The Possibilities of
Multivariate Time Series Analysis
as a Tool for Studying Forest Business Development
- with Case Studies from Westphalia-Lippe -

Master's Thesis

at the

Georg-August-University Göttingen,

Faculty of Forest Sciences and Forest Ecology

Author:	Jan Schick
Matr.No.:	11271067
First Examiner:	Prof. Dr. Bernhard Möhring
Second Examiner:	Prof. Dr. Helmut Herwartz €
Advisor:	Dr. Kai Husmann
Submitted on:	21.03.2019

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

Forest Economics, Göttingen University

€ Econometrics, Göttingen University

Forest Economics and Sustainable Land-use Planning, Göttingen University

Die Möglichkeiten der

Multivariaten Zeitreihenanalyse

als Werkzeug zur Untersuchung der

Entwicklung von Forstbetrieben

- mit Fallstudien aus Westfalen-Lippe -

Masterarbeit

an der

Georg-August-Universität Göttingen,

Fakultät für Forstwissenschaften und Waldökologie

Autor:	Jan Schick
Matr.Nr.:	11271067
Erstprüfer:	Prof. Dr. Bernhard Möhring
Zweitprüfer:	Prof. Dr. Helmut Herwartz €
Betreuer:	Dr. Kai Husmann
Abgabedatum:	21.03.2019

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

Forstökonomie, Universität Göttingen

€ Ökonometrie, Universität Göttingen

Forstökonomie und nachhaltige Landnutzungsplanung, Universität Göttingen

Contents

Zusammenfassung (German)

Sobald Daten über der Zeit, ergo in einer Zeitreihe erfasst werden, stößt die klassische Regression aufgrund der Abhängigkeit der Werte von ihren Vorgängern an ihre Grenzen. Solche Daten finden sich häufig in der Forstökonomie, die Anwendung von (multivariater) Zeitreihenanalyse in diesem Bereich scheint in Deutschland jedoch nicht sehr verbreitet. Ziel dieser Arbeit ist es, diese Situation zu verbessern, indem den Lesenden ein Einblick in das Thema gewährt wird sowie die mit der Analyse von Zeitreihen verbunden Möglichkeiten aufgezeigt werden.

Part I befasst sich mit den theoretischen Grundlagen. Abschnitt 1 behandelt grundlegende Definitionen und Konzepte sowie univariate Modelle. Der folgende Abschnitt 2 thematisiert elementare multivariate Modelle, wie das Vector Autoregressive Model (VAR) in 2.1, das Vector Error Correction Model (VECM) in 2.2, Impulse Response Functions (IRF) in 2.3 sowie die Forecast Error Variance Decomposition (FEVD) in 2.4. Die jeweiligen Abschnitte enthalten sowohl einen theoretisch-mathematischen Anteil, welcher die zugrundeliegenden Formeln erläutert, als auch angewandte Teile, in welchen Daten mit R simuliert und analysiert werden (gekennzeichnet mit `R:`).

Part II transferiert die Theorie auf praktische Beispiele in Form von drei vergleichsweise einfach gehaltenen Fallstudien:

Betriebserfolg und BIP (4.1): Mithilfe eines VECMs wird in einem ersten Minimalbeispiel der Frage nachgegangen, ob die Entwicklung des Umsatzes der Forstbetriebe mit der generellen wirtschaftlichen Entwicklung in Nordrhein-Westfalen (NRW) verknüpft ist.

Die verwendeten Daten sind einerseits besagter Umsatz in [€ / ha], entnommen dem "Betriebsvergleich Westfalen-Lippe" (BVGL, Dög et al. (2017)), andererseits das Bruttoinlandsprodukt (BIP) von NRW in [Mrd. €], bereitgestellt durch das Statistische Bundesamt (DESTATIS).

Es konnten keine Zusammenhänge in der kurzfristigen Entwicklung festgestellt werden, allerdings fand sich ein langfristiges Gleichgewicht zwischen den Variablen in Form eines linearen Zusammenhanges: $Umsatz - 0.178 \cdot BIP = 0$. Hierbei folgte jedoch eher der Umsatz der Entwicklung des BIP als umgekehrt.

Preisinduzierte Einschlagsänderungen (4.2): Es interessierte die Frage, ob die Betriebe des BVGL ihre Einschlagsaktivitäten an sich ändernde Preise anpassen. Es wurde je ein VAR pro Beratungsring des BVGL geschätzt, mit den Einschlagsmengen für Buche und Fichte als endogene sowie ihre aktuellen Stammholzpreisindizes (DESTATIS) als exogene Variablen.

Die geschätzten Parameter zeigten, dass die Fichtenernte bei steigendem Buchenpreis reduziert wurde, jedoch nicht umgekehrt. Allerdings war die Anpassungsgüte der Modelle sehr dürftig, sodass diese Ergebnisse weiterer Bestätigung bedürfen.

Wechselbeziehungen der Preise (4.3): Diese letzte Fallstudie untersuchte die Auswirkungen eines Schocks in den Preisen von Eiche, Buche, Fichte und Kiefer. Grundlage war der monatliche Stammholzpreis (DESTATIS). Als geeigneter Ansatz wurde eine IRF gewählt, zusätzlich wurde die FEVD berechnet.

Die Ergebnisse zeigten Wechselwirkungen verschiedener Stärken zwischen den Arten, wobei Kiefer und besonders Fichte stärkere und dauerhaftere Einflüsse zeigten als die Laubhölzer.

Abstract

Once data is recorded over time, ergo in a time series, classical regression methods have their limits due to the dependence of the values on their predecessors. Such data can be found frequently in forest economics, however, the application of (multivariate) time series analysis in this field does not seem to be used too widely in Germany. This work aims to improve this situation, by giving the reader an introduction into the topic, as well as conveying an impression of the possibilities connected with the analysis of time series.

Part I provides the theoretical basis. In section 1, basic definitions and concepts as well as univariate models are introduced. The following section 2 will deal with some basic multivariate models, such as the Vector Autoregressive Model (VAR) in 2.1, the Vector Error Correction Model (VECM) in 2.2 and Impulse Response Functions (IRF) in 2.3 as well as the Forecast Error Variance Decomposition (FEVD) in 2.4. The corresponding sections contain a theoretical-mathematical part, which explains underlying formulas, as well as a more applied sections where data is simulated and analyzed with R (denoted `R:`).

In part II, the theory will be applied to practical examples on the basis of three case studies, which were kept comparably simple:

Business Revenue and GDP (4.1): The question, whether the development of the forest businesses' revenue is connected to the general economic development in North Rhine-Westphalia (NRW) is tried to be answered with a VECM in a minimal starting example.

The data used is said revenue in [€ / ha] from the "Betriebsvergleich Westfalen-Lippe" (BVGL) Dög et al. (2017) as well as the Gross Domestic Product (GDP) of NRW [bn. €] obtained from the German Federal Statistical Office (DESTATIS).

No connections were found in the short run dynamics, however, there seemed to be a long-term equilibrium between the two variables in the form of a linear relationship with $revenue - 0.178 \cdot GDP = 0$. Here the connection from revenue to GDP was much stronger than vice versa.

Price-induced logging changes (4.2): It was of interest, whether the enterprises in the BVGL shift logging activities due to changing prices. The chosen model was one VAR for each consulting ring of the BVGL, with the logging for the beech and the spruce as endogenous and their current trunk wood price index, obtained from DESTATIS, as exogenous variables.

The estimated parameters indicated, that the Spruce harvest was decreased when the beech price increased, but not vice versa. However, since the fit of the model was rather poor, these results need further investigation.

Price Interdependencies (4.3): The last case study examined the effect of a shock in the prices of oak, beech, spruce and pine. The data was the monthly trunk wood price index, again obtained from DESTATIS. An IRF was chosen as an adequate approach, further the FEVD was computed.

The results showed interdependencies of varying strength between the species, with pine and especially spruce exerting larger and more consistent influences than the deciduous wood.

Introduction

Once data is recorded over time, it is called a time series and has to be analyzed with special models and methods, such that the characteristics of the data generating processes can be determined adequately (Lütkepohl, 2007). Such data is frequently present in forest sciences, for example in the "Betriebsvergleich Westfalen-Lippe" (Business Comparison Westphalia-Lippe) (Dög et al., 2017).

However, almost no application of (multivariate) time series analysis in the fields of forest economics in Germany could be found in the scope of this work. The exception being one scientific paper by Kolo and Tzanova (2017), whose aim was to forecast the German forest products trade.

In the U.S., on the contrary, time series analysis of topics of forest sciences, especially economics, has been used for more than 60 years. For example by Pringle (1954) (cited after Buongiorno (1996)) and McKillop (1967), who already used multivariate time series methods. Further examples from North-America are Alavalapati et al. (1997) and Shahi and Kant (2009). From Europe, the studies by Toppinen (1998), Hetemäki et al. (2004) and a collection of articles by Abildtrup et al. (1999) shall be mentioned here.

This work aims to improve said situation in Germany by providing a guideline on how to use multivariate time series analysis for applied work. It is targeted at researchers in the field of forest sciences, especially economics, whose prior knowledge in the field of time series analysis is sparse at the most, but have an otherwise substantial statistical background.

To convey impressions of theory and praxis, the work is organized in two parts: Part I has a more theoretical point of view and provides an introduction into the basic topics of multivariate time series analysis, accompanied by simulated examples. This part will be comparably detailed to make this work comprehensible for the target group.

Part II will apply the introduced models and methods to three case studies, which were kept comparably simple to generate understandable results. Moreover, the focus will lie on choosing the correct methodology for a given question and dataset. Thus enabling the reader to apply said methodology to own questions, while also giving an impression of the possibilities and limits connected with the analysis and hopefully inducing future work in this field.

For the case studies, data of the already mentioned "Betriebsvergleich Westfalen-Lippe" (Dög et al., 2017), among others, will be used. Its database holds annual economical key figures of up to 50 private forest enterprises continuously since 1969, therefore making it the oldest time series of its kind in Germany.

Part I starts with some basics of univariate time series analysis to give an introduction into the topic and to further define and illustrate underlying concepts. Later on, multivariate time series models will be dealt with: The Vector Autoregressive Model (VAR) is presented in 2.1, section 2.2 addresses the Vector Error Correction Model (VECM) for variables which follow a common trend. Impulse Response Functions (IRF), which trace the shock in a variable over time, are outlined in 2.3. Finally, 2.4 deals with the Forecast Error Variance Decomposition (FEVD).

The following part II contains applications of the models and methods introduced earlier. 4.1 uses a VECM to study connections between the development of forestry and the general development in North Rhine-Westphalia. Whether forest enterprises shift logging activities with price changes is investigated in 4.2 with a VAR approach, and, at last, 4.3 examines possible interdependencies in trunk wood prices with IRFs and a FEVD.

Part I
Theory

A basic assumption of conventional regression methods is, that the observed data is an independent and identically distributed random variable (i.i.d.). Usually this can be achieved via a suitable, randomized design of the sample, for example for the current diameter of trees, the number of earthworms in a defined amount of soil or the revenue of saw mills in one period. (e.g. Fahrmeir et al. (2013))

However, once the data is recorded over time, the assumption of independence usually fails, because the value of a variable in one period will usually influence the value in the next period(s). Therefore a certain amount of autocorrelation will be found and the measures are no longer independent. Such data recorded over time is called a *time series.* (e.g. Lütkepohl (2007), Shumway and Stoffer (2006))

At the beginning of this part some basic concepts of univariate time series analysis will be addressed as an introduction to the topic and to define some concepts, like *white noise* or *stationarity.* Further, the Autoregressive Model (AR) as a simple form of time series analysis and the basis of further models will be outlined.

In section 2 more sophisticated multivariate models will be dealt with. Such as the Vector Autoregressive Model in 2.1. Besides the basic formula and the assumptions, topics like stability, forecasting and diagnostic tests will be addressed. The Vector Error Correction Model is described in 2.2. Finally, 2.3 deals with (orthogonal) Impulse Response Functions and the Forecast Error Variance Decomposition is the topic of 2.4.

Each section will consist of a theoretical part, which covers the underlying concepts and equations, as well as more practical sections, where data will be simulated and analyzed with R (marked with `R:`). The analysis of real data will be dealt with in part II.

All R-scripts and graphics are also available on the enclosed CD and at GitHub: https://github.com/jan-schick/Masters_Thesis.

1 Univariate Time Series Analysis

Consider a univariate time series, meaning that for each point t in time there is only one value, denoted by y_t. Hence y_{t-k} is used for a value which was recorded k periods ago. An example for such a series would be the annual marginal return of a forest enterprise or a group of enterprises, or the wood price over time.

In such a time series it might occur, that y_t strongly depends on the past values, i.e. y_{t-1}, y_{t-2}, et cetera. This would be called and *autoregressive process* (Shumway and Stoffer, 2006). For example, the present value could be given by 0.9 times the last value plus an error term u_t:

$$y_t = 0.9 \cdot y_{t-1} + u_t \tag{1.1}$$

The error term basically contains all other, unobserved influences on y_t. (Lütkepohl, 2007, Shumway and Stoffer, 2006) Such a time series is depicted in figure 1.1 alongside the result of a simple linear regression of y over time. Note, that the time has been centered around 250 for clearer results.

The regression estimated a non-significant intercept of -0.066 and a highly significant slope of 0.0036 (p-value: $4.09 \cdot 10^{-6}$). However, since the error term is autocorrelated the used t-statistic is distorted (e.g. Fahrmeir et al. (2009)). Moreover, it is obvious, that this ordinary linear regression is of no benefit, because it couldn't grasp the underlying mechanism of the data generating algorithm.

In the following chapters more adequate methods to analyze time series data will therefore be discussed. The univariate part serves as an introduction into the whole topic and will mostly define basic concepts.

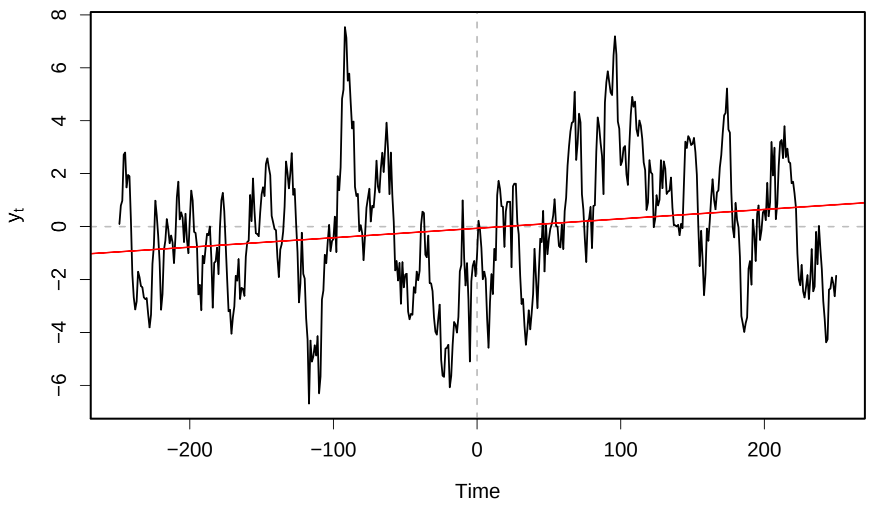


Figure 1.1: Using ordinary linear regression (red) on a time series generated by an autoregressive process (black).

1.1 White noise

First, briefly consider the error term u_t: A common assumption for the error term would be, that it is without autocorrelation, with time-invariant mean zero and variance σ^2. Such a time series is called *white noise*, or more compact: $wn(0, \sigma^2)$ (see fig. 1.2). A white noise series can for example be produced via drawing from a Gaussian distribution. (cp. Lütkepohl (2007), Pfaff (2008a), Shumway and Stoffer (2006))

A white noise error term in a model is assumed to imply that there are no informative components left unexplained by the model and the unexplained part itself therefore consists solely of random fluctuations. The assumption $u_t \sim wn(0, \sigma^2)$ is therefore quite crucial in the analysis. (Lütkepohl, 2007, Shumway and Stoffer, 2006)

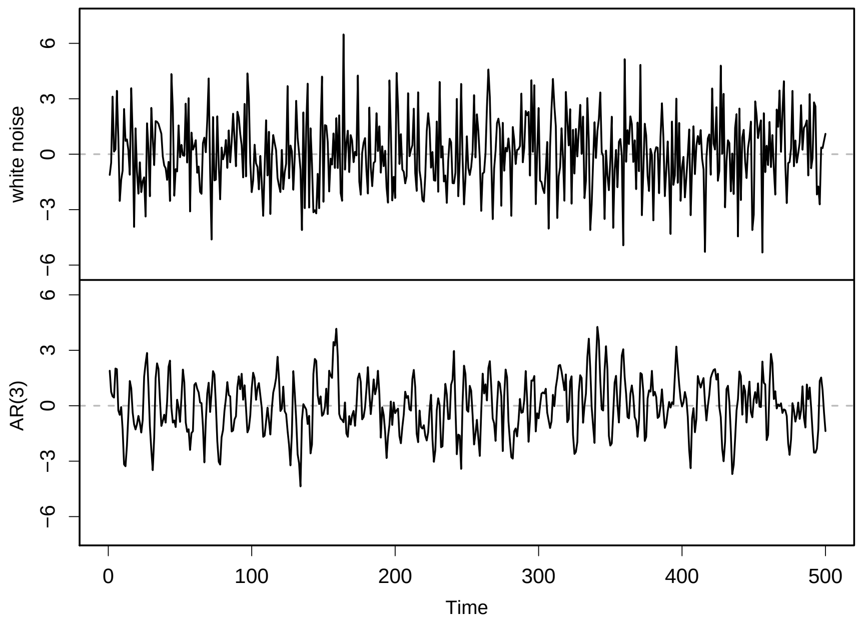


Figure 1.2: *White noise* with $\sigma^2 = 9$ (top) and an *AR(3)*-process with $a_1 =$ 0.9, $a_2 =$ -0.5, $a_3 =$ 0.05 and $u_t \sim wn(0,\ 1)$ (bottom).

1.2 Autoregressive Model

The basic idea behind the autoregressive model (AR) is, that a time series value at point t is influenced by its predecessors. A simple model, with a structure similar to (1.1), would be

$$y_t = a \cdot y_{t-1} + u_t \tag{1.2}$$

where y_t is the result of a times the last value plus a white noise error term (u_t). Because y_t only depends on the last value, the first *lag*, this is called a *first order autoregressive model*, or *AR(1)*. (Lütkepohl, 2007, Shumway and Stoffer, 2006) The order can in theory be increased arbitrarily, the resulting model of order p

$$y_t = a_1 \cdot y_{t-1} + ... + a_p \cdot y_{t-p} + u_t \tag{1.3}$$

is denoted with *AR(p)* and includes the last p lags. The basic difference to the ordinary regression in fig. 1.1 is, that each y_t is not regressed on t, but on the corresponding $y_{t-1}, \ldots, y_{t-p}$. (Lütkepohl, 2007, Shumway and Stoffer, 2006) For example, data generated by an *AR(3)*-process can be seen in fig. 1.2.

1.3 Stationarity and Integration

When dealing with time series the concept of stationarity is of great importance. Here, the definition of "weak stationarity" (Shumway and Stoffer (2006), p. 24) will be used: A time series y_t will be called *stationary*, or *stable*, if

1. the time series has a constant mean which does not depend on time t, and
2. the covariance of two points at time t and s depends only on the difference $|t-s|$ and not on t or s themselves.

An example of stationary data is provided by figure 1.2, non-stationary data can be seen in figure 1.3. There are stronger assumptions on stationarity which have a more theoretical viewpoint, but those will not be used for this work (see e.g. (Shumway and Stoffer, 2006)).

Whether an *AR(p)* process is stationary depends on its parameters. Based on equation (1.3), one can construct a polynomial of the form

$$z^p - a_1 z^{p-1} - ... - a_{p-1} z - a_p \tag{1.4}$$

and find the corresponding roots. If the modulus[1] of all roots is smaller than one, the process is stable. (Lütkepohl, 2007, Shumway and Stoffer, 2006)

If at least one of the moduli is exactly one and none is greater than one, the process is called a *random walk*. If there exists at least one modulus greater than one, the process is called *unstable*. Such processes can be seen in fig. 8.1 in the Appendix. (Lütkepohl, 2007, Shumway and Stoffer, 2006)

According to Box and Jenkins (1970) any non-stationary time series can be converted to a stationary form by differencing a sufficient number of times, e.g. y_t is non-stationary, $\Delta^d y_t$ is stationary for sufficiently large d.[2] Such a time series would be called *integrated of order* d, or short $I(d)$, where d is the smallest order difference needed for a stationary series. (Box and Jenkins, 1970)

[1] The modulus is needed for the occurrence of an imaginary unit i in the roots: $Mod(a + b * i) := \sqrt{a^2 + b^2}$. (Lütkepohl, 2007)

[2] Δ^d stands for the d^{th} order difference.

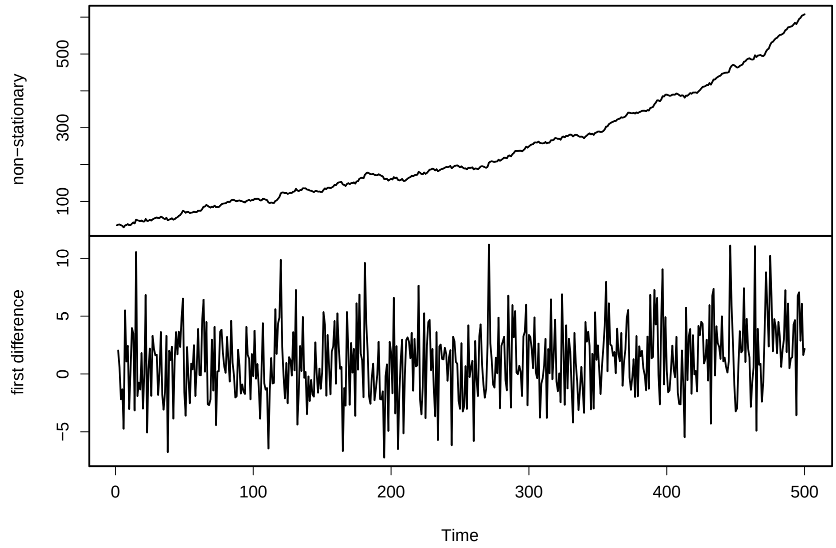


Figure 1.3: Non-stationary time series generated by an *AR(1)* process with $a = 1.005$ and $u_t \sim wn(0,\ 9)$ and its first order difference.

The stationarity of data can be tested with the Augmented Dickey-Fuller test (ADF), which uses the Null-Hypothesis of a unit root in the data. (Pfaff, 2008a) For the time series in figure 1.3 the first order difference was sufficient to reject said Null-Hypothesis.

Another option is the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test, which actually uses stationarity as the null hypothesis and further allows to test for trend-stationarity. (Kwiatkowski et al., 1992) Both tests are implemented, among others, in the package `tseries`, functions `adf.test()` and `kpss.test()`, respectively. (Trapletti and Hornik, 2018)

1.4 Simulations with R

Generating white noise in R is pretty straight forward: First a seed is set for reproducible results, afterwards $n = 500$ values are drawn from a Gaussian distribution with e.g. $\sigma = 2$ and stored as a time series:

```
set.seed(123)
wn <- as.ts(rnorm(500, 0, 2))
```

The code below simulates a stable $AR(p)$ process with a previously determined starting value at $t = 0$, here it is zero. Therefore a “burn-in” phase, i.e. discarding the first 20 values, was used, such that the process now starts at a random point. Therefore the initial vector needs to have 520 elements. Here an $AR(3)$ process is specified with a_1 = 0.9, a_2 = -0.5, a_3 = 0.05, the error term is white noise with variance 1.

```
set.seed(13)
y <- vector(length = 520, mode = "numeric")
for(i in seq(4, 520)){
    y[i] <- 0.9 * y[i-1] -0.5 * y[i-2] + 0.05 * y[i-3] + rnorm(1, 0, 1)
}
y <- as.ts(y[-1:-20])
```

The result of both calculations can be seen in fig. 1.2.

R provides the `ar()` command from the `stats` package (R Core Team, 2018) for the analyses of an *AR(p)* process. The lag order is chosen automatically according to the AIC. This command will not be discussed in detail in this work. However, when used on the time series y created above (via `ar(y)`) an *AR(2)* model with $a_1 = 0.8754$ and $a_2 =$ -0.4068 was estimated. a_3 was not estimated here, seemingly its influence on the the time series was too small. The variance of the error term was estimated as 1.105.

