Die approbierte Originalversion dieser Diplom-/Masterarbeit ist in der Hauptbibliothek der Technischen Universität Wien aufgestellt und zugänglich.
http://www.ub.tuwien.ac.at

The approved original version of this diploma or master thesis is available at the main library of the Vienna University of Technology.
http://www.ub.tuwien.ac.at/eng

TECHNISCHE UNIVERSITÄT WIEN
Vienna | Austria

Diplomarbeit

Pellets Price Analysis by Means of Cointegration

Ausgeführt am Institut für
Institut für Stochastik und Wirtschaftsmathematik
der Technischen Universität Wien

unter der Anleitung von
Ao. Univ. Prof. Wolfgang Scherrer

durch
Jakob Hruby
Servitengasse 13/5
1090 Wien

Wien, December 1, 2015 Jakob Hruby

Introduction

One of the most eminent themes of our times is the responsible use and distribution of resources and consequently the question for the appropriate sources for our energy production.
A very important sector in this area are the private consumers who make their choices mainly based on the technologies available and affordable. Nowadays, the households in Europe have an increasing variety of sources for their energy consumption consisting both of old, formerly market dominating resources and technolgies as well as new upcoming ones.
One of these very popular technolgies, especially in the German-speaking countries, are pellet furnaces for the climate-neutral heating of houses with wood pellets, a form of compressed waste generated as byproduct of the lumber industry.
This thesis aims at finding price connections between this form of fuel with its main resource wood, the fossil fuels gas and oil as well as key economic developments represented by the gross domestic product and consumer price index in Germany.
For these purposes we will seek longrun steady-state "price mixes" between those various time series by applying Johansen's method of trying to find cointegration relationships.
This work is structured in four parts: At first we will be introducing the data to be used. In the second part general basics about time series, in particular vector-autoregressive models for them, and a conceptual presentation of cointegration shall be given. The third chapter will be dealing with various tests and methods to prepare the original data for analytical usage, as well as the core technique applied in this work, the cointegration method developed by Johansen, in theory. Lastly, an analysis with the formerly introduced methods will take place, various models will be observed and refined and the results will be discussed.
All computations are being performed by RStudio, a user interface for the statistical programming language R. Both, precast packages as well as routines programmed by myself are being applied. The particular codes are provided at the end of this work.
I would particularly thank the C.A.R.M.E.N e.V., a charitable coordination office for renewable resources, for providing price indices on the matter of oil, gas and pellets prices. All other time series have been taken from the database of the German federal statistical office Statistisches Bundesamt. Data from German sources have been used due to a longer observation period than those from Austrian sources.

At this point I really want to thank everyone that has helped with the creation of this diploma thesis. First and foremost Ao.Univ.Prof. Wolfgang Scherrer for his intense and competent supervision and for always finding and making time to help me with improving and extending the content of this work.
I want to thank my Dad and Dr. Gerhard Dell for the initial advice.
Furthermore, I want to thank my Mum, who found the time to proofread these pages.
A very special thank you to Tanja, who was always there to support and motivate me.
Moreover, I want to thank all of my friends and colleagues, especially Dominik and Philipp, for backing me up.

In general I want to thank Tanja, my family, my brothers Lorenz and Tobias, and all of my friends and colleagues for believing in me and for always being there for me throughout all of my studies.

Contents

1 The Data

A base motivation for our analysis of wood pellets prices and their connections to other time series is given through the strongly changing composition of heating systems in newly built properties.

year	gas	oil	district heating	heat pump	electric	wood	others
2002	75.8	11.0	7.2	2.1	1.7	–	2.2
2003	74.3	12.0	7.0	2.8	1.2	–	2.7
2004	74.9	10.7	7.3	3.1	1.2	1.2	1.6
2005	74.0	6.4	8.6	5.4	1.2	3.0	1.4
2006	66.9	4.3	9.0	11.2	1.0	6.0	1.6
2007	65.6	3.2	10.2	14.3	1.3	3.0	2.4
2008	58.4	2.3	12.0	19.8	1.0	4.0	2.5
2009	50.9	1.9	13.1	23.9	0.8	5.0	4.4
2010	50.2	1.8	14.6	23.5	1.0	5.0	4.1
2011	50.1	1.5	16.3	22.6	0.9	5.6	2.6
2012	48.5	0.9	18.6	23.8	0.6	6.3	1.4
2013	48.3	0.8	19.8	22.5	0.7	6.4	1.5
2014	49.9	0.7	21.1	20.1	0.6	6.2	1.4

Table 1: Share of heating systems in 150 000 to 350 000 newly built German properties according to Bundesverband der Energie- und Wasserwirtschaft e.V. in [Bun15b]. Wood-based heating includes wood pellets and is listed among *others* until 2003.

Over the last decade we can observe a strong movement away from fossil energy forms to renweable new technologies such as heat pumps and woodbased heating, the latter one being fueled by the success of wood pellets heating systems. Though many newly built properties have switched to the still mainly fossil-based district heating systems, the netto growth of the share of renewable energies is nearly 25% in comparison to 2002.
Though the overall share of pellets and other renewable heating systems still lags behind in the grand scheme of things, especially in the southern regions such as Bavaria and in one or two family houses those forms of heating are becoming more and more market relevant (compare for example with [uSF14]). Especially for the matterbased wood pellets this raises the question whether there are underlying market interdependencies with other resources such as the formerly market dominating fossil fuels.

year	gas	oil	district heating	heat pump	electric	others
2002	46.0	31.9	12.4	–	4.5	5.2
2003	46.6	31.6	12.4	0.1	4.4	4.9
2004	47.2	31.2	12.4	0.1	4.3	4.8
2005	47.6	30.9	12.5	0.2	4.2	4.6
2006	48.0	30.5	12.5	0.3	4.1	4.6
2007	48.3	30.1	12.6	0.5	4.0	4.5
2008	48.5	29.8	12.6	0.7	3.8	4.6
2009	48.9	29.3	12.7	0.8	3.6	4.7
2010	49.0	28.9	12.8	1.0	3.4	4.9
2011	49.1	28.3	12.9	1.1	3.2	5.4
2012	49.2	27.8	13.1	1.2	3.1	5.6
2013	49.2	27.2	13.3	1.4	3.0	5.9
2014	49.3	26.8	13.5	1.5	2.9	6.0

Table 2: Share of heating systems in the entirety of 41 million German properties according to Bundesverband der Energie- und Wasserwirtschaft e.V. in [Bun15b]. Wood and wood pellets is listed among *others*.

1.1 Wood Pellets

Wood pellets are a form of pellet fuel made from industrial wood wastes like sawdust or sometimes also from storm-damaged timber [Wik15a]. In the last decade with the boom of biomass-based energy production and heating wood pellets have become one of the pioneer technologies in green tech in Europe. Especially on the private heating sector the amount of pellets users has exploded. Since the year 2000 the amount of pellet heating system has increased from an almost neglectable amount to a (for 2015 prognosed) amount of 400 000 in Germany (numbers taken from [Wik15a] and [Sta15]).
The lower price of wood pellets in comparison to fossil fuels as well as public fundings seem to be main reasons for this trend.
With the entrance of wood pellets as an important player on the market the question how the market influences its price has become more and more interesting.
Besides a disturbance of the price in the winter of 2006/07, caused by retailer misjudgement of the demand and the available stock, the pellets price has been very stable. In times of growing energy prices, its change has been small in comparison to others.
The relative price growth in the twelfth year between January 2002 and January 2014 is given by 56,89%, or 29.25% after eliminating the general German rate of inflation.
The data structure shows a small, periodic price fluctuation which, unsurprisingly, makes the product cheapest during summer time and most expensive in the winter.

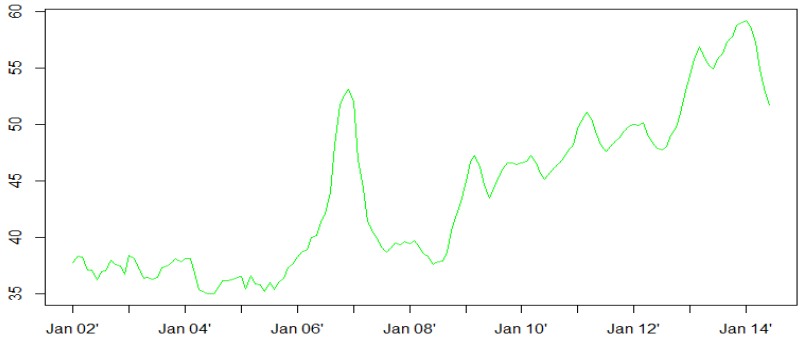


Figure 1: The German monthly pellets price in euro per megawatt hour, provided by CARMEN institute.

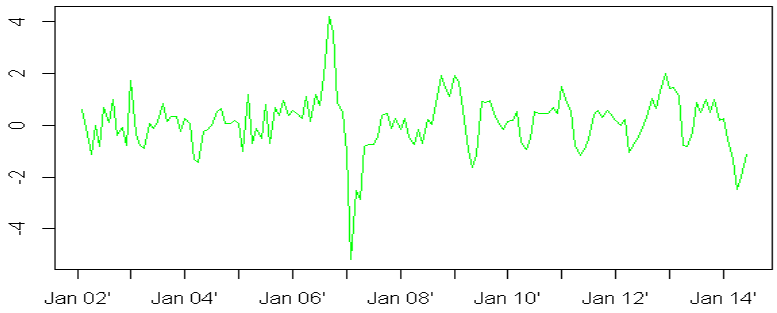


Figure 2: The growth rates of the German monthly pellets price in euro per megawatt hour.

1.2 Oil Fuel

When talking about energy prices one cannot help talking about oil, in this case in particular oil processed to be used for heating. Despite losing in popularity in the last decades, oil fuel still makes for a quarter of the German heating systems in households (see table 2) and thereby ranks second.
In general we can conclude that oil is considered one of the most important indices for energy market developments and its price changes usually have far reaching consequences, in particular on other energy demands and prices. It is therefore only natural when researching the price of a source of heating like wood pellets, to also have a look at the oil price.
The oil price is largely influenced by its availability on the global market. With its biggest suppliers in the Middle East going through various military and diplomatic conflicts the price is heavily dependent on political developments. Also the heavy speculations about its maximum production (the so called Hubbert peak [Hub56]) and longrun availability in times of vastly growing global energy demand can often take their toll on the price, e.g. in the years 2007-2008. A single reason for the supply-demand growth discrepancy in certain periods is often hard to find. It is usually a combination of reasons that can make the oil price fall or rise relatively quickly.
In late 2014 and going forward the policies of the Organization of the Petroleum Exporting Countries (OPEC) have caused a massive drop in prices for oil.

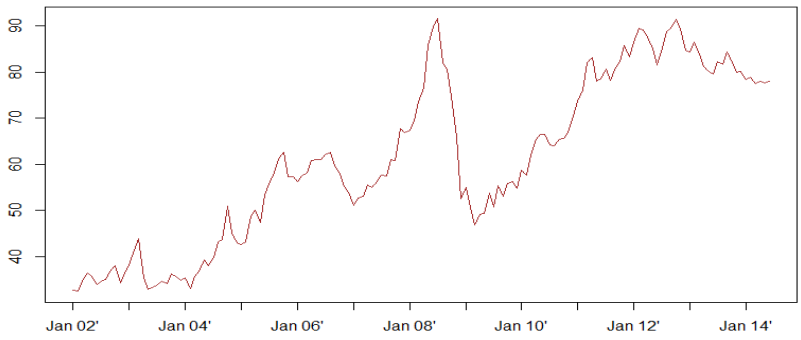


Figure 3: The German monthly oil price in euro per megawatt hour, provided by CARMEN institute.

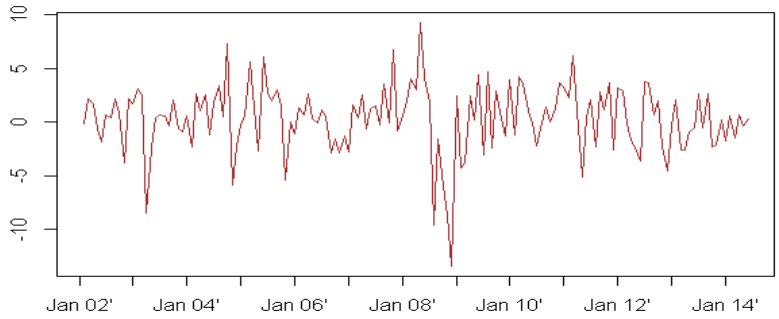


Figure 4: The growth rates of the German monthly oil price in euro per megawatt hour.

1.3 Gas

Gas-based heating is the most popular form in German households and had a market share of 49.3% in 2014 according to table 2. The acquisition of the heating system is cheap and requires the least amount of space, since no local storage is required. Additionally the external storage and delivery leads to a high efficiency because the fuel consumption can easily be regulated and adjusted to temporary needs.
Since the 1960s Germany has allowed an oil price link for gas. The main argument for it was to ensure that gas prices will rise when oil prices rise, which protects industry interests of the oil and gas producing firms. It generates windfall profits for them, i.e. a higher oil price leads to higher profits from selling gas, and also prevents that in competition with the gas prices the oil prices have to be lowered. Nowadays this link is controversial and seen as outdated, yet, in principle still remains (a more extensive read on the topic can be found in [Wik15b]). By comparison with other fossil energy forms the burning of gas leads to less enviromental pollution due to lower emissions of carbon dioxide and sulphur.

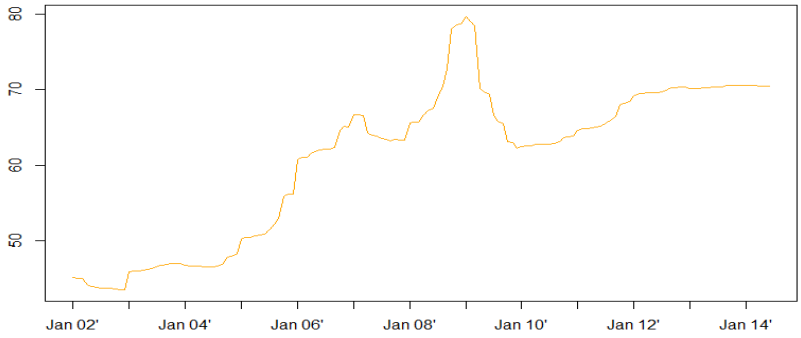


Figure 5: The German monthly gas price in euro per megawatt hour, provided by CARMEN institute.

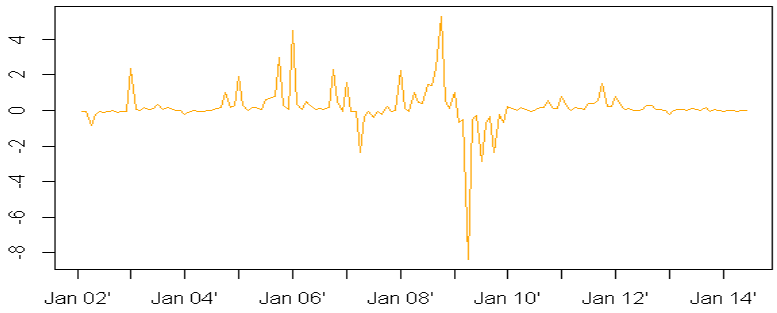


Figure 6: The growth rates of the German monthly gas price in euro per megawatt hour.

1.4 Wood Price

Wood as the natural resource should obviously be considered in one way or another as a factor in any price analysis of wood pellets. However, due to its versatile nature, wood has many purposes besides energy generation like as a building material for houses or it can be processed to paper and carton. We therefore should not expect a one-to-one correlation between wood and the wood pellets product, which is mainly a byproduct of named other products, generated from wood waste.
Additionally, the amount of wood produced can vary strongly every year due to weather conditions. For example the amount of wood produced reached its peak in 2007 after the storm *Kyrill* laid waste to German forests and lots of storm damaged timber had to be processed. This does obviously also impact prices in the shortrun.
Lastly, there are many forms of wood processed from different trees and used for different purposes. The given data is therefore a joint, aggregated index of all forms of wood generated in Germany.

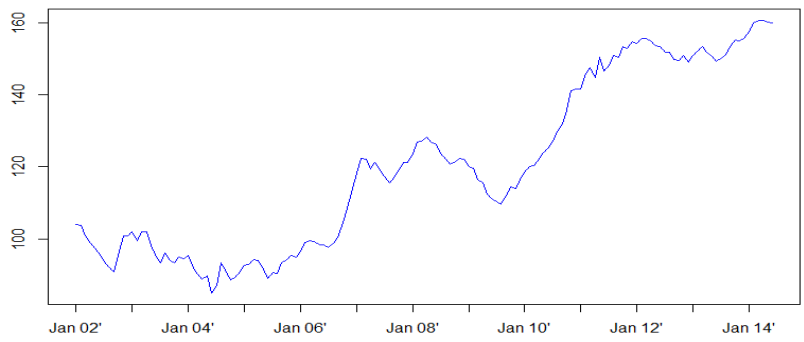


Figure 7: The German monthly wood price index as researched by Statistisches Bundesamt. 2000=100

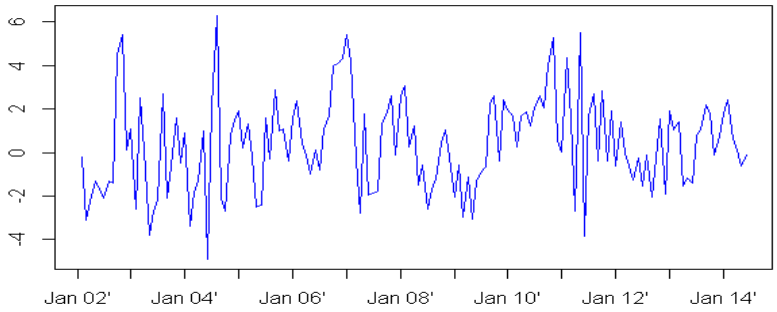


Figure 8: The growth rates of the German monthly wood price index.

1.5 The German GDP

The gross domestic product and its development are a general measure for changes in wealth and industrial progression or standstill of a country. Price developments often behave similar in one way or another to the development of GDP.
The German GDP, despite the general worldwide and European crisis, has shown a relatively stable growth rate in the past decade. In particular, Germany's growth rates have been above those of the EU-28 as well as the EURO-countries in the 5 years following the 2009 global financial crisis according to Eurostat ([Eur15]).
According to Statistisches Bundesamt, in today's German economy the biggest contributor to GDP are services with nearly 68.6%, followed by industrial production that makes for 25.9% and the construction industry with 4.8%. The sector agriculture, fishery and forestry only contributes with 0.8% (numbers taken from [Bun15a]).
The given time series as the only one in this work contains quarterly data, while all other time series hold monthly data. For working purposes it is therefore convenient to process a monthly estimate for the German GDP.
We can immediately see a strong, quarterly periodicity from the growth rates of the GDP.

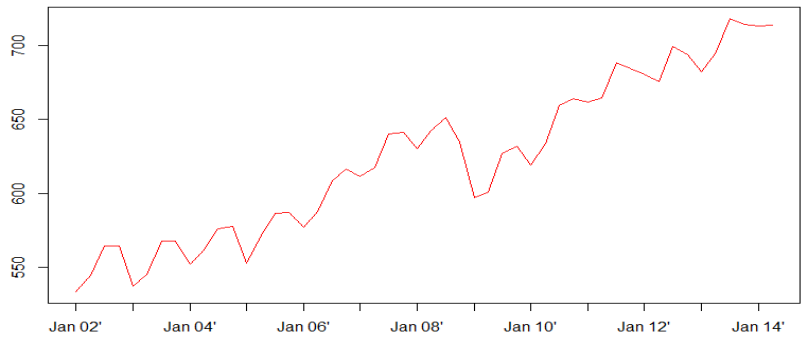


Figure 9: The German quartely GDP in billion euro as researched by Statistisches Bundesamt.

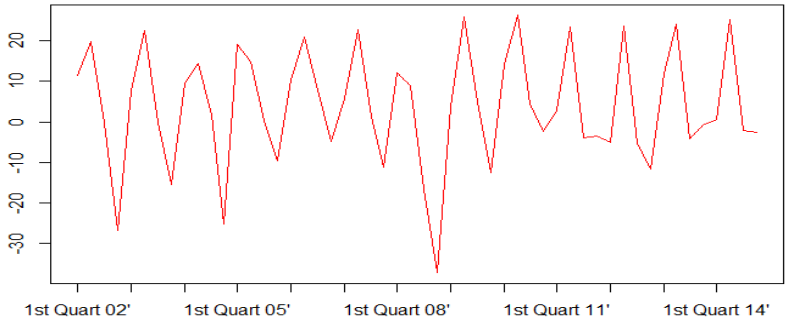


Figure 10: The growth rates of the German quarterly GDP.

1.6 The German CPI

The consumer price index is a measure for the general level of prices for goods purchased in a country. It is calculated from a mix of wares and services that is seen representative of the consumption in those countries. Especially the CPI's rate of change is of great interest as a measure for general inflation.
Historically the German CPI-growth has been fluctuating between one and four percent yearly since the mid-90s and has been relatively stable. However, in times of the worldwide and European regression the inflation rate has dropped to values close to zero percent.

The CPI in this work is being used as a measure for general inflation and applied indirectly by being factored out of the other time series.

