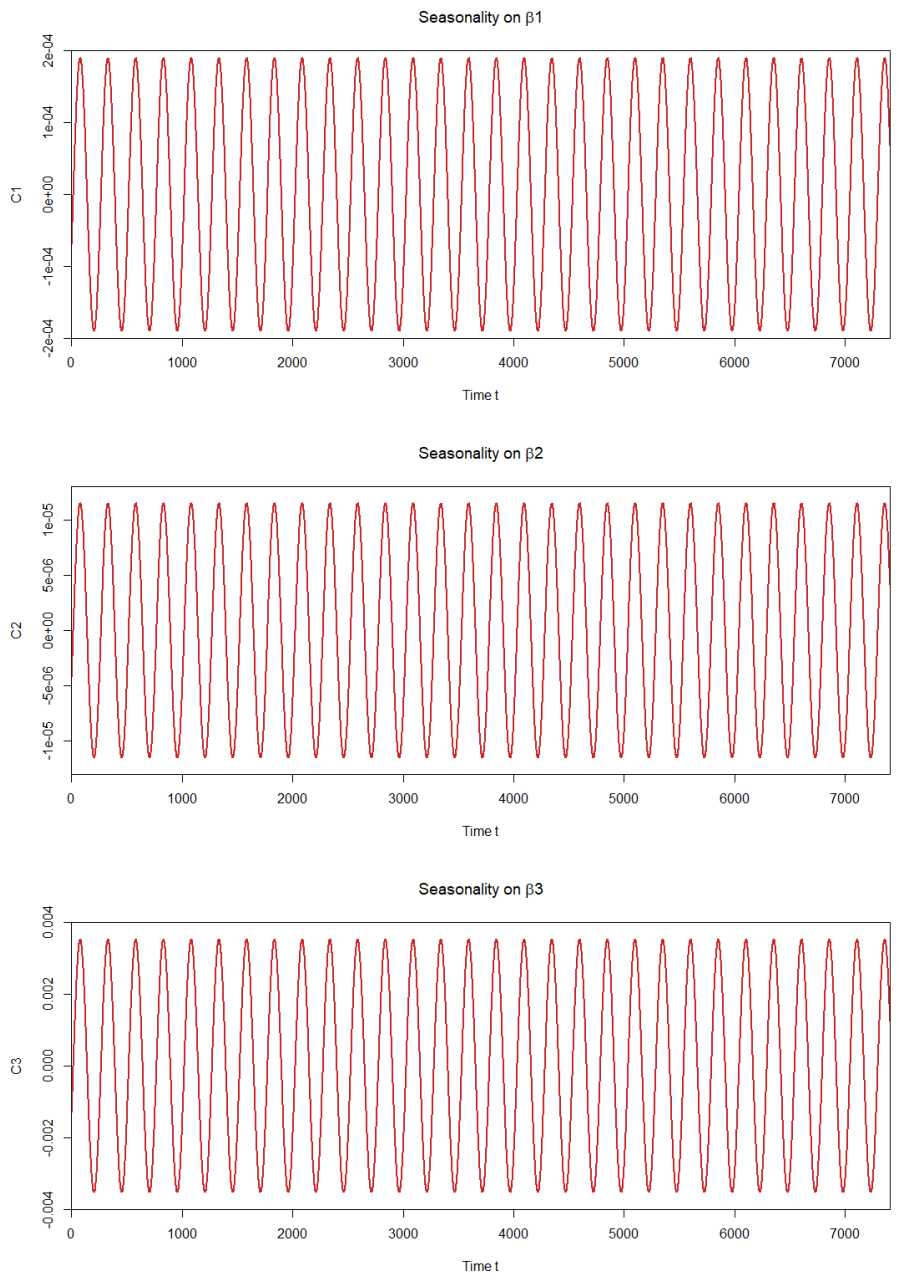


Figure 9.66: Seasonality on β_t for period $s = 253$

The β_t 's for the seasonal model behaves much the same way as for the three-

factor model, where the β_{1t} 's are mainly increasing throughout the sample size, with a part in the middle where there is a decrease in the β_{1t} 's. The β_{2t} 's and β_{3t} 's are changing sign several times over time. The seasonality on the β_{1t} 's have the biggest amplitude of the seasonality terms.