The unstable time series from figure 1.3 can be created with the same code already used above. The seed has to be changed to 1337, a_1 to 1.005, a_2 and a_3 to zero, thus creating a *AR(1)*, and the standard deviation of the white noise error term to 3. The difference was calculated via `diff(y)`. The calculation is not shown here, however, it can be found in the corresponding script on the enclosed CD / GitHub.

The Augmented Dickey-Fuller Test (`adf.test()`) from the package `tseries` (Trapletti and Hornik, 2018) was also used here. The test returned a test-statistic of 1.8358 and a p-value of >0.99 for the unstable series as well as -7.3692 and $p < 0.01$ for the first order difference, thus indicating that the series is *I(1)*.

The roots of the polynomial described in (1.4) were calculated with the function `polyroot(c(` `-`a_p`, ..., -`a_1`, 1))`. The modulus was obtained via `Mod()` from the `base` package (R Core Team, 2018).

2 Multivariate Time Series Analysis

Multivariate methods are based on their univariate counterparts, however, their equations have a higher level of complexity. The basic idea is, that a time series does not only depend on its own past values, but on those of other time series, recorded at the same time, as well. (Lütkepohl, 2007)

For example, the change in the price of firewood from one year to another, Δw_t, may depend on the last firewood price change, Δw_{t-1}, and the last change in the oil price, Δo_{t-1}. For example, say the changes could be described as

$$\begin{aligned}\Delta o_t &= 0.9 \cdot \Delta o_{t-1} + u_{o,t} \\ \Delta w_t &= 0.5 \cdot \Delta o_{t-1} + 0.8 \cdot \Delta w_{t-1} + u_{w,t}\end{aligned}$$

with Δo_t being a simple *AR(1)* process. $u_{o,t}$ and $u_{w,t}$ are white noise error terms with variances 1.0 and 0.1, respectively.

If one would not know the true relationship given above, and assumed that Δw_t only depends on itself, the analysis would fail: When using the `ar()` function on different simulated values for Δw_t, ARs of varying order with a variance distinctly larger than 0.1 were estimated and the true process remained uncovered. (cp. Lütkepohl (2007))

In practical application, most processes are assumed to be generated by systems of interdependent variables (e.g. Abildtrup et al. (1999), Toppinen and Kuuluvainen (2010)). Such as a price for a specific forest product, which at least depends on supply and demand. Therefore this chapter will deal with interdependent data of higher dimensions.

Let $\boldsymbol{y}_t$ denote a K-dimensional vector of different time series values y_i (with $i = 1, \ldots, K$) at point t. $\boldsymbol{y}_{t-k}$ stands for a vector of values sampled k periods ago, respectively. In the example given above, $\boldsymbol{y}_t$ would be

$$\boldsymbol{y}_t = \begin{bmatrix} \Delta o_t \\ \Delta w_t \end{bmatrix}$$

and could for example be analyzed with a Vector Autoregressive Model, which is presented in the next section. In 2.2, the Vector Error Correction Model for variables which follow a common trend will be dealt with. In 2.3 the Impulse Response Function, used to examine the influence of a shock in one variable, will be addressed. Finally, section 2.4 presents the Forecast Error Variance Decomposition (FEVD).

2.1 Vector Autoregressive Model (VAR)

A Vector Autoregressive Model of order p (*VAR(p)*) is the expansion of its univariate counterpart. Meaning, that the value of each $y_{i,t}$, stored in $\boldsymbol{y}_t$, depends on the values of $\boldsymbol{y}_{t-1}$ to $\boldsymbol{y}_{t-p}$, as in the example given above. The main assumptions of a *VAR(p)* are, that all time series stored in $\boldsymbol{y}_t$ are stationary with white noise error terms. (Lütkepohl, 2007)

To make this more clear, consider a general set of *VAR(1)* equations for a two dimensional vector

$$\begin{aligned} y_{1,t} &= a_{11}y_{1,t-1} + a_{12}y_{2,t-1} + u_{1,t} \\ y_{2,t} &= a_{21}y_{1,t-1} + a_{22}y_{2,t-1} + u_{2,t} \end{aligned} \tag{2.1}$$

where the coefficients a_{jk} show the influence of $y_{k,t-1}$ on $y_{j,t}$. $u_{i,t}$ denotes a white noise error term for $y_{i,t}$. The coefficients and data can then be arranged in vectors and a matrix, such that

$$\boldsymbol{y}_t = \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix}, \quad \boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad \boldsymbol{u_t} = \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix}$$

and therefore the equations above are represented by

$$\boldsymbol{y}_t = \boldsymbol{A} \cdot \boldsymbol{y}_{t-1} + \boldsymbol{u}_t \tag{2.2}$$

in a more compact form. (cp. Lütkepohl (2007)) In the example given in the introduction the vectors and the matrix would be as follows:

$$\boldsymbol{y}_t = \begin{bmatrix} \Delta o_t \\ \Delta w_t \end{bmatrix}, \quad \boldsymbol{A} = \begin{bmatrix} 0.9 & 0.0 \\ 0.5 & 0.8 \end{bmatrix}, \quad \boldsymbol{u_t} = \begin{bmatrix} u_{o,t} \\ u_{w,t} \end{bmatrix}$$

One result generated by such a process is shown in fig. 2.1, the used R-Code can be found in section 2.1.2.

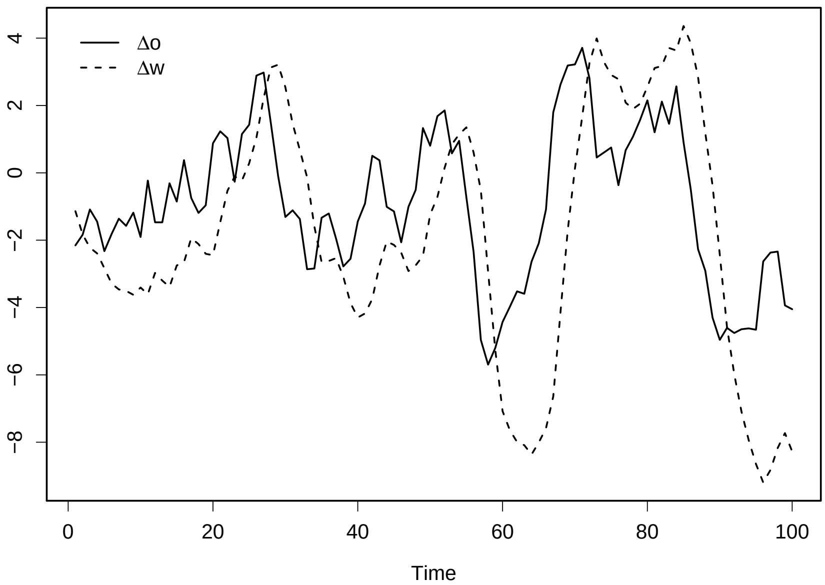


Figure 2.1: Simulated two-dimensional *VAR(1)* process for the changes in the oil price (Δo) and firewood price (Δw).

The *VAR(1)* model can be expanded arbitrarily, both in terms of dimension and lag. Also, an intercept, denoted ν, might be included. The result is the general form of the K-dimensional *VAR(p)* model

$$\boldsymbol{y}_t = \boldsymbol{\nu} + \boldsymbol{A}_1 \boldsymbol{y}_{t-1} + ... + \boldsymbol{A}_p \boldsymbol{y}_{t-p} + \boldsymbol{u}_t \tag{2.3}$$

with $\boldsymbol{y}_t$, $\boldsymbol{u}_t$ and $\boldsymbol{\nu}$ being $K \times 1$ vectors, the last one containing the intercepts, and the $\boldsymbol{A}_i$ $(i = 1, \dots, p)$ being fixed $K \times K$ matrices, one for each lag. $\boldsymbol{u}_t$ again represents the white noise error terms with time invariant covariance matrix $E(\boldsymbol{u}_t \boldsymbol{u}_t') = \boldsymbol{\Sigma}_u$. (Lütkepohl, 2007)

2.1.1 Deterministic terms and dummy variables

The R function which will be used later on, `VAR()` from the `vars` package (Pfaff, 2008b), allows for the inclusion of so called "exogenous" variables, leading to a modified form of (2.3):

$$\boldsymbol{y}_t = \boldsymbol{\nu} + \boldsymbol{A}_1 \boldsymbol{y}_{t-1} + ... + \boldsymbol{A}_p \boldsymbol{y}_{t-p} + \boldsymbol{C}\boldsymbol{D}_t + \boldsymbol{u}_t \tag{2.4}$$

where $\boldsymbol{C}$ is a $(K \times M)$ matrix of coefficients determining the impact of the M exogenous variables on the K values of $\boldsymbol{y}_t$. The corresponding values for time t of the exogenous variables, such as a constant, a trend, or (seasonal) dummy variables, are stored in $\boldsymbol{D}_t$. (Pfaff, 2008b)

It is therefore possible to analyze data with a trend, seasonal variation or any other extraordinary events, like storms, and still have a white noise error term $\boldsymbol{u}_t$. Further implying that the data itself does not need to be stationary as described in 1.3, but can be "trend-stationary" such that $\boldsymbol{z}_t := \boldsymbol{y}_t - (a + b \cdot t)$, with a as an intercept and b as a slope, resembles a stationary process. (Pfaff, 2008b)

2.1.2 `R`: Generating and analyzing VAR processes

Based on (2.2), the example from the beginnig of this chapter will be turned into a *VAR(1)* model with coefficients $a_{11} = 0.9$, $a_{12} = 0$, $a_{21} = 0.5$ and $a_{22} = 0.8$. The standard deviation of the white noise error terms was set to 1 and 0.1, respectively. Again, the first 20 values were used as a burn-in phase. Note, that in R a time series is stored within a column, the rows store the value at period t. Figure 2.1 displays the results.

A simple way to analyze a VAR is provided by the `VAR()` command from the `vars` package (Pfaff, 2008b). The results can be stored in an object of class `varest` which can be used for further analysis or plotting. `VAR()` uses linear models to estimate the matrices $\boldsymbol{A}_i$ as well as deterministic terms (cp. 2.1.10, 2.1.12). For the example given above said term was excluded via `type = "none"`.

```
set.seed(13)
d.m <- matrix(rep(0, 240),
              ncol = 2,
              dimnames = list(NULL, c("do", "dw")))
A1 <- matrix(c(0.9, 0,
               0.5, 0.8),
             nrow = 2, byrow = F)
for(i in seq(2, length(d.m[,1]))){
  d.m[i, ] <- d.m[i-1, ] %*% A1 + c(rnorm(1), rnorm(1, 0, 0.1))
}
d.m <- as.ts(d.m[-20:-1, ])

library(vars)
(var.model <- VAR(d.m, type = "none"))
Acoef(var.model)
```

The direct output of the `VAR()` function, shown in R Output 8.1, appendix, contains the estimated coefficients. For more details the `summary()` function can be used, this will show for example the p-values of the coefficients and the covariance-matrix of the residuals. By using `Acoef` one can directly retrieve the coefficient matrix, or matrices, respectively. Parameter estimations can be found in tab. 2.1.

	a_{11}	a_{12}	a_{21}	a_{22}	$\sigma_{\Delta o_t}$	$\sigma_{\Delta w_t}$
Estimated:	1.020	-0.092	0.503	0.798	0.975	0.107
Original:	0.9	0.0	0.5	0.8	1.0	0.1

Table 2.1: Estimated VAR coefficients and original values.

2.1.3 VAR(1) representation of a VAR(p) process

As stated e.g. by Lütkepohl (2007), any K-dimensional *VAR(p)* process can be written as a Kp-dimensional *VAR(1)* process of the form

$$\boldsymbol{Y}_t = \boldsymbol{\nu} + \boldsymbol{\mathcal{A}}\boldsymbol{Y}_{t-1} + \boldsymbol{U}_t \tag{2.5}$$

with

$$\underset{(Kp\times 1)}{\boldsymbol{Y}_t} := \begin{bmatrix} \boldsymbol{y}_t \\ \boldsymbol{y}_{t-1} \\ \vdots \\ \boldsymbol{y}_{t-p+1} \end{bmatrix}, \quad \underset{(Kp\times 1)}{\boldsymbol{\nu}} := \begin{bmatrix} \nu \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \underset{(Kp\times Kp)}{\boldsymbol{\mathcal{A}}} := \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{A}_2 & \dots & \boldsymbol{A}_{p-1} & \boldsymbol{A}_p \\ \boldsymbol{I}_K & 0 & \dots & 0 & 0 \\ 0 & \boldsymbol{I}_K & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \boldsymbol{I}_K & 0 \end{bmatrix}, \quad \underset{(Kp\times 1)}{\boldsymbol{U}_t} := \begin{bmatrix} \boldsymbol{u}_t \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

with $\boldsymbol{I}_K$ being the identity matrix of size K. This representation has no influence on the parameter estimation itself but is advantageous concerning computation and is therefore utilized by some R functions which will be used later on. (Lütkepohl, 2007, Pfaff, 2008a)

2.1.4 Stability of a VAR process

As mentioned at the beginning of this section, one assumption of a VAR is that all variables resemble stable processes. One way to check the stability of a VAR, is by looking at the matrix $\boldsymbol{\mathcal{A}}$ of its *VAR(1)* representation (see (2.5)). If the moduli of the eigenvalues of said coefficient matrix $\boldsymbol{\mathcal{A}}$ are smaller than one, the process is stable. (Lütkepohl, 2007, Pfaff, 2008b, Serepka, 2012) All R-packages used here have the *VAR(1)* representation of a *VAR(p)* process implemented and therefore evaluate the eigenvalues of $\boldsymbol{\mathcal{A}}$. (Pfaff, 2008a,b)

In the interest of completeness it should be noted, that when dealing directly with a K-dimensional *VAR(p)* process, and not its *VAR(1)* representation, the roots of the *reverse characteristic polynomial* would be of interest. They (or their modulus, where necessary) would all have to be *greater than one* in absolute value fo the process to be stable. (Lütkepohl, 2007) Or more formally, according to Lütkepohl (2007):

$$det(\boldsymbol{I}_K - \boldsymbol{A}_1 z - \cdots - \boldsymbol{A}_p z^p) \neq 0 \qquad \text{for } |z| \leq 1$$

2.1.5 Forecasting

Let $\boldsymbol{y}_t = (y_{1,t}, \ldots, y_{k,t})'$ be a time series generated by a stable *VAR(p)* process (see (2.3)), for which a **point forecast** h steps into the future shall be obtained. With $\boldsymbol{u}_t$ being independent white noise with mean zero (see 1.1), we know that when looking from point t into the future the expected value of $\boldsymbol{u}_t + h$, $E_t(\boldsymbol{u}_{t+h})$, is zero for $h > 0$. Hence it can be neglected. (Lütkepohl, 2007)

A prediction one step into the future, for example, is therefore obtained via

$$E_t(\boldsymbol{y}_{t+1}) = \nu + \boldsymbol{A}_1 \boldsymbol{y}_t + \cdots + \boldsymbol{A}_p \boldsymbol{y}_{t-p+1} \tag{2.6}$$

This can be expanded to a more general form:

$$E_t(\boldsymbol{y}_{t+h}) = \nu + \boldsymbol{A}_1 E_t(\boldsymbol{y}_{t+h-1}) + \cdots + \boldsymbol{A}_p E_t(\boldsymbol{y}_{t+h-p}) \tag{2.7}$$

Those equations can then be used to predict the values of $\boldsymbol{y}_{t+1}$ with (2.6), then (2.7) is used recursively for $\boldsymbol{y}_{t+2}, \ldots, \boldsymbol{y}_{t+h}$. (Lütkepohl, 2007)

For **interval forecasts** the assumed distribution of $\boldsymbol{y}_t$, or $\boldsymbol{u}_t$, respectively, is crucial. The common assumption would be a multivariate normal distribution for all $\boldsymbol{y}_t, \ldots, \boldsymbol{y}_{t+h}$, implying $\boldsymbol{u}_t \sim \mathcal{N}(0, \Sigma_u)$. Therefore the forecast errors $\boldsymbol{y}_{t+h} - \boldsymbol{y}_t(h)$ are $\mathcal{N}(0, \Sigma_y(h))$ distributed. (Lütkepohl, 2007)

With this assumption, according to Lütkepohl (2007), one can construct a $(1-\alpha)\,100\%$ forecast interval for the components of $\boldsymbol{y}_t$:

$$y_{k,t}(h) \pm z_{(\alpha/2)}\sigma_k(h) \tag{2.8}$$

$$\begin{aligned} \text{with} \quad y_{k,t}(h) &= h\text{-step forecast of the } k\text{-th element of } y_t \\ z_{(\alpha/2)} &= (\alpha/2)100\%\text{-quantile of the standard normal distribution} \\ \sigma_k(h) &= \text{the square root of the } k\text{-th diagonal element of } \Sigma_y(h) \end{aligned}$$

2.1.6 `R`: Forecasting

To better visualize the workings behind forecasting, a two-dimensional *VAR(2)* with $\boldsymbol{y}_t = (y_{1,t}, y_{2,t})'$ was simulated, where $y_{2,t}$ almost completely depends on $y_{1,t-2}$. Furthermore $y_{2,t}$ has been given a lower standard deviation than $y_{1,t}$ to create a bigger dependence on $y_{1,t}$. Again, a burn-in phase of 20 time steps was used.

The resulting time series $\boldsymbol{y}$ was once again analyzed using the `VAR()` command of the `vars` package (Pfaff, 2008b). The lag order was set to two and the deterministic term was excluded. Finally, the values of the following ten periods were predicted.

The generated time series as well as the first six predictions including the prediction intervals can be seen in fig. 2.2.

```
set.seed(5)
n <- 30
y <- matrix(vector(length = (n+20)*2,
                   mode = "numeric"),
             ncol = 2)
A1 <- matrix(c(0.5, 0.2,
               0, 0),
              nrow = 2, byrow = T)
A2 <- matrix(c(0, 0,
               0.95, 0),
              nrow = 2, byrow = T)
for(i in 3:(n+20)){
  y[i, ] <- A1 %*% y[i-1, ] + A2 %*% y[i-2, ] + c(rnorm(1), rnorm(1, 0, 0.1))
}
y <- as.ts(y[-1:-20, ])

library(vars)
var.1 <- VAR(y, p = 2, type = "none")
summary(var.1)
var.1.pred <- predict(var.1)
```

The coefficient matrices were estimated as follows, bold font indicates that the value had a significant difference ($\alpha = 0.05$) from zero:

$$\boldsymbol{A}_1 = \begin{bmatrix} \mathbf{0.776} & 0.351 \\ 0.012 & -0.001 \end{bmatrix}, \quad \boldsymbol{A}_2 = \begin{bmatrix} -0.096 & -0.173 \\ \mathbf{0.928} & 0.022 \end{bmatrix}$$

When looking at the predicted values and the intervals (fig. 2.2) one can clearly see the dependence of $y_{2,t}$ on $y_{1,t-2}$ leading to precise estimates. After two time steps however both series drift towards zero and uncertainty increases.

In terms of the prediction intervals it should be mentioned, that here the residual standard error of the linear models used in `VAR()`, estimating the equations for the series $\boldsymbol{y}_{1,t}$ and $\boldsymbol{y}_{2,t}$ (see 2.1.11 and 2.1.12), seems to have been used instead of the k-th diagonal element of the covariance matrix of the error term (cp. (2.8)).

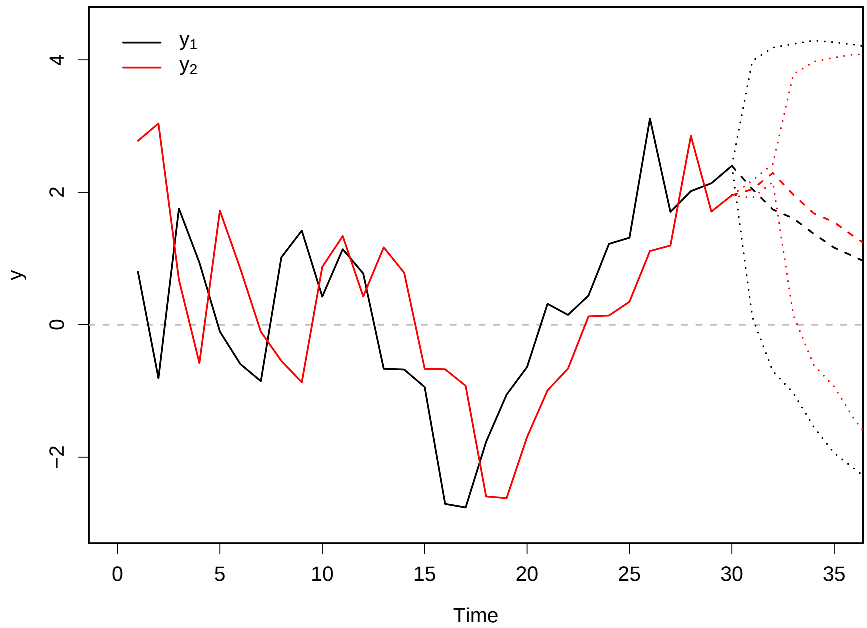


Figure 2.2: Two-dimensional *VAR(2)* process (solid lines) as well as the predicted values (dashed lines) and the confidence intervals for said values (dotted lines).

2.1.7 `R`: Restricted VAR

When looking at the estimated matrices above, one might want to exclude all non-significant coefficients from the estimation. This can be achieved with `restrict()` which will reestimate an object of class `varest`. If `method` is set to `ser`, a model will be estimated where only parameters with a t-value greater than the one specified in `thresh` will be included. The t-values for the parameters are obtained from the estimated linear models (see 2.1.11).

When in `manual` mode, a matrix of dimension $(K \times C)$, with $C = p \cdot K$ being the number of coefficients in each equation, needs to be provided. Said matrix can contain only ones and zeros, indicating which coefficients shall be estimated. (Pfaff, 2008b)

```
Acoef(restrict(var.1, method = "ser", thresh = 2.0))

res.mat <- matrix(c(1, 0, 0, 0,
                    0, 0, 1, 0),
                  nrow = 2, byrow = T)
Acoef(restrict(var.1, method = "manual", resmat = res.mat))
```

When using the t-value the matrices were estimated as

$$\boldsymbol{A}_1 = \begin{bmatrix} 0.8 & 0 \\ 0 & 0 \end{bmatrix}, \quad \boldsymbol{A}_2 = \begin{bmatrix} 0 & 0 \\ 0.936 & 0.022 \end{bmatrix}$$

with manual mode the result was

$$\boldsymbol{A}_1 = \begin{bmatrix} 0.8 & 0 \\ 0 & 0 \end{bmatrix}, \quad \boldsymbol{A}_2 = \begin{bmatrix} 0 & 0 \\ 0.946 & 0 \end{bmatrix}$$

2.1.8 Lag order selection

In practical application the lag order will usually be unknown. According to Pfaff (2008a) and Lütkepohl (2007), a suitable method for empirically determining the lag order is the use of information criteria, like those proposed by Akaike (AIC) (Akaike, 1981), Hannan and Quinn (HQ) (Hannan and Quinn, 1979, Quinn, 1980), Schwarz (SC a.k.a. BIC) (Schwarz, 1978) or the final prediction error (FPE) which minimizes the forecast MSE and is therefore useful if prediction is the objective (Lütkepohl, 2007).