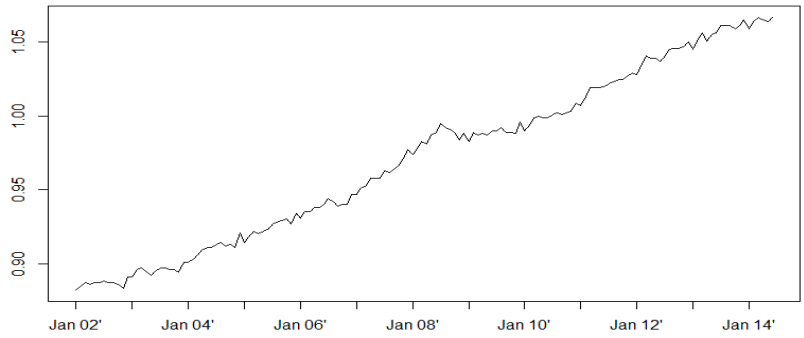


Figure 11: The German monthly CPI as researched by Statistisches Bundesamt. 2010=1

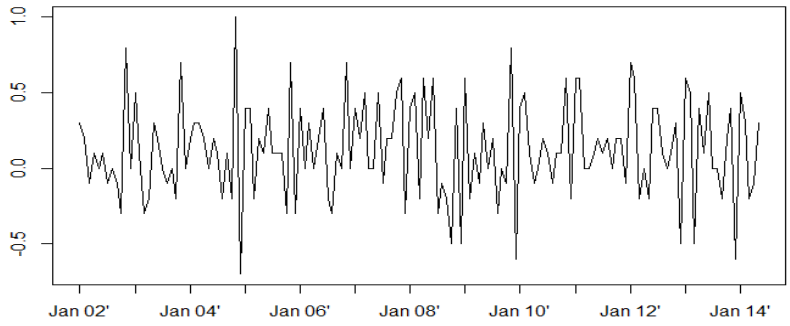


Figure 12: The growth rates of the German monthly CPI.

2 Time Series and the Concept of Cointegration

2.1 Basics about Time Series

Before explaining the methods used for the analysis of the data certain basic properties and definitions about time series need to be made (compare for example [Neu11], [DS10], [rJ95]). The most basic definition we will build upon in this work is the definition of a *stochastic process*, or more specifically, that of a *stochastic process in discrete time*.

Definition 2.1.1. Let $(\mathcal{S}, \mathcal{A}, P)$ be a probability space, then we call a collection of random variables $(X_t | t \in \mathbb{Z})$ with values in $\mathbb{R}^p$, $p \in \mathbb{N}$ a stochastic process in discrete time.

In the case $p > 1$ we call the process *multivariate*, in the case that $p = 1$ *univariate*. Multivariate processes can be denoted as a vector, i.e.

$$X_t = \begin{pmatrix} X_{1t} \\ \vdots \\ X_{pt} \end{pmatrix}.$$

If they exist, a stochastic process can be characterized by its first and second moments *mean* and *covariance*.

$$\begin{array}{rcll} \mu_{it} & = & \mathbb{E}(X_{it}), & \text{for } i = 1, \ldots, p \\ \gamma_{ij}(t,s) & = & \mathbb{E}(X_{it} - \mu_{it})(X_{js} - \mu_{js}), & \text{for } i, j = 1, \ldots, p \text{ and } s, t \in \mathbb{Z} \end{array}$$

Accordingly, in the multivariate case we also combine these to form a vector

$$\mu_t = \begin{pmatrix} \mu_{1t} \\ \vdots \\ \mu_{pt} \end{pmatrix}$$

and a matrix

$$\Gamma(t,s) = \begin{pmatrix} \gamma_{11}(t,s) & \ldots & \gamma_{1p}(t,s) \\ \vdots & \ddots & \vdots \\ \gamma_{p1}(t,s) & \ldots & \gamma_{pp}(t,s) \end{pmatrix},$$

which is called the *covariance function* of the process.

Definition 2.1.2. Given a discrete stochastic process X_t defined on $(\mathcal{S}, \mathcal{A}, P)$ and $\omega \in \Omega$, we call a finite sample of the trajectory $X_t(\omega)$ a time series.

An important property when handling time series is stationarity of the underlying process or its lack thereof.

Definition 2.1.3. We call a process in discrete time $(X_t|t \in \mathbb{Z})$ stationary if for its joint-distribution F it fulfills

$$F(x_1, \ldots, x_T) = F(x_{1+\tau}, \ldots, x_{T+\tau}) \quad \forall \tau \in \mathbb{Z}, \forall T \in \mathbb{N}.$$

However, for many purposes, the easier to handle concept of *stationarity in the weak sense* will be enough.

Definition 2.1.4 (Weak-sense Stationarity)**.** We call a process in discrete time $(X_t|t \in \mathbb{Z})$ weak-sense stationary if

$$\begin{array}{ll} \mathbb{E}(X_t X_t') < \infty & \forall t \in \mathbb{Z} \\ \mathbb{E}(X_t) = \mu_t = \mu & \forall t \in \mathbb{Z} \\ \Gamma(t,s) = \Gamma(t+r, s+r) & \forall r, s, t \in \mathbb{Z}, \end{array}$$

i.e. if the mean of the process is independent from the particular t and its covariance function only depends on the time difference between two observed variables. If there exists a deterministic function

$$f : \mathbb{Z} \to \mathbb{R}^p$$

so that the process $(Y_t|t \in \mathbb{Z})$ defined by $Y_t = X_t - f(t)$ is stationary we call X_t *trend stationary*.

Example 2.1.5 (White Noise Process). A stationary process with $\mu = 0$, covariance matrix $\Sigma > 0$ which fullfills $\Gamma(t,s) = 0$ for $t \neq s$ and $\Gamma(t,s) = \Sigma$ for $t = s$ is called a (multivariate) *white noise process*. White Noise Processes are important when modelling statistical errors and similar, unpredictable occurances.

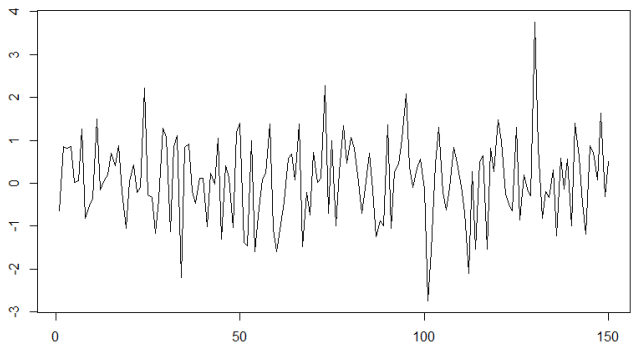


Figure 13: A trajectory of a one-dimensional white noise process generated by 150 standard normal distributed variables

Definition 2.1.6. A process represented by

$$X_t = \sum_{j=0}^{k} \beta_j \varepsilon_{t-j}, \quad \beta_j \in \mathbb{R}^{q \times p}$$

with (ε_t) being a white noise process is called a moving average process of order k (assuming $\beta_0 \neq 0$ and $\beta_k \neq 0$), in short MA(k).

Definition 2.1.7. A process represented by

$$X_t = \sum_{j=-\infty}^{\infty} \beta_j \varepsilon_{t-j}, \quad \beta_j \in \mathbb{R}^{q \times p}$$

with (ε_t) being a white noise process is called an infinite moving average process. The infinite sum has to be understood as the mean-square sense limit of its partial sums $\sum_{j=-N}^{N} \beta_j \varepsilon_{t-j}$, $N \in \mathbb{N}$.

Definition 2.1.8. A process represented by the difference equation

$$X_t = c + \sum_{i=1}^{k} \Pi_i X_{t-i} + \varepsilon_t, \quad \Pi_i \in \mathbb{R}^{p \times p}$$

where (ε_t) is a white noise process and c is a vector of constants is called an autoregressive process of order k or in short AR(k), or in case of $p > 1$ vector autoregressive process in short VAR(k).

Characterizing a VAR(p) process with the lag operator L, that is the unique, linear, unitary operator defined by the property

$$LX_t = X_{t-1}$$

gives the equation

$$c + \varepsilon_t = \Pi(L) X_t \tag{1}$$

with

$$\Pi(L) = (I_p - \Pi_1 L^1 - \cdots - \Pi_p L^p).$$

The z-transform of the later

$$\Pi(z) = I_p - \Pi_1 z^1 - \cdots - \Pi_k z^p$$

is called the *characteristic polynomial* of the process.

For many applications the basic VAR model proves to be insufficient. Only including a vector of constants c may be too restrictive to represent the data and its characteristics properly. Therefore, we give an alternative definition of a VAR model:

Definition 2.1.9. We define a p-dimensional vector autoregressive process with k lags, in short VAR(k), as a multivariate linear process of the form

$$X_t = \Pi_1 X_{t-1} + \cdots + \Pi_k X_{t-k} + \Phi D_t + \varepsilon_t, \quad t = 1, \ldots, T \tag{2}$$

where (ε_t) is a white noise process. The *deterministic term* $D_t : \mathbb{R} \to \mathbb{R}^q$ is known and may contain information such as constants, linear trends and seasonal dummy variables. The matrix Φ is an element of $\mathbb{R}^{p \times q}$.

A typical choice of $D_t = \begin{pmatrix} 1 \\ t \end{pmatrix}$, which assumes that the model follows a deterministic trend.

A further representation of a VAR(k) process is the following:

Definition 2.1.10. A vector autoregressive process in the so-called error correction form is given by

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Phi D_t + \varepsilon_t, \quad t = 1, \ldots, T. \tag{3}$$

The matrices Π and Γ_i, $i = 1, \ldots, k-1$ are defined as

$$\Pi = \sum_{i=1}^{k} \Pi_i - I \tag{4}$$

$$\Gamma_i = \sum_{j=i+1}^{k} \Pi_j. \tag{5}$$

Remark. There exists another frequently used vector error correction models, called the *longrun VECM*. The form showcased above is called the *transitory VECM*. Following the approach of Johansen as given in [rJ95] it is only necessary to define with the later.

When looking for a solution for 2.1.9 we are in particular interested in a (trend) stationary one. For that reason we need the following so-called stability condition.

Condition 2.1.11. For the unit roots, that are the solutions of

$$|\Pi(z)| := det(\Pi(z)) = 0$$

we demand that $|z| > 1$ or $z = 1$.

This leads us to the next result.

Theorem 2.1.12. *Assuming that 2.1.11 holds, (D_t) is bounded by a polynomial in t and $|A(1)| \neq 0$ the trend stationary solution to* 2.1.9 *takes the form*

$$X_t = \sum_{n=0}^{\infty} C_n(\varepsilon_{t-n} + \Phi D_{t-n}),$$

with the generating function $C_0(z) = \sum_{n=0}^{\infty} C_n z^n = \Pi(z)^{-1}$ being convergent for $|z| > 1 + \delta$ for some $\delta > 0$. The C_n are recursively defined as

$$\begin{aligned} C_0 &= I_p, \\ C_n &= \sum_{j=1}^{\min(k,n)} C_{n-j}\Pi_j, \quad n = 1, \ldots, t-1. \end{aligned}$$

This theorem and its proof can be found in [rJ95].

2.2 The Least-square Regression

The usual method to extract a VAR model from a given set of data vectors is called least-square regression and should be outlined here. Let X_t be a set of p-dimensional data vectors, $T \in \mathbb{N}$ the number of such data vectors and $k \in \mathbb{N}$ being the desired order of the process.
Given that we need k data points to form an equation of the VAR(k) form, this leads to the system of equations

$$\begin{aligned} X_{k+1} &= \Phi D_{k+1} + \Pi_1 X_k + \cdots + \Pi_k X_1 + \varepsilon_{k+1} \\ &\vdots \\ X_T &= \Phi D_T + \Pi_1 X_{T-1} + \cdots + \Pi_k X_{T-k} + \varepsilon_T \end{aligned}$$

with the coefficient matrices to be determined $\Phi, \Pi_1, \ldots, \Pi_k$. The criterion invoked is that the residuals ϵ_t should be minimal in the least-square sense. Rewriting the equations as

$$Y = BZ + U \tag{6}$$

with the notation

$$Y = (X_{k+1}, \ldots, X_T) = \begin{pmatrix} x_{1,k+1} & \cdots & x_{1,T} \\ \vdots & \vdots & \vdots \\ x_{p,k+1} & \cdots & x_{p,T} \end{pmatrix},$$

$$B = (\Phi, \Pi_1, \ldots, \Pi_k) =$$
$$= \begin{pmatrix} \phi_{1,1} & \cdots & \phi_{1,q} & \pi^1_{1,1} & \cdots & \pi^1_{1,p} & \cdots & \pi^k_{1,1} & \cdots & \pi^k_{1,p} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\ \phi_{p,1} & \cdots & \phi_{p,q} & \pi^1_{p,1} & \cdots & \pi^1_{p,p} & \cdots & \pi^k_{p,1} & \cdots & \pi^k_{p,p} \end{pmatrix},$$

$$Z = \begin{pmatrix} D_{k+1} & \cdots & D_T \\ X_k & \cdots & X_{T-1} \\ \vdots & \ddots & \vdots \\ X_1 & \cdots & X_{T-k} \end{pmatrix} = \begin{pmatrix} d_{1,k+1} & \cdots & d_{1,T} \\ \vdots & \ddots & \vdots \\ d_{q,k+1} & \cdots & d_{q,T} \\ x_{1,k} & \cdots & x_{1,T-1} \\ \vdots & \ddots & \vdots \\ x_{p,k} & \cdots & x_{p,T-1} \\ \vdots & \vdots & \vdots \\ x_{1,1} & \cdots & x_{1,T-k} \\ \vdots & \ddots & \vdots \\ x_{p,1} & \cdots & x_{p,T-k} \end{pmatrix},$$

$$U = \begin{pmatrix} \varepsilon_{1,k+1} & \cdots & \varepsilon_{1,T} \\ \vdots & \ddots & \vdots \\ \varepsilon_{p,k+1} & \cdots & \varepsilon_{p,T} \end{pmatrix}.$$

The formulation of the problem becomes finding the least-square-sense solution of (6), i.e. we seek a coefficient matrix $\hat{B}$ such that

$$\|Y - \hat{B}Z\| = \|u\| = \min_B \|Y - BZ\|.$$

This can be achieved by usage of the Moore-Penrose-Pseudoinverse Matrix of Z and is thus given by

$$\hat{B} = Z^+ Y$$

with

$$Z^+ = (Z'Z)^{-1}Z' \quad \text{with} \quad (Z'Z) > 0.$$

2.3 The Choice of Lag Length

Another important part of the modelling is the choice of a good lag length k. This means one has to find a good compromise between the often conflicting goals to find a model that has maximum explanatory power, whilst minimizing the complexity of the model, often measured in the amount of lags used in a regression model. It is usually derived by usage of an information criterium, e.g. Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) or Hannan-Quinn Criterion (HQ). The point of these criteria is to punish complicated models, for example when they use a greater lag length for only a small gain in fit. A comparison of those can be found for example in [Lü91].
Johansen claims in that context that *it is our experience that if a long lag length is required to get white noise residuals then it often pays to reconsider the choice of variables, and look around for another important explanatory variable to include in the information set. That is, rather than automatically increase the lag length, it is more fruitful in a multivariate context to increase the information set.* [rJ95]
Furthermore the residuals should be tested for correlation. It is suggested to use a Portmanteau Test such as the Ljung-Box or Box-Pierce test.

For the following definitions of the AIC, BIC and HQ criteria we define $H_1, \dots, H_n$ as VAR(k_i)-models with respective lag parameters $k_1, \dots, k_n$. The dimension shall be p for all cases and a q-dimensional deterministic term is given (see 2.1.9).
Furthermore for $(\hat{\varepsilon}_{i,t})$ being the estimated residuals of the i-th model, $i = 1, \dots, n$ with the property that $\hat{\varepsilon}_{i,t} = 0$, $\forall t \leq k_i, \forall i = 1, \dots, n$ we define their sample covariance matrix

$$\hat{\Sigma}_i = \frac{1}{T} \sum_{t=1}^{T} \hat{\varepsilon}_{i,t} \hat{\varepsilon}'_{i,t}.$$

Definition 2.3.1 (The AIC Criterion)**.** For testing models $H_1, \dots, H_n$ with respectively $k_1, \dots, k_n$ lag-parameters for goodness of fit we define the criterion

$$AIC(H_i) = \log \left| \hat{\Sigma}_i \right| + \frac{2}{T}(k_i p + q)p, \quad \text{for } i \in 1, \dots n.$$

We call the model with the lowest AIC-value the best fit according to the AIC-criterion amongst our models H_i, $i \in 1, \dots, n$.

Definition 2.3.2 (The BIC Criterion)**.** For testing models $H_1, \dots, H_n$ with respectively $k_1, \dots, k_n$ lag-parameters for goodness of fit we define the criterion

$$BIC(H_i) = \log \left| \hat{\Sigma}_i \right| + \frac{\log(T)}{T}(k_i p + q)p, \quad \text{for } i \in 1, \dots n.$$

We call the model with the lowest BIC-value the best fit according to the BIC-criterion amongst our models H_i, $i \in 1, \dots, n$.

Definition 2.3.3 (The HQ Criterion)**.** For testing models $H_1, \ldots, H_n$ with respectively $k_1, \ldots, k_n$ lag-parameters for goodness of fit we define the criterion

$$HQ(H_i) = \log \left| \hat{\Sigma}_i \right| + \frac{2 \log(\log(T))}{T} (k_i p + q) p, \quad \text{for } i \in 1, \ldots n.$$

We call the model with the lowest HQ-value the best fit according to the HQ-criterion amongst our models H_i, $i \in 1, \ldots, n$.

2.4 The Concept of Integrated and Cointegrated Time Series

It was shown that the standard regression analysis fails when dealing with relations between non-stationary time series. Very often the phenomenom of spurious correlation would occur, meaning the analysis would interprete relations where there are none. In that context, Granger and Newbold stated that *we would conclude that if a regression equation relating economic variables is found to have strongly autocorrelated residuals, equivalent to a low Durbin-Watson value, the only conclusion that can be reached is that the equation is mis-specified, whatever the value of R^2 observed.* [GN74]
For that purpose Engle and Granger and later on Johansen developed the concept of cointegrated time series, which is the idea that multiple non-stationary time series could be combined to form a stationary process. [EG87][rJ95]

Definition 2.4.1. We call a stochastic process X_t integrated of order 0, in short $I(0)$ if there exists a sequence of moving average parameters θ_i, $i \in \mathbb{N}$ with $\sum_{i=0}^{\infty} \theta_i^2 < \infty$ so that

$$X_t = \mathbb{E}(X_t) + \varepsilon_t + \sum_{\tau=1}^{\infty} \theta_\tau \varepsilon_{t-\tau}.$$

Definition 2.4.2. We call a p-dimensional process integrated of order d, for $d \in \mathbb{N}$, in short $I(d)$, if $(I_p - L)^d X_t$, L being the lag operator, is integrated of order 0.

Definition 2.4.3. Let X_t be a process of integration order $I(1)$. If there exists a $\beta \neq 0$ so that $\beta' X_t$ can be made stationary by a suitable choice of its initial distribution we call X_t cointegrated with the cointegration vector β.

Later on when using Johansen's method we will have $\beta \in \mathbb{R}^{p \times r}$ being a matrix of r linearly independent cointegration relations. We call r the *cointegration rank* of the cointegrated process X_t and the space spanned by the cointegration relations the *cointegration space*.
For defining a process as cointegrated, however, it is sufficient to find a single such relation.

As we can see, cointegration is multivariate by nature. Single time series cannot be cointegrated as that would imply stationarity to begin with, which stands in contrast to the prerequisition of being $I(1)$.
The fundamental result on the topic of cointegration is *Granger's representation theorem*. It shows that there is a decomposition for a vector-autoregressive cointegrated process in a random walk, a stationary process, a deterministic part and a part that depends on the initial values. (see [Han04] or [rJ95])

Given a VAR-process in error correction form as given by 2.1.10, $\Pi(z)$ being its characteristic polynomial with $\Pi(z) = I - \sum_{i=1}^{k} \Pi_i z^i, z \in \mathbb{C}$ we give conditions to ensure it being $I(1)$.

Condition 2.4.4.

- $\forall z \in \mathbb{C}$ that satisfy $\det(\Pi(z)) = 0$ we have that $|z| \geq 1$ or $z = 1$.
- The matrix Π from (4) is of reduced rank $r < p$, hence can be represented with $\Pi = \alpha\beta'$ with α, β being $p \times r$ matrices of full column rank r.
- $\alpha'_\perp \Gamma \beta_\perp$, with $\Gamma = I - \sum_{i=1}^{k-1} \Gamma_i$ is a matrix of full rank. $\alpha_\perp$ and $\beta_\perp$ are the orthogonal complements to the the matrices α and β.

Furthermore we bound the deterministic term D_t in 2.1.10 with the next condition.

Condition 2.4.5. There exist constants $a, b \in \mathbb{R}$ so that $|D_t| < a + |t|^b$.

This allows us to give the statement of the Granger representation theorem.

Theorem 2.4.6 (Granger's Representation Theorem)**.** *For an autoregressive process in error correction form as given in 2.1.10 under the conditions 2.4.4 and 2.4.5 we have the representation*

$$X_t = C \sum_{i=1}^{t} (\varepsilon_i + \Phi D_i) + C(L)(\varepsilon_t + \Phi D_t) + A.$$

A depends on initial values such that $\beta' A = 0$, $C = \beta_\perp(\alpha'_\perp \Gamma \beta_\perp)^{-1}\alpha'_\perp$ and $C(z)$ satisfies $A^{-1}(z) = C\frac{1}{1-z} + C(z), z \neq 1$, where the power series for $C(z)$ is convergent for $|z| < 1 + \delta$ for some $\delta > 0$.
This implies that X_t is a cointegrated $I(1)$-process.

2.5 Models for the Deterministic Component

In Error Correction Form we deal with a model that includes both levels and differences to describe the development of the process. As it turns out the trend and constant components of a deterministic or *drift term* D_t are similarily assigned in that model.
From the Granger Representation Theorem 2.4.6 we find that in general the process X_t has a *deterministic trend* of the form

$$C\Phi \sum_{i=1}^{t} D_i + C(L)\Phi D_t.$$

The Granger Representation Theorem shows that a constant drift term D_t will induce a linear trend in the process X_t and a linear drift term D_t will induce a quadratic trend. (see [rJ94] and [HJ00])
For the process X_t the theorem gives us

$$X_t = C \sum_{i=1}^{t} (\varepsilon_i + \Phi D_i) + C(L)(\varepsilon_t + \Phi D_t) + A. \tag{7}$$

Since

$$\begin{aligned} &\beta' C = \beta'(\beta_\perp(\alpha'_\perp \Gamma \beta_\perp)^{-1}\alpha'_\perp) = 0 \quad \text{and} \\ &\beta' A = 0 \end{aligned}$$

we have for the cointegration relations

$$\beta' X_t = \beta' C(L)\varepsilon_t + \beta' C(L)\Phi D_t. \tag{8}$$

The deterministic term ΦD_t is represented by the linear and constant terms $\pi + \delta t$, which we will split into components in α and $\alpha_\perp$ directions, i.e. we have

$$\begin{aligned} D_t &= \begin{pmatrix} 1 \\ t \end{pmatrix} \\ \Phi &= (\alpha\rho_1 + \alpha_\perp\gamma_1, \alpha\rho_2 + \alpha_\perp\gamma_2) \end{aligned}$$

with

$$\begin{aligned} \pi &= \alpha\rho_1 + \alpha_\perp\gamma_1 \\ \delta &= \alpha\rho_2 + \alpha_\perp\gamma_2. \end{aligned}$$

This leads to the five commonly used models for deterministic terms.