For the seasonal model it was possible to do the one-day-ahead forecasting. The one-day-ahead forecasting is very good for the seasonal model with $MSE = 2.13 \times 10^{-4}$. The forecasted observations are:

9.6.3 Hypothesis testing

For the three-factor model we have a log-likelihood value of $\log(L(Y_n|\psi_0)) = 122817.2$ and the seasonal model have a log-likelihood value of $\log(L(Y_n|\psi)) = 122830.5$. Comparing the three-factor model and the seasonal model to check if there is significant improvement by adding the seasonality term, we get the test statistic:

$$\begin{aligned} D &= -2 \log(\Lambda) \\ &= 2 \times (\log(L(Y_n|\psi)) - \log(L(Y_n|\psi_0))) \\ &= 2 \times (122830.5 - 122817.2) \\ &= 26.6 \end{aligned}$$

The degrees of freedom for the test statistic is $p - q = 12 - 8 = 4$. We have the test statistic $D = 26.6$, to see if the seasonal model has a better fit to the data we need to check if the value $D = 26.6$ is significant in a χ_4^2 -distribution. As mentioned in the hypothesis testing for the crude oil data, with a significance level of $\alpha = 0.05$ and 4 degrees of freedom, we need $D > 9.49$. We have $D = 26.6 > 9.49$, thus we can reject the null hypothesis and conclude that there is seasonality in the sugar data.

The AIC and BIC calculated for the three-factor model:

$$AIC = -245618$$

$$BIC = -245549$$

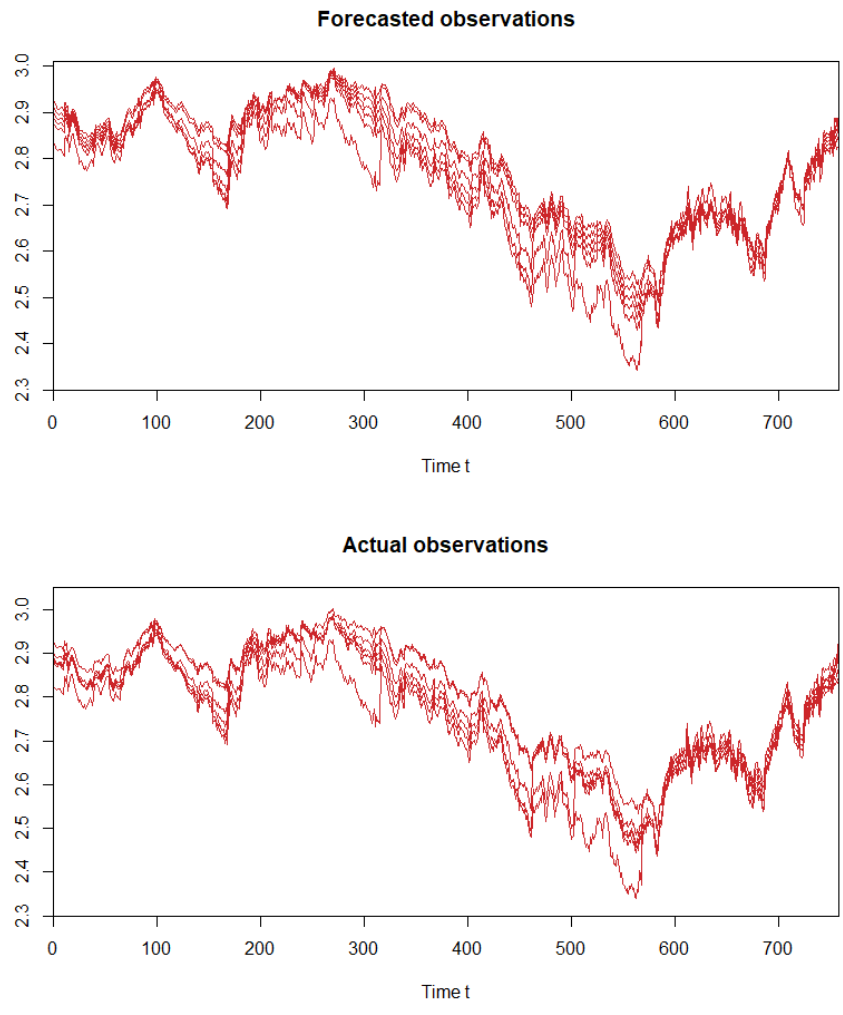


Figure 9.67: Forecasted observations of seasonal model and actual observations of the sugar data

and for the seasonal model:

$$AIC = -245637$$

$$BIC = -245532$$

these seems to favour different models, the AIC agrees with log-likelihood ratio test and prefers the seasonal model while the BIC disagrees with the ratio test and prefers the three-factor model. As with the AIC and BIC for the cotton data, this is probably due to the BIC giving a higher penalty to more complex models and thus it is more likely than the AIC to be choosing too small a model. Since we have $D = 26.6 > 9.49$ which gives a $p - value = 2.44 \times 10^{-5}$, we go with AIC and choose the seasonal model.