A detailed discussion about the different methods will not be held here, but can be found in Lütkepohl (2007). However, it should be mentioned, that the estimation for the lag order, $\hat{p}$, obtained by the criteria will be systematically different. For a sample size T of at least 16 the relation

$$\hat{p}(SC) \leq \hat{p}(HQ) \leq \hat{p}(AIC) \tag{2.9}$$

holds. (Lütkepohl, 2007, Pfaff, 2008a) It should further be noted, that the AIC and the FPE have to be used with care when dealing with cointegrated data, because they will not be consistent. HQ and SC, however, are both consistent criteria even when dealing with integrated data. (Lütkepohl, 2007, Paulsen, 1984)

The calculation of the four criteria described here for a given set of data are implemented in the function `VARselect()` from the package `vars`. (Pfaff, 2008b)

2.1.9 Moving Average Representation

As mentioned in 2.1.3, any K-dimensional *VAR(p)* process can be written as a *VAR(1)*. If said process $\boldsymbol{Y}_t$ (see (2.5)) is stable, it can be represented by

$$\boldsymbol{Y}_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \boldsymbol{\mathcal{A}}^i \boldsymbol{U}_{t-i}$$

with mean term $\boldsymbol{\mu}$. $\boldsymbol{Y}_t$, $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{U}_{t-i}$ are the same as in (2.5). This is also called the *moving average* (MA) *representation.* (Lütkepohl, 2007) By defining the $(K \times Kp)$ matrix $\boldsymbol{J} := [\boldsymbol{I}_K : 0 : \cdots : 0]$ the MA representation of $\boldsymbol{y}_t$ is obtained:

$$\begin{aligned}
\boldsymbol{y}_t = \boldsymbol{J}\boldsymbol{Y}_t &= \boldsymbol{J}\boldsymbol{\mu} + \sum_{i=0}^{\infty} \boldsymbol{J}\boldsymbol{\mathcal{A}}^i \boldsymbol{J}'\boldsymbol{J}\boldsymbol{U}_{t-i} \\
&= \underline{\boldsymbol{\mu}} + \sum_{i=0}^{\infty} \boldsymbol{\Phi}_i \boldsymbol{u}_{t-i} \qquad (2.10) \\
\text{with} \quad \underline{\boldsymbol{\mu}} &= \boldsymbol{J}\boldsymbol{\mu} \\
\boldsymbol{\Phi}_i &= \boldsymbol{J}\boldsymbol{\mathcal{A}}^i \boldsymbol{J}'
\end{aligned}$$

In other words, the $\boldsymbol{\Phi}_i$ represent the upper left hand $(K \times K)$ corner of the $\boldsymbol{\mathcal{A}}^i$ matrix. This will be used again when dealing with Impulse Response Functions and the Forecast Error Variance Decomposition in 2.3 and 2.4, respectively. (Lütkepohl, 2007)

2.1.10 Estimation of a Vector Autoregressive Progress

In this part the least squares (LS) estimator for a *VAR(p)*, as shown in Lütkepohl (2007), will be dealt with. First, it is assumed that a sample of size T for all K variables is available, ergo a time series $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T)$. To be able to regress $\boldsymbol{Y}$ over the last p lags, one also needs p presample values for each variable, $\boldsymbol{y}_{-p+1}, \ldots, \boldsymbol{y}_0$.[3] Those are then combined to matrices:

$$\begin{aligned}
\boldsymbol{Y} &:= (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T) && (K \times T) \\
\boldsymbol{y} &:= vec(\boldsymbol{Y}) && (KT \times 1) \\
\boldsymbol{Z}_t &:= \begin{bmatrix} 1 \\ \boldsymbol{y}_t \\ \vdots \\ \boldsymbol{y}_{t-p+1} \end{bmatrix} && ((Kp+1) \times 1) \\
\boldsymbol{Z} &:= (\boldsymbol{Z}_0, \ldots, \boldsymbol{Z}_{T-1}) && ((Kp+1) \times T)
\end{aligned}$$

with *vec* being the column stack operator. (Lütkepohl, 2007) Moreover, one can rearrange the components of the basic VAR equation, (2.1):

[3]Note, that the division in sample and presample is a technicality. Consider for example a time series $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_s)$ with $s > p$. Here $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$ would be the "presample" and $\boldsymbol{x}_{p+1}, \ldots, \boldsymbol{x}_s$ would resemble the "sample".

$$
\begin{array}{lll}
\boldsymbol{B} & := (\nu, \boldsymbol{A}_1, \ldots, \boldsymbol{A}_p) & (K \times (Kp+1)) \\
\boldsymbol{U} & := (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T) & (K \times T) \\
\boldsymbol{\beta} & := vec(\boldsymbol{B}) & ((K^2p+K) \times 1) \\
\boldsymbol{u} & := vec(\boldsymbol{U}) & (KT \times 1)
\end{array}
$$

For $t = 1, \ldots, T$, this allows a more compact notation of (2.1):

$$\boldsymbol{Y} = \boldsymbol{B}\boldsymbol{Z} + \boldsymbol{U} \tag{2.11}$$

Or in the general form of a linear model[4]

$$\boldsymbol{y} = (\boldsymbol{Z}' \otimes \boldsymbol{I}_K)\boldsymbol{\beta} + \boldsymbol{u} \tag{2.12}$$

with covariance matrix of $\boldsymbol{u}$ being $\boldsymbol{\Sigma_u} = \boldsymbol{I}_T \otimes \boldsymbol{\Sigma}_u$. (Lütkepohl, 2007) This allows to compute the LS estimator for $\boldsymbol{\beta}$,

$$\hat{\boldsymbol{\beta}} = ((\boldsymbol{Z}\boldsymbol{Z}')^{-1}\boldsymbol{Z} \otimes \boldsymbol{I}_K)\boldsymbol{y} \tag{2.13}$$

2.1.11 Estimation in the vars package

In the `vars` package, which will be used in this work, estimation is more straight forward: Let $\boldsymbol{Y} = \boldsymbol{y}_1, \ldots, \boldsymbol{y}_T$ be a K-dimensional time series for which a *VAR(p)* shall be estimated. As it can be seen in the source code, K separate linear models are estimated with the `lm()` function of the package `stats` (R Core Team, 2018). Those separately regress the values of each of the K series stored in $\boldsymbol{Y}$ for the periods $t = p+1, \ldots, T$ on the specified number of lags of all variables. Meaning, that initially an individual equation for each of the K series will be given instead of the $\boldsymbol{A}_i$ matrices. Furthermore an ordinary Least Squares approach is chosen here instead of the Estimated Generalized Least Squares (EGLS) estimator proposed by Lütkepohl (2007)[5]. (Pfaff, 2008b)

The `VAR()` command also allows the inclusion of deterministic terms in form of a constant, a trend or both, as shown in 2.1.1. When looking at the source code it can be seen, that for the estimation of a constant a column of ones is added to the design matrix of the linear model. A trend on the other hand is estimated via another extra column containing an integer sequence from one to the number of periods. (Pfaff, 2008b)

2.1.12 `R`: Estimation within vars

To demonstrate the workings behind the estimation of the VAR-parameters in the `VAR()` function, the *VAR(2)* time series `y` as well as the fitted model `var.1` generated in 2.1.6 is used

[4] $\otimes$ denotes the Kronecker product
[5] For more details see Lütkepohl (2007), p. 193 ff.

once more. With this series a linear model is formulated which regresses $y_{1,t}$ over the first two lags of all the other variables, which is the exact same mechanism as implemented in `VAR()`. Note, that in R the last value of a time series is the latest.

```
lm(y[3:30, 1] ~ y[2:29, 1] + y[2:29, 2] +
             y[1:28, 1] + y[1:28, 2] - 1)$coefficients
var.1$varresult$Series.1$coefficients
```

For the sake of completeness, the estimated coefficients of both functions can be seen in tab. 2.2, with $a_{i,j,k}$ denoting the element in the jth row and kth column of matrix A_i:

Coefficients:	$a_{1,1,1}$	$a_{1,1,2}$	$a_{2,1,1}$	$a_{2,1,2}$
`lm(...)`	0.77623940	0.35081395	-0.09563771	-0.17279360
`VAR(...)`	0.77623940	0.35081395	-0.09563771	-0.17279360

Table 2.2: Comparison of values estimated with `lm()` and `VAR()`.

2.1.13 Diagnostic tests

After the estimation of a VAR model, it is of interest, whether the residuals behave as expected. Meaning that they are normally distributed, homoscedastic and without autocorrelation. In the following the tests implemented in the package `vars` by Pfaff (2008b) will be discussed briefly. The mathematical details will not be dealt with in this work, however, a short outline can be found in Pfaff (2008a), pp. 28-34, a more detailed explanation can be found in Lütkepohl (2007), e.g. section 4.4.

Autocorrelation: First, the function `serial.test()` provides several tests for the autocorrelation[6] of the residuals, which can be chosen over the argument `type`:

1. `PT.asymptotic`: The so called *portmanteau* test, which uses the null hypothesis of no autocorrelation in the residuals, the test statistic is approximately χ^2-distributed. (Lütkepohl, 2007, Pfaff, 2008a)

2. `PT.adjusted`: A small sample correction for the above mentioned portmanteau test. (Lütkepohl, 2007, Pfaff, 2008a)

3. `BG`: An LM-statistic by Breusch (1978) and Godfrey (1978) which uses auxiliary regressions to test autocorrelation up to lag order h, the χ^2-distributed test-statistic also has the null hypothesis of no autocorrelation. (Pfaff, 2008a)

4. `ES`: A small-sample correction applied to the test statistic of the foregoing test by Edgerton and Shukur (1999). (Pfaff, 2008a)

[6] Also called "serial correlation".

Heteroscedasticity: The concept of Autoregressive conditional heteroscedasticity (ARCH) was introduced by Engle (1982), which assumes that the variance of the error term varies over time. To check whether such effects are present, an ARCH test can be conducted with the function `arch.test()`. The Null hypothesis of the χ^2-distributed test statistic is homoscedasticity, ergo no ARCH effects, for the last q lags (Engle, 1982, Lütkepohl, 2007, Pfaff, 2008a). The test is applicable to uni- and multivariate time series as well. (Pfaff, 2008a)

Normality: To test the residuals for univariate and multivariate normality, the test developed by Jarque and Bera (1987) can be used. There are not only test statistics for normality, but also for skewness and kurtosis. The respective null hypothesis is always a regular normal distribution. (Jarque and Bera, 1987, Lütkepohl, 2007, Pfaff, 2008a)

Application of these functions can be found in part II.

2.2 Vector Error Correction Model (VECM)

In the last section one basic assumption was, that all included time series are stationary (or *I(0)*). Although it is possible to convert any time series integrated of order d (or *I(d)*) to a stationary one by differencing a sufficient number of times (see 1.3), some characteristics may be lost in the process. (Lütkepohl, 2007)

This is especially true if the variables follow a common trend, which will be lost completely when taking differences (see fig. 2.3). A common trend is also frequently present in the "Betriebsvergleich Westfalen-Lippe" (Dög et al., 2017), particularly when looking at the same feature for different tree species or consulting rings (see 3.1), respectively.

Therefore the following sections will address possibilities to account for common trends in the data. First, some basics regarding the concept of *cointegration* will be considered. The rest of this section will primarily deal with the *Vector Error Correction Model (VECM)* and its properties as well as the estimation of the related parameters.

2.2.1 Cointegrated processes

If two or more variables follow a common trend, a state of equilibrium in form of a linear relationship between the values can be assumed. Take for example the time series depicted in

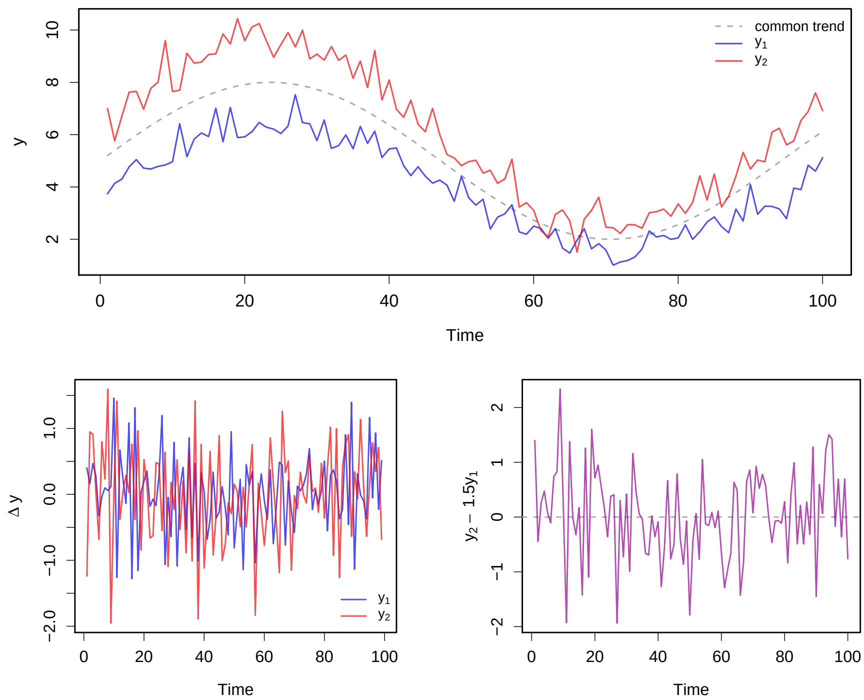


Figure 2.3: Two variables (y_1 and y_2) with a common trend (top), their first order differences (bottom left) and the deviations from the equilibrium (bottom right).

fig. 2.3, where both series follow a common trend and $y_{2,t}$ is on average 50 % larger than $y_{1,t}$. Because of the latter relationship, the difference $z_t := y_{2,t} - 1.5 \cdot y_{1,t}$ resembles a stationary process with an expected value of zero.

Such a linear relationship yielding a stationary process may also be found in systems with more than two variables. To put it in general terms: If K variables with a common trend are stored in $\boldsymbol{y_t} = (y_{1,t}, \ldots, y_{K,t})'$, the long-run equilibrium may be described as $\boldsymbol{\beta}'\boldsymbol{y_t} = \beta_1 y_{1,t} + \cdots + \beta_K y_{K,t} = 0$ with $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)'$. (Lütkepohl, 2007)

As indicated before, this relationship may not always hold precisely because of unexplained variation. In this case we'll have $\boldsymbol{\beta}'\boldsymbol{y_t} = \boldsymbol{z}_t$ with $\boldsymbol{z}_t$ denoting the deviations from the equilibrium. If the variables of $\boldsymbol{y_t}$ move together, a stationary $\boldsymbol{z}_t$ is quite plausible. Therefore, although the components of $\boldsymbol{y}_t$ may be integrated, there exists a linear combination $\boldsymbol{\beta}'\boldsymbol{y}_t$ which is stationary, as in fig. 2.3. Integrated variables which fulfill this condition are called *cointegrated*. (Lütkepohl, 2007)

More formally, a process will be called cointegrated if at least one variable $y_{i,t}$ $(i = 1, \ldots, K)$ is integrated of order d with $d > 0$ and there exists a linear combination $\boldsymbol{\beta}'\boldsymbol{y}_t$ with $\boldsymbol{\beta} \neq 0$ whose order of integration is less than d. Note, that this includes the possibility for a non-stationary $\boldsymbol{\beta}'\boldsymbol{y}_t$. $\boldsymbol{\beta}$ is then called the *cointegrating vector*, $\boldsymbol{\beta}'\boldsymbol{y}_t$ is called a *cointegration relation*. (Johansen, 1995, Lütkepohl, 2007)

Although this differs from the original definition given by Engle and Granger, where it is required that "all components of [$\boldsymbol{y}_t$] are *I(d)*" (Engle and Granger (1987), p. 3), and therefore integrated of the same order d, the definition used here is more suitable for practical applications. (Lütkepohl, 2007)

Note, that the cointegrating vector $\boldsymbol{\beta}$ is not unique. If it is multiplied with any nonzero constant, another cointegrating vector is returned. A solution for this is, to normalize $\boldsymbol{\beta}$ to the first variable, meaning that $\boldsymbol{\beta}$ is multiplied by the reciprocal of its first element, β_1^{-1}. Thus implicitly setting the first element of $\boldsymbol{\beta}$, β_1 to one. (Johansen, 1995, Lütkepohl, 2007) The issue of non-uniqueness will be addressed further in section 2.2.3.

Moreover there might be more than one cointegration relation, for example when a stationary combination of the first and the last two elements of a four-dimensional $\boldsymbol{y}_t$ exists. In this case there would be two cointegrating vectors, one with zeros in the first two positions, and one with zeros in the last two positions. Lütkepohl (2007)

2.2.2 The VECM representation

The basic idea behind a *Vector Error Correction Model (VECM)* is, that an equilibrium between two or more variables exists and that the changes in the variable are at least partially driven by the deviation from said equilibrium. (Johansen, 1995, Lütkepohl, 2007)

Take for example the price of spruce wood in some forest enterprise, $y_{e,t}$, and the mean price in Germany, $y_{g,t}$, stored in $\boldsymbol{y}_t = (y_{e,t}, y_{g,t})'$. We assume, for example, that the enterprise has better than average wood quality and can therefore charge 30% more than the mean. Ergo the equilibrium would be given by $y_{e,t} = \beta_1 y_{g,t}$ with $\beta_1 = 1.3$. At some point in time, random fluctuation may cause $y_{e,t}$ to be lower than $y_{g,t}$, but in the long term the price will tend to equilibrium. (cp. Johansen (1995), Lütkepohl (2007))

Therefore the changes in $y_{e,t}$ could be modeled as an adjustment towards the equation given above. It furthermore seems reasonable to assume, that said adjustment will not happen instantaneously but gradually, with speed $\alpha_e = 0.5$ for example. To put it in mathematical terms

$$\Delta y_{e,t} = \alpha_e (y_{e,t-1} - \beta_1 y_{g,t-1}) + u_{e,t}$$

with $u_{e,t}$ as a white noise error term for the unexplained variation. Moreover, if the enterprise is based in Germany the changes will also influence the mean price, ergo there might be a similar relation for $y_{g,t}$

$$\Delta y_{g,t} = \alpha_g (y_{e,t-1} - \beta_1 y_{g,t-1}) + u_{g,t}$$

with a possibly much smaller adjustment speed α_g. It furthermore seems reasonable, that the current changes are influenced by the changes in the last period. The terms given above are therefore expanded to

$$\begin{aligned} \Delta y_{e,t} &= \alpha_e (y_{e,t-1} - \beta_1 y_{g,t-1}) + \gamma_{ee,1} \Delta y_{e,t-1} + \gamma_{eg,1} \Delta y_{g,t-1} + u_{e,t} \\ \Delta y_{g,t} &= \alpha_g (y_{e,t-1} - \beta_1 y_{g,t-1}) + \gamma_{ge,1} \Delta y_{e,t-1} + \gamma_{gg,1} \Delta y_{g,t-1} + u_{g,t} \end{aligned} \tag{2.14}$$

with γ_{ee} for the effect of the last price changes in the enterprise itself and γ_{eg} for the effect of a change in the average price in Germany on the enterprise. Interpretation of $\gamma_{ge,1}$ and $\gamma_{gg,1}$ is analogue. (cp. Johansen (1995), Lütkepohl (2007))

When arranging the data and coefficients in vectors and matrices, such as

$$\boldsymbol{y}_t := \begin{bmatrix} y_{e,t} \\ y_{g,t} \end{bmatrix}, \quad \boldsymbol{\alpha} := \begin{bmatrix} \alpha_e \\ \alpha_g \end{bmatrix}, \quad \boldsymbol{\beta} := \begin{bmatrix} 1 \\ -\beta_1 \end{bmatrix}, \quad \boldsymbol{\Gamma}_1 := \begin{bmatrix} \gamma_{ee,1} & \gamma_{eg,1} \\ \gamma_{ge,1} & \gamma_{gg,1} \end{bmatrix}, \quad \boldsymbol{u}_t := \begin{bmatrix} u_{e,t} \\ u_{g,t} \end{bmatrix},$$

the equations 2.14 can be expressed as

$$\Delta \boldsymbol{y}_t = \boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{y}_{t-1} + \boldsymbol{\Gamma}_1 \Delta \boldsymbol{y}_{t-1} + \boldsymbol{u}_t \tag{2.15}$$

in a much more compact form (cp. Lütkepohl (2007)). Note, that when $\Delta \boldsymbol{y}_t$ is replaced by $\boldsymbol{y}_t - \boldsymbol{y}_{t-1}$ and $\Delta \boldsymbol{y}_{t-1}$ by $\boldsymbol{y}_{t-1} - \boldsymbol{y}_{t-2}$, respectively, this term can be rearranged to

$$\boldsymbol{y}_t = \underbrace{(\boldsymbol{I}_K + \boldsymbol{\Gamma}_1 + \boldsymbol{\alpha}\boldsymbol{\beta}')}_{\boldsymbol{A}_1} \boldsymbol{y}_{t-1} - \underbrace{\boldsymbol{\Gamma}_1}_{\boldsymbol{A}_2} \boldsymbol{y}_{t-2} + \boldsymbol{u}_t \tag{2.16}$$

hence a cointegrated series can be generated by a possibly unstable VAR process, in this case with lag order $p = 2$.[7] Note, that when moving from VECM to VAR the lag order is

[7] The transformation from VAR to VECM is dealt with in more detail in Lütkepohl (2007).

increased by one due to the fact, that VECMs work with differences, VARs with the absolute values. (Lütkepohl, 2007)

The example used in (2.15) can be expanded and generalized to a K-dimensional VECM with lag order $p-1$,[8] where $\boldsymbol{y}_t = (y_{1,t}, \ldots, y_{K,t})$, $\boldsymbol{u}_t = (u_{1,t}, \ldots, u_{K,t})$ and there are $p-1$ $\boldsymbol{\Gamma}_i$ matrices of dimension $(K \times K)$ containing the coefficients for each lag.

Furthermore the number of linearly independent cointegration relations (see 2.2.1) may not be limited to one, but can be $r \leq K$. In this case $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ would be $(K \times r)$ matrices of rank r ($rk(\boldsymbol{\beta}) = rk(\boldsymbol{\alpha}) = r$). $\boldsymbol{\alpha}$ will be called the *loading matrix*, $\boldsymbol{\beta}$ will be referred to as the *cointegration matrix*. Both matrices can be combined to the $(K \times K)$ matrix $\boldsymbol{\Pi} := \boldsymbol{\alpha}\boldsymbol{\beta}'$, also of rank r. (Johansen, 1995, Lütkepohl, 2007)

This yields the general VECM representation of a *VAR(p)* process:

$$\begin{aligned} \Delta\boldsymbol{y}_t &= \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{y}_{t-1} + \boldsymbol{\Gamma}_1\Delta\boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1}\Delta\boldsymbol{y}_{t-p+1} + \boldsymbol{u}_t \\ &= \boldsymbol{\Pi}\boldsymbol{y}_{t-1} + \boldsymbol{\Gamma}_1\Delta\boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1}\Delta\boldsymbol{y}_{t-p+1} + \boldsymbol{u}_t \end{aligned} \tag{2.17}$$

Note, that if the cointegration rank $r = rk(\boldsymbol{\Pi})$ is zero, $\boldsymbol{\Pi} = 0$ and therefore $\Delta\boldsymbol{y}_t$ has a stable *VAR(p-1)* representation. For $r = K$ all elements of $\boldsymbol{y}_t$ have to be stable, hence the data generating process behind $\boldsymbol{y}_t$ is a stable *VAR(p)*. (Johansen, 1995, Johansen and Jusélius, 1990, Lütkepohl, 2007) Furthermore, it is possible to convert any VECM to its VAR representation, details can be found in Lütkepohl (2007), section 6.3.