Case 1 $\rho_1 = \rho_2 = \gamma_1 = \gamma_2 = 0$, hence no drift term is present. We have no growth or intercepts in the cointegration relations. This is a rare case, since usually some intercept is needed to account for the starting measure at X_0. Only if the measuring starts at zero or if the different levels cancel in the cointegration relations these restrictions should be used.

Case 2 $\rho_2 = \gamma_1 = \gamma_2 = 0$, but $\rho_1 \neq 0$, hence the constant part of the drift term is present in the direction of α. Linear deterministic trends are not present in the data. Since in this form $C\Phi = 0$ we also only reproduce a constant in both (7) and (8). This constant cancels in when taking differences and thus we have $\mathbb{E}(\Delta X_t) = 0$.

Case 3 $\rho_2 = \gamma_2 = 0$, but $\rho_1 \neq 0$ and $\gamma_1 \neq 0$ and thus the constant term π is unrestricted. $C\Phi$ takes the form $C\Phi = (C\alpha_\perp\gamma_1, 0)$, which again produces a constant in (7) and (8), but additionally from $C\Phi \sum_{i=1}^{t} D_i$ we get a linear trend in X_t. Having $\mathbb{E}(\Delta X_t) = C\alpha_\perp\gamma_1 \neq 0$ is consistent with these restrictions.

Case 4 We only restrict $\gamma_2 = 0$, but $\rho_1 \neq 0$, $\gamma_1 \neq 0$ and $\rho_2 \neq 0$. Since just as in Case 3 we have $C\Phi = (C\alpha_\perp\gamma_1, 0)$ we also get a constant and a linear trend in (7) again (this time caused by both of the last terms in the representation). However, we now also get a linear trend and a constant in (8). For $\mathbb{E}(\Delta X_t) \neq 0$ applies again.

Case 5 There are no restrictions on the variables π and δ. We have $\rho_1 \neq 0$, $\gamma_1 \neq 0$, $\rho_2 \neq 0$ and $\gamma_2 \neq 0$. This additionally produces a quadratic trend through the sum in (7) in X_t. Usually it is adviced against this case, unless it is plausible that the data actually contain quadratic trends.

3 Methods, Tests and Tools

This section deals with the theory to derive the lag parameter k for our Cointegration Model, as well as all the matrices associated with cointegration Π, Φ, Γ_i for $i = 1 \dots k-1$ and the covariance of the errors Σ. But first of all we want to check whether Cointegration is really necessary or if we are dealing with stationary time series anyway. For that purpose we have to briefly consider a unit root test.

3.1 Data Interpolation for Quarterly Data

Since not all time series are of similar frequency it is often necessary to either work with less information than available in common models of these time series, or to estimate missing data points based on the existing, lower frequency data to achieve the same periodicity. Since in our case all our time series except for one are monthly, it would be a waste of information to perform calculation based on the quarterly periodicity of the German GDP.
For that, we assume that a time series $x_t, t = 1, \dots, T$ with $T \in 3\mathbb{N}$ to be completed is a random walk process, $x_t \sim RW$, i.e.

$$x_t = x_{t-1} + \varepsilon_t.$$

For the variance of the process we therefore have

$$\mathbb{E}x_t x_s = \min(t, s) \text{ , i.e.}$$
$$\mathbb{E}xx' = \sigma^2 \begin{pmatrix} 1 & 1 & 1 & \cdots \\ 1 & 2 & 2 & \cdots \\ 1 & 2 & 3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \text{ with } x = \begin{pmatrix} x_1 \\ \vdots \\ x_T \end{pmatrix}$$

and σ^2 being the variance of the white noise process ε_t.
Since we treat our original time series as incomplete, monthly time series for which we only have the quarterly data points

$$Sx = \begin{pmatrix} x_1 \\ x_4 \\ \vdots \\ x_{T^*} \end{pmatrix} \text{ with } T^* = T - 2$$

we have for x_t

$$x_t = \begin{pmatrix} c_1 & c_2 & \dots & c_{T/3} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_4 \\ \vdots \\ x_{T^*} \end{pmatrix} + u_t = c(Sx) + u_t. \tag{9}$$

By least-square regression we get for the vector of parameters c to be estimated

$$c = \mathbb{E}(x_t(Sx_t)')\mathbb{E}((Sx_t)(Sx_t)')^{-1}$$

and thus, can interpolate our data through (9).

3.2 Estimations for Seasonality and Deterministic Trends

There are various ways to estimate seasonal trends. The method used in this work is the Holt-Winters approach for triple exponential smoothing. ([Hol57], [Win60])
For a time series x_t, $t = 1, \ldots, T$ with a seasonal length L three sequences are calculated:

$s_t \ldots$ a smoothed component
$b_t \ldots$ a trend component
$c_t \ldots$ a seasonal component.

The initial $2L$-values for the smoothing component s_t are initialized with the data-values x_t. The trend component is initialized by

$$b_1 = \frac{1}{L}\left(\frac{x_{L+1} - x_1}{L} + \frac{x_{L+2} - x_2}{L} + \cdots + \frac{x_{L+L} - x_L}{L}\right).$$

The first L-values of the seasonal component c_t are initialized by

$$c_t = \frac{1}{N}\sum_{i=1}^{N}\frac{x_{L(i-1)+t}}{A_i}$$

with N being the overall amount of complete L-length cycles in the original time series x_t and A_i being the average value of x in the i-th cycle, hence

$$A_i = \frac{1}{L}\sum_{j=1}^{L} x_{L(i-1)+j}, \text{ for } i = 1, \ldots, N.$$

The rest of the values for s_t, b_t and c_t can be derived by the recursions

$$\begin{aligned} s_t &= \alpha\frac{x_t}{c_{t-L}} + (1+\alpha)(s_{t-1} + b_{t-1}) \\ b_t &= \beta(s_t - s_{t-1}) + (1-\beta)\beta_{t-1} \\ c_t &= \gamma\frac{x_t}{s_t} + (1-\gamma)c_{t-L}. \end{aligned}$$

The smoothing parameters $\alpha, \beta, \gamma \in (0, 1)$ are chosen following a criterium like least-square error estimation.

A forecast of the form

$$F_{t+m} = (s_t + mb_t)c_{t-L+1+(m-1) \mod L}$$

can be generated, which uses the data values $x_1 \dots x_t$ for a forecast for the value $x_{t+m}, m > 0$.

Our time series x_t can now be deseasonalized by factoring out the determined seasonal components c_t.

3.3 The Augmented Dickey-Fuller Unit Root Test

The Augmented Dicker-Fuller Unit Root Test ([DF79]) is necessery because for cointegration we want our original data series to be $I(1)$. Hence, the differenced time series should be stationary. Furthermore it helps us determining the validity of a hypothesis on the cointegration rank r by testing for actual stationarity of resulting steady-state relations.
To conduct an ADF test we first estimate autoregressive models for each time series in our data of the form

$$\Delta x_t = \alpha + \beta t + \gamma x_{t-1} + \sum_{i=1}^{i=k-1} \delta_i \Delta x_{t-i} + \varepsilon_t.$$

The parameters α and β can be set to 0 if no trend should be assumed. The amounts of lags k used in the model is found by fitting various such models and testing them for goodness of fit, for example by usage of the AIC-criterion.

The unit root test is then carried out under the null hypothesis $\gamma = 0$ against the alternative hypothesis $\gamma < 0$. The test statistic used is derived as $\frac{\hat{\gamma}}{\hat{\sigma}}$ for the estimated parameter $\hat{\gamma}$ and the estimated standard error $\hat{\sigma}$. The idea is that in case of an integrated series x_{t-1} should turn out to be irrelevant in the model, hence, the $\gamma = 0$ hypothesis is not rejected.

3.4 The Ljung-Box Test

The Ljung-Box test belongs to the so called Portmanteau tests and it tests the null hypothesis that the data are independently distributed against the alternative hypothesis that they exhibit serial correlation. We will apply it to see whether an estimated vector autoregressive or vector error correction model represents the data properly. In particular, we test whether the residuals of the fitted model are actually independently distributed.
The test statistic for the Ljung-Box test is

$$Q = n(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{n-k}$$

with n being the sample size of the data, $\hat{\rho}_k^2$ being the sample autocorrelation at lag k and h being the number of lags being tested.
The null hypthesis is rejected at significance level α when Q exceeds the α-quantile of the χ^2-distribution with h degrees of freedom, , i.e. when $Q > \chi^2_{1-\alpha,h}$.

3.5 The Reduced Rank Regression

We start off with a mathematical result that we will need later.

Lemma 3.5.1. *Let $M, N \in \mathbb{R}^{p\times p}$, $p \in \mathbb{N}$ be two symmetric, positive definite matrices. Then the function*

$$f(V) = |V'MV| \,/\, |V'NV|$$

is maximized among all $p \times r$ matrices by $\hat{V} = (v_1, \ldots, v_r)$ and the maximal value is $\prod_{i=1}^{r}\lambda_i$, with $v_i \in \mathbb{R}^p$, $\lambda_1, \ldots, \lambda_r > 0$ being solutions of the general eigenvalue problem

$$\lambda N v = M v.$$

$|\cdot|$ is the determinant of a matrix.

Proof. A proof for this Lemma can be found in [rJ95] □

We follow the guidance and notation of Johansen from [rJ95]. We start with the Error Correction Form of a VAR Process given in 2.1.10 by

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1}\Gamma_i \Delta X_{t-i} + \Phi D_t + \varepsilon_t, \quad t = 1, \ldots, T.$$

For the purpose of cointegration we shall define matrices $\alpha \in \mathbb{R}^{p\times r}$ and $\beta \in \mathbb{R}^{p\times r}$ such that we have

$$\Pi = \alpha\beta'.$$

Furthermore, define

$$\begin{aligned} Z_{0t} =& \Delta X_t \\ Z_{1t} =& X_{t-1} \\ Z_{2t} =& (\Delta X_{t-1}, \ldots, \Delta X_{r-k+1}, D_t)' \\ \Psi =& (\Gamma_1, \ldots, \Gamma_{k-1}, \Phi) \end{aligned}$$

such that 2.1.10 becomes

$$Z_{0t} = \alpha\beta' Z_{1t} + \Psi Z_{2t} + \varepsilon_t, t = 1, \ldots, T. \tag{10}$$

This is a linear regression model with unrestriced parameters Ψ and the coefficient matrix to Z_{1t} being of reduced rank. Therefore, we can apply reduced rank regression as developed by Anderson [And51].
The log-likelihood function is given, apart from a constant, by

$$\log L(\Psi, \alpha, \beta, \Sigma) = -\frac{1}{2}T\log|\Sigma| - \frac{1}{2}\sum_{t=1}^{T}(Z_{0t} - \alpha\beta' Z_{1t} - \Psi Z_{2t})'\Sigma^{-1}(Z_{0t} - \alpha\beta' Z_{1t} - \Psi Z_{2t})$$

and thus, the first order condition for estimating Ψ is

$$\begin{aligned} \frac{d\left(\log L(\Psi, \alpha, \beta, \Sigma)\right)}{d\Psi} =& 0 \\ \sum_{t=1}^{T}(Z_{0t} - \alpha\beta' Z_{1t} - \hat{\Psi} Z_{2t})Z_{2t}' =& 0. \end{aligned} \tag{11}$$

We denote the product moment matrices

$$M_{ij} = T^{-1}\sum_{t=1}^{T} Z_{it}Z_{jt}' \text{ for } i, j = 0, 1, 2$$

which have the symmetry-property

$$M_{ij} = M_{ji}' \text{ for } i, j = 0, 1, 2.$$

This transforms (11) into

$$M_{02} = \alpha\beta' M_{12} + \hat{\Psi} M_{22} \tag{12}$$

with

$$\hat{\Psi}(\alpha, \beta) = M_{02}M_{22}^{-1} - \alpha\beta' M_{12}M_{22}^{-1}.$$

We can now define

$$\begin{aligned} R_{0t} =& Z_{0t} - M_{02}M_{22}^{-1}Z_{2t} \\ R_{1t} =& Z_{1t} - M_{12}M_{22}^{-1}Z_{2t}, \end{aligned}$$

which are the residuals we would obtain by regressing Z_{0t} and Z_{1t} on Z_{2t} (i.e. ΔX_t and X_{t-1} on the lagged differences $\Delta X_{t-1}, \ldots, \Delta X_{r-k+1}$ and D_t).
The concentrated likelihood function is

$$\log L(\alpha, \beta, \Sigma) = -\frac{1}{2}T\log(\Sigma) - \frac{1}{2}\sum_{t=1}^{T}(R_{0t} - \alpha\beta' R_{1t})'\Sigma^{-1}(R_{0t} - \alpha\beta' R_{1t}).$$

From a regression equation for the residuals

$$R_{0t} = \alpha\beta' R_{1t} + \hat{\varepsilon}_t \tag{13}$$

we would get the same likelihood, hence the parameters Ψ can be eliminated by regression and what remains in (13) is a reduced rank regression as investigated by [And51].
Finally, we define S_{ij} by

$$S_{ij} = T^{-1}\sum_{t=1}^{T} R_{it}R_{jt}' = M_{ij} - M_{j2}M_{22}^{-1}M_{2j}, \text{ for } i, j = 0, 1$$

which, for a fixed β, provides us with an easy way to estimate α and Σ through regression of R_{0t} on βR_{1t} and leads to

$$\begin{aligned} \hat{\alpha}(\beta) =& S_{01}\beta(\beta' S_{11}\beta)^{-1}, \\ \hat{\Sigma}(\beta) =& S_{00} - S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10} = S_{00} - \hat{\alpha}(\beta)(\beta' S_{11}\beta)\hat{\alpha}(\beta)' \end{aligned}$$

What is left to do is to find is a good estimator for β.
The contracted likelihood function is given by

$$L_{max}^{-2/T}(\beta, \hat{\alpha}(\beta), \hat{\Sigma}(\beta)) = L_{max}^{-2/T}(\beta) = \left|\hat{\Sigma}(\beta)\right| = \left|S_{00} - S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10}\right|. \tag{14}$$

Examining this expression and using the invertibility of S_{00} and $\beta' S_{11}\beta$, we have

$$\begin{aligned}
|S_{00} - S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta S_{10}| &= \left|S_{00} - S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta S_{10}\right| |\beta' S_{11}\beta| \,/\, |\beta' S_{11}\beta| = \\
&= \left| \begin{array}{cc} I & S_{01}\beta \\ 0 & \beta' S_{11}\beta \end{array} \right| \left| \begin{array}{cc} S_{00} - S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10} & 0 \\ (\beta' S_{11}\beta)^{-1}\beta' S_{10} & I \end{array} \right| \Big/ |\beta' S_{11}\beta| = \\
&= \left| \begin{array}{cc} S_{00} & S_{01}\beta \\ \beta' S_{10} & \beta' S_{11}\beta \end{array} \right| \Big/ |\beta' S_{11}\beta| = \\
&= \left| \begin{array}{cc} S_{00} & 0 \\ \beta' S_{10} & I \end{array} \right| \left| \begin{array}{cc} I & S_{00}^{-1} S_{01}\beta \\ 0 & \beta' S_{11}\beta - \beta' S_{10} S_{00}^{-1} S_{01}\beta \end{array} \right| \Big/ |\beta' S_{11}\beta| = \\
&= |S_{00}| \left|\beta' S_{11}\beta - \beta' S_{10} S_{00}^{-1} S_{01}\beta\right| \,/\, |\beta' S_{11}\beta| = \\
&= |S_{00}| \left|\beta'(S_{11} - S_{10} S_{00}^{-1} S_{01})\beta\right| \,/\, |\beta' S_{11}\beta| \,. \qquad (15)
\end{aligned}$$

Since $S_{11} - S_{10}S_{00}^{-1}S_{01}$ and S_{11} are both positive finite, symmetric matrices, we can apply 3.5.1, i.e. (15) and therefore, the maximum-likelihood function (14) is maximized by solving the general eigenvalue problem

$$\left|\lambda S_{11} - S_{10} S_{00}^{-1} S_{01}\right| = 0$$

for eigenvectors v_i and eigenvalues λ_i, for $i = 1, \ldots, r$ and putting

$$\hat{\beta} = (v_1, \ldots, v_r)$$

with the maximization value

$$L_{max}^{-2/T} = |S_{00}| \prod_{i=1}^{r} (1 - \lambda_i). \qquad (16)$$

The symmetry and positive finity properties of $S_{11} - S_{10}S_{00}^{-1}S_{01}$ and S_{11} (and the use of a normalization) also provide a simultanous diagonalization for these matrices

$$v_j' S_{11} v_i = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases} \qquad (17)$$

and

$$v_j' S_{10} S_{00}^{-1} S_{01} v_i = \begin{cases} \lambda_i & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases} .$$

Therefore, our choice of $\hat{\beta}$ as the r largest eigenvectors gives us the r-dimensional cointegration space. It shall be noted that we have solved all models $H(r)$, $r = 0, \ldots, p$ with the same eigenvalue calculation.
For the choice of the best hypothesis $H(r)$ we derive the *TRACE-Statistic* through the likelihood ratio test by dividing the expression in (16) for r by the same expression for $r = p$ and taking the logarithm

$$TRACE(r) = -T \sum_{i=r+1}^{p} \log(1 - \hat{\lambda}_j). \qquad (18)$$

3.6 The Varimax Method

So far our matrices α and β are not identified. To see this, suppose that O is an orthogonal $p \times p$-matrix and $\beta^* = \beta O$. Then β^* also fulfills (17) from the last section as we have

$$(\beta^*)' S_{11} \beta^* = O' \beta' S_1 1 \beta O = OIO = I.$$

Therefore, and to allow for a better interpretation of the achieved cointegration relations a so called varimax rotation can be performed. The idea behind applying this method is to rotate the orthogonal basis of the cointegration (sub-)space in a way that it aligns with the coordinates.
The $p \times r$ matrix β containing our cointegration relations shall be called loading matrix A in this context with its members being referred to as loadings a_{ij}.
Then the formal varimax-criterion suggested by Kaiser ([Kai58]) is the maximization of the sum of variances of squared loadings in columns across all columns

$$v^* = \sum_{j=1}^{r} v_j^* = \sum_{j=1}^{r} \frac{p \sum_{i=1}^{p} (a_{ij}^2)^2 - \left(\sum_{i=1}^{p} a_{ij}^2 \right)^2}{p^2}.$$

The original method to determine a solution for the orthogonal rotation matrix O that achieves that AO minimizes the criterion shall be outlined.
The criterion will always be applied to two factors at a time until the process converges.
The orthogonal rotation can be represented as

$$\begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_p & y_p \end{pmatrix} \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} = \begin{pmatrix} X_1 & Y_1 \\ X_2 & Y_2 \\ \vdots & \vdots \\ X_p & Y_p \end{pmatrix}$$

with x_i and y_i, $i = 1 \dots p$ being the present loadings of the two factors, X_i and Y_i being the desired loadings and ϕ being the angle of rotation to be found.
We also have the useful relations

$$\begin{aligned} X_i &= x_i \cos\phi + y_i \sin\phi \\ Y_i &= -x_i \sin\phi + y_i \cos\phi \\ \frac{\mathrm{dX_i}}{\mathrm{d}\phi} &= Y_i \\ \frac{\mathrm{dY_i}}{\mathrm{d}\phi} &= -X_i. \end{aligned}$$

The varimax-criterion is given by finding the maximum for

$$p^2 v^* = p\sum_{i=1}^{p}(X_i^2)^2 - \left(\sum_{i=1}^{p} X_i^2\right)^2 + p\sum_{i=1}^{p}(Y_i^2)^2 - \left(\sum_{i=1}^{p} Y_i^2\right)^2.$$

Taking first derivatives and setting them to zero yields

$$p\sum_{i=1}^{p} X_i Y_i (X_i^2 - Y_i^2) - \sum_{i=1}^{p} X_i Y_i \sum_{i=1}^{p}(X_i^2 - Y_i^2) = 0.$$

With some algebraic manipulation we end up with

$$\tan 4\phi = \frac{2\left(p\sum_{i=1}^{p}(x_i^2 - y_i^2)(2x_i y_i) - \sum_{i=1}^{p}(x_i^2 - y_i^2)\sum_{i=1}^{p}(2x_i y_i)\right)}{p\left(\sum_{i=1}^{p}((x_i^2 - y_i^2) - (2x_i y_i)^2)\right) - \left(\left(\sum_{i=1}^{p}(x_i^2 - y_i^2)\right)^2 - \left(\sum_{i=1}^{p}(2x_i y_i)\right)^2\right)}$$

which gives us the maximized ϕ under the sufficiency conditions for the second derivative. These are summarized in the table below for the numerator and denominator of the right side of the above equation.

		sign of numerator	
		+	-
sign of denominator	+	0° to +22.5°	0° to −22.5°
	-	+22.5° to +45°	−22.5° to −45°

Additionally it was suggested that the maximization should be applied to the normalized loadings, therefore, dividing the loadings first by their squared communalities

$$h_i = \sum_{j=1}^{r} a_{ij}^2$$

which leads to the normalized varimax criterion

$$v^* = \sum_{j=1}^{r} v_j^* = \sum_{j=1}^{r} \frac{p\sum_{i=1}^{p}(\frac{a_{ij}^2}{h_i^2})^2 - \left(\sum_{i=1}^{p}\frac{a_{ij}^2}{h_i}\right)^2}{p^2}.$$

With H^2 being the $p \times p$ diagonal matrix of communalities we act upon $H^{-1}A$ and get the resulting matrix $B^* = H^{-1}AO$ with O being the $r \times r$ orthogonal-transformation matrix determined by the varimax criterion. Then the matrix to be interpreted is the matrix $B = HB^*$.