10 Conclusion

Throughout this thesis we have built an understanding of the theory behind a linear Gaussian state space model. The Kalman filter equations have been derived for the use of estimating parameters and forecasting observations. The parameters in a three-factor Nelson-Siegel model and a three-factor Nelson-Siegel model with seasonality has been estimated for several different data sets. Those data sets consists of prices and maturities on future contracts of commodities such as, crude oil, natural gas, copper, cotton, gold and sugar. From our results we can see that both the three-factor model and the seasonal model are very good fits for the different commodities with the two question marks being the natural gas and cotton data as the estimates of β_{1t} 's aren't as expected for those two commodities, following from this is that the β_{2t} 's and the β_{3t} 's for these commodities are probably compensating for the level factor, β_{1t} 's, when it comes to the forecasting of the observations. This could suggest that the model is not to great for those data sets and it could be the basis for further work to find out why the estimates for the β_{1t} 's are so far from what we expected them to look like.

From the log-likelihood ratio tests on these model regarding the different data sets we can conclude that three of the six data sets have seasonality, those are the data sets for the commodities natural gas, cotton and sugar. The AIC and BIC for the cotton and sugar data are choosing different models, the AIC for both commodities agree with the log-likelihood ratio test, but the BIC chooses the three-factor model in both cases. Which model to choose are up for debate as it is highly debated which of the two information criterions that are the best, thus in this thesis we side with the AIC.

The results of the forecasting are very good and for most of the commodities the results from the forecasting of the three-factor model is indistinguishable from the forecasting of the seasonal model. Excluding the sugar data, where the forecasting of the observations with the three-factor model was not possible, the commodity with the largest difference in the forecasting performance is natural

gas. Although both the three-factor model and the seasonal model performed very well for the natural gas data, the seasonal model do have almost half the *MSE* value as the three-factor model.

In this thesis we see that the Kalman filter is very good at estimating parameters for the three-factor and the seasonal model. It is also very good at forecasting the observations of the different commodities. Although there are some areas, such as the reason for the unexpected results in estimating the β_t 's for the natural gas and cotton data.

References

- Aic vs bic <https://www.methodology.psu.edu/resources/aic-vs-bic/>. [13.06.2019].
- Akaike information criterion https://en.wikipedia.org/wiki/Akaike_information_criterion. [08.06.2019].
- Bayesian information criterion https://en.wikipedia.org/wiki/Bayesian_information_criterion. [08.06.2019].
- Delta method https://en.wikipedia.org/wiki/Delta_method. [29.01.19].
- Dynamic factor models <http://bactra.org/notebooks/factor-models.html>. [08.06.2019].
- Latent variable models https://en.wikipedia.org/wiki/Latent_variable_model. [21.03.19].
- Likelihood-ratio test https://en.wikipedia.org/wiki/Likelihood-ratio_test. [13.06.2019].
- Local independence https://en.wikipedia.org/wiki/Local_independence. [21.03.19].
- Wilks theorem https://en.wikipedia.org/wiki/Wilks%27_theorem. [13.06.2019].
- Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*. John Wiley and Sons Ltd.
- Cai, L. (2012, April). Latent variable modeling. *Shanghai archives of psychiatry* 24, 118–120.
- Christensen, J. H. E., F. X. Diebold, and G. D. Rudebusch (2009). An arbitrage-free generalized nelson-siegel term structure model. *The Econometrics Journal* 12(3), C33–C64.

de Rezende, R. B. and M. S. Ferreira (2013). Modeling and forecasting the yield curve by an extended nelson-siegel class of models: A quantile autoregression approach. *Journal of Forecasting* 32, 111–123.

James Durbin, S. J. K. (2012). *Time Series Analysis by State Space Methods*. Oxford University Press.

Shumway, R. H. and D. S. Stoffer (2017). *Time Series Analysis and Its Applications* (4 ed.). Springer Texts in Statistics. Springer International Publishing.

Svensson, L. E. (1994, September). Estimating and interpreting forward interest rates: Sweden 1992 - 1994. Working Paper 4871, National Bureau of Economic Research.