2.2.3 Non-uniqueness of the cointegration matrix

As already mentioned in 2.2.1, the cointegration relations are not unique. The same problem arises with $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$. Take any arbitrary, nonsingular $(r \times r)$ matrix $\boldsymbol{Q}$, for example. When defining $\boldsymbol{\alpha}^* = \boldsymbol{\alpha}\boldsymbol{Q}'$ and $\boldsymbol{\beta}^* = \boldsymbol{\beta}\boldsymbol{Q}^{-1}$, $\boldsymbol{\Pi} = \boldsymbol{\alpha}^*\boldsymbol{\beta}^{*\prime} = \boldsymbol{\alpha}\boldsymbol{Q}'\boldsymbol{Q}^{-1\prime}\boldsymbol{\beta}' = \boldsymbol{\alpha}\boldsymbol{\beta}'$ still holds. (Lütkepohl, 2007)

However, when $\boldsymbol{Q}$ is chosen appropriately, unique solutions for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ can be obtained. Because $rk(\boldsymbol{\beta}) = r$, it is clear that there must be r linearly independent rows in $\boldsymbol{\beta}$. By arranging the variables in $\boldsymbol{y}_t$ in an appropriate way, one can ensure linear independence of the first r rows and with that a nonsingular upper $(r \times r)$ submatrix of $\boldsymbol{\beta}$. (Lütkepohl, 2007)

If $\boldsymbol{Q}$ is set to the inverse of said submatrix, i.e. $\boldsymbol{Q} = (\boldsymbol{J}\boldsymbol{\beta})^{-1}$ with $\boldsymbol{\beta}$ of dimensions $(K \times r)$, $r = rk(\boldsymbol{\beta})$, $r < K$ and $(r \times K)$ matrix $\boldsymbol{J} = [\boldsymbol{I}_r : 0]$, one obtains

$$\boldsymbol{\beta}^* = \boldsymbol{\beta}\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{I}_r \\ \widetilde{\boldsymbol{\beta}}^*_{(K-r)} \end{bmatrix} \tag{2.18}$$

with $\widetilde{\boldsymbol{\beta}}^*_{(K-r)}$ denoting the last $K-r$ rows of $\boldsymbol{\beta}\boldsymbol{Q}$ with dimension $((K-r) \times r)$. (Lütkepohl, 2007) Consider for example a three-dimensional process $\boldsymbol{y}_t = (y_{1,t}, y_{2,t}, y_{3,t})$ with cointegration rank two. This would yield

[8]The lag order is expressed as $p-1$ to keep the connection to the VAR

$$\boldsymbol{\beta}^* = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \widetilde{\beta}_1^* & \widetilde{\beta}_2^* \end{bmatrix}$$

from (2.18) with the $\widetilde{\beta}_i^*$ being the two elements of $\widetilde{\boldsymbol{\beta}}^*_{(K-r)}$. It is obvious, that hereby a nonzero coefficient for $y_{1,t}$ alongside a zero coefficient for $y_{2,t}$ in the first column, and vice versa in the second one, is implicitly assumed. This means, that in the subsystems $(y_{1,t}, y_{3,t})'$ and $(y_{2,t}, y_{3,t})'$ a cointegration relation must be present and $(y_{1,t}, y_{2,t})'$ must not be cointegrated. To verify this, one can check the subsystems separately. (Lütkepohl, 2007)

At this point one should reconsider the definition of cointegration used for this work (see 2.2.1), which allows the presence of stationary variables in a cointegrated system. If a variable of $\boldsymbol{y}_t$, say $y_{k,t}$, is stationary, then a cointegrating vector $\boldsymbol{\beta}$ with a one at position k and zeros elsewhere, such that $\boldsymbol{\beta}'\boldsymbol{y}_t = y_{k,t}$, can be found. Furthermore, if there are n $I(0)$ variables, then there are n such, linearly independent vectors. Therefore the upper r subvectors of $\boldsymbol{y}_t$ must contain all stationary variables. This furthermore implies, that there can only be a maximum of r stationary variables in $\boldsymbol{y}_t$. (Lütkepohl, 2007)

2.2.4 Deterministic terms

Until this point only stochastic terms were included in the models, such as the one described by (2.17). In the following the integration of deterministic terms will briefly be described in the general form of

$$\boldsymbol{y}_t = \boldsymbol{\mu}_t + \boldsymbol{x}_t$$

with $\boldsymbol{\mu}_t$ denoting some K-dimensional vectors containing the deterministic terms, and $\boldsymbol{x}_t$ as a time series which can be represented as a VECM in the form of (2.17). (Lütkepohl, 2007)

First, there could be a nonzero intercept, ergo $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0$ and therefore $\boldsymbol{x}_t = \boldsymbol{y}_t - \boldsymbol{\mu}_0$ as well as $\Delta\boldsymbol{x}_t = \Delta\boldsymbol{y}_t$ follows. When inserted in (2.17), one gets

$$\begin{aligned} \Delta\boldsymbol{y}_t &= \boldsymbol{\alpha\beta}'(\boldsymbol{y}_{t-1} - \boldsymbol{\mu}_0) + \boldsymbol{\Gamma}_1\Delta\boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1}\Delta\boldsymbol{y}_{t-p+1} + \boldsymbol{u}_t \\ &= \boldsymbol{\alpha\beta}^{o\prime}\begin{bmatrix} \boldsymbol{y}_{t-1} \\ 1 \end{bmatrix} + \boldsymbol{\Gamma}_1\Delta\boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1}\Delta\boldsymbol{y}_{t-p+1} + \boldsymbol{u}_t \\ &= \boldsymbol{\Pi}^o\boldsymbol{y}^o_{t-1} + \boldsymbol{\Gamma}_1\Delta\boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1}\Delta\boldsymbol{y}_{t-p+1} + \boldsymbol{u}_t \end{aligned} \tag{2.19}$$

$$\begin{array}{llll} \text{with} & \boldsymbol{\beta}^{o\prime} & := [\boldsymbol{\beta}' : \boldsymbol{\tau}'] & (r\times(K+1)) \\ & \boldsymbol{\tau}' & := -\boldsymbol{\beta}'\boldsymbol{\mu}_0 & (r\times 1) \\ & \boldsymbol{y}^{o\prime}_{t-1} & := [\boldsymbol{y}'_{t-1} : 1] & ((K+1)\times 1) \\ & \boldsymbol{\Pi}^o & := [\boldsymbol{\Pi} : \boldsymbol{\nu}_0] & (K\times(K+1)) \\ & \boldsymbol{\nu}_0 & := -\boldsymbol{\Pi}\boldsymbol{\mu}_0 = \boldsymbol{\alpha\tau}' & (K\times 1) \end{array}$$

showing that a simple constant can be included into the cointegration relations. (Lütkepohl, 2007) Another option is to write the model with an overall intercept term as

$$\Delta \boldsymbol{y}_t = \boldsymbol{\nu}_0 + \boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{y}_{t-1} + \boldsymbol{\Gamma}_1 \Delta \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1} \Delta \boldsymbol{y}_{t-p+1} + \boldsymbol{u}_t \tag{2.20}$$

although $\boldsymbol{\nu}_0$ can no longer be arbitrary but has to fulfill the condition $\boldsymbol{\nu}_0 = -\boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{\mu}_0$ to ensure that no trend is generated in the data. (Lütkepohl, 2007)

A linear trend in the mean, ergo $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t$ can also be absorbed into the matrices by using $\boldsymbol{x}_t = \boldsymbol{y}_t - \boldsymbol{\mu}_0 - \boldsymbol{\mu}_1 t$ and $\Delta \boldsymbol{x}_t = \Delta \boldsymbol{y}_t - \boldsymbol{\mu}_1$. At this point only the result will be shown, more details can be found in Lütkepohl (2007), p. 257:

$$\Delta \boldsymbol{y}_t = \boldsymbol{\nu}^* + \boldsymbol{\alpha}[\boldsymbol{\beta}' \;:\; \boldsymbol{\eta}'] \begin{bmatrix} \boldsymbol{y}_{t-1} \\ t-1 \end{bmatrix} + \boldsymbol{\Gamma}_1 \Delta \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1} \Delta \boldsymbol{y}_{t-p+1} + \boldsymbol{u}_t \tag{2.21}$$

$$\text{with} \quad \begin{aligned} \boldsymbol{\nu}^* &:= -\boldsymbol{\Pi}\boldsymbol{\mu}_0 + (\boldsymbol{I}_K - \boldsymbol{\Gamma}_1 - \cdots - \boldsymbol{\Gamma}_{p-1})\boldsymbol{\mu}_1 \quad && (K \times 1) \\ \boldsymbol{\eta}' &:= -\boldsymbol{\beta}'\boldsymbol{\mu}_1 && (r \times 1) \end{aligned}$$

Here the intercept term $\boldsymbol{\nu}^*$ depends on $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$ as well as the other parameters, but can in theory take any value from $\mathbb{R}^K$. (Lütkepohl, 2007)

If the linear trend appears in the variables and not the cointegration relations, the intercept term $\boldsymbol{\nu}$ will generate the trends. This situation is represented by the model

$$\Delta \boldsymbol{y}_t = \boldsymbol{\nu} + \boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{y}_{t-1} + \boldsymbol{\Gamma}_1 \Delta \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1} \Delta \boldsymbol{y}_{t-p+1} + \boldsymbol{u}_t$$

with unrestricted intercept term $\boldsymbol{\nu}$. This requires a cointegrating rank smaller than K, because otherwise the process has a stable *VAR(p)* representation without a linear trend. (Lütkepohl, 2007)

Finally, one might include d fixed, non-stochastic variables such as dummies for seasonal influences or extraordinary occasions, or any arbitrary data in a vector $\boldsymbol{D}_t$ of dimension $(d \times 1)$, specific for each period t. (cp. Johansen (1995), Pfaff (2008a)) Such as

$$\Delta \boldsymbol{y}_t = \boldsymbol{\nu} + \boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{y}_{t-1} + \boldsymbol{\Gamma}_1 \Delta \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1} \Delta \boldsymbol{y}_{t-p+1} + \boldsymbol{\phi}\boldsymbol{D}_t + \boldsymbol{u}_t$$

where $\boldsymbol{\phi}$ is a $(K \times d)$ matrix of coefficients. Note, that Johansen (1995) uses $\boldsymbol{D}_t$, instead of $\boldsymbol{\nu}$, to incorporate constants and trends in the data, resulting in

$$\Delta \boldsymbol{y}_t = \boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{y}_{t-1} + \boldsymbol{\Gamma}_1 \Delta \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1} \Delta \boldsymbol{y}_{t-p+1} + \boldsymbol{\phi}\boldsymbol{D}_t + \boldsymbol{u}_t \tag{2.22}$$

2.2.5 Estimation and the Johansen test

This section is closely related to the `ca.jo()` function from the `urca` package. (Pfaff, 2008a) Said function will conduct the Johansen test for cointegration (Johansen, 1988, 1991, Johansen and Jusélius, 1990) and estimate initial $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ matrices. For the test the quantiles proposed by Osterwald-Lenum (1992) will be used. (Pfaff, 2008a)

The Johansen test is designed to determine the cointegration rank r of a K-dimensional system of variables with $t = 1, \ldots, T$ observations each, stored in $\boldsymbol{y}_t$. Those variables can be $I(0)$ or $I(1)$, and have a VECM representation with lag order $p-1$. Note, that the cointegration rank corresponds to the rank of $\boldsymbol{\Pi}$, and therefore to that of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ as well. (Johansen, 1995, Lütkepohl, 2007, Pfaff, 2008a)

There are two statistics available which can be used to test a set of Hypothesis:

(i) The *trace statistic* has the null Hypothesis that there is a maximum of r cointegrating vectors.

(ii) The *maximal eigenvalue statistic*, which tests the Hypothesis of r versus $r+1$ cointegrating vectors.

The Hypothesis will be tested sequentially from $r = 0$ to $r = K - 1$. The first null Hypothesis which can not be rejected at a given significance level determines the cointegration rank. (Johansen, 1995, Pfaff, 2008a)

Cheung and Lai (1993) examined, that the trace statistic is more robust to both kurtosis and skewness in the distribution of the error term than the maximal eigenvalue statistic. An exemplary result of such a test is shown in table 2.3, p. 28.

For the calculation of the test statistic a number of steps is required. Initially, a set of auxiliary ordinary least squares (OLS) regressions will be estimated: $\Delta\boldsymbol{y}_t$ is regressed on $\Delta\boldsymbol{y}_{t-1}, \ldots, \Delta\boldsymbol{y}_{t-p+1}$, and $\boldsymbol{D}_t$ in the sense of (2.22), i.e. all used non-stochastic variables, e.g. dummies, a constant, a trend, et cetera. The corresponding K series of residuals are stored in $\boldsymbol{R}_{0,t}$. Furthermore, the same set of regressors will be used for $\boldsymbol{y}_{t-p}$, giving the second K series of residuals, termed $\boldsymbol{R}_{1,t}$. (Johansen, 1995, Pfaff, 2008a)

These residuals are used to calculate the product moment matrices

$$\hat{\boldsymbol{S}}_{i,j} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{R}_{i,t}\boldsymbol{R}'_{j,t}$$

with i, $j = 0, 1$. (Johansen, 1995, Pfaff, 2008a) The $\hat{\boldsymbol{S}}_{i,j}$ will then be used to determine the eigenvalues of

$$|\boldsymbol{\lambda}\hat{\boldsymbol{S}}_{1,1} - \hat{\boldsymbol{S}}_{1,0}\hat{\boldsymbol{S}}_{0,0}^{-1}\hat{\boldsymbol{S}}_{0,1}| = 0 \tag{2.23}$$

which are the basis of the trace statistic

$$-2\ln(Q) = -T\sum_{i=r+1}^{p} \ln(1 - \hat{\lambda}_i) \tag{2.24}$$

where $\hat{\lambda}_{r+1}, \ldots, \hat{\lambda}_p$ resemble the $p - r$ smallest eigenvalues of (2.23).[9] (Johansen, 1995, Pfaff, 2008a)

According to Pfaff (2008a) the maximal eigenvalue statistic is computed by

$$-2\ln(Q; r|r+1) = -T\ln(1 - \hat{\lambda}_{r+1}) \tag{2.25}$$

After determining the cointegration rank r, the cointegration matrix $\boldsymbol{\beta}$ can be estimated. For this purpose the $\hat{\boldsymbol{S}}_{1,1}$ is decomposed into $\hat{\boldsymbol{S}}_{1,1} = \boldsymbol{C}\boldsymbol{C}'$ with $\boldsymbol{C}$ being a $(K \times K)$, non-singular matrix. (Pfaff, 2008a) Therefore (2.23) can be written as

$$|\boldsymbol{\lambda}\boldsymbol{I} - \boldsymbol{C}^{-1}\hat{\boldsymbol{S}}_{1,0}\hat{\boldsymbol{S}}_{0,0}^{-1}\hat{\boldsymbol{S}}_{0,1}\boldsymbol{C}'^{-1}| = 0 \tag{2.26}$$

where the resulting eigenvectors to the corresponding eigenvalues will be stored in $\boldsymbol{e}_i$. Then $\hat{\boldsymbol{v}}_i$ is defined as $\boldsymbol{C}'^{-1}\boldsymbol{e}_i$, giving the columns of $\hat{\boldsymbol{\beta}}$ with

$$\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{v}}_1, \ldots, \hat{\boldsymbol{v}}_r)$$

as an estimate for the cointegration matrix. (Pfaff, 2008a) The loading matrix $\boldsymbol{\alpha}$ is then estimated as

$$\hat{\boldsymbol{\alpha}} = -\hat{\boldsymbol{S}}_{0,1}\hat{\boldsymbol{\beta}}$$

which leads to $\hat{\boldsymbol{\Pi}} = \hat{\boldsymbol{\alpha}}\hat{\boldsymbol{\beta}}'$. Note, that the initial estimates for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are of dimensions $K \times K$ and need to be restricted (see 2.2.6). (Pfaff, 2008a) The variance-covariance matrix of the K-dimensional error term $\boldsymbol{u}_t$, as in (2.17), is given by

$$\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{S}}_{0,0} - \boldsymbol{\alpha}\boldsymbol{\alpha}'$$

For more details see Pfaff (2008a), p. 78 ff, or Johansen (1995).

2.2.6 `R`: Generating and analyzing cointegrated data

In the following the example given in 2.2.2 will be expanded. For this the mean wood price in Germany, `yg`, as well as the wood price in two different enterprises, `ye1` and `ye2` (denoted $\boldsymbol{y}_g$, $\boldsymbol{y}_{e1}$ and $\boldsymbol{y}_{e2}$ in the figures and tables), are exemplary simulated. It is assumed, that there are long term equilibria between the two prices and the mean price, given by `ye1` $= 1.2$ `yg` and `ye2` $= 0.9$ `yg`. For the sake of simplicity there are no autoregressive components included, the innovations are white noise processes with a smaller variance for `yg` than for `ye1` and `ye2`. All data will be collected in `y`, the results are displayed in fig. 2.4.

[9] Note, that there seems to be a misprint in Pfaff (2008a), p. 81. Here the formula is written without the ln() on the right hand side. This differs from Johansen (1991), Johansen (1995) as well as the source code of `ca.jo()`, line 232.

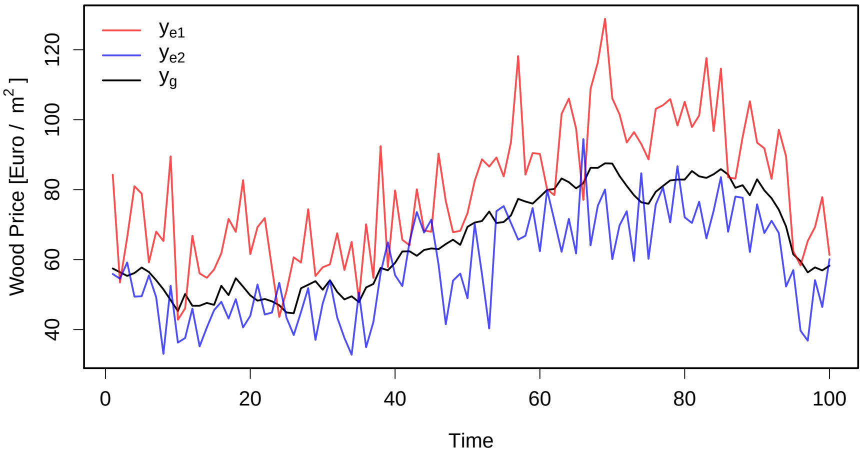


Figure 2.4: Exemplary simulation of cointegrated data: Wood price in Germany (y_g) and in two enterprises (y_{e1}, y_{e2}), with long term equilibria $y_{e1} = 1.2y_g$ and $y_{e2} = 0.9y_g$.

```
set.seed(2048)
yg      <- cumsum(rnorm(100, 0, 3)) + 60
ye1     <- 1.2 * yg + rnorm(100, 0, 10)
ye2     <- 0.9 * yg + rnorm(100, 0, 8)
y       <- data.frame(ye1, ye2, yg)
```

The data will be analyzed with functions implemented in the package `urca` (Pfaff, 2008a). First, the Johansen procedure is conducted (see 2.2.5), implemented in `ca.jo()`, to determine the cointegration rank and retrieve initial estimates for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. By setting `type = "eigen"`, the maximal eigenvalue test statistic, the default, was used. Deterministic terms were excluded with `ecdet = "none"` and `K = 2` specifies the number of lags in the VAR representation. The correct number would have been zero, however, two resembles the minimal value accepted by the function. The option `spec = "transitory"` is used, such that $\boldsymbol{\Pi y}_{t-1}$ is estimated instead of $\boldsymbol{\Pi y}_{t-p}$. This won't have any effect on $\boldsymbol{\Pi}$ itself, only on the $\boldsymbol{\Gamma}_i$. (Pfaff, 2008a)

The last option was chosen primarily for compliance of the calculations with (2.17). The test-results are displayed in table 2.3.

```
library(urca)
vecm <- ca.jo(y,
              type = "eigen",
              ecdet = "none",
              K = 2,
              spec = "transitory")
summary(vecm)
```

	test	10pct	5pct	1pct
r <= 2	1.14	6.50	8.18	11.65
r <= 1	42.31	12.91	14.90	19.19
r = 0	51.18	18.90	21.07	25.75

Table 2.3: Output of the test statistic for the Johansen procedure

The resulting test statistic indicates at all given levels of significance, that there are two cointegration relations, as expected. In the next step the matrices $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\Gamma}$ are estimated via a restricted least squares approach implemented in the function `cajorls()`, which uses the methodology shown in 2.2.3 to restrict $\hat{\boldsymbol{\beta}}$. The argument $r = 2$ is used for the presence of two cointegration relations.

```
(vecm.r2 <- cajorls(vecm, r = 2))
summary(vecm.r2$rlm)
```

The direct output of the function contains two elements, `rlm` and `beta`, shown in tab. 2.4. The first one displays the $\hat{\boldsymbol{\alpha}}$ matrix in the first $r = 2$ rows, followed by a constant. The latter one is needed for an unbiased estimation of the autoregressive part of the model, ergo $\hat{\boldsymbol{\Gamma}}$, and can not be interpreted directly.[10] Because the lag order of the VAR representation was set to two, the lag order in the VECM is $p - 1 = 1$ (cp. 2.2.2). Thus one $\hat{\boldsymbol{\Gamma}}$ matrix can be found in the last $K = 3$ rows.

$\hat{\boldsymbol{\beta}}$ is restricted as described in 2.2.3 by multiplying the first $r = 2$ eigenvectors, stored in `vecm@V`, with the inverse of the upper left hand $(r \times r)$, here (2×2), submatrix of `vecm@V`. In the given example $\hat{\boldsymbol{\beta}}$ provides close estimates to the original equilibria of `ye1` - 1.2 `yg` $= 0$ and `ye2` - 0.9 `yg` $= 0$.

The linear model also estimated standard errors for the data, which were close to the original values, shown in tab. 2.5.

The cointegration relations can be obtained by multiplying `y` with $\hat{\boldsymbol{\beta}}$. Further, `y` was also multiplied with $\boldsymbol{R}_{1,t}$, stored in `vecm@RK`, explanations can be found below. The results are displayed in fig. 2.5 and seem to resemble stationary processes, as expected.

```
beta <- vecm.r2$beta
EC <- as.matrix(y) %*% beta

R1t <- vecm@RK
EC.R1t <- R1t %*% beta
```

[10] Personal conversation via email with Dr. Pfaff, 19th December 2018. Further thanks should be expressed towards Alexander Lange and Lennart Empting, Statistics and Econometrics, Göttingen, whose expertise was of great help.

	ye1.d	**ye2.d**	**yg.d**	
ect1	-1.114	0.038	-0.013	$\hat{\boldsymbol{\alpha}}'$
ect2	0.208	-1.011	0.104	
constant	3.436	1.463	-0.099	
ye1.dl1	0.076	-0.105	0.011	$\hat{\boldsymbol{\Gamma}}'$
ye2.dl1	-0.018	-0.006	-0.045	
yg.dl1	0.620	0.552	0.086	

	ect1	**ect2**	
ye1.l1	1.0	0.0	$\hat{\boldsymbol{\beta}}$
ye2.l1	0.0	1.0	
yg.l1	-1.159	-0.866	

Table 2.4: Summarized output of `cajorls()` with `$rlm` on the left and `$beta` on the right. The matrices are labeled according to (2.17).

	$\boldsymbol{y}_{e1}$	$\boldsymbol{y}_{e2}$	$\boldsymbol{y}_{g}$
$\hat{\boldsymbol{\sigma}}$	10.460	8.642	2.597
$\boldsymbol{\sigma}$	10.000	8.000	3.000

Table 2.5: Estimated Standard Errors for the data simulated in 2.2.6.

The $\boldsymbol{R}_{1,t}$ matrix: In statistical literature one may find additional multiplication of $\boldsymbol{R}_{1,t}$ (see 2.2.5) with $\hat{\boldsymbol{\beta}}$ (e.g. Johansen (1995), Pfaff (2008a)). Because this will also be used in part II and further information is hard to find, these results will be examined closer here.

As mentioned in 2.2.5, $\boldsymbol{R}_{1,t}$ holds the residuals of a regression of $\boldsymbol{y}_{t-p}$ on the lagged differences of $\boldsymbol{y}_t$ and $\boldsymbol{D}_t$ as in (2.22). Hence the short run effects as well as any non-stochastic influence on the data has been accounted for. (Johansen, 1995, Pfaff, 2008a). For the example given above, `ye1`, `ye2` and the corresponding rows of $\boldsymbol{R}_{1,t}$ are displayed in fig. 2.6, the multiplication with $\hat{\boldsymbol{\beta}}$ is depicted in fig. 2.5.