4 Modelling, Variations, Tests and Results

As a first step – before starting the analysis of the data – we factor the average German inflation rate (represented by the CPI) out of the other five time series. Purely for the cause of a better graphical representation we also divide each of the time series by their mean-value.
Furthermore, a deseasonalization seems to be in line since pellets, oil, gas, wood and the GDP all undergo cyclic changes through the course of a year. We use the Holt-Winters triple exponential smoothing method with a period of 12, $\alpha = \beta = \gamma = 0.05$ as parameters.

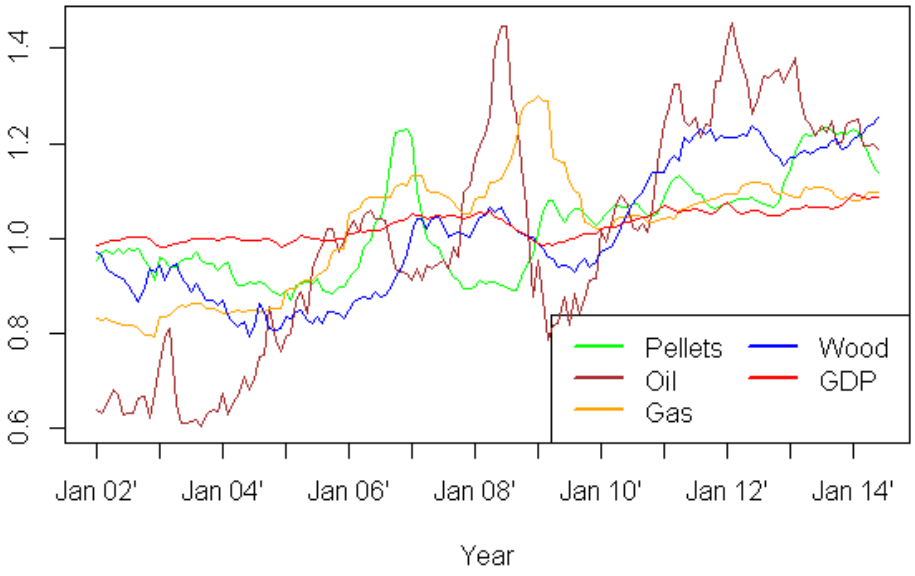


Figure 14: A combined image of all the processed time series.

On the resulting time series we perform the Kwiatkowski–Phillips–Schmidt–Shin and the Augemented-Dickey-Fuller Test to test for level and trend stationarity as well as for unit roots. The results are summarized in the next table.

	p-value KPSS (level)	p-value KPSS (trend)	p-value ADF (trend)
Pellets	< 0.01	< 0.01	0.07422
Oil	< 0.01	0.02046	0.0892
Gas	< 0.01	< 0.01	0.4426
Wood	< 0.01	< 0.01	0.2501
GDP	< 0.01	0.01396	0.2888

Table 3: KPSS and ADF values for the processed time series.

The KPSS null hypotheses implying level and trend stationarity are rejected for all time series. The ADF test does not reject that a unit root is present for gas, wood and oil prices and neither strongly reject it for pellets and oil. These results confirm that we should not invoke a stationary model and instead assume we are dealing with non-stationary, integrated data.

4.1 The General Model

We run a first analysis of the data calculating some basic statistics.

Name	mean	standard deviation	min	max
Pellets levels	1.021853	0.1055988	0.8704	1.2330
Oil levels	1.005213	0.2444876	0.6047	1.4540
Gas levels	1.022721	0.1239374	0.7944	1.3000
Wood levels	1.011749	0.139298	0.7939	1.2550
GDP levels	1.027516	0.03044649	0.9812	1.0940
Pellets increases	0.001263853	0.02090173	-0.10950	0.07677
Oil increases	0.003674951	0.2444876	0.04971851	0.14460
Gas increases	0.001786795	0.01742856	-0.1201000	0.0741500
Wood increases	0.001912349	0.139298	0.01682073	0.05024
GDP increases	0.0006683408	0.03044649	0.00481478	0.01516

Table 4: Some descriptive statistics for the prices and their increases

The purpose of these statistics is to help us find the proper calibrations for the reduced rank regression procedure later.
We, therefore, first consider the Vector Error Correction Model given in 2.1.10 by

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Phi D_t + \varepsilon_t, \quad t = 1, \ldots, T.$$

Before performing the Johansen analysis we have to determine what form of deterministic term we want to use and what lag k fits our data. Therefore, we estimate various ECF through linear regression as described in section 2.2 for lags

$k = 2, \ldots, 12$ with and without a constant and a trend.
The tabulated results are

	best k	AIC	BIC	HQ
ECM	2	-41.65	-40.65	-41.25
ECM with constant	2	-41.73	-40.62	-41.28
ECM with constant and trend	2	-41.75	-40.54	-41.26

Table 5: Information criterions and best lag k according to all criterions in the general model.

In all cases and by all criterions the best fitting lag k seems to be the minimum two. However, while the AIC suggests the usage of constant and trend, the BIC suggests the usage of no deterministic term at all and the HQ suggests only using a constant. We will further conduct a Ljung-Box test to see which of the models features the smallest error correlation at the residues.
In all cases the test results for pellets, oil, gas and wood are satisfying, while the residues for GDP seem to show correlation, especially when performing the test for higher lag-parameters. The reason for this is most likely that the monthly GDP index has been estimated from originally quarterly data. The model without a constant or trend, which is the one supported by the BIC, shows the worst correlation values. Of the remaining two there is no clear better model and therefore, we choose, mainly for reasons of simplicity, the model with a constant but no trend.
It shall be noted that, since there is no clear rule which lag-parameter to test for in the Ljung-Box test tables for lags from $1, \ldots, 20$ are given. Literature on the topic often gives no clear directive, but possible choices include values around 20 as a universal standard or $\ln(T-k)$, which in our case becomes $\ln(148) = 4.997212$.
Finally, for our deterministic term we need to consider section 2.5, that is in which form to include it into the Johansen procedure. Since we have no linear deterministic trends in the data indicated by $\mathbb{E}(\Delta X_T) \approx 0$ we choose Case 2, i.e. we restrict the constant term used in the model to lie in the cointegration space.
Given the above results, for our VECM we reach

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Phi D_t + \varepsilon_t \Leftrightarrow$$
$$\Leftrightarrow \Delta X_t = \alpha(\beta', \rho) \begin{pmatrix} X_{t-1} \\ 1 \end{pmatrix} + \Gamma \Delta X_{t-1} + \varepsilon_t.$$

With the given model we perform the reduced rank regression as described by Johansen. To determine whether and to which degree cointegration is present, we calculate the TRACE-statistic. Calculation is performed by the R-package *urca*. The critical values are also taken from the package.

	test	10pct	5pct	1pct
$r \leq 4$	1.72	7.52	9.24	12.97
$r \leq 3$	8.91	17.85	19.96	24.60
$r \leq 2$	29.42	32.00	34.91	41.07
$r \leq 1$	63.96	49.65	53.12	60.16
$r = 0$	114.63	71.86	76.07	84.45

Table 6: TRACE statistics and critical values for the general model derived by the Johansen test.

The hypothesis of no cointegration, the $r = 0$ case, as well as for a cointegration rank of $r \leq 1$ are discarded using 1%-Quantiles. Hypothesis with cointegration rank $r > 2$ do not make it into any meaningful confidence intervalls. Therefore we deduce a cointegration of $r = 2$ for our model, implying that there are two stationary linear combinations of our data.
The normalized eigenvector cointegration relations β that the Johansen procedure yields are given as

$Pellets$	1.0000	1.0000	1.0000	1.0000	1.0000
Oil	−0.4350	0.3970	2.4037	0.1769	0.7824
Gas	0.6584	−0.2652	−0.3890	−1.5645	−2.0331
$Wood$	0.4815	−1.2932	0.4750	0.8447	0.9115
GDP	−3.3255	1.8562	−22.4195	−5.6427	−2.4183
$constant$	1.6526	−1.7282	19.5126	5.4094	1.5461

and the respective loading matrix α as

$Pellets$	−0.0211	−0.0038	0.0157	−0.0065	−0.0029
Oil	−0.0849	−0.0635	−0.0384	0.0063	−0.0087
Gas	−0.0661	0.0317	0.0063	0.0069	0.0006
$Wood$	0.0077	0.0817	−0.0017	−0.0036	−0.0014
GDP	0.0049	0.0052	0.0019	0.0039	−0.0006.

Notably the constant has become one of our variables in the cointegration space by choice of our deterministic term. From the cointegration relations β we can see for the first $r = 2$ columns what a price steady-state mix for prices could look like. I.e we should have that

$$\begin{aligned}\xi_{1t} = Pellets - 0.4350 \cdot Oil + 0.6584 \cdot Gas + 0.4815 \cdot Wood \\ - 3.3255 \cdot GDP + 1.6526 \\ \xi_{2t} = Pellets + 0.3970 \cdot Oil - 0.2652 \cdot Gas - 1.2932 \cdot Wood \\ + 1.8562 \cdot GDP - 1.7282,\end{aligned}$$

with ξ_{1t} and x_{2t} being stationary processes. We plot the resulting graphs for those two

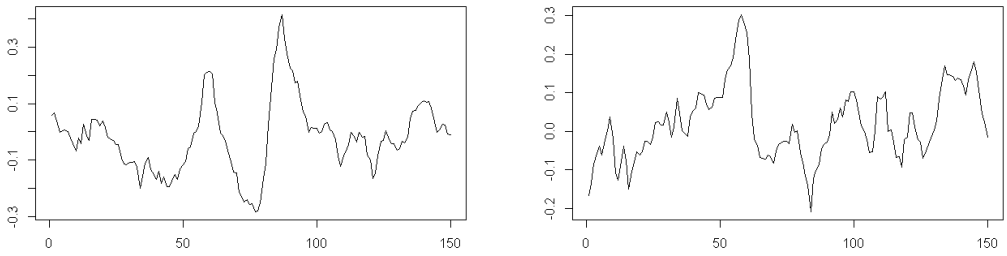


Figure 15: The two steady-state relations in the general model.

and then we test these time series for actual stationarity and unit roots with the KPSS and ADF tests.

	p-value KPSS (Level)	p-value KPSS (Trend)	p-value ADF
ξ_{1t}	> 0.1	> 0.1	0.01453
ξ_{2t}	0.04532	< 0.01	0.04338

Table 7: KPSS and ADF test values for the steady-state relations in the general model.

The results for the first series are satisfying. The null hypotheses of the two KPSS tests, level stationarity and trend stationarity, are both not rejected. The null hypothesis of the ADF test, the presence of a unit root is rejected.
The second relation, however, is less satisfying. The low test values of the KPSS indicate that we are dealing with some form of non-stationarity and suggest that we might have to discard the hypothesis of two cointegration relations in our models and thus, lower the cointegration rank r to one. However, the low ADF value still discards the notion of a unit root.
Furthermore, we want to repeat the above analysis of the Johansen results after transforming the $r = 2$ cointergration relations from the β-matrix into a – hopefully – easier to interprete form through application of the Varimax procedure. The resulting β^* after Varimax transformation is given as

$$\begin{array}{cc} Pellets & 0.0 & 1.411 \\ Oil & -0.589 & 0.0 \\ Gas & 0.670 & 0.235 \\ Wood & 1.215 & -0.654 \\ GDP & -3.724 & -0.798 \\ constant & 2.382 & -0.209 \end{array}$$

or written as (hopefully) stationary time series as

$$\begin{aligned}\zeta_{1t} = -0.589 \cdot Oil + 0.670 \cdot Gas + 1.215 \cdot Wood \\ - 3.724 \cdot GDP + 2.382 \\ \zeta_{2t} = 1.411 \cdot Pellets - 0.235 \cdot Gas - 0.654 \cdot Wood \\ - 0.798 \cdot GDP - 0.209.\end{aligned}$$

This is an interesting result, stating that we have two steady-state relations with only 4 of our original variables. In the one relation pellets, in the other one oil is discarded from the price mix.

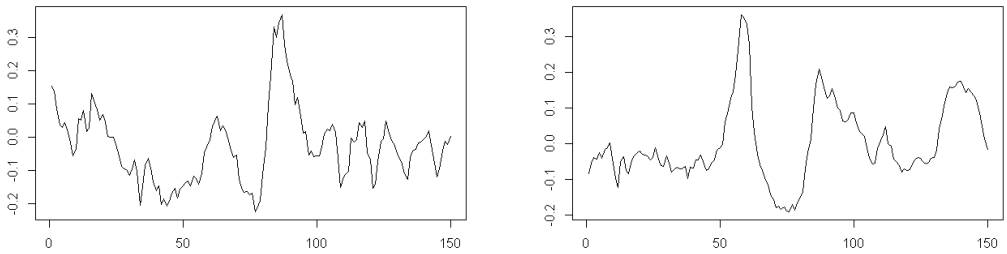


Figure 16: The two steady-state relations in the general model after a varimax transformation was performed.

The according KPSS and ADF test values are:

	p-value KPSS (Level)	p-value KPSS (Trend)	p-value ADF
ζ_{1t}	> 0.1	< 0.01	0.02206
ζ_{2t}	0.03795	> 0.1	0.01712

Table 8: KPSS and ADF test values for the steady-state relations in the general model after a varimax transformation was performed.

Again we have mixed results. For both relations the KPSS values imply some form of non-stationarity, while the unit root hypothesis is rejected again by the ADF test.

We now take a look at the covariance matrix of the errors

$$\hat{\Sigma} = 10^{-4} \begin{pmatrix} 3.1 & -0.9 & 0.5 & 0.3 & 0.2 \\ -0.9 & 22.2 & -0.9 & 0.1 & 0.1 \\ 0.5 & -0.9 & 2.2 & 0.2 & 0.2 \\ 0.3 & 0.1 & 0.2 & 2.2 & 0.1 \\ 0.2 & 0.1 & 0.2 & 0.1 & 0.2 \end{pmatrix}$$

estimated from

$$\hat{\varepsilon}_t = \Delta X_t - \hat{\alpha}(\hat{\beta}', \rho) \begin{pmatrix} X_{t-1} \\ 1 \end{pmatrix} + \hat{\Gamma} \Delta X_{t-1}$$

with $\hat{\alpha}, \hat{\beta} \in \mathbb{R}^{r \times p}$ being the first r-columns of α, β respectively.
Furthermore, we investigate the error plots as well as the autocovariance function (ACF).

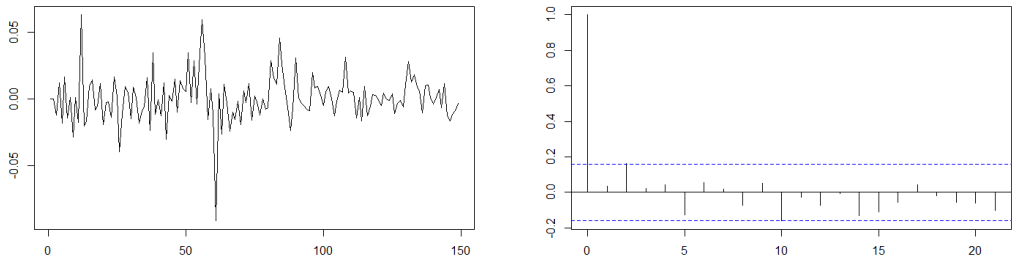


Figure 17: Residuals $\hat{\varepsilon}_{Pellets}$ and their ACF in the general model.

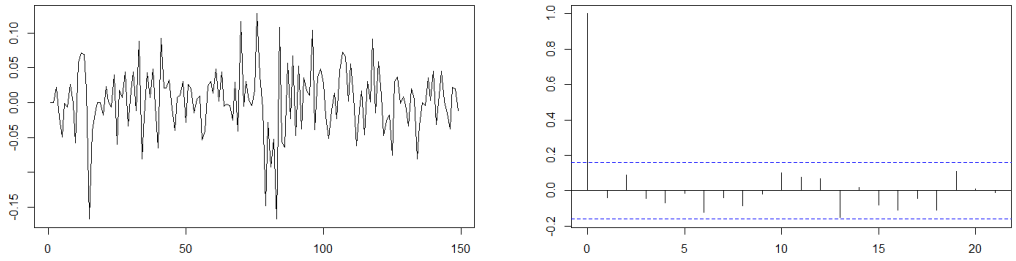


Figure 18: Residuals $\hat{\varepsilon}_{Oil}$ and their ACF in the general model.

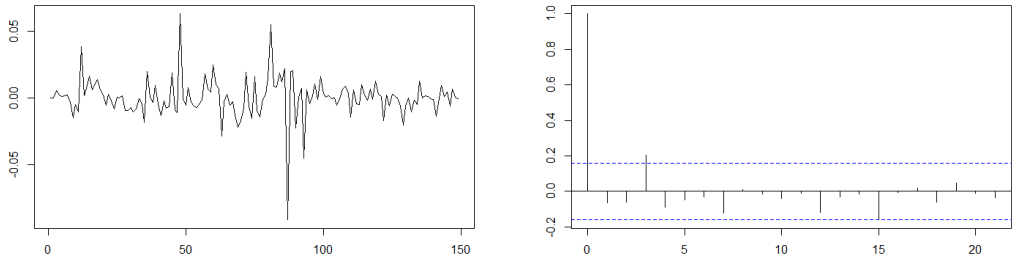


Figure 19: Residuals $\hat{\varepsilon}_{Gas}$ and their ACF in the general model.

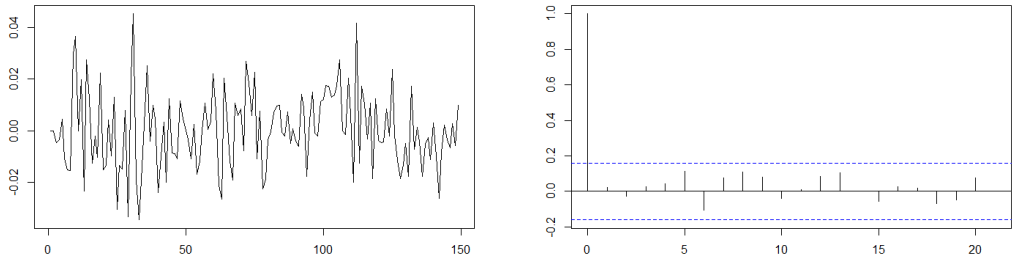


Figure 20: Residuals $\hat{\varepsilon}_{Wood}$ and their ACF in the general model.

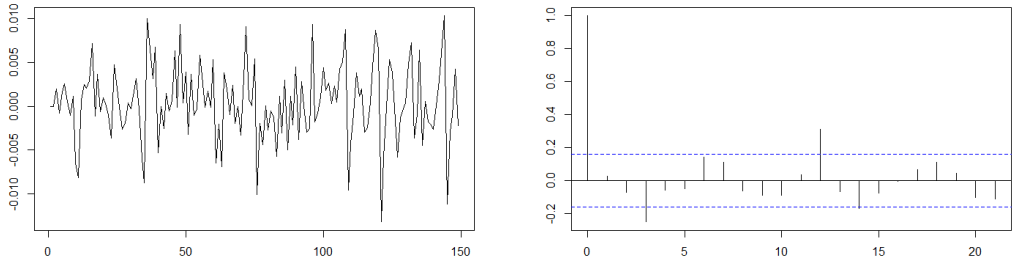


Figure 21: Residuals $\hat{\varepsilon}_{GDP}$ and their ACF in the general model.

The results support the thesis that our residuals are indeed white noise and thus, the chosen r and the consequent model approximate the (differenced) data in satisfying fashion. We, furthermore, derive some statistics for the errors, the Median Absolute Deviation, the Root Mean Square Error and the relative error $\frac{Var(\hat{\varepsilon})}{Var(\hat{X})}$.

	Pellets	Oil	Gas	Wood	GDP
MAD	0.014	0.040	0.009	0.015	0.004
RMSE	0.018	0.047	0.015	0.015	0.004
relative error	0.700	0.896	0.709	0.772	0.776

Table 9: Statistics for the errors in the general model.

Lastly, we attempt a one-step forecast for values $T + m$ with $m = 1, \dots 6$. For the purpose of comparing the forecast we introduce the extended time series X^* that is the original time series X extended by six data points. Again, we process the data as described at the start of chapter. Note that we do not recalculate a new deseasonalization, since thereby we would use the future data $X_{T+1}, \dots X_{T+6}$ for re-processing our dataset, but instead use the factors $c_{t-L+1+(m-1) \mod L}$ as described in section 3.2 for $t = T$ and $L = 12$.
We take the prediction errors and get the statistics

	Pellets	Oil	Gas	Wood	GDP
MAD	0.005	0.030	0.008	0.009	0.003
RMSE	0.009	0.073	0.007	0.007	0.004
relative error	0.020	1.396	0.173	0.127	0.518

.

Table 10: Statistics for the prediction errors in the general model.

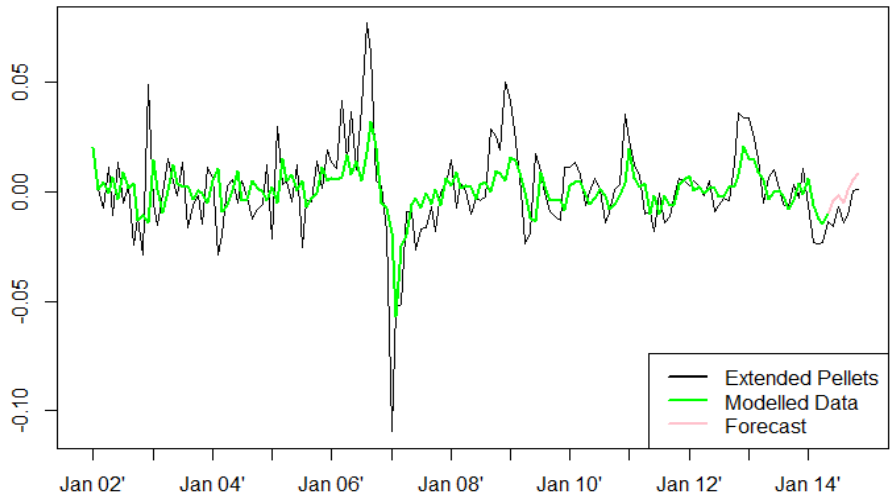


Figure 22: The real pellets price growth rates, the modelled ones and the forecast extending the data period in the general model.

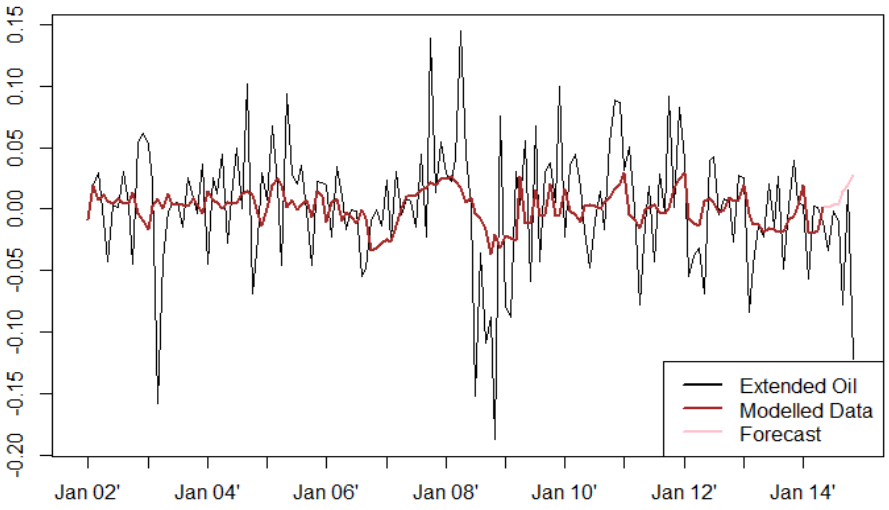


Figure 23: The real oil price growth rates, the modelled ones and the forecast extending the data period in the general model.

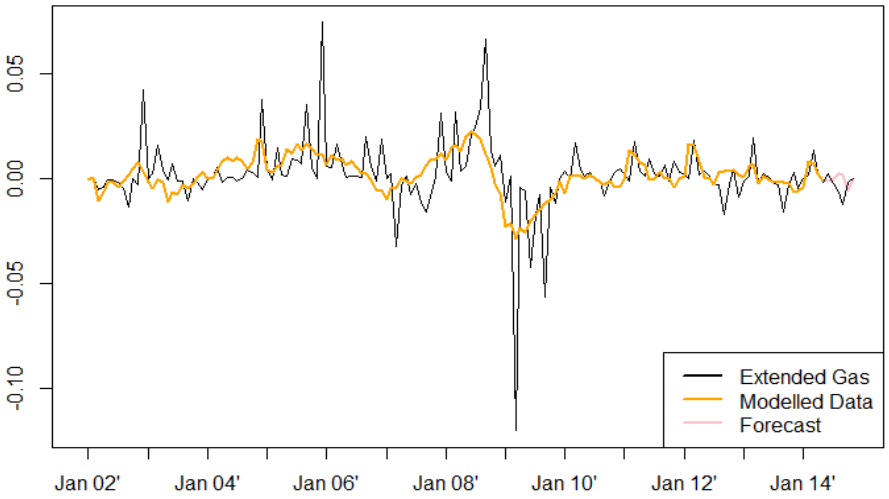


Figure 24: The real gas price growth rates, the modelled ones and the forecast extending the data period in the general model.

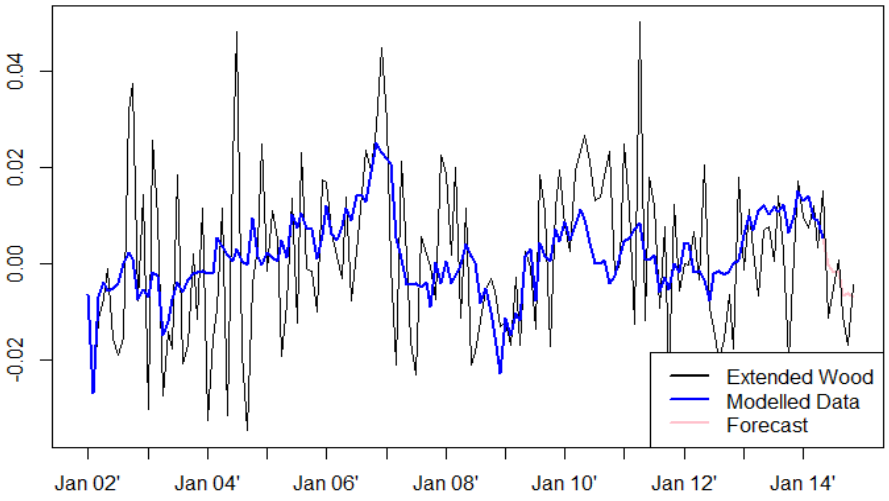


Figure 25: The real wood price growth rates, the modelled ones and the forecast extending the data period in the general model.

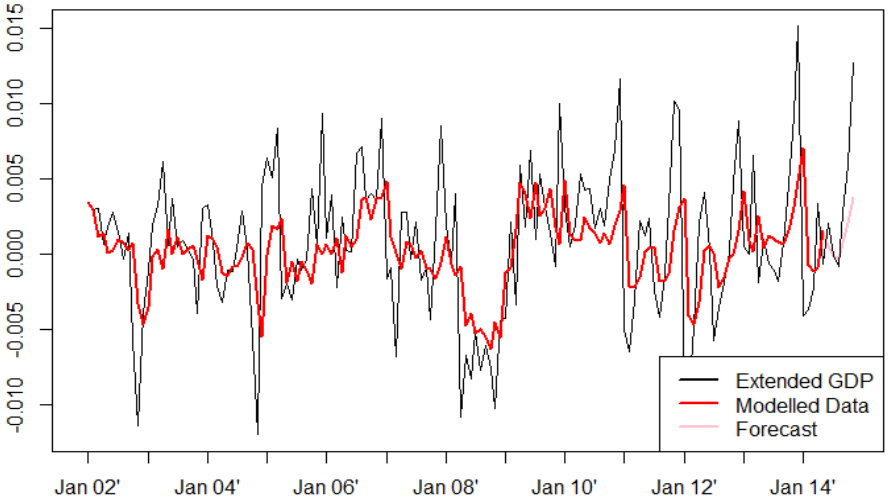


Figure 26: The real GDP growth rates, the modelled ones and the forecast extending the data period in the general model.

4.2 Discussion of the Information Set

Now that we've run the analysis with a model including all our data we can draw first conclusions. The TRACE test and the post-stationarity analysis have shown that this model has cointegration rank $r \leq 2$, hence a maximum of 2 cointegration relations can be reached. Given that the original data is not stationary, we could reach up to $r = 4$ cointegration relations with our original set containing pellets, oil, gas, wood and GDP indices. The question arises whether all of these time series are needed in the analysis and whether some of them are even holding the model back from achieving better results.
As a first determination we perform the Johansen procedure on all pairs of our data, determining the lag-parameter k by lowest AIC-value and using a constant restricted to the cointegration space without running a deeper analysis into the case determination since we are only interested in finding rough connections here. The according 10%−, 5%− and 1%− quantiles of the TRACE test are reached when the test statistic exceeds the following critical values

quantile	10%	5%	1%
critical value	17.85	19.96	24.60.

Table 11: The criticial values of the TRACE-test under the hypothesis r=0.

The test values and the used lags are given as

	Pellets		Oil		Gas		Wood		GDP	
Pellets	–		17.95	3	12.32	4	17.09	3	17.55	2
Oil	17.95	3	–		11.88	6	12.62	3	11.79	2
Gas	12.32	4	11.88	6	–		8.20	4	14.75	4
Wood	17.09	3	12.62	3	8.20	4	–		23.76	2
GDP	17.55	2	11.79	2	14.75	4	23.76	2	–	

Table 12: The TRACE-test values for pairs of data.

The test value for pellets with oil, wood and GDP all get close to or in the case of oil even exceeds the 10%-quantile critical value. Furthermore, the result suggests a strong relation of wood prices and GDP. Gas prices show no connection to the other time series.
In terms of used lags the gas prices also seem to be an outlier. The VECM pairs including was prices all require higher lags $k = 4$ or $k = 6$ to describe the data best according to the AIC criterion, while all other models only have a lag of $k = 2$ or $k = 3$.
A similar analysis dropping more than one variable and working only with triples results in $r \leq 1$ for all models. We decide to run the in-depth analysis as given in the previous chapter for the general model for a model only containing the pellets, oil, wood and GDP indices, hoping to reach a similar cointegration rank $r \approx 2$ as in the general model without gas.

4.3 The Model without Gas

We start off again with the vector error correction model given in 2.1.10 by

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Phi D_t + \varepsilon_t, \quad t = 1, \dots, T.$$

Our lag determination k and the according usage of a deterministic term work as previously. We estimate the various forms of VECM through linear regression as described in section 2.2 for lags $k = 1, \dots, 12$ with and without a constant and a trend.

	best k	AIC	BIC	HQ
ECM	2	-33.26	-32.62	-33.00
ECM with constant	2	-33.35	-32.63	-33.06
ECM with constant and trend	2	-33.36	-32.56	-33.04

Table 13: Information criterions and best lag k according to all criterions in the model without gas.

In all cases and by all criterions the best fitting k seems to be the minimum two, similar to the general model. The minimum BIC and HQ are achieved when including a constant, but not a trend, while the AIC suggests using a trend as well. We further conduct the Ljung-Box test to check for error correlation in all three models. In all three models the results for pellets, oil and wood suggest no correlation, while, once again, for certain greater lag-calibrations of the used Box-Ljung Test the residuals of the GDP seem to be correlated. The VECM with a constant but without a trend (and lag 2) has the cleanest results and therefore, we will use it.
We next consider the 5 Cases from section 2.5 for our inclusion of the deterministic term in the Johansen procedure. We use the same argumentation as previously: Since we have no linear deterministic trends in the data indicated by $\mathbb{E}(\Delta X_T) \approx 0$ we choose Case 2, i.e. we restrict the constant term used in the model to lie in the cointegration space.

Our VECM once again becomes

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Phi D_t + \varepsilon_t \Leftrightarrow$$
$$\Leftrightarrow \Delta X_t = \alpha(\beta', \rho) \begin{pmatrix} X_{t-1} \\ 1 \end{pmatrix} + \Gamma \Delta X_{t-1} + \varepsilon_t.$$

Now that we have determined all the parameters we can perform the reduced rank regression and give the according results like TRACE-statistic, the possible

cointegration relations described by the normalized eigenvector matrix β and their loading matrix α.

	test	10pct	5pct	1pct
$r \leq 3$	2.19	7.52	9.24	12.97
$r \leq 2$	19.83	17.85	19.96	24.60
$r \leq 1$	41.16	32.00	34.91	41.07
$r = 0$	74.62	49.65	53.12	60.16

Table 14: TRACE statistics and critical values for the general model derived by the Johansen test.

The TRACE test gives us the result we hoped for rejecting hypothesis on the cointegration rank for $r = 0$ and $r \leq 1$. Additionally, to the retention of these two statements, the test suggests that there could be an additional third steady state relation possible by rejecting the hypothesis $r \leq 2$ at the 10%-quantile and only barely not reaching the critical value at 5%. It is in any case keeping the possibility of $r = 3$ in mind in the further analysis.
The normalized eigenvector cointegration relations β deduced by the reduced rank regression are

$$\begin{array}{rrrrr} Pellets & 1.0000 & 1.0000 & 1.0000 & 1.0000 \\ Oil & 0.2297 & 1.0553 & -1.2970 & 0.1693 \\ Wood & -0.9453 & 1.2326 & 1.6852 & 1.6047 \\ GDP & 0.6251 & -17.1226 & -1.9944 & 4.3744 \\ constant & -0.9243 & 14.2479 & 0.6240 & -8.1284 \end{array}$$

and the respective loading matrix α as

$$\begin{array}{rrrrr} Pellets & -0.0110 & 0.0162 & -0.0191 & -0.0012 \\ Oil & -0.1059 & -0.0618 & 0.0110 & -0.0039 \\ Wood & 0.0897 & -0.0083 & -0.0081 & -0.0004 \\ GDP & 0.0097 & 0.0040 & 0.0033 & -0.0004. \end{array}$$

Once again we have one extra relation created by including the constant as part of the cointegration space in our choice of the deterministic term. The up to 3 possible cointegration relations for steady state price mixes are given as the processes

$$\begin{aligned} \xi_{1t} &= Pellets + 0.2297 \cdot Oil - 0.9453 \cdot Wood \\ &\quad + 0.6251 \cdot GDP - 0.9243 \\ \xi_{2t} &= Pellets + 1.0553 \cdot Oil + 1.2326 \cdot Wood \\ &\quad - 17.1226 \cdot GDP + 14.2479 \\ \xi_{3t} &= Pellets - 1.2970 \cdot Oil + 1.6852 \cdot Wood \\ &\quad - 1.9944 \cdot GDP + 0.6240. \end{aligned}$$

The according plots are given

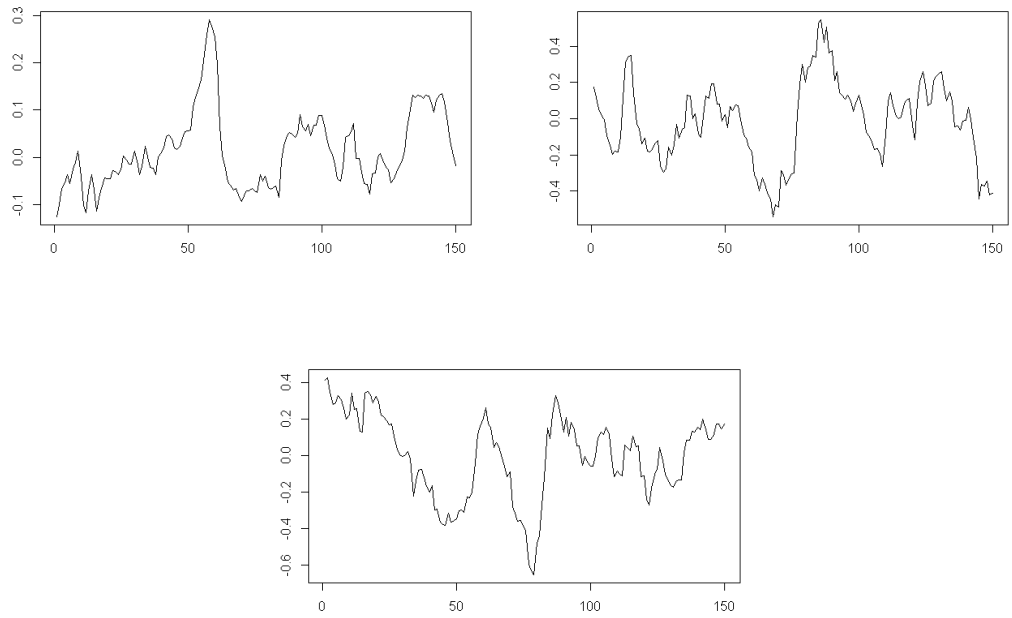


Figure 27: The three steady-state relations in the model without gas.

and then we test these time series for actual stationarity and unit roots with the KPSS and ADF tests.

	p-value KPSS (Level)	p-value KPSS (Trend)	p-value ADF
ξ_{1t}	0.0232	0.02083	0.03803
ξ_{2t}	> 0.1	0.03517	0.4176
ξ_{3t}	0.03628	< 0.01	0.09535

Table 15: KPSS and ADF test values for the steady-state relations in the model without gas.

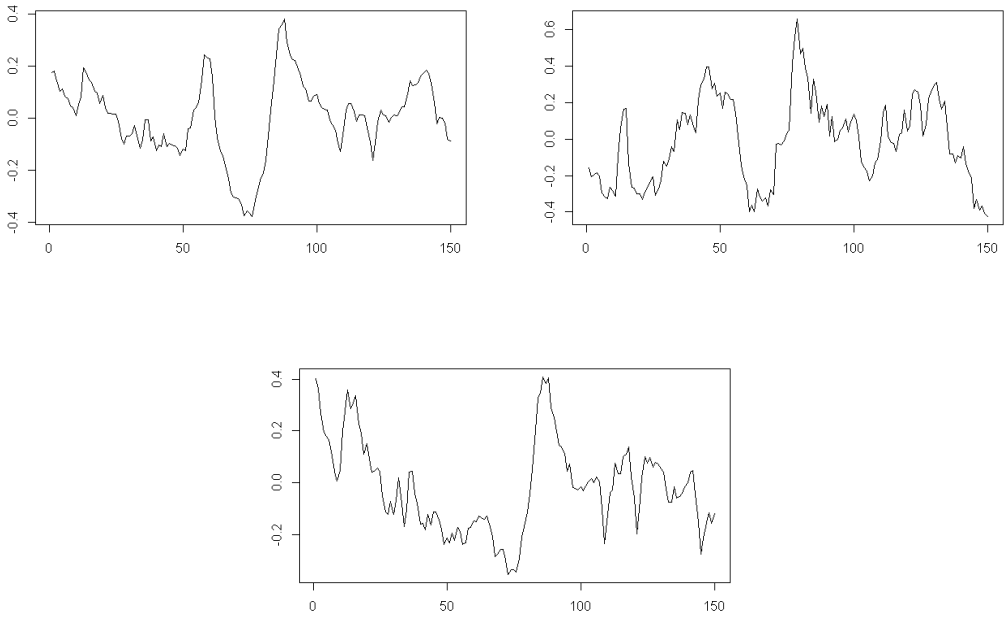
The KPSS p-values for the most part suggest that the null hypothesis of stationarity or trend stationarity should be rejected. At least the ADF test for the first relation also rejects the null of a unit root.
We apply the Varimax transformation to the 3 relations and attain for the transformed relations β^*

$Pellets$	1.661	0.128	0.473
Oil	0.0	1.664	-0.278
$Wood$	0.583	-0.275	2.199
GDP	-7.155	-11.287	-10.907
$constant$	5.126	10.062	8.759

i.e.

$$\zeta_{1t} = 1.661 \cdot Pellets + 0.583 \cdot Wood - 7.155 \cdot GDP + 5.126$$
$$\zeta_{2t} = 0.128 \cdot Pellets + 1.664 \cdot Oil - 0.275 \cdot Wood - 11.287 \cdot GDP + 10.062$$
$$\zeta_{3t} = 0.473 \cdot Pellets - 0.278 \cdot Oil + 2.199 \cdot Wood - 10.907 \cdot GDP + 8.759$$

and their graphs



.

Figure 28: The three steady-state relations in the model without gas after a varimax transformation was performed.

Only one of the variables has been rotated to 0 and also the KPSS and ADF test statistics do not support this as a good model. In fact, the results are even less probable to be stationary than the ones before the transformation.

	p-value KPSS (Level)	p-value KPSS (Trend)	p-value ADF
ζ_{1t}	> 0.1	< 0.01	0.04031
ζ_{2t}	0.07799	< 0.01	0.4098
ζ_{3t}	0.05234	< 0.01	0.2056

Table 16: KPSS and ADF test values for the steady-state relations in the model without gas after a varimax transformation was performed.

Since the choice of rank r matters for the results of our Varimax rotation, we repeat it under the hypthesis of $r = 2$ instead of the previous one stating $r = 3$.

Here the relations β^* are given as

$Pellets$	1.409	0.126
Oil	0.853	0.662
$Wood$	0.0	1.552
GDP	-10.506	-13.535
$constant$	8.432	11.522.

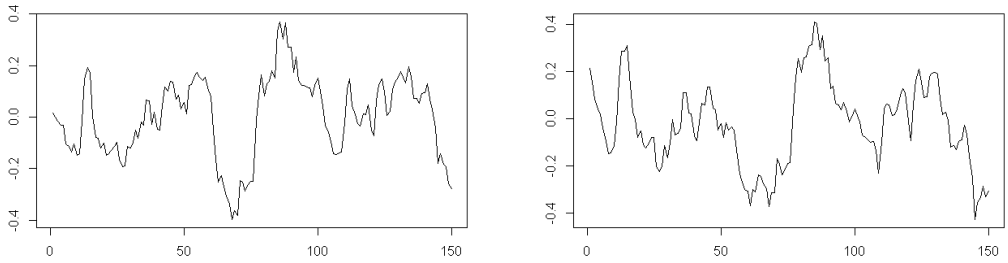


Figure 29: The two steady-state alternative in the model without gas after a varimax transformation was performed.

or written in process form as

$$\eta_{1t} = 1.409 \cdot Pellets + 0.853 \cdot Oil - 10.506 \cdot GDP + 8.432$$
$$\eta_{2t} = 0.126 \cdot Pellets + 0.662 \cdot Oil - 1.552 \cdot Wood - 13.535 \cdot GDP + 11.522$$

with one variable rotated to 0. This time the results of the stationarity tests are much better, at least for the KPSS tests of the first relation ζ_{1t}.

	p-value KPSS (Level)	p-value KPSS (Trend)	p-value ADF
η_{1t}	> 0.1	> 0.1	0.2983
η_{2t}	> 0.1	0.01431	0.4164

Table 17: KPSS and ADF test values for the steady-state alternatives in the model without gas after a varimax transformation was performed.

With these results we tend towards the cointegration rank $r = 2$ in this model without gas. Based on that hypothesis we will now also perform tests on the errors and make a prediction for the time series that exceeds the original time period.
The covariance matrix of the errors is given by

$$\hat{\Sigma} = 10^{-4} \begin{pmatrix} 2.9 & -0.2 & 0.4 & 0.1 \\ -0.2 & 21.1 & -0.1 & 0.2 \\ 0.4 & -0.1 & 2.2 & 0.1 \\ 0.1 & 0.2 & 0.1 & 0.2 \end{pmatrix}$$

again, estimated from

$$\hat{\varepsilon}_t = \Delta X_t - \hat{\alpha}(\hat{\beta}', \rho) \begin{pmatrix} X_{t-1} \\ 1 \end{pmatrix} + \hat{\Gamma} \Delta X_{t-1}$$

with $\hat{\alpha}, \hat{\beta} \in \mathbb{R}^{r \times p}$ being the first r-columns of α, β respectively. The according error plots and their ACF are investigated.