In this example, the multiplication of $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{R}_{1,t}$ was not explicitly necessary, since only the cointegration relations and stochastic terms were present in the data generating process, thus the multiplication of `y` and $\hat{\boldsymbol{\beta}}$ already resulted in a stationary series. However, in a more complex practical example, as given in 4.1, the cointegration relations can be observed more clearly when using $\boldsymbol{R}_{1,t}$ instead of `y`, as shown in fig. 4.2. (cp. Johansen (1995), Pfaff (2008a))

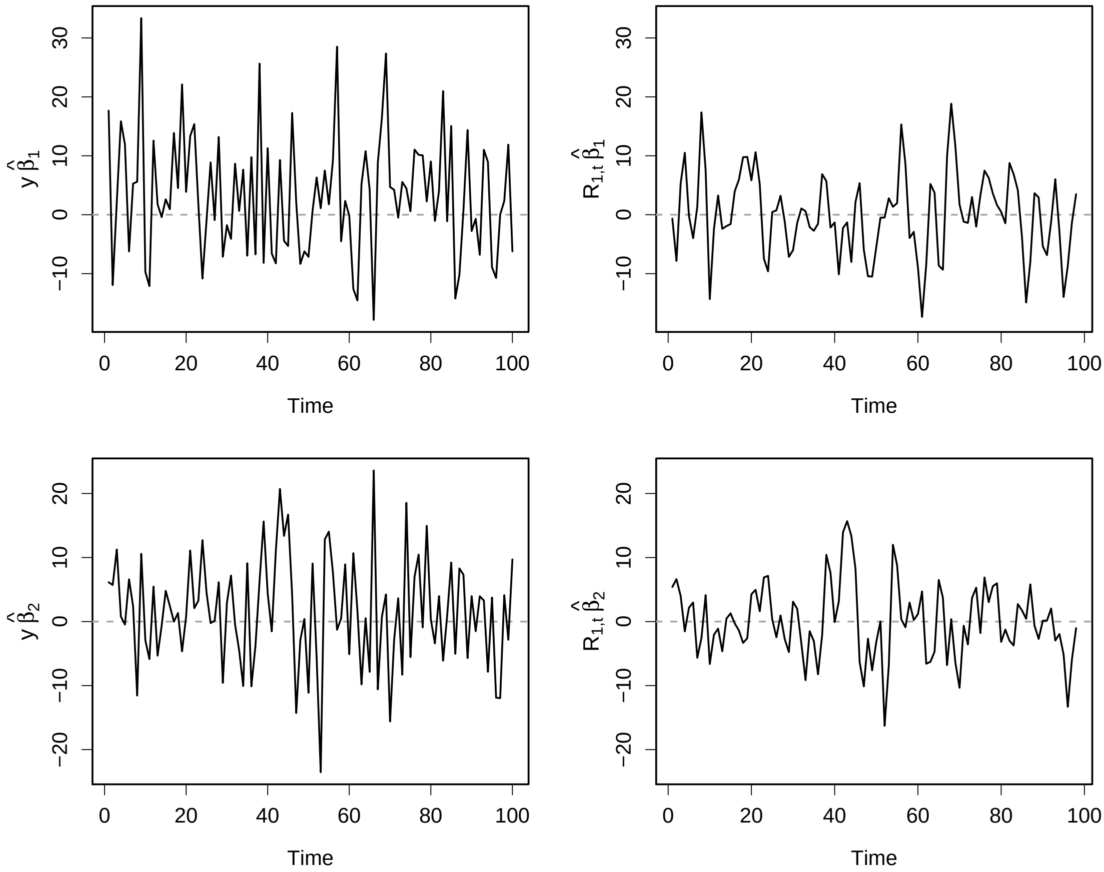


Figure 2.5: Cointegration relations obtained by $\boldsymbol{y}\hat{\boldsymbol{\beta}}$ and $\boldsymbol{R}_{1,t}\hat{\boldsymbol{\beta}}$.

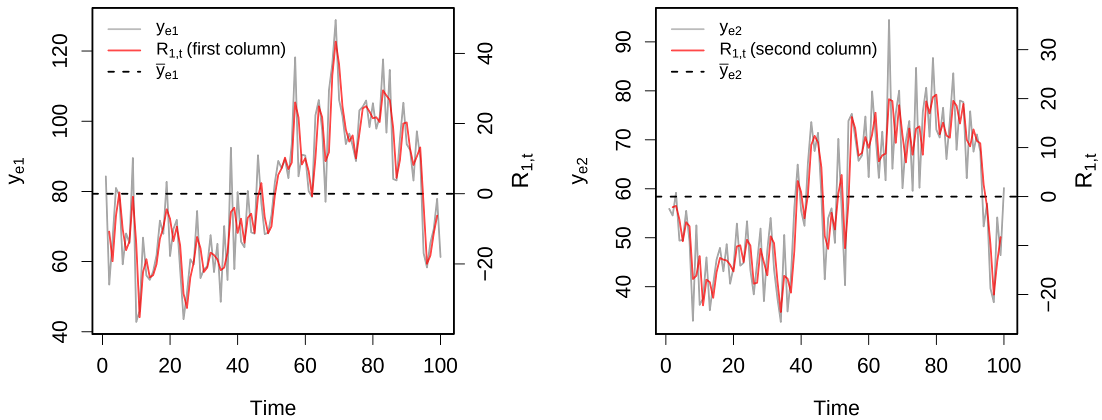


Figure 2.6: Comparison of $\boldsymbol{y}_{e1}$ and $\boldsymbol{y}_{e2}$ (grey) with the respective columns of $\boldsymbol{R}_{1,t}$ (red). Dashed lines indicate the arithmetic mean of $\boldsymbol{y}_{e1}$ and $\boldsymbol{y}_{e2}$, respectively.

2.3 Impulse Response Functions (IRF)

The basic idea behind Impulse Response Functions (IRF) is to trace the effect of a shock in one variable of a K-dimensional *VAR(p)* process at time point $t = 0$ over the next h time steps. To be able to do this, first all values of $\boldsymbol{y}_t := (y_{1,t}, \dots, y_{K,t})'$ prior to time point $t = 0$ are assumed to be zero. Then a shock in $y_{k,0}$ is induced by setting $\boldsymbol{u}_0$ to a vector with a one at position k and zeros elsewhere ($\boldsymbol{u}_0 = (0, \dots, 1, \dots, 0)'$). Moreover it is assumed that no further shocks occur, meaning that $\boldsymbol{u}_1 = \dots = \boldsymbol{u}_h = 0$. Finally the mean $\boldsymbol{\nu}$ (or $\boldsymbol{\mu}$) is usually not considered in the analysis, because one is only interested in the change, not the absolute value of the variable. (Lütkepohl, 2007)

Consider for example a *VAR(1)* process with

$$\boldsymbol{y}_t = \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix}, \quad \boldsymbol{A}_1 = \begin{bmatrix} 0.5 & 0.2 \\ 0.3 & 0.6 \end{bmatrix}, \quad \boldsymbol{u}_t = \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix}$$

in which shock in the first variable will be induced and traced over the next two periods:

$$\begin{aligned}
\boldsymbol{y}_0 &= \begin{bmatrix} u_{1,0} \\ u_{2,0} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\
\boldsymbol{y}_1 &= \boldsymbol{A}_1 \boldsymbol{y}_0 = \begin{bmatrix} 0.5 \\ 0.3 \end{bmatrix} \\
\boldsymbol{y}_2 &= \boldsymbol{A}_1 \boldsymbol{y}_1 = \boldsymbol{A}_1^2 \boldsymbol{y}_0 = \begin{bmatrix} 0.31 \\ 0.33 \end{bmatrix}
\end{aligned}$$

It can be seen, that after i time steps a unit shock in the kth variable results in $\boldsymbol{y}_i$ being the kth column of $\boldsymbol{A}_1^i$. (Lütkepohl, 2007)

Recall that in the MA representation (see 2.1.9) of a *VAR(1)*, $\boldsymbol{A}_1^i$ corresponds to $\boldsymbol{\Phi}_i$. The latter therefore contains the impulse responses of the system. This result also holds for higher order *VAR(p)* processes. Here, the impulse responses coincide with the upper left hand $(K \times K)$ block of $\boldsymbol{\mathcal{A}}^i$, which also equals $\boldsymbol{\Phi}_i$.[11] (Lütkepohl, 2007)

This means, that a the value at row j and column k of $\boldsymbol{\Phi}_i$, denoted $\phi_{j,k,i}$, shows what effect a unit shock in variable k, i periods ago, has on variable j, given that no further shocks occur. If $\phi_{j,k,i}$ is zero for all $i = 1, 2, \dots$, then there is no effect of k on j. To check whether this holds, it is sufficient to compute the first $p(K-1)$ $\boldsymbol{\Phi}_i$ matrices. (Lütkepohl, 2007)

An alternative to a unit shock would be a shock the size of one standard deviation of the variable of interest, to account for the different scales and furthermore get a more "realistic" shock. (Lütkepohl, 2007)

[11] For further details see Lütkepohl (2007), p. 52

2.3.1 Orthogonal Impulses

The methods presented above implicitly assume, that a shock in $y_{l,t}$ can happen isolated and is therefore independent of one in $y_{m,t}$ (with $l \neq m$). Connoting that the off-diagonal elements of the covariance matrix of the error term $\boldsymbol{u}_t$, $\boldsymbol{\Sigma}_u$, are zero. (Lütkepohl, 2007, Pfaff, 2008b)

If this assumption does not hold, which is usually the case, orthogonalized impulse responses which utilize a transformed MA representation (see 2.1.9) can be considered. First, a Choleski decomposition of $\boldsymbol{\Sigma}_u = \boldsymbol{P}\boldsymbol{P}'$ is derived. This matrix $\boldsymbol{P}$ multiplied with its inverse is then inserted into equation (2.10):

$$\boldsymbol{y}_t = \sum_{i=0}^{\infty} \boldsymbol{\Phi}_i \boldsymbol{P}\boldsymbol{P}^{-1}\boldsymbol{u}_{t-i} \tag{2.27}$$

which is then written as

$$\boldsymbol{y}_t = \sum_{i=0}^{\infty} \boldsymbol{\Theta}_i \boldsymbol{w}_{t-i} \tag{2.28}$$

$$\text{with} \quad \begin{aligned} \boldsymbol{\Theta}_i &= \boldsymbol{\Phi}_i \boldsymbol{P} \\ \boldsymbol{w}_{t-i} &= \boldsymbol{P}^{-1}\boldsymbol{u}_{t-i} \end{aligned}$$

where the components of the transformed error term $\boldsymbol{w}_t = (w_{1,t}, \ldots, w_{K,t})'$ are uncorrelated and have a variance of one, ergo $\boldsymbol{\Sigma}_w = I_K$. This further means, that a unit shock in $\boldsymbol{u}_0$ now equals a shock the size of one standard deviation. (Lütkepohl, 2007, Pfaff, 2008b)

The matrices $\boldsymbol{\Theta}_i$ can then be interpreted like the $\boldsymbol{\Phi}_i$: The j,k-th component represents the effect of a unit shock in variable k on j after i periods, ergo the (orthogonal) impulse responses. If one wants to check whether there is no effect of k on j, ergo $\theta_{j,k,i} = 0$ for $i = 1, 2, \ldots$, it is again sufficient to check the values of $\theta_{j,k,i}$ for $i = 1, \ldots, p(K-1)$. (Lütkepohl, 2007)

A problem when dealing with orthogonal impulse responses is, that the obtained IRFs depend on the order in which the variables were used in the model, e.g. which series is in first or second position, ergo $y_{1,t}$ or $y_{2,t}$ (see 2.3.2 and fig. 2.7). Because $\boldsymbol{P}$ is a lower triangular matrix, a shock in variable $y_{s,t}$ with $s < K$ can only have an instantaneous effect on the variables $y_{s+1,t}, \ldots, y_{K,t}$. Therefore, only a shock in the first variable can directly influence all other variables. Hence the variables need to be causally ordered. (Lütkepohl, 2007, Pfaff, 2008b) Moreover, said ordering has to be done "manually" by the researcher as it cannot be determined by statistical methods. Another problem is the non-uniqueness of the $\boldsymbol{P}$-matrix, which leads to non-unique results. (Lütkepohl, 2007)

A solution to these aspects would be a decomposition of $\boldsymbol{\Sigma}_u = \boldsymbol{P}\boldsymbol{P}'$ which is non-triangular and unique. This approach is implemented in the concept of structural VARs, or SVARs (see 6). However, choosing a decomposition method to obtain such a $\boldsymbol{P}$, which moreover fits the intended analysis, is far from trivial and will not be part of this work. (Lütkepohl, 2007, Pfaff, 2008a) More details can be found in Lütkepohl (2007), chapter nine, and Kilian and Lütkepohl (2017).

2.3.2 R: Impulse response function

Consider once more the *VAR(2)* process from 2.1.6, which was analyzed with the `VAR()` command. The results were stored in an object called `var.1`. The IRFs can then be calculated using the `irf()` function, which is also included in the `vars` package, and requires an object of the class `varest`. (Pfaff, 2008b)

First, the ordinary (`ortho = F`) and the orthogonal (`ortho = T`) IRFs were calculated for the existing model and stored in objects. The argument `impulse = ...` selects, in which series the shock will occur. By default a shock for every series will be computed. (Pfaff, 2008b)

Afterwards a second VAR (`var.2`) with switched series was estimated to demonstrate the effect of differently ordered variables. The IRFs were calculated analogously. Before plotting, the results stored in `var.2.irf` and `var.2.irf.o` were switched back, such that the order of the graphs was the same in all plots. The results can be seen in figure 2.7.

```
library(vars)
var.1.irf   <- irf(var.1, ortho = F, impulse = "Series.2")
var.1.irf.o <- irf(var.1, ortho = T, impulse = "Series.2")

var.2 <- VAR(y[, 2:1], type = "none", p = 2)
var.2.irf   <- irf(var.2, ortho = F, impulse = "Series.2")
var.2.irf.o <- irf(var.2, ortho = T, impulse = "Series.2")
```

The results of the ordinary impulse response functions (IRF) were equal for both VARs,[12] and clearly showed the inner workings of the VAR: A shock in series two will not directly affect the series itself in the first lag, but series one. After two lags the shock will be transferred back to series two, as one would expect.

The orthogonal IRFs on the other hand were completely different, as it can be seen in fig. 2.7 b). When in the “correct” order, with series one being in the first position, the shape is similar to the ordinary impulse response, although the scale is reduced to less than $^1/_{10}$th. With the switched, non-causally order, the shape is completely different.

[12] Minor deviations up to $3.9 \cdot 10^{-16}$ were found and assumed to be numerical errors.

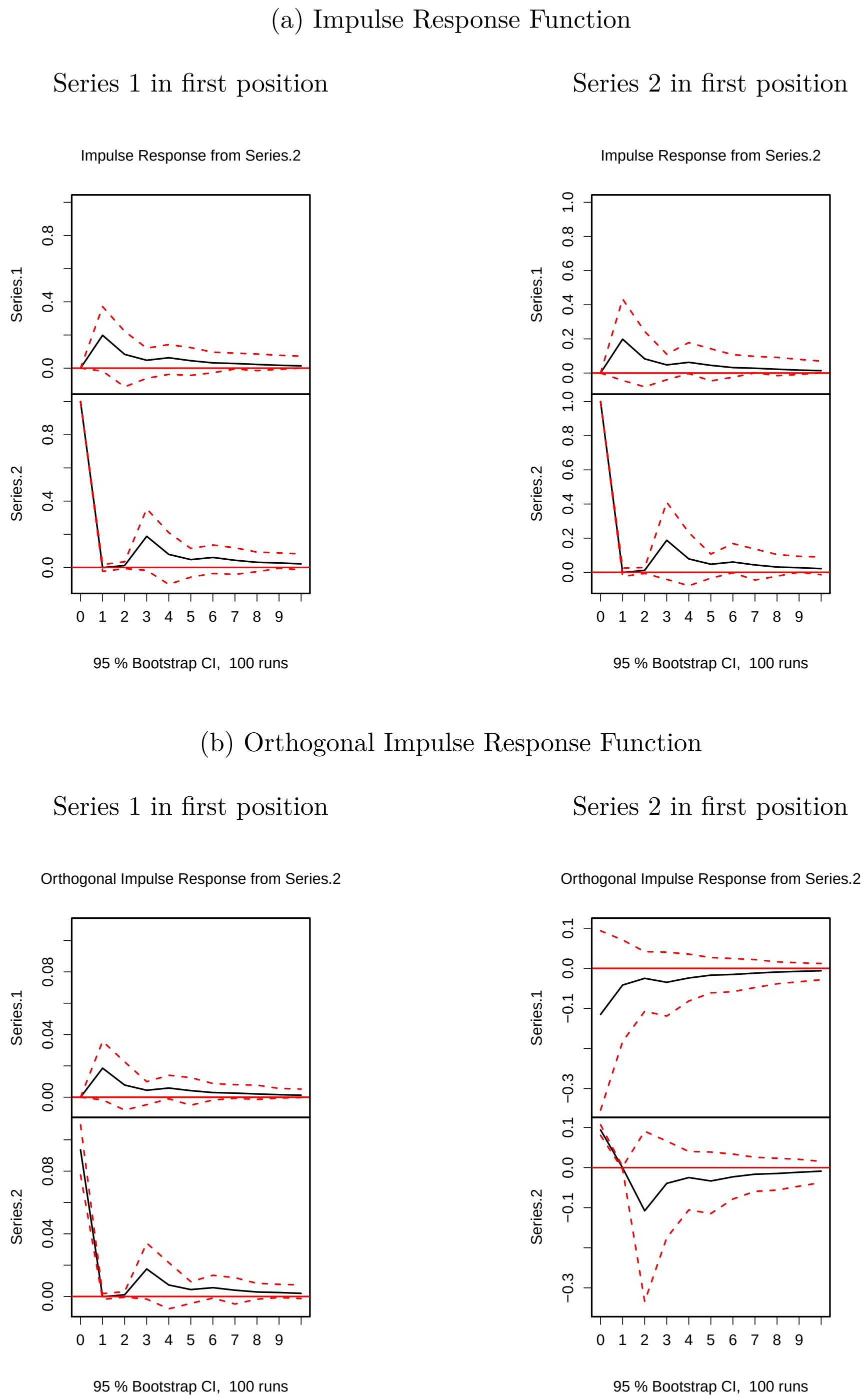


Figure 2.7: Response functions of an impulse in Series 2 of the *VAR(2)* used in 2.1.2 created with the `irf()` function. In the left hand plots the order of the series was as in 2.1.2, in the ones on the right hand the order was switched (via using `VAR(y[, 2:1], ...)`).

2.4 Forecast Error Variance Decomposition (FEVD)

When forecasting a VAR-process (see 2.1), the error of the h-step forecast of $\boldsymbol{y}_t$, denoted $\boldsymbol{y}_{t+h} - \boldsymbol{y}_t(h)$, obviously has a variance. Because of the interdependencies between the variables in a VAR-process, the fluctuation of the forecasted jth component of $\boldsymbol{y}_t$, denoted $y_{j,t}$, may not only be due to its own variance, but also due to the one of other variables. Forecast Error Variance Decomposition (FEVD) can be used to examine, how said variance of $y_{j,t}(h)$ is influenced by all used variables. To put it in other words, it is of interest what proportion of the forecast error variance of $y_{j,t}$ is accounted for by innovations in $y_{k,t}$. (Lütkepohl, 2007)

The basis of the FEVD is again a MA-representation, transformed in such a way that the errors are uncorrelated white noise processes, as given in (2.28) with $\boldsymbol{\Sigma}_w = I_K$. Considering such a process, the error of the optimal h-step forecast is according to Lütkepohl (2007) given by

$$\boldsymbol{y}_{t+h} - \boldsymbol{y}_t(h) = \sum_{i=0}^{h-1} \boldsymbol{\Theta}_i \boldsymbol{w}_{t+h-i} \tag{2.29}$$

with variance

$$\boldsymbol{\Sigma}_y(h) = \sum_{i=0}^{h-1} \boldsymbol{\Theta}_i \boldsymbol{\Theta}_i'$$

Due to the uncorrelated error terms $w_{k,t}$ $(k = 1, \ldots, K)$ with a unit variance, the Mean Squared Error (MSE) of $y_{j,t}(h)$, ergo of a h-step forecast for $\boldsymbol{y}_t$'s j-th component, can be written as

$$E(y_{j,t+h} - y_{j,t}(h))^2 = MSE(y_{j,t}(h)) = \sum_{k=1}^{K} (\theta^2_{j,k,0} + \cdots + \theta^2_{j,k,h-1}) \tag{2.30}$$

with $\theta_{m,n,i}$ denoting the m,n-th element of $\boldsymbol{\Theta}_i$. (Lütkepohl, 2007) By division of each of the k terms in parentheses in (2.30) through the whole $MSE(y_{j,t}(h))$, ergo

$$\frac{\theta^2_{jk,0} + \cdots + \theta^2_{jk,h-1}}{MSE(y_{j,t}(h))} \tag{2.31}$$

the contributions of variable k to the variance of the forecast error of an h-step forecast of variable j is obtained. However, since the FEVD has the same basis as the orthogonal IRF, similar criticism as in 2.3.1, concerning e.g. the ordering of the variables, applies. (Lütkepohl, 2007)

2.4.1 `R`: Forecast Error Variance Decomposition

Again, the *VAR(2)* process analyzed in 2.1.6 is used as an example. The FEVD is implemented in the R function `fevd()` from the `vars` package by Pfaff (2008b). Its only argument is `n.ahead`, which specifies h, the number of steps to be forecasted, and defaults to ten. The results are saved in an object called `var.1.fevd`, and are displayed in fig. 2.8.

```
var.1.fevd <- fevd(var.1,
                   n.ahead = 10)
```

The results are as expected: Since series one depends only (recursively) on itself, series two has no influence on the forecasts. Series two on the other hand almost fully depends on series one after two lags. Hence the variance of the forecast is determined mostly by itself in the first two lags, after that it is almost fully determined by series one.

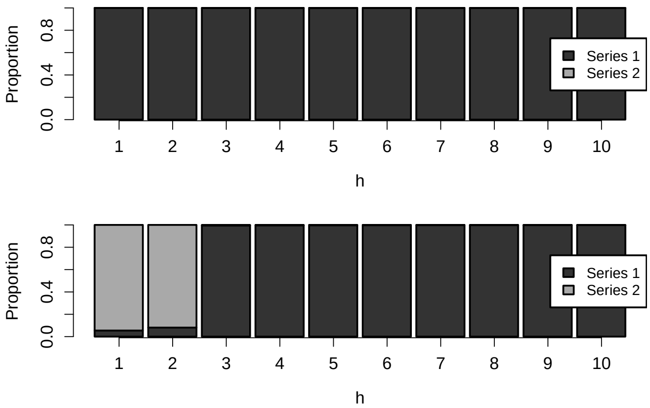


Figure 2.8: FEVD estimated in 2.4.1. Series one is displayed in the upper graphic, series two in the lower one.

Part II
Application

Until now, this work was only dealing with theoretical concepts and simulated data where the true results were known in advance. The following sections will apply the models and methods introduced in part I to "real" data, where the true data generating process is unknown. However, the effort was made to only use variables in which at least some characteristics of the estimated models can be seen or anticipated beforehand, while still being at least somewhat interesting and relevant.

Section 3 will give a short introduction to the sources of information used for this work, i.e. the business comparison Westphalia-Lippe (BVGL, see Dög et al. (2017)) and the German Federal Statistical Office. The former annually collects key figures of up to 50 private forest enterprises since 1969, making it the oldest continuous time series of its kind in Germany.

Said data will then be used to conduct three exemplary case studies, each trying to answer a specific question with an adequate modeling approach from part I. Section 4.1 investigates for possible connections between forestry and the general economic development with a VECM. In 4.2 a VAR is used to analyze, whether logging activities are shifted due to a change in prices. Finally, the last case study in section 4.3 will examine possible interdependencies in price changes of different tree species with impulse response functions.

In each case study, the used data will be presented first. Afterwards the modeling approach is described, followed by an extensive check for model adequacy and the results, which will be discussed briefly. At the end of each case study further literature applying the same or similar techniques will be given for the interested reader.

All R-scripts and graphics are also available on the enclosed CD and at GitHub: https://github.com/jan-schick/Masters_Thesis. However, the data of the BVGL (see 3.1) is highly sensitive. Therefore the respective values have been distorted by adding white noise to the data before uploading or burning, respectively. Further, the GDP used in 4.1 could not be made available publicly and therefore has to be retrieved manually by the reader. Thus the results of 4.1 and 4.2 can not be reproduced with the supplied material.