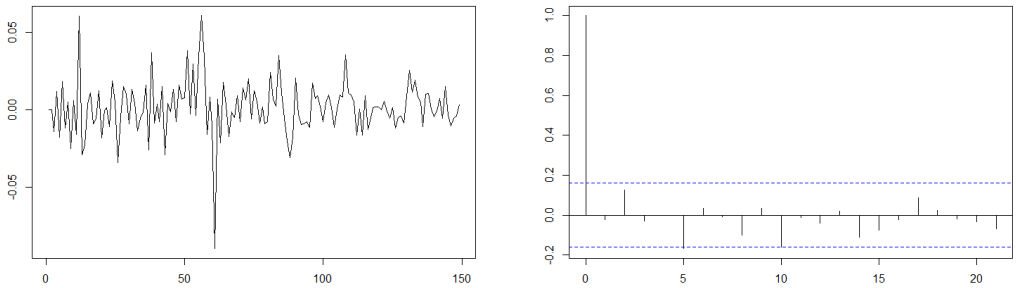


Figure 30: Residuals $\hat{\varepsilon}_{Pellets}$ and their ACF in the gasless model.

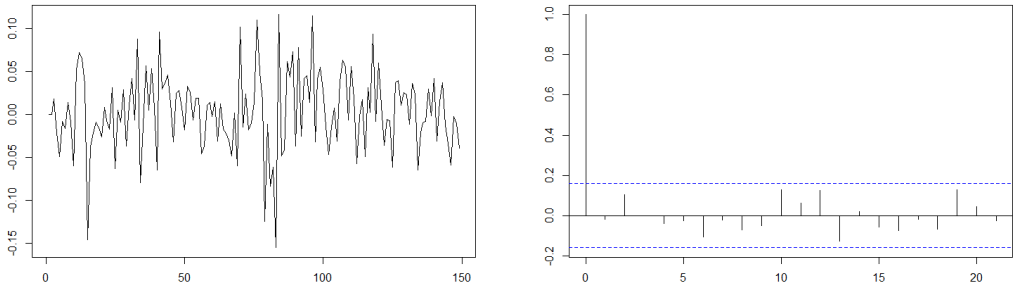


Figure 31: Residuals $\hat{\varepsilon}_{Oil}$ and their ACF in the gasless model.

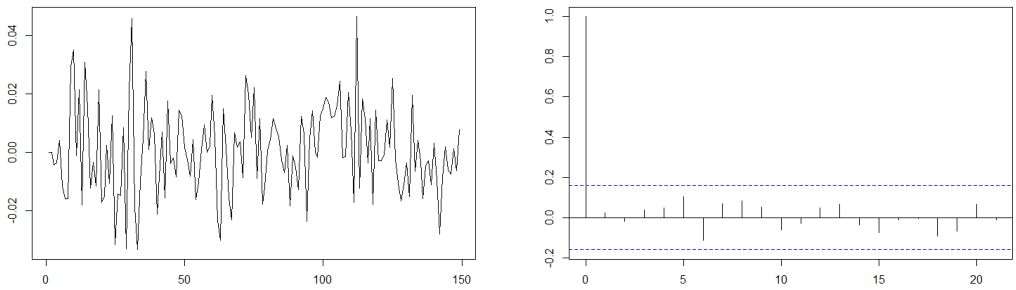


Figure 32: Residuals $\hat{\varepsilon}_{Wood}$ and their ACF in the gasless modell.

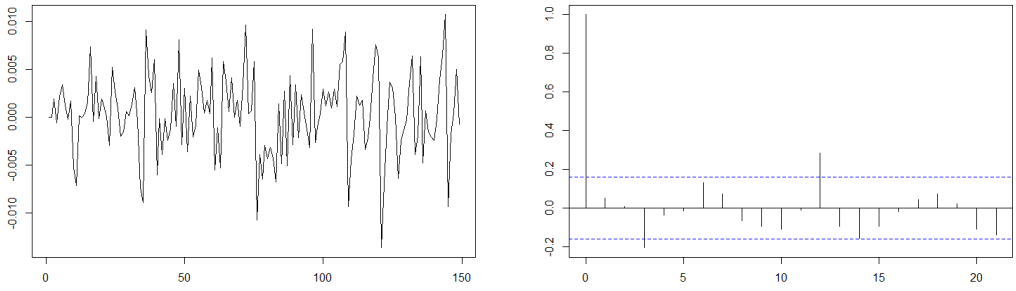


Figure 33: Residuals $\hat{\varepsilon}_{GDP}$ and their ACF in the gasless model.

Again, the ACF for GDP as well as the covariance matrix show slight dependencies (unsurprisingly given the underlying interpolation in that time series), otherwise the graphs and the covariance matrix support the hypothesis of white noise residuals.

	Pellets	Oil	Wood	GDP
MAD	0.013	0.041	0.016	0.003
RMSE	0.017	0.046	0.015	0.004
relative error	0.672	0.850	0.790	0.785

Table 18: Statistics for the errors in the gasless model.

Finally, we also attempt a one-step forecast for time values $T + m$ with $m = 1, \ldots 6$. Therefore, just like in the general model we introduce the extended time series X^* (in this model obviously without gas), which is the the original time series X extended by six data points and process it for the purpose of comparing to our forecast.
For the prediction errors we derive the same statistics as for the modelled data

	Pellets	Oil	Wood	GDP
MAD	0.006	0.021	0.007	0.003
RMSE	0.005	0.099	0.010	0.006
relative error	0.055	1.435	0.111	0.725

Table 19: Statistics for the prediction errors in the gasless model.

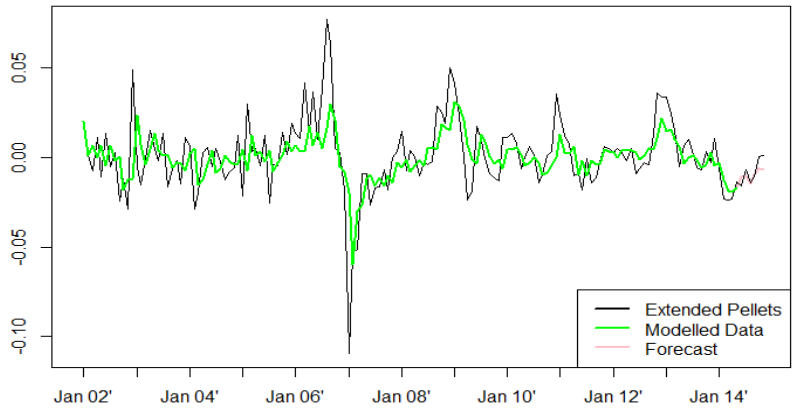


Figure 34: The real pellet sprice growth rates, the modelled ones and the forecast extending the data period in the gasless model.

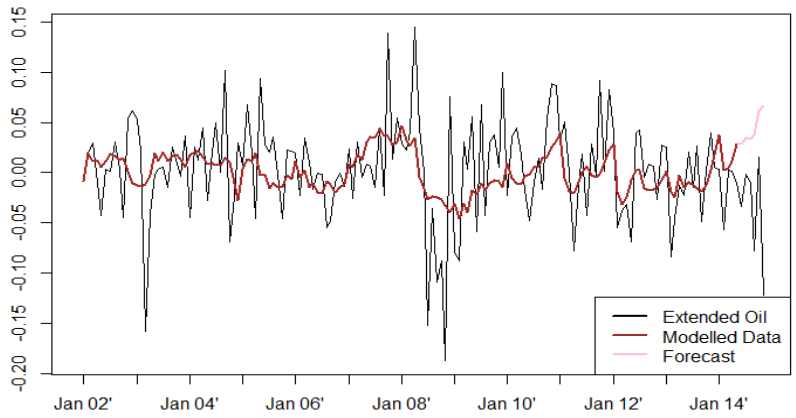


Figure 35: The real oil price growth rates, the modelled ones and the forecast extending the data period in the gasless model.

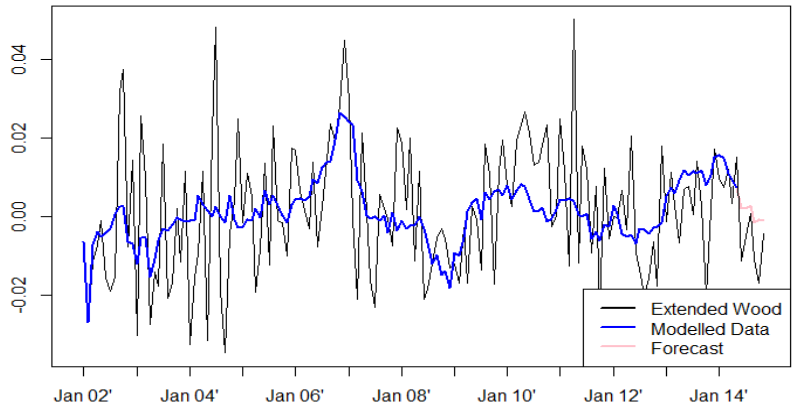


Figure 36: The real wood price growth rates, the modelled ones and the forecast extending the data period in the gasless model.

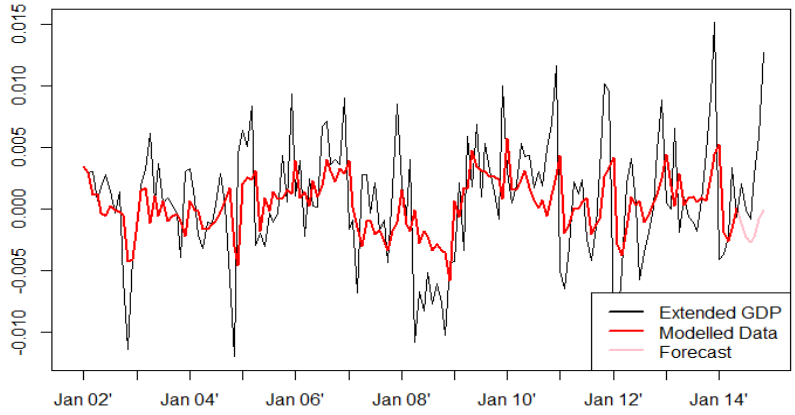


Figure 37: The real GDP growth rates, the modelled ones and the forecast extending the data period in the gasless model.

4.4 Conclusions

The investigated time series seem to be indeed integrated without a deterministic trend. The results from the various investigated models show us that we can find multiple steady-state relations according to our strongest criterion, the TRACE test, from the data. Their actual stationarity is however of questionable nature according to the performed KPSS and ADF tests. Also a graphical investigation showcases that often characteristic phases within the data can be hard to eliminate. The characteristic up and downswings – particulary of oil and pellets – seem to be reproduced in some of our cointegration relations. However, a certain degree of cointegration seems to be present in the time series.
In the derived relations we see a clear trend that the parameter for GDP often greatly overshadows other parameters, sometimes even by a factor of 100. At this point it must be noted that the work was done with indices for the prices and their movements that were derived by dividing the research data through the respective average for the investigated time period. Therefore, as can be seen in figure 14, the actual levels of the used time series are close to each other in size. This means we can safely say from the observation on the parameter for the GDP that in basically all models it seems to be by far the most influential factor. On the other hand, the connection between gas and the other parameters has been proven to be so weak that removing it from the information set used can actually improve the results in one way or another.
The derived statistics in both models, the general one and the one without gas, seem to be pretty similar for the residuals as well as for the prediction errors. Significant price swings, for example the price drop of oil beginning in July 2014 (and ongoing to this day), cannot be properly predicted (see figure 23 and figure 35), otherwise the statistics let us conclude that both models produce reasonable forecasts and models. We can conclude that the inclusion of gas does not grant significant advantages for the prediction of pellets prices, while we have found significant relations with oil and wood prices, as well as GDP.

List of Figures

List of Tables

A References

[And51] T.W. Anderson. *Estimating linear restrictions on regression coefficients for multivariate normal distributions.* Annals of Mathematics Statistics, 1951.

[Bun15a] Statistisches Bundesamt. *Anteil der Wirtschaftsbereiche am Bruttoinlandsprodukt (BIP)* in Deutschland im Jahr 2014.* 2015. [Online; accessed 16-November-2015].

[Bun15b] Bundesverband der Energie- und Wasserwirtschaft e.V. Bdew zur weltleitmesse ish: Erdgas weiterhin beliebtester heizenergieträger. https://bdew.de/, 2015. Anhang zur BDEW Pressemeldung.

[DF79] D.A. Dickey and W.A. Fuller. *Distribution of the Estimators for Autoregressive Time Series with a Unit Root.* Journal of the American Statistical Association, 1979.

[DS10] M. Deistler and W. Scherrer. *Stationary Processes and Time Series Analysis; lecture notes.* 2010.

[EG87] Robert F Engle and Clive W. J. Granger. *Co-integration and error correction: Representation, estimation and testing.* Econometrica 55, 1987.

[Eur15] Eurostat. Real gdp growth rate - volume, 2015. [Online; accessed 16-November-2015].

[GN74] Clive Granger and Paul Newbold. *Spurious Regressions in Econometrics.* Journal of Econometrics 2, 1974.

[Han04] Peter Reinhard Hansen. *Granger's representation theorem: A closed-form expression for I(1) processes.* 2004.

[HJ00] David F. Hendry and Katarina Juselius. *Explaining Cointegration Analysis: Part II.* Nuffield College, Oxford, OX1 1NF. Department of Economics, University of Copenhagen, Denmark, 2000.

[Hol57] Charles C. Holt. *Forecasting Trends and Seasonals by Exponentially Weighted Averages.* Carnegie Institute of Technology, 1957.

[Hub56] M. King Hubbert. *Nuclear Energy and the Fossil Fuels.* 1956.

[Kai58] Henry F. Kaiser. *The Varimax Criterion for Analytic Rotation in Factor Analysis.* 1958.

[Lü91] H. Lütkepohl. *Introduction to Multiple Time Series Analysis.* Springer-Verlag, New York, 1991.

[Neu11] Klaus Neusser. *Zeitreihenanalyse in den Wirtschaftswissenschaften.* Springer-Verlag, 2011.

[rJ94] Søren Johansen. *The role of the constant and linear terms in cointegration analysis of non-stationary variables.* Econometric Reviews, 13: 205–29., 1994.

[rJ95] Søren Johansen. *Likelihood-Based Interference In Cointegrated Vector Auto-Regressive Models.* Oxford University Press, New York, 1995.

[Sta15] Statistisches Bundesamt. Anzahl der pelletheizungen in deutschland in den jahren 2012 bis 2015. http://de.statista.com/statistik/daten/studie/171886/umfrage/anzahl-der-pelletheizungen-in-deutschland/, 2015. [Online; accessed 16-November-2015].

[uSF14] Melanie Zenker und Stefan Friedrich. Der holzpelletmarkt in bayern, 2014.

[Wik15a] Wikipedia. Holzpellet — Wikipedia, the free encyclopedia, 2015. [Online; accessed 16-November-2015].

[Wik15b] Wikipedia. Ölpreisbindung — Wikipedia, the free encyclopedia, 2015. [Online; accessed 16-November-2015].

[Win60] P. R. Winter. *Forecasting Sales by Exponentially Weighted Moving Averages.* Carnegie Institute of Technology, 1960.

B Program Codes for R

B.1 R Libraries

```
#Package used for vector autoregressive modelling
library(vars)
#Package that contains matrix operations
library(Matrix)
#Package for cointegration processing such as the Johanson
    Approach
library(urca)
#Package for calculation of generalized Eigenvectors and -
    values
library(geigen)
#Package that contains the ADF and KPSS Tests
library(tseries)
#Package that contains the LjungBox Test
library(portes)
# Package that includes the calculation of RMSE
library(hydroGOF)
```

B.2 Functions

Quarterly to Monthly

```
q2m <- function(data, rand) {
  Q=length(data)
  T = Q*3
  Sx=data
  sig=var(diff(data,1))[1]

  S=matrix(0, ncol=T, nrow=Q)
  for (i in 1:Q) {
    S[i,(1+3*(i-1))]=1
  }

  Var = matrix(0, ncol=T, nrow=T)
  for (i in 1:T) {
    Var[i,i:T]=Var[i,i:T]+i
    Var[i:T,i]=t(Var[i,i:T])
  }
  Var=Var
  VarS = Var
  for (i in T:1) {
    if ((i%%3) != 1) {
      VarS=VarS[-i,]
      VarS=VarS[,-i]
    }
  }
  c=Var%*%t(S)%*%solve(VarS)
  monthly=vector(length=Q)
  monthly=c%*%data
  if (rand==TRUE) {
    monthly=monthly+rnorm(T,0,3*sig)
  }
  return(monthly)
}
```

Triple Exponential Smoothing

```
tripexp <- function(data, L, alpha=0.05, beta=0.05, gamma
   =0.05) {
  T=length(data)
  N=floor(T/L)
  A=numeric(N)
  for (j in 1:N) {
    sum=0
    for (i in 1:L) {
      sum=sum+data[L*(j-1)+i]
    }
    A[j]=sum/L
  }
  s=numeric(T)
  b=numeric(T)
  c=numeric(T)
  s[1:(2*L)]=data[1:(2*L)]

  sum=0
  for (i in 1:L) {
    sum=sum+(data[L+i]-data[i])/L
  }
  b[1]=1/L*sum
  for (i in 1:L) {
    sum=0
    for (j in 1:N) {
      sum=sum+data[L*(j-1)+i]/A[j]
    }
    c[i]=1/N*sum
  }
  for (t in 2:L) {
    b[t]=beta*(s[t]-s[t-1])+(1-beta)*b[t-1]
  }
  for (t in (L+1):(2*L)) {
    b[t]=beta*(s[t]-s[t-1])+(1-beta)*b[t-1]
    c[t]=gamma*data[t]/s[t]+(1-gamma)*c[t-L]
  }
  for (t in (2*L+1):T) {
    s[t]=alpha*(data[t]/c[t-L])+(1-alpha)*(s[t-1]+b[t-1])
    b[t]=beta*(s[t]-s[t-1])+(1-beta)*b[t-1]
    c[t]=gamma*data[t]/s[t]+(1-gamma)*c[t-L]
  }
  return(c)
}
```

VECM Estimation and Information Criteria

```
myVARJ <- function(data,k,const=c(TRUE, FALSE),trend=c(
    TRUE, FALSE), rvalue=c("Info", "eps")) {
  T=dim(data)[1]
  p=dim(data)[2]
  ddata=rbind(0,diff(data,1))
  Y=t(ddata[(k+1):T,])
  Z=matrix(ncol=(T-k),nrow=const+trend+(p*k))
  Z[1:p,]=t(data[k:(T-1),])
  for (i in 1:(k-1)) {
    Z[(i*p+1):((i+1)*p),]=t(ddata[(k+1-i):(T-i),])
  }
  if (const==TRUE) {
    Z[k*p+1,]=1
  }
  if (trend==TRUE) {
    Z[k*p+2,]=(k+1):T
  }
  B=Y%*%t(Z)%*%solve((Z%*%t(Z)))
  estim = matrix(0,nrow=p, ncol=T)
  estim[,1:k]=t(ddata[1:k,])
  estim[,(k+1):T]=B%*%Z
  eps=matrix(nrow=p,ncol=T)
  eps[,1:k]=0
  eps[,(k+1):T]=t(ddata[(k+1):T,])-estim[,(k+1):T]
  sum = 0
  for (i in (k+1):T) {
    sum = sum + (eps[,i])%*%t(eps[,i])
  }
  Sigk = 1/(T-k)*sum
  AIC = log(det(Sigk)) + 2/T*(k*p+const+trend)*p
  BIC = log(det(Sigk)) + log(T)/T*(k*p+const+trend)*p
  HQ = log(det(Sigk)) + (2*log(log(T)))/T*(k*p+const+trend
      )*p
  if (rvalue=="Info") {
    return(cbind(k, AIC, BIC, HQ))
  } else {
    return(eps)
  }
}
```