3 Used Data

3.1 Forest Businesses in Westphalia-Lippe

The main dataset used in this work is from the “Betriebsvergleich Westfalen-Lippe”[13] (BVGL), in which economical key figures of up to 50 private forest enterprises in the region of Westphalia-Lippe, in western Germany, are collected annually. (Dög et al., 2017)

The numbers are being gathered since 1969, making it the oldest time series of its kind in Germany. The businesses manage an average area of 1,619 ha with varying main tree species: In the North-Western part of the region, the “Münsterland”, mainly pines are grown. The dominant species in the “Sauerland”, located in the South, is the Spruce, whereas in the eastern Part, “Ostwestfalen”, deciduous timber, mainly beech, is of great importance. (Dög et al., 2017)

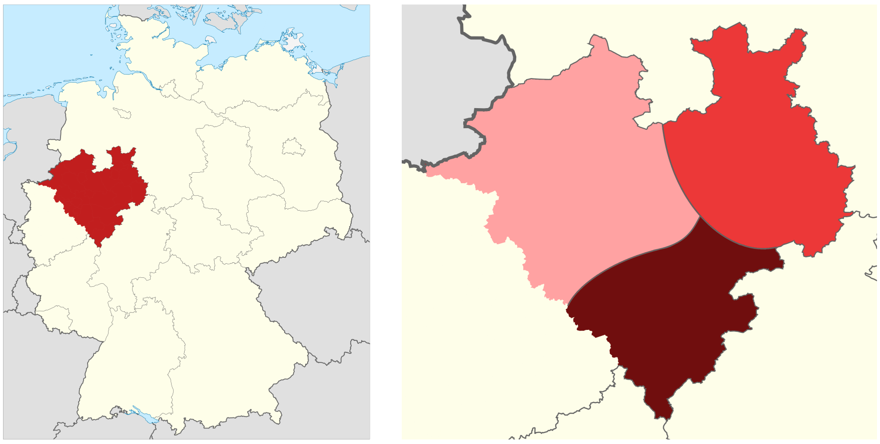


Figure 3.1: Location of Westphalia-Lippe in Germany (left) and the roughly sketched regions (right): The “Münsterland” in the North-West, the “Sauerland” in the South and “Ostwestfalen” in the East.
(Source: wikimedia.org, modified)

The main goal of the BVGL is, to capture the situation and the development of the enterprises over the years as comprehensive as possible. Therefore all kinds of data is being gathered, such as the revenue, how much area is occupied by certain tree species or the numbers of hours worked per hectare, resulting in a total of over 1000 variables.

The data is mainly stored in one database, where all variables are collected in several tables. At the beginning of this work, only data until 2015 was present, later years still have to be transferred. However, there exist Excel-spreadsheets which are created annually and store main key figures of the enterprises, here the data was available for one more year.

To account for the differing natural resources of the enterprises, each one is seen to be part of a “consulting ring”[14], categorized by the economically most relevant tree species for the

[13] German, translation: Business comparison Westphalia-Lippe
[14] Translated from the German word “Beratungsring”

respective enterprise: (i) *Spruce*: Mainly in the Sauerland. (ii) *Deciduous wood*: Mostly beech, primarily present in Ostwestfalen. (iii) *Pine*: Predominantly located in the Münsterland. The distribution of the tree species in the rings can be seen in fig. 3.3. (Dög et al., 2017)

The data is reported annually on a voluntary basis by a changing number of enterprises. When exemplary analyzing the table "HOLZ" (German: wood) of the corresponding database, a total of 83 businesses was found, 45 of which submitted their data from 2006 to 2015 on average (see fig. 3.2).

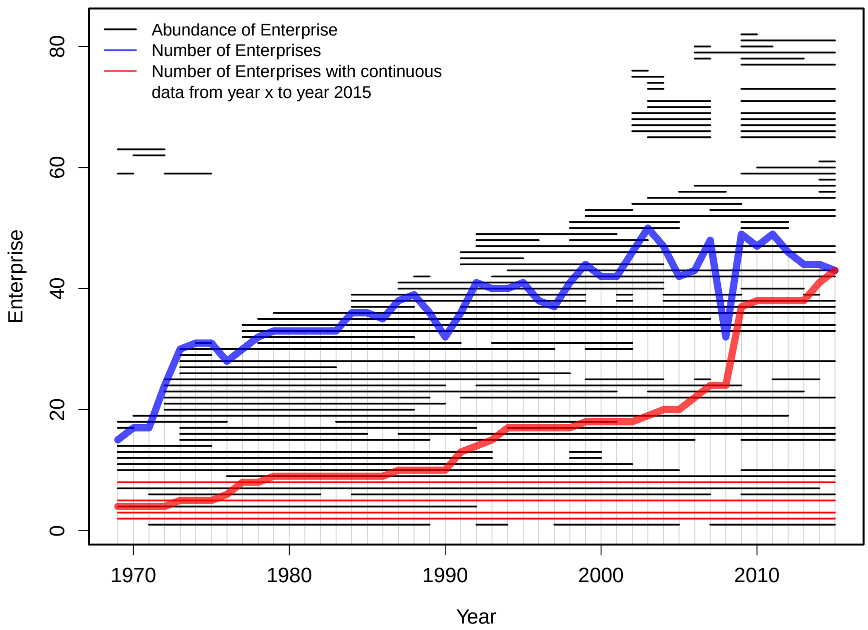


Figure 3.2: Abundance of Enterprises in the table "HOLZ" (German: wood) of the "Betriebsvergleich Westfalen-Lippe" (Dög et al., 2017), red thin lines indicate businesses which were continuously present from the beginning. The number for the Enterprises on the y-axis was assigned randomly.

Thus an issue when dealing with the data, or any data gathered over such a timespan, is the possible discontinuity of the reported values. Although the overall number of enterprises in the table mentioned above is quite high, 15 in 1969 and $\geq$30 since 1977, only four enterprises submitted data continuously from the beginning of the BVGL, eight since 1977. Changes in the data may hence be induced by a changing number of enterprises. Furthermore, the submitted data may not always have been complete. Moreover, some variables were dropped, changed, or newly included over time or the methods of elicitation have been changed.

Therefore the decision was made to only use values averaged or somehow aggregated over all enterprises or the consulting rings, respectively, for the analysis and to disregard the individual values. Yet still, one has to be careful when interpreting the data and should not jump to conclusions.

Nevertheless, in the yearly meetings with involved enterprises the aggregated data was confirmed in representing the overall development of forestry in the related region. Hence the aggregated data can in general be rated as valid.

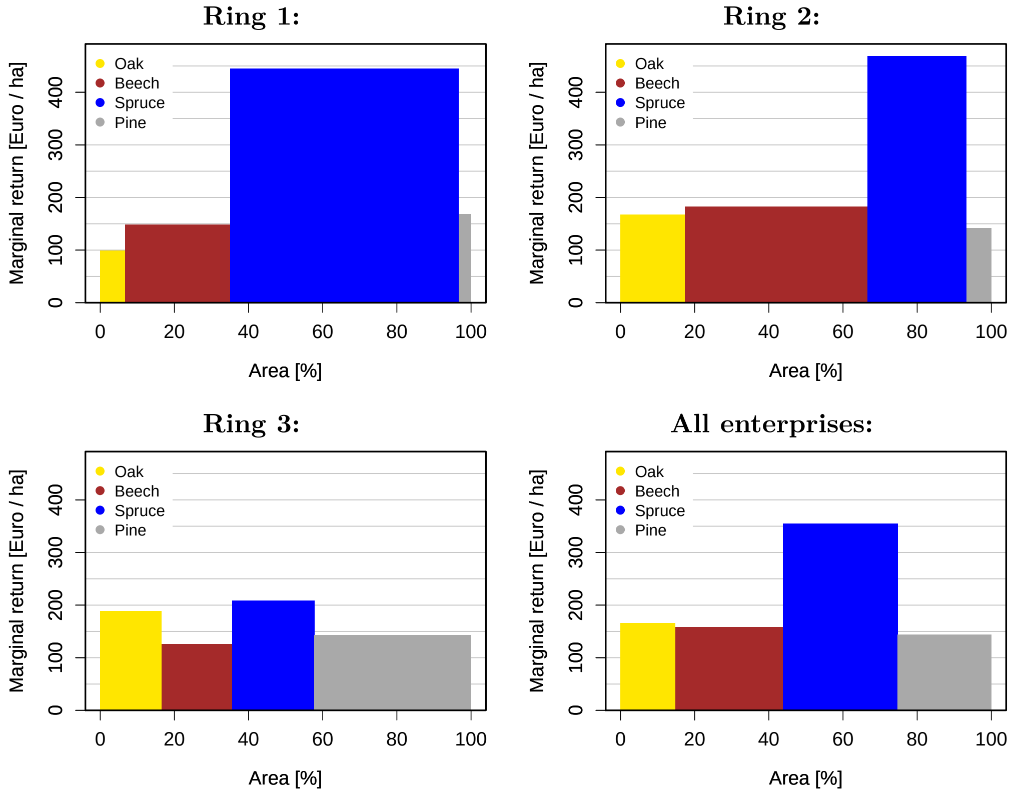


Figure 3.3: Marginal return and percentage of the area covered by the different tree species for the three rings as well as all enterprises. Other tree species than the ones shown were neglected.

3.2 DESTATIS

To be able to compare the development of forest businesses not only within the enterprises itself, but with other branches as well, further data will be necessary. The German Federal Statistical Office, Statistisches Bundesamt (DESTATIS), publicly provides key figures of the Federal Republic of Germany and its states in the GENESIS database, which will be used for this purpose.

4 Case Studies

In this section three case studies will be provided, which shall illustrate the possibilities of the models and methods presented in part I. The main objective, however, will not be to generate results as new and precise as possible, but to use comparably simple datasets to generate understandable and comprehensible results. This is done to follow the overall aim of the work, i.e. it should enable the reader to develop own questions, or to see which already existing ones could be answered with multivariate time series analysis. If one is interested in more advanced work, there will be references to scientific articles, which used similar methods, at the end of each case study in a section called *Further applications*.

4.1 Business Revenue and GDP (VECM)

The first case study which shall be considered here, is the analysis of the mean revenue per hectare achieved by the forest businesses and the gross domestic product of North Rhine-Westphalia (NRW). The main question will be:

$\mathcal{Q}_1$: Is the development of the forest businesses' revenue connected to the overall economic development in NRW?

This is meant to be an introductory example with a minimal dataset of two time series, where at least some characteristics can directly be anticipated. The main purpose is, to give a stepwise introduction into the practical application of some of the methods introduced in part I, while possibly still providing some new and interesting information.

4.1.1 Used data

The revenue of the forest enterprises was taken from the BVGL, more precisely variable K032 from the Excel-spreadsheet (see 3.1), which contains the enterprises revenue per hectare, averaged over all businesses. The averaged value per hectare was chosen primarily because the overall number of enterprises fluctuates over time (see fig. 3.2), therefore a cumulated revenue over all businesses or an averaged total revenue per enterprise would lead to misinterpretations. Furthermore the data obviously does not contain all enterprises in NRW. Therefore the expansion of one business in terms of surface area, and the hereby induced higher total revenue, does not necessarily result in a decreasing area and revenue of other businesses as it would be the case for the GDP. The averaged monetary return of the surveyed enterprises was interpreted as an auxiliary variable for the total value of return of all forest enterprises in NRW.

The GDP itself was chosen as an indicator for the general development of all branches in NRW. The data was obtained from statista.com.[15]

The revenues were available from 1969 to 2016, the GDP from 1970 to 2017. Therefore both time series were shortened by one value, leaving 47 values for the years 1970 to 2016 stored in an R time series object called `gr.data`. It should be noted, that this sample is comparably

[15] "Bruttoinlandsprodukt von Nordrhein-Westfalen von 1970 bis 2017 (in Millionen Euro)", 25.10.18

small. However, since the BVGL is the oldest and therefore longest time series of its kind in Germany, larger samples are not available.

Two big winter storm events, Vivian / Wiebke (1990) and Kyrill (2007), were clearly visible in the data (see fig. 4.1). Storms are the largest threat to European forests, from 1950 to 2000 they caused 53 % of the occurring damage (Schelhaas et al., 2003). Especially Kyrill had an extraordinary impact on NRW, half of the damages in Germany occurred in said federal state (Kropp et al., 2009).

The trees which fell due to the storms or were heavily damaged by them need to be sold quickly because of the threat of a bark beetle outbreak (e.g. Schmidt et al. (2010)). This crucially increases the wood harvest and therefore the revenue of the enterprises. In the following years further effects should be considered, for example because the enterprises may not have been able to remove all damaged trees in one year, or because costs are generated by the reforestation of the affected stands, removal of fallen trees, et cetera (cp. Baur et al. (2003)).

To account for this, a (47×6) matrix called `gr.storm` was created with one dummy variable for each storm event as well as the two following years, respectively.

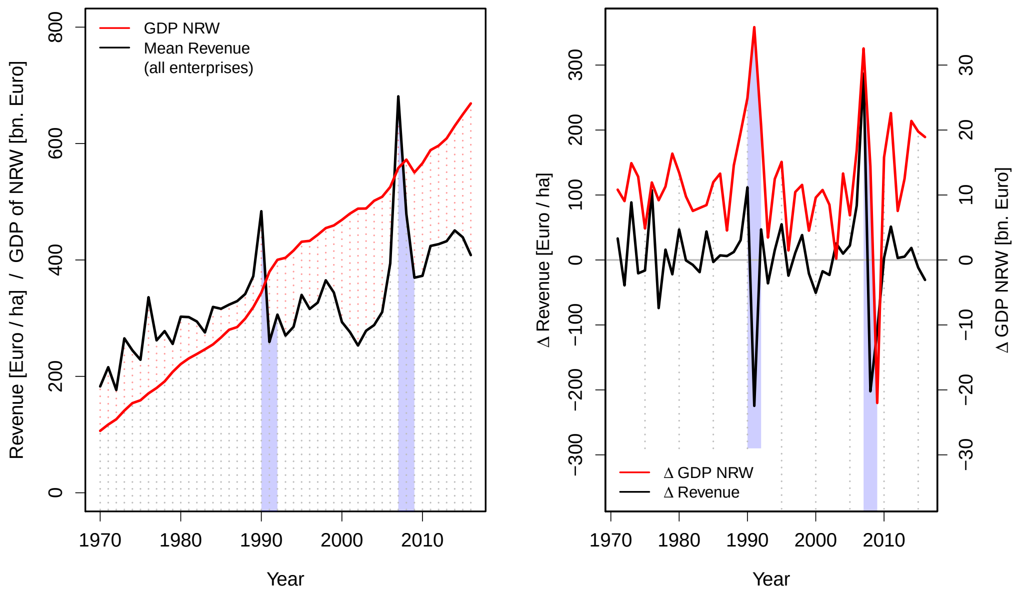


Figure 4.1: Development of the Revenue of all enterprises in the "Betriebsvergleich Westfalen-Lippe" (Dög et al., 2017) and the GDP in North Rhine-Westphalia (NRW) (left) and first order differences (right). Blue background indicates years, for which an effect of the storm events Vivian/Wiebke (1990 to 1992) and Kyrill (2007 to 2009) was considered.

4.1.2 Modeling approach

In the following, the main objective is to select a model to use. To that end the data will be analyzed concerning the order of integration, the lag order and cointegration relations such that a modeling approach adequate for the data can be chosen. Said model will then be estimated.

Order of integration: When looking at the data (see fig. 4.1) it is obvious, that at least one non-stationary time series, the GDP, is present. This is also reflected in the results of the ADF test, which uses stationarity as the alternative Hypothesis, and the KPSS test, where stationarity is the Null Hypothesis, (see 1.3) with $p = 0.29$ and $p < 0.01$[16], respectively.

Moreover the KPSS test can also use the Null Hypothesis of trend stationary data via setting `null = "trend"`. This was also deemed unlikely with $p = 0.029$.

The first differences, however, appear to be stationary (see fig. 4.1). Both tests did not reject the Hypothesis of Stationarity at $\alpha = 0.1$. The GDP therefore seems to be *I(1)*.

The revenue of the enterprises was tested in the same way, the results can be found in table 4.1.[17] The test results indicate, that this series is *I(1)* as well. The hypothesis of trend stationarity, however, could not be rejected.

Variable:	GDP			Revenue		
Test:	**ADF**	**KPSS** *(level)*	**KPSS** *(trend)*	**ADF**	**KPSS** *(level)*	**KPSS** *(trend)*
original data	0.2867	<0.0100	0.0292	0.1045	<0.0100	>0.1000
first differences	0.0925	>0.1000		<0.0100	>0.1000	

Table 4.1: p-Values of different tests for stationarity of the data used in 4.1.

Lag order selection: After having established, that the two series are *I(1)*, the next step would be to test for cointegration. However, the Johansen procedure (see 2.2.5) requires the lag length, which will therefore be determined first (cp. Lütkepohl (2007)).

For the lag order selection the function `VARselect()` from the `vars` package was used on the data. The argument `type = "both"` was used to include the possibility for a constant and a trend in the data, the storm dummy variables were included as exogenous variables (see 2.1.1). The maximal lag length (`lag.max`) was set to five, as it seems reasonable that the influence of the current revenue on the one in more than five years is considerably small. Furthermore, with a sample size of 47, even five $\boldsymbol{\Gamma}_i$ matrices with two parameters relevant for each series will considerably reduce the degrees of freedom. An even larger number of lags would result in a saturated model. The resulting output of the function can be seen in tab. 4.2. The lag order was set to one, as indicated by all criteria.

[16] The function `kpss.test()` from the `tseries` package interpolates the p-values from Table 1 of Kwiatkowski et al. (1992) which only covers p-values from 0.1 to 0.01. If a value is outside of this interval, 0.1 or 0.01, respectively, is displayed as a p-value alongside a warning message. (Trapletti and Hornik, 2018)

[17] According to Trapletti and Hornik (2018), `adf.test()` interpolates the p-values from Table 4.2, p. 103 of Banerjee et al. (1993). For values outside the table, the closest existing one is displayed with a warning message.

```
library(vars)
VARselect(gr.data,
          type = "both",
          lag.max = 5,
          exogen = gr.storm)
```

Lag order	AIC	HQ	SC	FPE
1	**10.29919**	**10.60248**	**11.12665**	**30267.22**
2	10.42499	10.78894	11.41794	34811.47
3	10.48391	10.90853	11.64236	37679.54
4	10.41868	10.90396	11.74262	36299.74
5	10.48056	11.02650	11.96999	40089.06

Table 4.2: Partial output of `VARselect()`: Different criterions for lag order selection for the data used in 4.1, bold values indicate the respective minimum.

Cointegration test: Because the series are both *I(1)*, it is also possible that they are cointegrated (see 2.2.1) as well. In this case a VECM would be necessary to capture the long run equilibrium between the variables (see 2.2.2).

To check for the presence of cointegration relations the Johansen test (see 2.2.5) was computed with the `ca.jo()` function from the package `urca` (Pfaff, 2008a). Said function was run three times, once for every option for the deterministic term (`ecdet = "none"`, `"const"` and `"trend"`). The results were converted to a VAR with `vec2var()`, also from the package `urca`, such that the BIC could be extracted. The model with no deterministic term showed the smallest BIC, hence the corresponding option was chosen.

`K = 2` specifies the lag order in the corresponding VAR (see 2.2.2). The lag order of two is not in unison with the lag order determined previously, yet it is the minimal value accepted by the function (Pfaff, 2008a). The dummy variables are set with `dumvar`, `spec = "transitory"` is chosen such that the earlier introduced equation (2.17) is used for the estimation (cp. 2.2.6). Further calculations will therefore be congruent to part I.

If not specified otherwise with argument `type`, the maximal eigenvalue statistic (see 2.2.5) is used for testing. Because of the small sample, the trace statistic was also computed to get clearer results. Said results are shown in tab. 4.3, both indicate the presence of one cointegration relation at a significance level of 5 %, ergo $r = 1$.

```
library(urca)
gr.vecm <- ca.jo(gr.data,
                 ecdet = "none",
                 K = 2,
                 dumvar = gr.storm,
                 spec = "transitory")
summary(gr.vecm)
```

Maximal eigenvalue statistic:

	test	**10pct**	**5pct**	**1pct**
r <= 1	0.24	6.50	8.18	11.65
r = 0	18.07	12.91	14.90	19.19

Trace statistic:

	test	**10pct**	**5pct**	**1pct**
r <= 1	0.24	6.50	8.18	11.65
r = 0	18.31	15.66	17.95	23.52

Table 4.3: Test results obtained from `ca.jo()`: Results of the Johansen test, maximal eigenvalue and trace statistic, for the data used in 4.1.

VECM estimation: After having established, that the suitable model for the data would be a VECM, the parameters were estimated via a restricted least squares approach (cp. 2.2.3 and 2.2.6). This was done with the function `cajorls()`, `r = 1` specifies the number of cointegration relations. The results are displayed in tab. 4.4, interpretation is analogue to tab. 2.4 from section 2.2.6.

```
(gr.vecm.r1 <- cajorls(gr.vecm,
                       r = 1))
summary(gr.vecm.r1$rlm)
```

4.1.3 Model adequacy

Before the model is interpreted and discussed, its basic properties should be checked. Recall, that one main goal of the analysis with a VECM is, to find a linear combination of the variables in the form of $\boldsymbol{y}\boldsymbol{\beta}$[18] whose order of integration is smaller than the one of the original data.

[18] Recall, that R stores a time series in a column. To keep the connection to R, the notation here is different from the more theoretical section 2.2 and from some textbooks, e.g. Lütkepohl (2007).

Coefficients for the Revenue:

	Estimate	Std. Error	t value	Pr(>|t|)	
ect1	-0.251	0.133	-1.891	0.067	.
constant	65.718	33.588	1.957	0.058	.
Vivian.Wiebke	123.818	37.267	3.322	0.002	**
Vivian.Wiebke.l1	-166.03	43.018	-3.86	<0.001	***
Vivian.Wiebke.l2	-49.27	66.586	-0.74	0.464	
Kyrill	311.163	37.765	8.239	<0.001	***
Kyrill.l1	-59.659	62.88	-0.949	0.349	
Kyrill.l2	-142.658	57.261	-2.491	0.018	*
revenue.dl1	-0.273	0.175	-1.561	0.127	
gdp.nrw.dl1	0.484	0.941	0.514	0.61	

Coefficients for the GDP:

	Estimate	Std. Error	t value	Pr(>|t|)	
ect1	0.045	0.019	2.399	0.022	*
constant	0.369	4.801	0.077	0.939	
Vivian.Wiebke	11.436	5.326	2.147	0.039	*
Vivian.Wiebke.l1	20.304	6.148	3.302	0.002	**
Vivian.Wiebke.l2	4.156	9.517	0.437	0.665	
Kyrill	21.544	5.398	3.991	<0.001	***
Kyrill.l1	-2.301	8.987	-0.256	0.799	
Kyrill.l2	-46.351	8.184	-5.664	<0.001	***
revenue.dl1	-0.034	0.025	-1.378	0.177	
gdp.nrw.dl1	-0.008	0.135	-0.058	0.954	

$\hat{\beta}$ matrix:

	ect1
revenue.l1	1
gdp.nrw.l1	-0.178

Table 4.4: Summarized results of `cajorls()`: Estimates for the parameters of the autoregressive part and the $\hat{\beta}$ matrix of the VECM estimated in 4.1.2.

Therefore the data was multiplied with the above estimated $\hat{\boldsymbol{\beta}}$, the result is displayed in fig. 4.2. Obviously, the storm events are still clearly visible. Therefore $\hat{\boldsymbol{\beta}}$ was also multiplied with one set of residuals form the auxiliary regression of the Johansen procedure, $\boldsymbol{R}_{1,t}$ (see 2.2.5), stored in `gr.vecm@RK`, which accounts for short-run dynamics, the used dummy variables and the mean. Hence deviations from the long-run equilibrium ($\hat{\boldsymbol{\beta}}$) are visible more clearly. Because of the storm dummy variables, the model shows a perfect fit for the years 1990 to 1992 and 2007 to 2009, resulting in corresponding values of zero in $\boldsymbol{R}_{1,t}$ as well.

```
gr.beta <- gr.vecm.r1$beta
gr.EC <- gr.data %*% gr.beta
gr.EC.R1t <- gr.vecm@RK %*% gr.beta
```

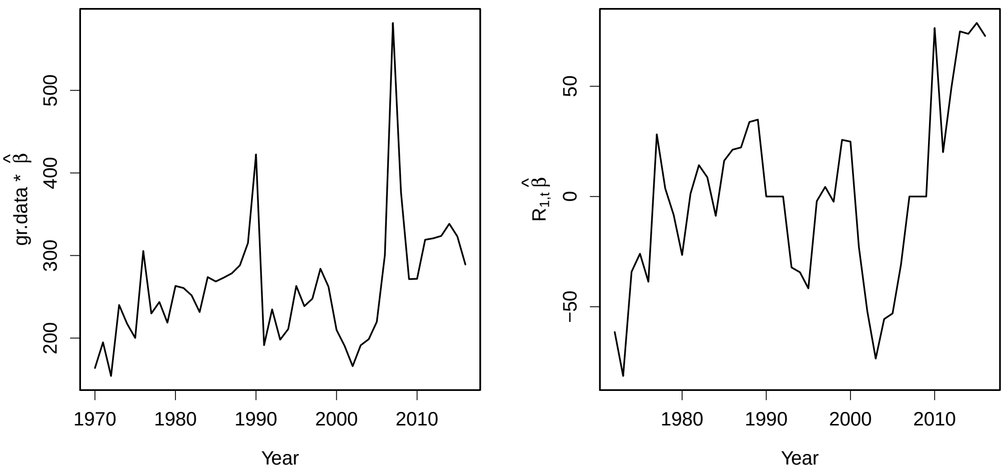


Figure 4.2: The estimated cointegration relations for the data used in 4.1.