Alternative Cointegration Program and TRACE

```
TRACE <- function(r,vect, T) {
  n=length(vect)
  sum=0
  for (i in (r+1):n) {
    sum=sum-T*log(1-vect[i])
  }
  return(sum)
}

cointegrate <- function(x, k,const=c(TRUE, FALSE),trend=c(
   TRUE,FALSE), cconst=c(TRUE,FALSE), ctrend=c(TRUE,FALSE)
   ) {
  T = dim(x)[1]
  p = dim(x)[2]

  Z0=diff(x,lag=1)[k:(T-1),]

  Z1=x[k:(T-1),]
  ######################################
  LaggedL=list()
  for (i in 1:(k-1)) {
    LaggedL[[i]]=x[(k-i):(T-i),]
  }
  Z2=diff(LaggedL[[1]],1)
  if (k>2) {
    for (i in 2:(k-1)) {
      Z2 = cbind(Z2, diff(LaggedL[[i]],1))
    }
  }
  if (trend == TRUE) {
    Z2 = cbind(k:(T-1), Z2)
  }
  if (const==TRUE) {
    Z2 = cbind(numeric(T-k)+1,Z2)
  }
  if (cconst == TRUE) {
    Z1 = cbind(Z1,numeric(T-k)+1)
  }
  if (ctrend==TRUE) {
    Z1 = cbind(Z1,k:(T-1))
  }

  T1=T-k
```

```
M00=Z0[1,]%*%t(Z0[1,])
M01=Z0[1,]%*%t(Z1[1,])
M11=Z1[1,]%*%t(Z1[1,])
M02=Z0[1,]%*%t(Z2[1,])
M12=Z1[1,]%*%t(Z2[1,])
M22=Z2[1,]%*%t(Z2[1,])
for (i in 2:(T1)) {
  M00 = M00 + Z0[i,]%*%t(Z0[i,])
  M01 = M01 + Z0[i,]%*%t(Z1[i,])
  M11 = M11 + Z1[i,]%*%t(Z1[i,])
  M02 = M02 + Z0[i,]%*%t(Z2[i,])
  M12 = M12 + Z1[i,]%*%t(Z2[i,])
  M22 = M22 + Z2[i,]%*%t(Z2[i,])
}
M00=(1/(T1))*M00
M01=(1/(T1))*M01
M11=(1/(T1))*M11
M02=(1/(T1))*M02
M12=(1/(T1))*M12
M22=(1/(T1))*M22


R0 = matrix(ncol=p, nrow=(T1))
R1 = matrix(ncol=p+cconst+ctrend, nrow=(T1))
for (i in 1:(T1)) {
  R0[i,]=Z0[i,]-(M02%*%solve(M22)%*%Z2[i,])
  R1[i,]=Z1[i,]-(M12%*%solve(M22)%*%Z2[i,])
}

S00=R0[1,]%*%t(R0[1,])
S01=R0[1,]%*%t(R1[1,])
S10=R1[1,]%*%t(R0[1,])
S11=R1[1,]%*%t(R1[1,])
for (i in 2:(T1)) {
  S00 = S00 + R0[i,]%*%t(R0[i,])
  S01 = S01 + R0[i,]%*%t(R1[i,])
  S10 = S10 + R1[i,]%*%t(R0[i,])
  S11 = S11 + R1[i,]%*%t(R1[i,])
}
S00=(1/(T1))*S00
S01=(1/(T1))*S01
S10=(1/(T1))*S10
S11=(1/(T1))*S11
```

```
  eig=geigen(S10%*%solve(S00)%*%S01,S11)
  eig$values=eig$values[(p+cconst+ctrend):1]
  eig$vectors[,1:(p+cconst+ctrend)]=eig$vectors[,(p+cconst
     +ctrend):1]

  #### normalize the Eigenvectors
  for (i in 1:(p+cconst+ctrend)) {
    eig$vectors[,i]=eig$vectors[,i]/eig$vectors[1,i]
  }

  beta = eig$vectors
  alpha = S01%*%beta%*%solve(t(beta)%*%S11%*%beta)
  Omega = S00-alpha%*%(t(beta)%*%S11%*%beta)%*%t(alpha)
  Psi = M02%*%solve(M22)-alpha%*%t(beta)%*%M12%*%solve(M22
     )

  ReturnList=list()
  ReturnList[[1]]=beta
  ReturnList[[2]]=alpha
  ReturnList[[3]]=Omega
  ReturnList[[4]]=Psi

  print("TRACE Statistic")
  for (r in 0:(p-1)) {
    print(TRACE(r,eig$values, T1))
  }
  return(ReturnList)
}
```

B.3 Data Reading and Basic Data Processing

```
#Reading Data from sources
BIPRead = read.csv2("...PATH/BIP.csv", header=TRUE)
BIP = numeric(dim(BIPRead)[1]*3)
for (i in 1:dim(BIPRead)[1])
{
  for (j in ((i-1)*3+1):(i*3))
  {
    BIP[j]=BIPRead[i,3]/3
  }
}
HolzRead = read.csv2("...PATH/Rohholz.csv", header=TRUE)
PelletsRead = read.csv2("...PATH/Daten_Brennstoffe_02-14.
    csv", header=TRUE) #Contains Pellets, Oil and Gas Data
VPIRead = read.csv2("...PATH/VPI.csv", header=FALSE)
VPIRead[,3]=VPIRead[,3]

# Calibration of variables
T = 150 #Datalength, starts Jan 2002 and goes to Jun 2014
p=5 #Dimension of Data
per = 12 #Periodicity

BIPmonthly=q2m(BIPRead[,3],FALSE) #Data Estimation from
    quarterly BIP to monthly BIP

#Data/mean factoring out inflation
#Monthly Data
InfRate = VPIRead[1:T,3]/100
data = matrix(nrow=T, ncol=p)
data[,1] = (PelletsRead$Holzpellets[1:T]*(1/InfRate))/mean
    (PelletsRead$Holzpellets[1:T])
data[,2] = (PelletsRead$Heizol[1:T]*(1/InfRate))/mean(
    PelletsRead$Heizol[1:T])
data[,3] = (PelletsRead$Erdgas[1:T]*(1/InfRate))/mean(
    PelletsRead$Erdgas[1:T])
data[,4] = (HolzRead[1:T,6]*(1/InfRate))/mean(HolzRead[1:T
    ,6])
#Quarterly Data
data[,5] = (BIPmonthly[1:T]*(1/InfRate))/mean(BIPmonthly
    [1:T])
```

```
# Original Plots
plot(PelletsRead$Holzpellets[1:T],type="l",xlab="", ylab="
    ", xaxt="n", col="green")
axis(1, at=seq(1,150, 12), labels=c("Jan 02'","Jan 03'","
    Jan 04'","Jan 05'", "Jan 06'", "Jan 07'", "Jan 08'", "
    Jan 09'", "Jan 10'", "Jan 11'", "Jan 12'", "Jan 13'", "
    Jan 14'"))
plot(PelletsRead$Heizol[1:T],type="l",xlab="", ylab="",
    xaxt="n", col="brown")
axis(1, at=seq(1,150, 12), labels=c("Jan 02'","Jan 03'","
    Jan 04'","Jan 05'", "Jan 06'", "Jan 07'", "Jan 08'", "
    Jan 09'", "Jan 10'", "Jan 11'", "Jan 12'", "Jan 13'", "
    Jan 14'"))
plot(PelletsRead$Erdgas[1:T],type="l",xlab="", ylab="",
    xaxt="n", col="orange")
axis(1, at=seq(1,150, 12), labels=c("Jan 02'","Jan 03'","
    Jan 04'","Jan 05'", "Jan 06'", "Jan 07'", "Jan 08'", "
    Jan 09'", "Jan 10'", "Jan 11'", "Jan 12'", "Jan 13'", "
    Jan 14'"))
plot(HolzRead[1:T,6],type="l",xlab="", ylab="", xaxt="n",
    col="blue")
axis(1, at=seq(1,150, 12), labels=c("Jan 02'","Jan 03'","
    Jan 04'","Jan 05'", "Jan 06'", "Jan 07'", "Jan 08'", "
    Jan 09'", "Jan 10'", "Jan 11'", "Jan 12'", "Jan 13'", "
    Jan 14'"))
plot(BIPRead[1:50,3],type="l",xlab="", ylab="", xaxt="n",
    col="red")
axis(1, at=seq(1,50, 4), labels=c("Jan 02'","Jan 03'","Jan
     04'","Jan 05'", "Jan 06'", "Jan 07'", "Jan 08'", "Jan
    09'", "Jan 10'", "Jan 11'", "Jan 12'", "Jan 13'", "Jan
    14'"))
plot(VPIRead[1:T,3]/100,type="l",xlab="", ylab="", xaxt="n
    ")
axis(1, at=seq(1,150, 12), labels=c("Jan 02'","Jan 03'","
    Jan 04'","Jan 05'", "Jan 06'", "Jan 07'", "Jan 08'", "
    Jan 09'", "Jan 10'", "Jan 11'", "Jan 12'", "Jan 13'", "
    Jan 14'"))

#plots of the data/mean without inflation
plot(data[,1],type="l",col="green", ylim=c(min(data),max(
    data)), xlab="Year", ylab="", xaxt="n")
axis(1, at=seq(1,150, 12), labels=c("Jan 02'","Jan 03'","
    Jan 04'","Jan 05'", "Jan 06'", "Jan 07'", "Jan 08'", "
    Jan 09'", "Jan 10'", "Jan 11'", "Jan 12'", "Jan 13'", "
    Jan 14'"))
```

```
legend("bottomright", legend=c("Pellets","Oil","Gas","Wood
    ","GDP"), lwd=2, col=c("green","brown","orange", "blue"
    , "red"), ncol=2)
lines(data[,2],col="brown")
lines(data[,3],col="orange")
lines(data[,4],col="blue")
lines(data[,5],col="red")

# creation of deseasonalized data; renaming of data
    necessary for later use of the urca-function ca.jo
data[,1]=data[,1]*1/tripexp(data[,1],12,0.05,0.05,0.05)
data[,2]=data[,2]*1/tripexp(data[,2],12,0.05,0.05,0.05)
data[,3]=data[,3]*1/tripexp(data[,3],12,0.05,0.05,0.05)
data[,4]=data[,4]*1/tripexp(data[,4],12,0.05,0.05,0.05)
data[,5]=data[,5]*1/tripexp(data[,5],12,0.05,0.05,0.05)
a=data[,1]
b=data[,2]
c=data[,3]
d=data[,4]
e=data[,5]

# time series of first differences
ad=diff(a,1)
bd=diff(b,1)
cd=diff(c,1)
dd=diff(d,1)
ed=diff(e,1)

#KPSS and ADF test results
kpss.test(a,"Level")
kpss.test(b,"Level")
kpss.test(c,"Level")
kpss.test(d,"Level")
kpss.test(e,"Level")
kpss.test(a,"Trend")
kpss.test(b,"Trend")
kpss.test(c,"Trend")
kpss.test(d,"Trend")
kpss.test(e,"Trend")
adf.test(a)
adf.test(b)
adf.test(c)
adf.test(d)
adf.test(e)
```

```
#Descriptive Statistics for the data
mean(a)
mean(b)
mean(c)
mean(d)
mean(e)
sd(a)
sd(b)
sd(c)
sd(d)
sd(e)
min(a)
min(b)
min(c)
min(d)
min(e)
max(a)
max(b)
max(c)
max(d)
max(e)
#Descriptive Statistics for the first differences
mean(ad)
mean(bd)
mean(cd)
mean(dd)
mean(ed)
sd(ad)
sd(bd)
sd(cd)
sd(dd)
sd(ed)
min(ad)
min(bd)
min(cd)
min(dd)
min(ed)
max(ad)
max(bd)
max(cd)
max(dd)
max(ed)
```

B.4 Estimations for the General Model

```
#choice of information set for general model
x=cbind(a,b,c,d,e)
#determining AIC, BIC and HQ for lag k=2...12 in an VECM
    without constant or trend
for (k in 2:12) {
  print(myVARJ(x,k,FALSE,FALSE, "Info"))
}
#determining AIC, BIC and HQ for lag k=2...12 in an VECM
    with constant
for (k in 2:12) {
  print(myVARJ(x,k,TRUE,FALSE, "Info"))
}
#determining AIC, BIC and HQ for lag k=2...12 in an VECM
    with trend
for (k in 2:12) {
  print(myVARJ(x,k,TRUE,TRUE, "Info"))
}

# choice of lag k
k=2

#LjungBox tests for the residuues of the model without
    constant or trend
LjungBox(myVARJ(x,k,FALSE,FALSE, "eps")[1,],lags=1:20)
LjungBox(myVARJ(x,k,FALSE,FALSE, "eps")[2,],lags=1:20)
LjungBox(myVARJ(x,k,FALSE,FALSE, "eps")[3,],lags=1:20)
LjungBox(myVARJ(x,k,FALSE,FALSE, "eps")[4,],lags=1:20)
LjungBox(myVARJ(x,k,FALSE,FALSE, "eps")[5,],lags=1:20)

#LjungBox tests for the residuues of the model with
    constant
LjungBox(myVARJ(x,k,TRUE,FALSE, "eps")[1,],lags=1:20)
LjungBox(myVARJ(x,k,TRUE,FALSE, "eps")[2,],lags=1:20)
LjungBox(myVARJ(x,k,TRUE,FALSE, "eps")[3,],lags=1:20)
LjungBox(myVARJ(x,k,TRUE,FALSE, "eps")[4,],lags=1:20)
LjungBox(myVARJ(x,k,TRUE,FALSE, "eps")[5,],lags=1:20)

#LjungBox tests for the residuues of the model with trend
LjungBox(myVARJ(x,k,TRUE,TRUE, "eps")[1,],lags=1:20)
LjungBox(myVARJ(x,k,TRUE,TRUE, "eps")[2,],lags=1:20)
LjungBox(myVARJ(x,k,TRUE,TRUE, "eps")[3,],lags=1:20)
LjungBox(myVARJ(x,k,TRUE,TRUE, "eps")[4,],lags=1:20)
LjungBox(myVARJ(x,k,TRUE,TRUE, "eps")[5,],lags=1:20)
```

```
#processing of Johanson with a constant in cointegration
    space
coco=ca.jo(x,type="trace",spec="transitory", ecdet="const"
    ,season = NULL, dumvar = NULL,K=k)
summary(coco)

#alternative programm for Johanson with a constant in
    cointegration space
mycoco=cointegrate(x,k,FALSE, FALSE, TRUE, FALSE)
mycoco[[1]]
mycoco[[2]]

# choice of number of cointegration relations r
r=2

#calculation and plotting of the cointegration relations
coint=matrix(ncol=r, nrow=T)
beta=slot(coco,"V")[,1:r]
for (i in 1:T) {
    coint[i,]=t(beta)%*%c(x[i,],1)
}
plot(coint[,1],type="l", ylab="", xlab="")
plot(coint[,2],type="l", ylab="", xlab="")

#KPSS and ADF Tests for the cointegration relations
kpss.test(coint[,1],"Level")
kpss.test(coint[,1],"Trend")
adf.test(coint[,1])
kpss.test(coint[,2],"Level")
kpss.test(coint[,2],"Trend")
adf.test(coint[,2])

#Varimax transformation of the cointegration relations
beta=varimax(slot(coco,"V")[,1:r], normalize=TRUE)$
    loadings
beta

#calculation and plotting of the varimax processed
    cointegration relations
for (i in 1:T) {
    coint[i,]=t(beta)%*%c(x[i,],1)
}
plot(coint[,1],type="l", ylab="", xlab="")
plot(coint[,2],type="l", ylab="", xlab="")
```

```
#KPSS and ADF Tests for the varimax processed
    cointegration relations
kpss.test(coint[,1],"Level")
kpss.test(coint[,1],"Trend")
adf.test(coint[,1])
kpss.test(coint[,2],"Level")
kpss.test(coint[,2],"Trend")
adf.test(coint[,2])

#AR model and predictions
ddata=rbind(0,diff(x,1))

alpha=slot(coco,"W")[,1:r]
beta=slot(coco,"V")[,1:r]
gamma=slot(coco,"GAMMA")

ARdata=matrix(nrow=150,ncol=5)
ARdata[1:3,]=ddata[1:3,]
for (i in 0:146) {ARdata[3+1+i,]=alpha%*%t(beta)%*%c(x[3+i
    ,],1)+gamma%*%ddata[3+i,]}

error = matrix(nrow=T-1, ncol=p)
error[,1]=ddata[2:150,1]-ARdata[2:150,1]
error[,2]=ddata[2:150,2]-ARdata[2:150,2]
error[,3]=ddata[2:150,3]-ARdata[2:150,3]
error[,4]=ddata[2:150,4]-ARdata[2:150,4]
error[,5]=ddata[2:150,5]-ARdata[2:150,5]

sum = 0
for (i in (k):(T-1)) {
    sum = sum + (error[i,])%*%t(error[i,])
}
Sigk = 1/(T-k)*sum

plot(error[,1],type="l", xlab="",ylab="")
plot(error[,2],type="l", xlab="",ylab="")
plot(error[,3],type="l", xlab="",ylab="")
plot(error[,4],type="l", xlab="",ylab="")
plot(error[,5],type="l", xlab="",ylab="")
acf(error[,1],xlab="",ylab="", main="")
acf(error[,2],xlab="",ylab="", main="")
acf(error[,3],xlab="",ylab="", main="")
acf(error[,4],xlab="",ylab="", main="")
acf(error[,5],xlab="",ylab="", main="")
```

```
mad(error[,1])
mad(error[,2])
mad(error[,3])
mad(error[,4])
mad(error[,5])
rmse(ARdata[4:150,1],ddata[4:150,1])
rmse(ARdata[4:150,2],ddata[4:150,2])
rmse(ARdata[4:150,3],ddata[4:150,3])
rmse(ARdata[4:150,4],ddata[4:150,4])
rmse(ARdata[4:150,5],ddata[4:150,5])
var(error[,1])/var(ddata[2:150,1])
var(error[,2])/var(ddata[2:150,2])
var(error[,3])/var(ddata[2:150,3])
var(error[,4])/var(ddata[2:150,4])
var(error[,5])/var(ddata[2:150,5])

#Remaking data without seasonalization to remake the
    seasonalization indices
#Monthly Data
InfRate = VPIRead[1:T,3]/100
noseasondata = matrix(nrow=T, ncol=p)
noseasondata[,1] = (PelletsRead$Holzpellets[1:T]*(1/
    InfRate))/mean(PelletsRead$Holzpellets[1:T])
noseasondata[,2] = (PelletsRead$Heizol[1:T]*(1/InfRate))/
    mean(PelletsRead$Heizol[1:T])
noseasondata[,3] = (PelletsRead$Erdgas[1:T]*(1/InfRate))/
    mean(PelletsRead$Erdgas[1:T])
noseasondata[,4] = (HolzRead[1:T,6]*(1/InfRate))/mean(
    HolzRead[1:T,6])
#Quarterly Data
noseasondata[,5] = (BIPmonthly[1:T]*(1/InfRate))/mean(
    BIPmonthly[1:T])


predcchoice=T-per+1+(1:6-1)%%12
predcs=matrix(nrow=6, ncol=p)
predcs[,1]=tripexp(noseasondata[,1],12,0.05,0.05,0.05)[
    predcchoice]
predcs[,2]=tripexp(noseasondata[,2],12,0.05,0.05,0.05)[
    predcchoice]
predcs[,3]=tripexp(noseasondata[,3],12,0.05,0.05,0.05)[
    predcchoice]
predcs[,4]=tripexp(noseasondata[,4],12,0.05,0.05,0.05)[
    predcchoice]
```

```
predcs[,5]=tripexp(noseasondata[,5],12,0.05,0.05,0.05)[
    predcchoice]


#Data for prediction purposes
#Monthly Data
futureInfRate = VPIRead[(T+1):(T+6),3]/100
future = matrix(nrow=6, ncol=p)
future[,1] = (PelletsRead$Holzpellets[(T+1):(T+6)]*(1/
    futureInfRate))/mean(PelletsRead$Holzpellets[1:T])
future[,2] = (PelletsRead$Heizol[(T+1):(T+6)]*(1/
    futureInfRate))/mean(PelletsRead$Heizol[1:T])
future[,3] = (PelletsRead$Erdgas[(T+1):(T+6)]*(1/
    futureInfRate))/mean(PelletsRead$Erdgas[1:T])
future[,4] = (HolzRead[(T+1):(T+6),6]*(1/futureInfRate))/
    mean(HolzRead[1:T,6])
#Quarterly Data
future[,5] = (BIPmonthly[(T+1):(T+6)]*(1/futureInfRate))/
    mean(BIPmonthly[1:T])

# creation of deseasonalized future data
future[,1]=future[,1]*1/predcs[,1]
future[,2]=future[,2]*1/predcs[,2]
future[,3]=future[,3]*1/predcs[,3]
future[,4]=future[,4]*1/predcs[,4]
future[,5]=future[,5]*1/predcs[,5]


extended=rbind(x,future)

forecast=matrix(nrow=6,ncol=5)
for (i in 1:6) {forecast[i,]=alpha%*%t(beta)%*%c(extended
    [150-1+i,],1)+gamma%*%diff(extended,1)[150-1+i,]}

plot(149:155,c(ARdata[T,1],forecast[,1]),col="pink",lwd=2,
    type="l", xlab="",ylab="",xaxt="n",xlim=c(0,155),ylim=c
    (min(forecast[,1],diff(extended,1)[,1]),max(forecast
    [,1],diff(extended,1)[,1])))
lines(diff(extended,1)[,1])
lines(ARdata[2:T,1],col="green", lwd=2)
axis(1, at=seq(1,150, 12), labels=c("Jan 02'","Jan 03'","
    Jan 04'","Jan 05'", "Jan 06'", "Jan 07'", "Jan 08'", "
    Jan 09'", "Jan 10'", "Jan 11'", "Jan 12'", "Jan 13'", "
    Jan 14'"))
legend("bottomright", legend=c("Extended Pellets","
```

```
    Modelled Data", "Prognosis"), lwd=2, col=c("black","
    green","pink"), ncol=1)


plot(149:155,c(ARdata[T,2],forecast[,2]),col="pink",lwd=2,
    type="l", xlab="",ylab="",xaxt="n",xlim=c(0,155),ylim=c
    (min(forecast[,2],diff(extended,1)[,2]),max(forecast
    [,2],diff(extended,1)[,2])))
lines(diff(extended,1)[,2])
lines(ARdata[2:T,2],col="brown", lwd=2)
axis(1, at=seq(1,150, 12), labels=c("Jan 02'","Jan 03'","
    Jan 04'","Jan 05'", "Jan 06'", "Jan 07'", "Jan 08'", "
    Jan 09'", "Jan 10'", "Jan 11'", "Jan 12'", "Jan 13'", "
    Jan 14'"))
legend("bottomright", legend=c("Extended Oil","Modelled
    Data", "Prognosis"), lwd=2, col=c("black","brown","pink
    "), ncol=1)

plot(149:155,c(ARdata[T,3],forecast[,3]),col="pink",lwd=2,
    type="l", xlab="",ylab="",xaxt="n",xlim=c(0,155),ylim=c
    (min(forecast[,3],diff(extended,1)[,3]),max(forecast
    [,3],diff(extended,1)[,3])))
lines(diff(extended,1)[,3])
lines(ARdata[2:T,3],col="orange", lwd=2)
axis(1, at=seq(1,150, 12), labels=c("Jan 02'","Jan 03'","
    Jan 04'","Jan 05'", "Jan 06'", "Jan 07'", "Jan 08'", "
    Jan 09'", "Jan 10'", "Jan 11'", "Jan 12'", "Jan 13'", "
    Jan 14'"))
legend("bottomright", legend=c("Extended Gas","Modelled
    Data", "Prognosis"), lwd=2, col=c("black","orange","
    pink"), ncol=1)

plot(149:155,c(ARdata[T,4],forecast[,4]),col="pink",lwd=2,
    type="l", xlab="",ylab="",xaxt="n",xlim=c(0,155),ylim=c
    (min(forecast[,4],diff(extended,1)[,4]),max(forecast
    [,4],diff(extended,1)[,4])))
lines(diff(extended,1)[,4])
lines(ARdata[2:T,4],col="blue", lwd=2)
axis(1, at=seq(1,150, 12), labels=c("Jan 02'","Jan 03'","
    Jan 04'","Jan 05'", "Jan 06'", "Jan 07'", "Jan 08'", "
    Jan 09'", "Jan 10'", "Jan 11'", "Jan 12'", "Jan 13'", "
    Jan 14'"))
legend("bottomright", legend=c("Extended Wood","Modelled
    Data", "Prognosis"), lwd=2, col=c("black","blue","pink"
    ), ncol=1)
```

```
plot(149:155,c(ARdata[T,5],forecast[,5]),col="pink",lwd=2,
    type="l", xlab="",ylab="",xaxt="n",xlim=c(0,155),ylim=c
    (min(forecast[,5],diff(extended,1)[,5]),max(forecast
    [,5],diff(extended,1)[,5])))
lines(diff(extended,1)[,5])
lines(ARdata[2:T,5],col="red", lwd=2)
axis(1, at=seq(1,150, 12), labels=c("Jan 02'","Jan 03'","
    Jan 04'","Jan 05'", "Jan 06'", "Jan 07'", "Jan 08'", "
    Jan 09'", "Jan 10'", "Jan 11'", "Jan 12'", "Jan 13'", "
    Jan 14'"))
legend("bottomright", legend=c("Extended GDP","Modelled
    Data", "Prognosis"), lwd=2, col=c("black","red","pink")
    , ncol=1)

prederror=diff(extended,1)[150:155,]-forecast
mad(prederror[,1])
mad(prederror[,2])
mad(prederror[,3])
mad(prederror[,4])
mad(prederror[,5])
rmse(forecast[,1],diff(extended,1)[150:155,1])
rmse(forecast[,2],diff(extended,1)[150:155,2])
rmse(forecast[,3],diff(extended,1)[150:155,3])
rmse(forecast[,4],diff(extended,1)[150:155,4])
rmse(forecast[,5],diff(extended,1)[150:155,5])
var(prederror[,1])/var(ddata[2:150,1])
var(prederror[,2])/var(ddata[2:150,2])
var(prederror[,3])/var(ddata[2:150,3])
var(prederror[,4])/var(ddata[2:150,4])
var(prederror[,5])/var(ddata[2:150,5])
```