Both the ADF and the KPSS test show no strong evidence for a stationary $\boldsymbol{R}_{1,t}\hat{\boldsymbol{\beta}}$. However, the graphical display indicates that this might be due to the low values at the start and the high ones at the end, respectively (see fig. 4.2). Therefore the tests were run again on reduced datasets, where one to five values have sequentially been deselected from both the start and the end, respectively. Especially the results of the KPSS-test, which actually uses the null hypothesis of stationarity (see 1.3), reaffirmed the assumption pretty early. The ADF test took five values deselected from start and end, respectively, to reject the null hypothesis of a unit root. The corresponding p-values are reported in tab. 4.5. Therefore the test results indicating non-stationarity for the full $\boldsymbol{R}_{1,t}\hat{\boldsymbol{\beta}}$ were assumed to be caused by the marginal outliers, which again were assumed to be due to unfortunate, but random fluctuations. Hence the assumption of a reduced order of integration in the cointegration relation was seen to be fulfilled.

Another additional option for checking model adequacy is to plot the fitted values over the real ones. For educational purposes, the fitted values will be calculated manually according to (2.17): First, the estimated coefficients for the dummy variables and $\boldsymbol{\Gamma}$ (see tab. 4.4) were extracted and stored in `gr.vecm.coef`, the $\boldsymbol{\alpha}$ matrix was stored in `gr.alpha`.

The initial VECM estimation obtained by `ca.jo()` contains the matrices `ZK` and `Z1`. The first one consists of the first lags of the original data (ergo $\boldsymbol{y}_{t-1}$), the second one contains a column of ones for the constant, the dummy variables, and the first lagged differences of the data (ergo $\Delta\boldsymbol{y}_{t-1}$). By multiplying the different objects as in (2.17), the fitted values of the VECM were obtained and stored in `y.pred`. The results are displayed in fig. 4.3.

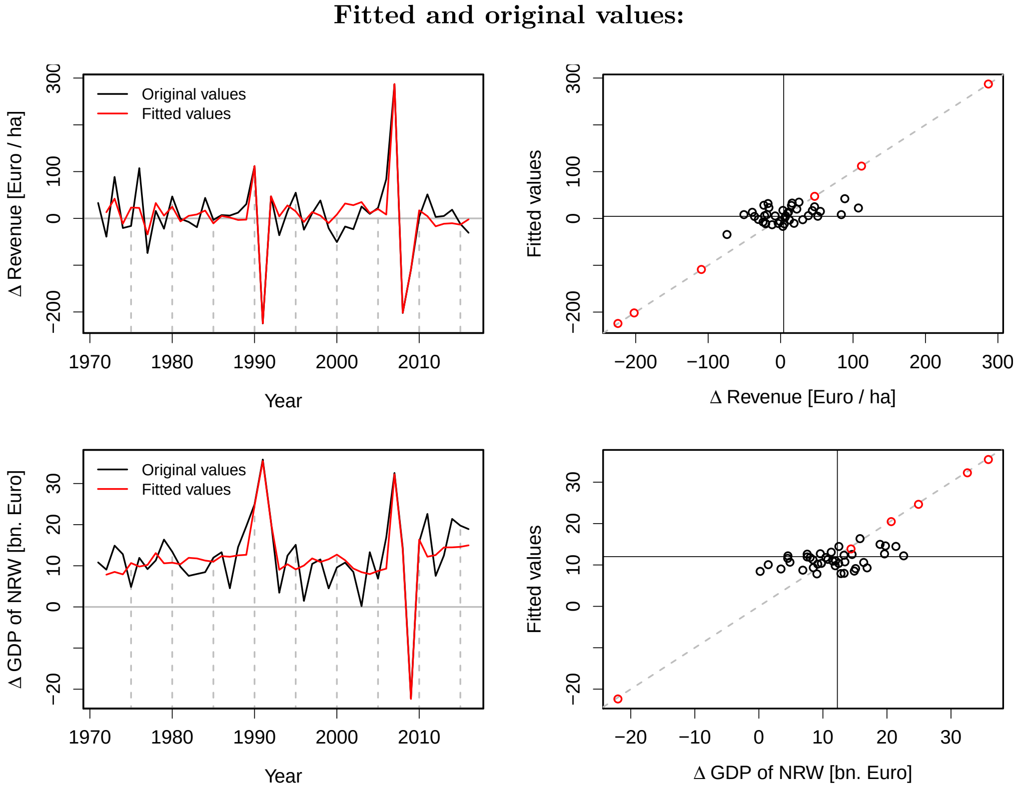


Figure 4.3: Fitted and real values for first differences of the data used in 4.1 plotted over time (left) and against each other (right). On the right the black lines indicate the means of the fitted and the original values, respectively, and red markers indicate the years covered by the storm dummy variables (1990 to 1992 and 2007 to 2009). The dashed gray line has intercept zero and slope one.

	ADF	**KPSS**
full dataset	0.473	0.043
reduction:		
-1	0.481	0.096
-2	0.711	>0.100
-3	0.595	>0.100
-4	0.405	>0.100
-5	0.057	>0.100

Table 4.5: p-values for the ADF and KPSS test on $\boldsymbol{R}_{1,t}\hat{\boldsymbol{\beta}}$, for the original dataset (first row) and a series of reduced sets, where one to five values have been deselected from both the start and the end, respectively.

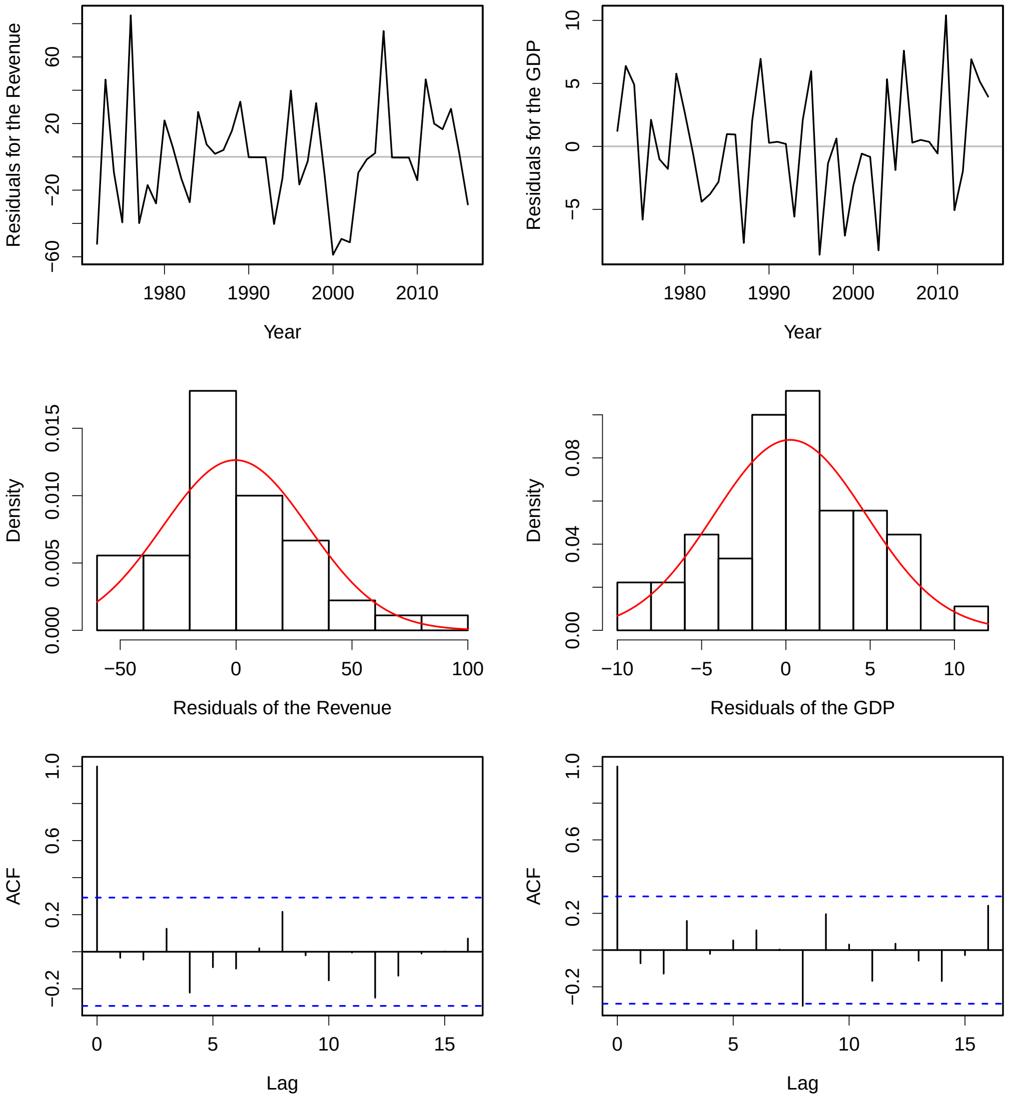


Figure 4.4: Residuals for the VECM estimated in 4.1.2, for the Revenue (left column) and the GDP (right column). The top row displays the residual values over time. In the middle a graphical check for normality in the form of a histogram plus a fitted Gaussian distribution (red) can be found, followed by an autocorrelation function at the bottom generated with the R function `acf()`.

Via subtraction from the differenced original data, conveniently stored in `gr.vecm@Z0`, the residuals of the fit were obtained, shown in fig. 4.4.

One of the assumptions of the VECM was, that the error term, which is represented by the residuals, is white noise (see 2.2.2). The graphical display in fig. 4.4 indicates no violation of this assumption, the values seem to be approximately normally distributed with mean zero and without autocorrelation, as expected. This impression was reaffirmed by several tests: Besides the ADF and KPSS tests for stationarity, the univariate Jarque-Bera test for normality (R function `jarque.bera.test()`) (Jarque and Bera, 1987), and the Ljung-Box test (`Box.test()` (R Core Team, 2018)) were computed, the latter one has the null hypothesis of no autocorrelation in the data (Ljung and Box, 1978). Results are shown in tab. 4.6.

```
gr.vecm.coef <- gr.vecm.r1$rlm$coefficients[-1, ]
gr.alpha <- as.matrix(gr.vecm.r1$rlm$coefficients[1, ])

y.pred <- gr.vecm@ZK %*% (gr.alpha %*% t(gr.beta)) + gr.vecm@Z1 %*% gr.vecm.coef
y.pred <- ts(y.pred,
             start = 2016 - length(y.pred[, 1]) + 1,
             end = 2016)
res <- gr.vecm@Z0 - y.pred
```

	Revenue	**GDP**
ADF	0.045	0.281
KPSS	>0.100	>0.100
Ljung-Box	0.821	0.617
Jarque-Bera	0.416	0.823

Table 4.6: Test results (p-values) for stationarity (ADF, KPSS), autocorrelation (Box-Ljung) and normality (Jarque-Bera) for the residuals of the VECM estimated in 4.1.2.

To obtain a measure for the quality of the fit, Pearson's correlation coefficient was calculated for the fitted and observed values (cp. fig. 4.3). Because of the essentially perfect fit combined with a high leverage, the values of the years covered by the storm dummy variables were excluded from the calculations. Results are shown in tab. 4.7.

			Correlation	**Mean**
Δ	**Revenue**		0.441	4.280
Δ	**Revenue**	(fit)		4.566
Δ	**GDP**		0.401	12.259
Δ	**GDP**	(fit)		12.009
	Revenue		0.760	335.460
	Revenue	(fit)		335.460
	GDP		0.991	390.834
	GDP	(fit)		390.834

Table 4.7: Key figures of the original and fitted values for the VECM estimation. Years 1990 to 1992 and 2007 to 2009 were excluded for the correlation.

Furthermore, the fitted values were compared with the original data. For easier computations the function `vec2var()` was used to convert the initial VECM estimation to the corresponding VAR representation (cp. (2.16)), and stored in `gr.vec2var`. Thus the fitted values will be given at the scale of the original data instead of the first order differences, shown in fig. 4.5. Means and correlation were also calculated, the latter one again without the values for the years 1990 to 1992 and 2007 to 2009. Results can be found in tab. 4.7.

```
(gr.vec2var <- vec2var(gr.vecm, r = 1))
```

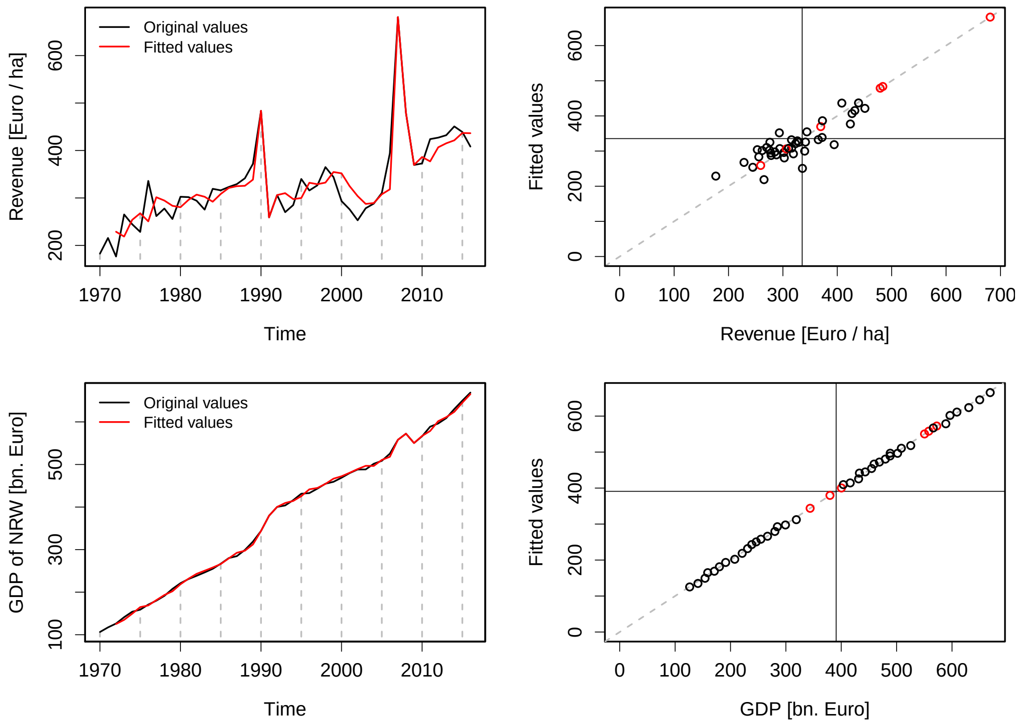


Figure 4.5: Fitted and real values for the data used in 4.1 plotted over time (left) and against each other (right). On the right the black lines indicate the means of the fitted and the original values, respectively, and red markers indicate the years covered by the storm dummy variables (1990 to 1992 and 2007 to 2009). The dashed gray line has intercept zero and slope one.

4.1.4 Results and discussion

The aim of this case study was to provide a simple example how a multivariate time series, such as the BVGL, can be analyzed, and to figure out which model is appropriate for said analysis. The results should be used to investigate, whether the development of the forest businesses' revenue is connected to the general development in NRW (cp. $\mathcal{Q}_1$). As it was said at the beginning of the section, some characteristics could already be anticipated from fig. 4.1, such as a slower increase of the revenue compared to the GDP, or the similar development in the storm years (1990 to 1992 and 2007 to 2009).

In section 4.1.2 it was determined, that each time series is *I(1)* and the series are cointegrated as well. Therefore a VECM was deemed to be the correct modeling approach. A VAR for the differenced data would also have been possible, yet it wouldn't have been able to capture the long run equilibrium.

The lag order and the deterministic term were selected according to information criteria, the final estimates for the parameters can be seen in tab. 4.4. The influence of the two winter storm events which were clearly visible in the data, Vivian/Wiebke (1990) and Kyrill (2007), were included as dummy variables for the year of the event as well as the two following years to capture the aftermath of the storms.

Whether this approach has been adequate, was evaluated in section 4.1.3. It can be seen in tab. 4.5, that $\hat{\boldsymbol{\beta}}$ did not directly yield a stationary time series, according to the tests, as it would be expected. However, when neglecting the first and last two values the KPSS test indicated stationarity, hence the test-results of the full dataset was seen as an effect of the marginal outliers. Therefore the cointegration relation $\hat{\boldsymbol{\beta}}$ was assumed to yield a series integrated of a smaller order than the original data, as the definition implies (see 2.2.1).

When looking at the predicted values for the differences in fig. 4.3, it is clear that the model could not capture the yearly changes of the data. This came at no surprise since all parameters of the $\hat{\boldsymbol{\Gamma}}$ matrix, ergo the autoregressive part, were relatively close to zero and / or clearly insignificant at $\alpha = 5\,\%$ (see tab. 4.4). Thus indicating no distinct connection between the changes in revenue and / or the GDP.

However, fig. 4.3 also shows, that the model could capture the general development of the changes of the data. The means were quite identical to the original ones (see tab. 4.7). The overall trend of the original dataset is also represented quite well by the model (see fig. 4.5). Furthermore, the residuals were white noise, as assumed in 2.2.2 (see fig. 4.4 and tab. 4.6). Thus indicating a certain validity.

When looking at the autoregressive parameters and storm dummies in tab. 4.4, it was striking that only the storms seem to have a significant impact on the data. For the revenue of the enterprises this is as expected: In the year of the storm the value increases due to the large amount of wood that can be harvested. In the following years the revenue will decrease because of the money which has to be spent to compensate the occurred damage (cp. Baur et al. (2003)).

For the GDP this connection is not so obvious at first glance. For the development in the year 1990 influences from the unification of the Federal Republic of Germany and the German Democratic Republic in the same year have to be considered. Said effects may have led to the strictly positive parameters for Vivian / Wiebke, and the following years, for the GDP (see tab. 4.4). Because of the impact of this event, a comparison of the parameters for the revenue and the GDP would not be trivial and be beyond the scope of this work. Hence a connection to the forest enterprises via the storm event Vivian / Wiebke could neither be affirmed nor denied here.

For the year 2007, on the contrary, no obvious explanation or interfering effects for the similar development of the revenue and the GDP (see 4.1) have been found in the scope of this work. However, the decrease in the year 2009 can be explained by the Great Recession which followed the financial crisis of 2007 to 2008 and had worldwide effects. (e.g. Jenkins et al. (2013)) Again, although a connection of some sort between forestry and other branches in NRW via the storm Kyrill may be present, it could not be deduced here because of possible interfering effects.

In the following the $\hat{\boldsymbol{\beta}}$ matrix from tab. 4.4, which stores the estimated equilibrium between the two variables, shall be examined more closely: The value of -0.178 indicates, that an increase of one in the GDP lead to an increase of 0.178 in the revenue. However, the latter one was measured in Euro per hectare, the GDP on the other hand in billions of Euros with no explicit connection to an area. To still make those valuables comparable, it should be noted that the revenue of the last used year, 2016, was 408 € / ha, the GDP on the other hand was at 668 bn. € (cp. fig. 4.1). Therefore an increase of 0.178 € / ha in the revenue would equal 0.044 %, whereas one bn. € would equal 0.149 % for the GDP. This shows, that the enterprises revenue develops at a much slower rate than the GDP, as anticipated.

Finally, the adjustment speed to the equilibrium described above, the $\hat{\boldsymbol{\alpha}}$ matrix, should be considered. Its values can be seen in tab. 4.4 under `ect1`. For the GDP, the adjustment speed of 0.045 is considerably slow, indicating that it was not influenced crucially by the revenue of the forest enterprises, as one would expect. For the revenue itself, the value of -0.251 and hence the adjustment to changes in the GDP is substantially larger.[19] Thus indicating a stronger dependence of the revenue on the GDP than vice versa, again, as expected.

The question behind this case study was to investigate a possible connection between the revenue of the enterprises and the GDP. When looking at the autoregresssive part ($\hat{\boldsymbol{\Gamma}}$) of the model, no significant parameters, ergo no significant connection between the short-term development of the variables was found. However, the Johansen test (see tab. 4.3) and the estimated cointegration relation ($\hat{\boldsymbol{\beta}}$, tab. 4.4) indicated, that the two variables follow a common linear trend upwards, although the slope is substantially smaller for the revenue. However, the adjustment speed $\hat{\boldsymbol{\alpha}}$ (`ect1`, tab. 4.4) was much smaller for the GDP than for the revenue. Therefore a connection for the long run development was found in the form of a linear relationship, where the revenue orients itself towards the GDP. Hence $\mathcal{Q}_1$ can be partially confirmed.

4.1.5 Further applications

The Vector Error Correction Model is quite common in applied work, since an equilibrium exists or is assumed between many economic variables. This is also the model used in the only

[19] The negative value is due to the fact, that the GDP is larger than the revenue. Say e.g., that the revenue was 15.8 and the GDP 100, a multiplication with $\hat{\beta}$ would yield -2. The respective signs in $\hat{\boldsymbol{\alpha}}$ correct this, such that the GDP will de- and the revenue increase towards the equilibrium.

scientific paper found in the scope of this work, which deals with time series analysis of forestry and was written by a German researcher: Kolo and Tzanova (2017) used said model to forecast export and import prices and quantities of raw timber in Germany.

In other countries, or topics, these models are used far more frequently. For example, Alavalapati et al. (1997) examined how the Canadian wood pulp price is influenced by the U.S. pulp price, the use of pulp in Canada and the currency exchange rate between the U.S. and Canada. Further research in North America with VECMs was conducted by Shahi and Kant (2009), to investigate cointegration relationships for different categories of coniferous lumber products, and by Parajuli and Zhang (2016) for price discovery of softwood lumber in the United States.

Hetemäki and Mikkola (2005) combined multiple models, amongst which was also a VECM, to forecast Germany's paper imports. Another example from Finland is given by Toppinen (1998), who investigated the cointegration relations in the Finnish sawlog market. To forecast the demand of said market as well as lumber exports, based among others on imports from Germany, was also done, among others, with a VECM (Hetemäki et al., 2004). Further, Kuuluvainen et al. (2018) studied cointegration relations between the prices for domestic and imported sawlogs and pulpwood in Finland.

4.2 Price-induced logging changes (VAR)

The second case study combines topics of forestry and economics: An advantage of forestry is, that the trees, which usually represent the economic basis of the enterprise, need not to be harvested in a certain month or even year. Assuming no extraordinary circumstances, such as storms or bark beetle gradations, of course. Furthermore there are usually at least two tree species in relevant quantities within one enterprise. Therefore it is possible to choose, within limits, which kind of wood to cut down in a certain year.

It would therefore be reasonable, that an enterprise will direct logging activities to the species which can achieve the highest price, such that the revenue is maximized. This leads to the following question:

Q_2: Do the enterprises shift logging activities in accordance with a price change for certain tree species?

This can be seen as a part, or rather result, of the utilization of the portfolio theory. Said theory basically states, that by diversifying the portfolio, and hence the risk, the return can be stabilized by decreasing its variance (see e.g. Markowitz (1952, 2008)). Concerning forestry, Beinhofer (2010) found, that optimally diversified stands can make a substantial difference in the monetary outcome.