B.5 Calculations Performed to Further Discuss the Information Set

```
x=cbind(a,b)
#determining AIC, BIC and HQ for lag k=2...12 in an VECM
    with constant; automatically determine best k by
    smallest AIC
k0=2
Crit0 = myVARJ(x,k0,TRUE,FALSE, "Info")[2]
for (k in 3:12) {
  Crit = myVARJ(x,k,TRUE,FALSE, "Info")[2]
  if (Crit<Crit0) {
    Crit0=Crit
    k0=k
  }
}
k=k0
coco=ca.jo(x,type="trace",spec="transitory", ecdet="const"
    ,season = NULL, dumvar = NULL,K=k)
summary(coco)

x=cbind(a,c)
#determining AIC, BIC and HQ for lag k=2...12 in an VECM
    with constant; automatically determine best k by
    smallest AIC
k0=2
Crit0 = myVARJ(x,k0,TRUE,FALSE, "Info")[2]
for (k in 3:12) {
  Crit = myVARJ(x,k,TRUE,FALSE, "Info")[2]
  if (Crit<Crit0) {
    Crit0=Crit
    k0=k
  }
}
k=k0
coco=ca.jo(x,type="trace",spec="transitory", ecdet="const"
    ,season = NULL, dumvar = NULL,K=k)
summary(coco)
```

```
x=cbind(a,d)
#determining AIC, BIC and HQ for lag k=2...12 in an VECM
    with constant; automatically determine best k by
    smallest AIC
k0=2
Crit0 = myVARJ(x,k0,TRUE,FALSE, "Info")[2]
for (k in 3:12) {
   Crit = myVARJ(x,k,TRUE,FALSE, "Info")[2]
   if (Crit<Crit0) {
     Crit0=Crit
     k0=k
   }
}
k=k0
coco=ca.jo(x,type="trace",spec="transitory", ecdet="const"
    ,season = NULL, dumvar = NULL,K=k)
summary(coco)

x=cbind(a,e)
#determining AIC, BIC and HQ for lag k=2...12 in an VECM
    with constant; automatically determine best k by
    smallest AIC
k0=2
Crit0 = myVARJ(x,k0,TRUE,FALSE, "Info")[2]
for (k in 3:12) {
   Crit = myVARJ(x,k,TRUE,FALSE, "Info")[2]
   if (Crit<Crit0) {
     Crit0=Crit
     k0=k
   }
}
k=k0
coco=ca.jo(x,type="trace",spec="transitory", ecdet="const"
    ,season = NULL, dumvar = NULL,K=k)
summary(coco)
```

```
x=cbind(b,c)
#determining AIC, BIC and HQ for lag k=2...12 in an VECM
    with constant; automatically determine best k by
    smallest AIC
k0=2
Crit0 = myVARJ(x,k0,TRUE,FALSE, "Info")[2]
for (k in 3:12) {
   Crit = myVARJ(x,k,TRUE,FALSE, "Info")[2]
   if (Crit<Crit0) {
      Crit0=Crit
      k0=k
   }
}
k=k0
coco=ca.jo(x,type="trace",spec="transitory", ecdet="const"
    ,season = NULL, dumvar = NULL,K=k)
summary(coco)

x=cbind(b,d)
#determining AIC, BIC and HQ for lag k=2...12 in an VECM
    with constant; automatically determine best k by
    smallest AIC
k0=2
Crit0 = myVARJ(x,k0,TRUE,FALSE, "Info")[2]
for (k in 3:12) {
   Crit = myVARJ(x,k,TRUE,FALSE, "Info")[2]
   if (Crit<Crit0) {
      Crit0=Crit
      k0=k
   }
}
k=k0
coco=ca.jo(x,type="trace",spec="transitory", ecdet="const"
    ,season = NULL, dumvar = NULL,K=k)
summary(coco)
```

```
x=cbind(b,e)
#determining AIC, BIC and HQ for lag k=2...12 in an VECM
    with constant; automatically determine best k by
    smallest AIC
k0=2
Crit0 = myVARJ(x,k0,TRUE,FALSE, "Info")[2]
for (k in 3:12) {
  Crit = myVARJ(x,k,TRUE,FALSE, "Info")[2]
  if (Crit<Crit0) {
    Crit0=Crit
    k0=k
  }
}
k=k0
coco=ca.jo(x,type="trace",spec="transitory", ecdet="const"
    ,season = NULL, dumvar = NULL,K=k)
summary(coco)

x=cbind(c,d)
#determining AIC, BIC and HQ for lag k=2...12 in an VECM
    with constant; automatically determine best k by
    smallest AIC
k0=2
Crit0 = myVARJ(x,k0,TRUE,FALSE, "Info")[2]
for (k in 3:12) {
  Crit = myVARJ(x,k,TRUE,FALSE, "Info")[2]
  if (Crit<Crit0) {
    Crit0=Crit
    k0=k
  }
}
k=k0
coco=ca.jo(x,type="trace",spec="transitory", ecdet="const"
    ,season = NULL, dumvar = NULL,K=k)
summary(coco)
```

```
x=cbind(c,e)
#determining AIC, BIC and HQ for lag k=2...12 in an VECM
    with constant; automatically determine best k by
    smallest AIC
k0=2
Crit0 = myVARJ(x,k0,TRUE,FALSE, "Info")[2]
for (k in 3:12) {
  Crit = myVARJ(x,k,TRUE,FALSE, "Info")[2]
  if (Crit<Crit0) {
    Crit0=Crit
    k0=k
  }
}
k=k0
coco=ca.jo(x,type="trace",spec="transitory", ecdet="const"
    ,season = NULL, dumvar = NULL,K=k)
summary(coco)

x=cbind(d,e)
#determining AIC, BIC and HQ for lag k=2...12 in an VECM
    with constant; automatically determine best k by
    smallest AIC
k0=2
Crit0 = myVARJ(x,k0,TRUE,FALSE, "Info")[2]
for (k in 3:12) {
  Crit = myVARJ(x,k,TRUE,FALSE, "Info")[2]
  if (Crit<Crit0) {
    Crit0=Crit
    k0=k
  }
}
k=k0
coco=ca.jo(x,type="trace",spec="transitory", ecdet="const"
    ,season = NULL, dumvar = NULL,K=k)
summary(coco)
```

B.6 Estimations for the Model without Gas

```
#choice of information set for model without gas
x=cbind(a,b,d,e)
#determining AIC, BIC and HQ for lag k=2...12 in an VECM
    without constant or trend
for (k in 2:12) {
  print(myVARJ(x,k,FALSE,FALSE, "Info"))
}
#determining AIC, BIC and HQ for lag k=2...12 in an VECM
    with constant
for (k in 2:12) {
  print(myVARJ(x,k,TRUE,FALSE, "Info"))
}
#determining AIC, BIC and HQ for lag k=2...12 in an VECM
    with trend
for (k in 2:12) {
  print(myVARJ(x,k,TRUE,TRUE, "Info"))
}

# choice of lag k
k=2

#LjungBox tests for the residuues of the model without
    constant or trend
LjungBox(myVARJ(x,k,FALSE,FALSE, "eps")[1,],lags=1:20)
LjungBox(myVARJ(x,k,FALSE,FALSE, "eps")[2,],lags=1:20)
LjungBox(myVARJ(x,k,FALSE,FALSE, "eps")[3,],lags=1:20)
LjungBox(myVARJ(x,k,FALSE,FALSE, "eps")[4,],lags=1:20)

#LjungBox tests for the residuues of the model with
    constant
LjungBox(myVARJ(x,k,TRUE,FALSE, "eps")[1,],lags=1:20)
LjungBox(myVARJ(x,k,TRUE,FALSE, "eps")[2,],lags=1:20)
LjungBox(myVARJ(x,k,TRUE,FALSE, "eps")[3,],lags=1:20)
LjungBox(myVARJ(x,k,TRUE,FALSE, "eps")[4,],lags=1:20)

#LjungBox tests for the residuues of the model with trend
LjungBox(myVARJ(x,k,TRUE,TRUE, "eps")[1,],lags=1:20)
LjungBox(myVARJ(x,k,TRUE,TRUE, "eps")[2,],lags=1:20)
LjungBox(myVARJ(x,k,TRUE,TRUE, "eps")[3,],lags=1:20)
LjungBox(myVARJ(x,k,TRUE,TRUE, "eps")[4,],lags=1:20)
```

```
#processing of Johanson with a constant in cointegration
    space
coco=ca.jo(x,type="trace",spec="transitory", ecdet="const"
    ,season = NULL, dumvar = NULL,K=k)
summary(coco)

#alternative programm for Johanson with a constant in
    cointegration space
mycoco=cointegrate(x,k,FALSE, FALSE, TRUE, FALSE)
mycoco[[1]]
mycoco[[2]]

# choice of number of cointegration relations r
r=3

#calculation and plotting of the cointegration relations
coint=matrix(ncol=r, nrow=T)
beta=slot(coco,"V")[,1:r]
for (i in 1:T) {
    coint[i,]=t(beta)%*%c(x[i,],1)
}
plot(coint[,1],type="l", ylab="", xlab="")
plot(coint[,2],type="l", ylab="", xlab="")
plot(coint[,3],type="l", ylab="", xlab="")

#KPSS and ADF Tests for the cointegration relations
kpss.test(coint[,1],"Level")
kpss.test(coint[,1],"Trend")
adf.test(coint[,1])
kpss.test(coint[,2],"Level")
kpss.test(coint[,2],"Trend")
adf.test(coint[,2])
kpss.test(coint[,3],"Level")
kpss.test(coint[,3],"Trend")
adf.test(coint[,3])

#Varimax transformation of the cointegration relations
beta=varimax(slot(coco,"V")[,1:r], normalize=TRUE)$
    loadings
beta
```

```
#calculation and plotting of the varmiax processed
    cointegration relations
for (i in 1:T) {
    coint[i,]=t(beta)%*%c(x[i,],1)
}
plot(coint[,1],type="l", ylab="", xlab="")
plot(coint[,2],type="l", ylab="", xlab="")
plot(coint[,3],type="l", ylab="", xlab="")

#KPSS and ADF Tests for the varimax processed
    cointegration relations
kpss.test(coint[,1],"Level")
kpss.test(coint[,1],"Trend")
adf.test(coint[,1])
kpss.test(coint[,2],"Level")
kpss.test(coint[,2],"Trend")
adf.test(coint[,2])
kpss.test(coint[,3],"Level")
kpss.test(coint[,3],"Trend")
adf.test(coint[,3])

#Reducing the cointegration rank in this model to r=2 for
    alternative varimax rotations
r=2

#Varimax transformation of the cointegration relations
beta=varimax(slot(coco,"V")[,1:r], normalize=TRUE)$
    loadings
beta

#calculation and plotting of the varmiax processed
    cointegration relations
for (i in 1:T) {
    coint[i,]=t(beta)%*%c(x[i,],1)
}
plot(coint[,1],type="l", ylab="", xlab="")
plot(coint[,2],type="l", ylab="", xlab="")
```

```
#KPSS and ADF Tests for the varimax processed
    cointegration relations
kpss.test(coint[,1],"Level")
kpss.test(coint[,1],"Trend")
adf.test(coint[,1])
kpss.test(coint[,2],"Level")
kpss.test(coint[,2],"Trend")
adf.test(coint[,2])

#AR model and prediction in the gasless model
ddata=rbind(0,diff(x,1))

alpha=slot(coco,"W")[,1:r]
beta=slot(coco,"V")[,1:r]
gamma=slot(coco,"GAMMA")

ARdata=matrix(nrow=150,ncol=4)
ARdata[1:3,]=ddata[1:3,]
for (i in 0:146) {ARdata[3+1+i,]=alpha%*%t(beta)%*%c(x[3+i
    ,],1)+gamma%*%ddata[3+i,]}


error = matrix(nrow=T-1, ncol=4)
error[,1]=ddata[2:150,1]-ARdata[2:150,1]
error[,2]=ddata[2:150,2]-ARdata[2:150,2]
error[,3]=ddata[2:150,3]-ARdata[2:150,3]
error[,4]=ddata[2:150,4]-ARdata[2:150,4]

sum = 0
for (i in (k):(T-1)) {
    sum = sum + (error[i,])%*%t(error[i,])
}
Sigk = 1/(T-k)*sum

plot(error[,1],type="l", xlab="",ylab="")
plot(error[,2],type="l", xlab="",ylab="")
plot(error[,3],type="l", xlab="",ylab="")
plot(error[,4],type="l", xlab="",ylab="")
acf(error[,1],xlab="",ylab="", main="")
acf(error[,2],xlab="",ylab="", main="")
acf(error[,3],xlab="",ylab="", main="")
acf(error[,4],xlab="",ylab="", main="")

mad(error[,1])
mad(error[,2])
```

```
mad(error[,3])
mad(error[,4])
rmse(ARdata[4:150,1],ddata[4:150,1])
rmse(ARdata[4:150,2],ddata[4:150,2])
rmse(ARdata[4:150,3],ddata[4:150,3])
rmse(ARdata[4:150,4],ddata[4:150,4])
var(error[,1])/var(ddata[2:150,1])
var(error[,2])/var(ddata[2:150,2])
var(error[,3])/var(ddata[2:150,3])
var(error[,4])/var(ddata[2:150,4])

#Remaking data without seasonalization to remake the
    seasonalization indices
#Monthly Data
InfRate = VPIRead[1:T,3]/100
noseasondata = matrix(nrow=T, ncol=p)
noseasondata[,1] = (PelletsRead$Holzpellets[1:T]*(1/
    InfRate))/mean(PelletsRead$Holzpellets[1:T])
noseasondata[,2] = (PelletsRead$Heizol[1:T]*(1/InfRate))/
    mean(PelletsRead$Heizol[1:T])
noseasondata[,3] = (HolzRead[1:T,6]*(1/InfRate))/mean(
    HolzRead[1:T,6])
#Quarterly Data
noseasondata[,4] = (BIPmonthly[1:T]*(1/InfRate))/mean(
    BIPmonthly[1:T])


predcchoice=T-per+1+(1:6-1)%%12
predcs=matrix(nrow=6, ncol=4)
predcs[,1]=tripexp(noseasondata[,1],12,0.05,0.05,0.05)[
    predcchoice]
predcs[,2]=tripexp(noseasondata[,2],12,0.05,0.05,0.05)[
    predcchoice]
predcs[,3]=tripexp(noseasondata[,3],12,0.05,0.05,0.05)[
    predcchoice]
predcs[,4]=tripexp(noseasondata[,4],12,0.05,0.05,0.05)[
    predcchoice]


#Data for prediction purposes
#Monthly Data
futureInfRate = VPIRead[(T+1):(T+6),3]/100
future = matrix(nrow=6, ncol=4)
future[,1] = (PelletsRead$Holzpellets[(T+1):(T+6)]*(1/
    futureInfRate))/mean(PelletsRead$Holzpellets[1:T])
```

```
future[,2] = (PelletsRead$Heizol[(T+1):(T+6)]*(1/
    futureInfRate))/mean(PelletsRead$Heizol[1:T])
future[,3] = (HolzRead[(T+1):(T+6),6]*(1/futureInfRate))/
    mean(HolzRead[1:T,6])
#Quarterly Data
future[,4] = (BIPmonthly[(T+1):(T+6)]*(1/futureInfRate))/
    mean(BIPmonthly[1:T])

# creation of deseasonalized future data
future[,1]=future[,1]*1/predcs[,1]
future[,2]=future[,2]*1/predcs[,2]
future[,3]=future[,3]*1/predcs[,3]
future[,4]=future[,4]*1/predcs[,4]

extended=rbind(x,future)

forecast=matrix(nrow=6,ncol=4)
for (i in 1:6) {forecast[i,]=alpha%*%t(beta)%*%c(extended
    [150-1+i,],1)+gamma%*%diff(extended,1)[150-1+i,]}

plot(149:155,c(ARdata[T,1],forecast[,1]),col="pink",lwd=2,
    type="l", xlab="",ylab="",xaxt="n",xlim=c(0,155),ylim=c
    (min(forecast[,1],diff(extended,1)[,1]),max(forecast
    [,1],diff(extended,1)[,1])))
lines(diff(extended,1)[,1])
lines(ARdata[2:T,1],col="green", lwd=2)
axis(1, at=seq(1,150, 12), labels=c("Jan 02'","Jan 03'","
    Jan 04'","Jan 05'", "Jan 06'", "Jan 07'", "Jan 08'", "
    Jan 09'", "Jan 10'", "Jan 11'", "Jan 12'", "Jan 13'", "
    Jan 14'"))
legend("bottomright", legend=c("Extended Pellets","
    Modelled Data", "Prognosis"), lwd=2, col=c("black","
    green","pink"), ncol=1)

plot(149:155,c(ARdata[T,2],forecast[,2]),col="pink",lwd=2,
    type="l", xlab="",ylab="",xaxt="n",xlim=c(0,155),ylim=c
    (min(forecast[,2],diff(extended,1)[,2]),max(forecast
    [,2],diff(extended,1)[,2])))
lines(diff(extended,1)[,2])
lines(ARdata[2:T,2],col="brown", lwd=2)
axis(1, at=seq(1,150, 12), labels=c("Jan 02'","Jan 03'","
    Jan 04'","Jan 05'", "Jan 06'", "Jan 07'", "Jan 08'", "
    Jan 09'", "Jan 10'", "Jan 11'", "Jan 12'", "Jan 13'", "
    Jan 14'"))
legend("bottomright", legend=c("Extended Oil","Modelled
```

```
    Data", "Prognosis"), lwd=2, col=c("black","brown","pink
    "), ncol=1)

plot(149:155,c(ARdata[T,3],forecast[,3]),col="pink",lwd=2,
    type="l", xlab="",ylab="",xaxt="n",xlim=c(0,155),ylim=c
    (min(forecast[,3],diff(extended,1)[,3]),max(forecast
    [,3],diff(extended,1)[,3])))
lines(diff(extended,1)[,3])
lines(ARdata[2:T,3],col="blue", lwd=2)
axis(1, at=seq(1,150, 12), labels=c("Jan 02'","Jan 03'","
    Jan 04'","Jan 05'", "Jan 06'", "Jan 07'", "Jan 08'", "
    Jan 09'", "Jan 10'", "Jan 11'", "Jan 12'", "Jan 13'", "
    Jan 14'"))
legend("bottomright", legend=c("Extended Wood","Modelled
    Data", "Prognosis"), lwd=2, col=c("black","blue","pink"
    ), ncol=1)

plot(149:155,c(ARdata[T,4],forecast[,4]),col="pink",lwd=2,
    type="l", xlab="",ylab="",xaxt="n",xlim=c(0,155),ylim=c
    (min(forecast[,4],diff(extended,1)[,4]),max(forecast
    [,4],diff(extended,1)[,4])))
lines(diff(extended,1)[,4])
lines(ARdata[2:T,4],col="red", lwd=2)
axis(1, at=seq(1,150, 12), labels=c("Jan 02'","Jan 03'","
    Jan 04'","Jan 05'", "Jan 06'", "Jan 07'", "Jan 08'", "
    Jan 09'", "Jan 10'", "Jan 11'", "Jan 12'", "Jan 13'", "
    Jan 14'"))
legend("bottomright", legend=c("Extended GDP","Modelled
    Data", "Prognosis"), lwd=2, col=c("black","red","pink")
    , ncol=1)

prederror=diff(extended,1)[150:155,]-forecast
mad(prederror[,1])
mad(prederror[,2])
mad(prederror[,3])
mad(prederror[,4])
rmse(forecast[,1],diff(extended,1)[150:155,1])
rmse(forecast[,2],diff(extended,1)[150:155,2])
rmse(forecast[,3],diff(extended,1)[150:155,3])
rmse(forecast[,4],diff(extended,1)[150:155,4])
var(prederror[,1])/var(ddata[2:150,1])
var(prederror[,2])/var(ddata[2:150,2])
var(prederror[,3])/var(ddata[2:150,3])
var(prederror[,4])/var(ddata[2:150,4])
```