The portfolio theory is mentioned here, as such questions regarding changing logging activities due to the price are filed under this keyword in the department of forest economics in Göttingen. Thus the connection of this case study to former research may be more obvious.[20]

4.2.1 Used data

Two datasets were used for the analysis: The first one contains the annual wood harvest of the enterprises per hectare to obtain a measurement for the logging activity. To account for the differing natural resources and therefore possible strategies of the consulting rings (see 3.1), the data was calculated separately for each ring. The rings will be abbreviated in this section, *Spruce* as ring one, *Deciduous wood* will be ring two and ring three stands for *Pine*.

The database of the BVGL holds the harvested amounts of oak, beech, spruce and pine[21] in each economic year[22] per enterprise as well as the basal area, which is planted with said species[23]. Those values were summed up for each ring. By division values in [m^3 / ha] were obtained. Because the data had to be generated from the database, data was available only until the year 2015. Furthermore ring three reported zero values in 1969, therefore said year was excluded, leaving a time series from 1970 to 2015, displayed in fig. 4.7.

Because of the small sample size of 46 values the decision was made, not to use all of the available variables, but two time series to prevent severely reduced degrees of freedom in the

[20]For more information on the use of the portfolio theory in forestry, the interested (German-speaking) reader is referred to Beinhofer (2010).

[21]Table “HOLZ”, columns “Gesamteinschlag <*tree species*> WJ” with <*tree species*> = Eiche, Buche, Fichte, Kiefer (German: oak, beech, spruce, pine), WJ = Wirtschaftsjahr (German: economical year).

[22]For forest enterprises in Germany: 1[st] October to 30[th] September, §8 EStDV.

[23]Table “Altersklassen”, columns “<*tree species*>GES” (<*tree species*> = EICHE, BUCHE, FICHTE, KIEFER).

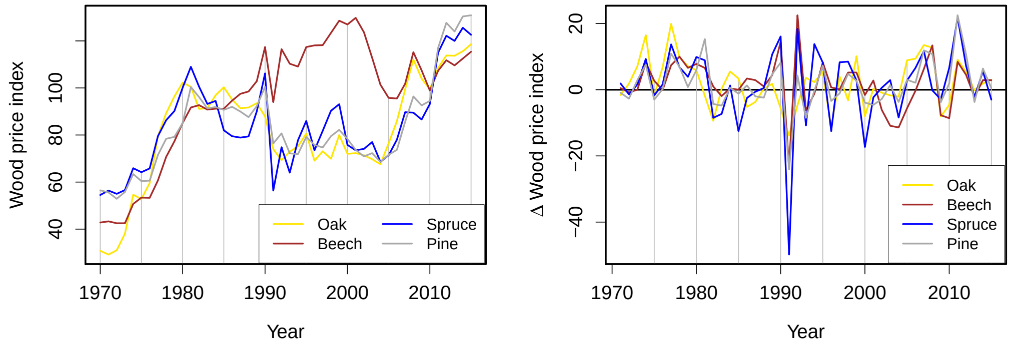


Figure 4.6: Wood price index of the national forest enterprises obtained from DESTATIS in absolute values (left) and first differences (right).

model. The first series chosen was spruce, since this species generated the highest marginal return in all enterprises and is therefore of great importance. (Dög et al., 2017) Moreover, it is seen as a leading price in forestry. To include deciduous timber in the analysis, beech was chosen as the second series, since it inhabits far larger areas in the enterprises than the oak. (Dög et al., 2017)

To be able to capture the current adaption strategies for each year, the choice was made to work with the first differences of the data. The values were stored in three separate objects called `pt.data.r`i`.d` ($i = 1, 2, 3$), one for each ring.

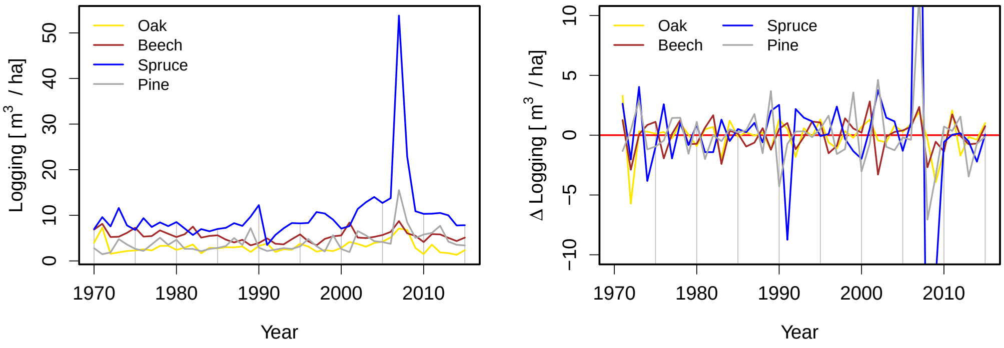
When searching for an answer to $\mathcal{Q}_2$, a wood price is obviously needed. A total yearly revenue for the different species is available in the database of the BVGL, which can be used to calculate an average wood price. However, the variable is only reported since 1985. Hence the decision was made, not to use this data as it would considerably shorten the sample.

The only suitable data available came again from DESTATIS: Table 61231-0003 of the GENESIS database holds a price index for trunk wood for the national forest enterprises for each economic year, incidentally for the same tree species (fig. 4.6). Reference point is the calendar year 2010 ($\triangleq 100$).

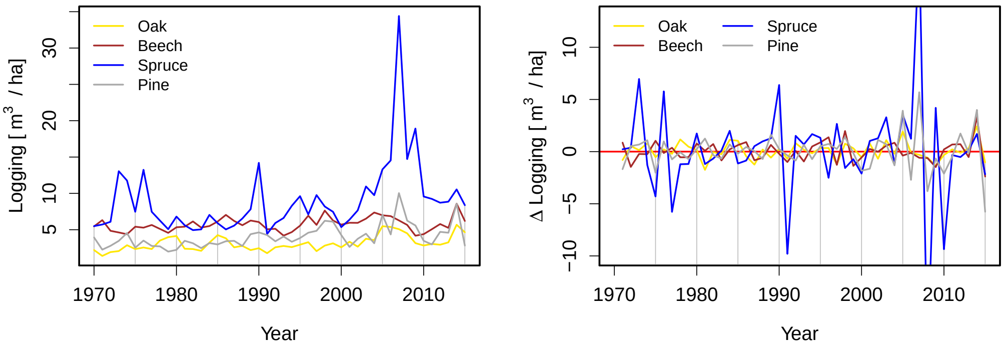
It should be noted, that in the following the complete harvest will be compared with the price index for trunk wood. However, the trunk wood price is seen to have the biggest influence on all wood prices. Furthermore, when a tree is cut down not only trunk wood but a wider range of timber will be co-produced. This is due to the fact, that (i) a tree consists of wood with varying diameters and quality, not all of which can be sold as trunk wood and (ii) not all harvested trees will contain wood classified as trunk wood. Hence this comparison was seen as valid.

The fact that the price is given as an index was also not seen to be too problematic. Direct connections from Euro to m^3 will not be feasible, however, a possible change in logging activity with a changing price can still be observed. Furthermore it is obvious, that a national price will not necessarily represent the data generating process in NRW. It was yet assumed, that the national price may not capture every detail of the one in NRW, but will still show the same

Ring 1:



Ring 2:



Ring 3:

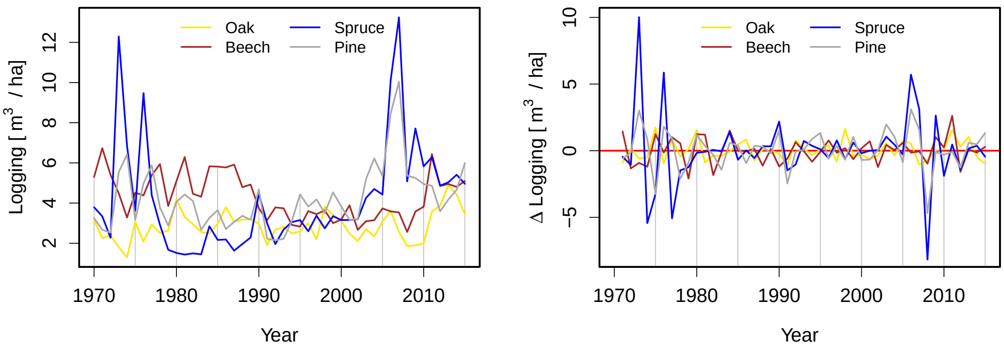


Figure 4.7: Annual wood harvest of the three rings used in the BVGL (cp Dög et al. (2017)) in absolute value (left) and first differences (right). Note, that the first differences for the first two rings do not show the values for the years 2007 to 2009, as the display of those extreme outliers would have made the graphics non-readable.

general development. Because in this variable the changes are again of higher interest than the absolute values, and to be in unison with the logging data, the price was also differenced.

In addition a matrix with dummy variables for the storm years was created with the same methodology and reasoning as in 4.1.1. Ergo a (46×6) matrix with zeros and ones for each storm year and the two following ones, respectively.

The differenced price and the dummy matrix were stored together in an object called `pt.exo.d`. The reason behind this being, that the decision was made to also use the price as an exogenous variable due to two main arguments: (i) The development of the price itself doesn't need to be modeled for $\mathcal{Q}_2$, and (ii) it was assumed, that the price is not crucially determined by the harvest, with the reason that the price was acquired on a national basis whereas the harvest is specifically for Westphalia-Lippe. Hence modeling it would only lead to an unnecessary reduction in degrees of freedom and be of no benefit for the analysis. Further it should be noted, that the price was not lagged, such that the current price changes can directly influence the harvest.

4.2.2 Modeling approach

Order of integration: When looking at the differenced data in fig. 4.7, the series appear to be stationary, with a possible exception of the storm years due to outliers. To reaffirm this impression, the usual tests were conducted (ADF and KPSS). The resulting p-values are displayed in tab. 4.8, they also indicate no strong violation of stationarity. Thus it was deduced that all series are *I(0)*, including the storm years. Hence a VAR seemed adequate for the analysis.

Had the tests indicated non-stationary data, one could have also used the moduli of the eigenvalues of $\boldsymbol{\mathcal{A}}$ (see 2.1.3) given in the summary of an object estimated with `VAR()` as an indicator. If they are smaller than one, the process is stable. Here the (possibly) interfering effects from the storm years would have been accounted for by the model, because of the used dummy variables for said storm years.

		ADF	KPSS
Ring 1	**Beech**	<0.010	>0.100
	Spruce	<0.010	>0.100
Ring 2	**Beech**	0.029	>0.100
	Spruce	0.015	>0.100
Ring 3	**Beech**	<0.010	>0.100
	Spruce	0.021	>0.100

Table 4.8: Stationarity test results (p-values) for the data used in 4.2.

Lag order selection: For the VAR estimation a suitable lag length needs to be determined, for which the function `VARselect()` from the package `vars` (Pfaff, 2008b) is utilized once again. Because the processes are not integrated, as in 4.1, all information criteria can be used. Relation (2.9), however, still holds.

With respect to the small sample size, maximal lag order was again set to five. Because the means of the data, displayed in tab. 4.9, were close the zero and a trend was not visible, the choice was made to exclude the deterministic terms entirely.[24] As said before, the changes in the wood prices as well as the storm dummy variables were included as exogenous data. Because none of the data is specific for a consulting ring, one dataset was used for all analysis. The results are displayed in tab. 4.10.

```
VARselect(pt.data.r1.d, lag.max = 5, type = "none", exogen = pt.exo.d)
VARselect(pt.data.r2.d, lag.max = 5, type = "none", exogen = pt.exo.d)
VARselect(pt.data.r3.d, lag.max = 5, type = "none", exogen = pt.exo.d)
```

	Δ Beech	**Δ Spruce**
Ring 1	-0.039	0.019
Ring 2	0.017	0.064
Ring 3	-0.004	0.026

Table 4.9: Means of the data used in 4.2 (stored in `pt.data.r`i.d, $i = 1, 2, 3$).

Based on those results, the lag order (p) was set to three for the first ring, as most indicators point to this value. For ring two, $p = 2$ was chosen, as it is closest to the mean of the lag orders chosen by the criteria ($\hat{p}$). Further an increase to $p = 3$ would lead to higher values for all criteria, especially the FPE. For the last ring p was also set to three, as a lag of four (the mean of $\hat{p}$) would lead to an increase in all criteria. When looking at the table, a lag of five would be just as good, however, it would lead to a decrease of the degrees of freedom. Hence $p = 3$ was chosen.

VAR estimation: The models were then estimated with parameters set in unison with the explanations above and stored in three separate objects, called `pt.var.r`i ($i = 1, 2, 3$). The output of the summaries is displayed in R Output 4.1 below as well as R Output 8.2 and 8.3 in the appendix.

[24] Further, when the models were estimated with deterministic terms, they were always close to zero and never had a significant influence.

Ring 1:

	1	2	3	4	5
AIC(n)	1.213	1.16	**1.054**	1.237	1.296
HQ(n)	1.518	1.527	**1.481**	1.726	1.846
SC(n)	**2.057**	2.174	2.236	2.588	2.816
FPE(n)	3.436	3.315	**3.05**	3.787	4.2

Ring 2:

	1	2	3	4	5
AIC(n)	2.084	2.044	2.157	2.014	**1.952**
HQ(n)	**2.389**	2.41	2.584	2.502	2.502
SC(n)	**2.928**	3.057	3.339	3.365	3.472
FPE(n)	8.213	**8.019**	9.196	8.234	8.095

Ring 3:

	1	2	3	4	5
AIC(n)	1.061	0.873	0.549	0.593	**0.453**
HQ(n)	1.366	1.24	**0.976**	1.082	1.003
SC(n)	1.905	1.887	**1.731**	1.944	1.973
FPE(n)	2.952	2.488	1.842	1.989	**1.807**

Table 4.10: Partial Output from `VARselect()` for all three consulting rings. Bold font indicates the respective minimal value.

```
pt.var.r1 <- VAR(pt.data.r1.d, p = 3, type = "none", exogen = pt.exo.d)
summary(pt.var.r1)

pt.var.r2 <- VAR(pt.data.r2.d, p = 2, type = "none", exogen = pt.exo.d)
summary(pt.var.r2)

pt.var.r3 <- VAR(pt.data.r3.d, p = 3, type = "none", exogen = pt.exo.d)
summary(pt.var.r3)
```

`R Output 4.1:` Var results for ring one from section 4.2.

```
VAR Estimation Results:
=========================
Endogenous variables: Beech.harvest.d, Spruce.harvest.d
Deterministic variables: none
Sample size: 42
Log Likelihood: -112.427
```

```
Roots of the characteristic polynomial:
0.6631 0.6631 0.6053 0.6053 0.4013 0.4013
Call:
VAR(y = pt.data.r1.d, p = 3, type = "none", exogen = pt.exo.d)


Estimation results for equation Beech.harvest.d:
================================================
Beech.harvest.d = Beech.harvest.d.l1 + Spruce.harvest.d.l1 + Beech.harvest.d.l2 +
    Spruce.harvest.d.l2 + Beech.harvest.d.l3 + Spruce.harvest.d.l3 + Beech.price.d +
    Spruce.price.d + Vivian.Wiebke + Vivian.Wiebke.l1 + Vivian.Wiebke.l2 + Kyrill + Kyrill.l1
    + Kyrill.l2

                  Estimate Std. Error t value Pr(>|t|)
Beech.harvest.d.l1 -0.62239  0.17774 -3.502 0.00157 **
Spruce.harvest.d.l1 -0.10893 0.09742 -1.118 0.27301
Beech.harvest.d.l2 -0.46200  0.15063 -3.067 0.00476 **
Spruce.harvest.d.l2 0.06086 0.04322  1.408 0.17007
Beech.harvest.d.l3 -0.18066  0.14895 -1.213 0.23533
Spruce.harvest.d.l3 -0.03194 0.02676 -1.194 0.24258
Beech.price.d      0.05439   0.03961  1.373 0.18063
Spruce.price.d    -0.02568   0.02686 -0.956 0.34714
Vivian.Wiebke     -0.16498   1.17678 -0.140 0.88951
Vivian.Wiebke.l1   1.01156   1.60260  0.631 0.53303
Vivian.Wiebke.l2  -2.31685   1.52144 -1.523 0.13903
Kyrill             3.22036   1.09538  2.940 0.00651 **
Kyrill.l1          2.70226   4.19731  0.644 0.52494
Kyrill.l2         -6.42524   4.50424 -1.426 0.16478
---
Signif. codes: 0 ‘***‘ 0.001 ‘**‘ 0.01 ‘*‘ 0.05 ‘.‘ 0.1 ‘ ‘ 1


Residual standard error: 1.014 on 28 degrees of freedom
Multiple R-Squared: 0.591, Adjusted R-squared: 0.3866
F-statistic: 2.89 on 14 and 28 DF, p-value: 0.008151


Estimation results for equation Spruce.harvest.d:
=================================================
Spruce.harvest.d = Beech.harvest.d.l1 + Spruce.harvest.d.l1 + Beech.harvest.d.l2 +
    Spruce.harvest.d.l2 + Beech.harvest.d.l3 + Spruce.harvest.d.l3 + Beech.price.d +
    Spruce.price.d + Vivian.Wiebke + Vivian.Wiebke.l1 + Vivian.Wiebke.l2 + Kyrill + Kyrill.l1
    + Kyrill.l2

                   Estimate Std. Error t value Pr(>|t|)
Beech.harvest.d.l1  0.54523   0.23392  2.331 0.027186 *
Spruce.harvest.d.l1 -0.14261  0.12821 -1.112 0.275460
Beech.harvest.d.l2  0.44127   0.19824  2.226 0.034245 *
Spruce.harvest.d.l2 0.01454   0.05688  0.256 0.800111
Beech.harvest.d.l3  0.43685   0.19603  2.228 0.034059 *
Spruce.harvest.d.l3 -0.06641  0.03522 -1.886 0.069752 .
```

```
52 Beech.price.d      -0.14202   0.05213 -2.724 0.010981 *
53 Spruce.price.d      0.06402   0.03535  1.811 0.080876 .
54 Vivian.Wiebke       4.60925   1.54873  2.976 0.005958 **
55 Vivian.Wiebke.l1   -8.58958   2.10915 -4.073 0.000346 ***
56 Vivian.Wiebke.l2    2.80359   2.00234  1.400 0.172451
57 Kyrill             39.74968   1.44160 27.573 < 2e-16 ***
58 Kyrill.l1         -25.14931   5.52400 -4.553 9.41e-05 ***
59 Kyrill.l2         -17.71684   5.92793 -2.989 0.005776 **
60 ---
61 Signif. codes: 0 ‘***‘ 0.001 ‘**‘ 0.01 ‘*‘ 0.05 ‘.‘ 0.1 ‘ ‘ 1
62
63
64 Residual standard error: 1.334 on 28 degrees of freedom
65 Multiple R-Squared: 0.9826, Adjusted R-squared: 0.974
66 F-statistic: 113.2 on 14 and 28 DF, p-value: < 2.2e-16
67
68
69
70 Covariance matrix of residuals:
71                 Beech.harvest.d Spruce.harvest.d
72 Beech.harvest.d          1.0221          -0.4256
73 Spruce.harvest.d        -0.4256           1.7078
74
75 Correlation matrix of residuals:
76                 Beech.harvest.d Spruce.harvest.d
77 Beech.harvest.d          1.0000          -0.3222
78 Spruce.harvest.d        -0.3222           1.0000
```

4.2.3 Model adequacy

In the following, the analysis will mainly be shown for ring one (R Output 4.1) to save space. However, model adequacy was still checked for the other two models with analogous commands and methods, graphics can be found in the appendix.

First of all, it can be seen in R Output 4.1, line 8, that all roots are distinctly smaller than one, hence no unstationary time series will be generated by the estimated VAR. Same goes for the other two VARs. The next check is the display of fitted over residual values, shown for ring one in fig. 4.8, for the other rings in figures 8.2 and 8.3 in the appendix. Further Pearsons correlation coefficient for the two was estimated, as well as the slope and the intercept of a simple linear model to obtain an objective measure of the direction of the relationship between fitted and original values. Both were computed without the modeled storm years (1990 to 1992 and 2007 to 2009), results are in tab. 4.11. Ideally, correlation and slope should be close to one, the intercept should be zero.

As it can be seen, both the beech and the spruce show a positive correlation with their fitted values of roughly the same strength in each ring. However, the estimated slopes are all substantially smaller than one, indicating that the fit of the model is not perfect. This is reaffirmed when looking at the plot of fitted and original values over time (fig. 4.8, 8.2, 8.3).

The models clearly could not grasp the yearly changes in the harvest of neither beech nor spruce, hence their explanatory power has to be questioned. The intercepts on the other hand were close to zero, as expected for data which is basically centered around zero (cp. tab. 4.9).

		Ring 1	**Ring 2**	**Ring 3**
Beech	**correlation**	0.690	0.508	0.761
	intercept	0.075	-0.063	0.004
	slope	0.472	0.253	0.580
Spruce	**correlation**	0.647	0.516	0.664
	intercept	-0.204	-0.237	-0.012
	slope	0.399	0.257	0.442

Table 4.11: Correlation of fitted and original values, plus the intercept and the slope of a linear model (see fig. 4.8, 8.2 and 8.3) in 4.2. The storm years (1990 to 1992 and 2007 to 2009) were excluded due to the perfect fit.

To check the assumptions of the VAR, the function `serial.test()` was used first, to test for autocorrelation in the residuals (see 2.1.13). Because of the small sample size, the respectively corrected tests will be used. Further a maximum lag order for the autocorrelation had to be chosen, here it was set to five, due to the small sample size, furthermore a longer horizon was seen to be unrealistic. Results can be seen in tab. 4.12. Moreover the `plot()` function was used on `serial.test()` to get a graphical impression of the residuals behavior, results for ring one can be seen in fig. 4.9, rings two and three are shown in figures 8.4 and 8.5 in the appendix. Because the correlation of the residuals in ring one seemed a little high (-0.3172), the cross-correlation function (CCF) was also computed (fig. 4.10). For the sake of completeness, the CCFs for the other two rings can be found in fig. 8.6 in the appendix.

For the first two rings, the test results are clear: Both the tests and the graphical analysis indicate no violation of the assumptions. For the last ring the portmanteau test indicates a possible violation, however, the Edgerton-Shukur test reports a high p-value. For further investigations, the autocorrelation functions (ACF) were used (see fig. 8.5, appendix). Only the squared residuals show some irregular properties, therefore the assumptions were seen to be fulfilled.

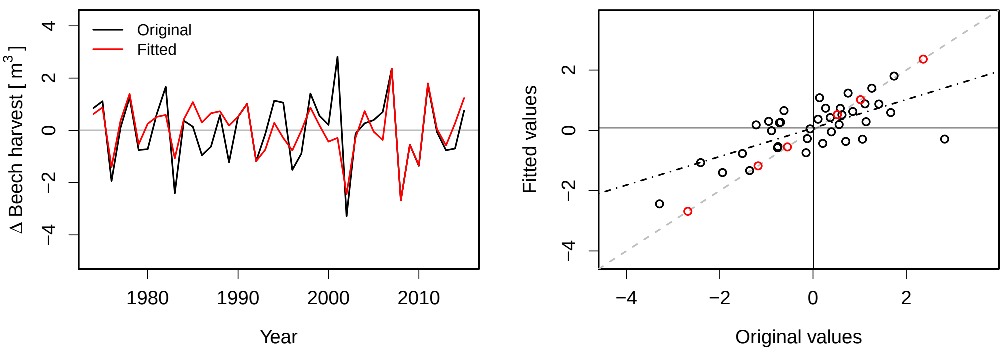
```
serial.test(pt.var.r1, type = "PT.adjusted", lags.pt = 5)
serial.test(pt.var.r1, type = "ES", lags.bg = 5)
```

The next assumption to check was the one of homoscedasticity. This is done on the one hand with an ARCH test (see 2.1.13), implemented in `arch.test()`. To be safe, the multivariate and the univariate tests were conducted via setting `multivariate.only = FALSE`. The results can be seen in tab. 4.13. On the other hand, the plots of the residuals in figures 4.9, 8.4 and 8.5 were examined. There seems to be no strong evidence of heteroscedasticity present in the data, neither in the graphical display nor in the test results.

```
arch.test(pt.var.r1, multivariate.only = FALSE)
```

To check the residuals for normality the test by Jarque and Bera (1987) implemented in the function `normality.test()` was used. To also obtain results for the univariate series, the

Beech:



Spruce:

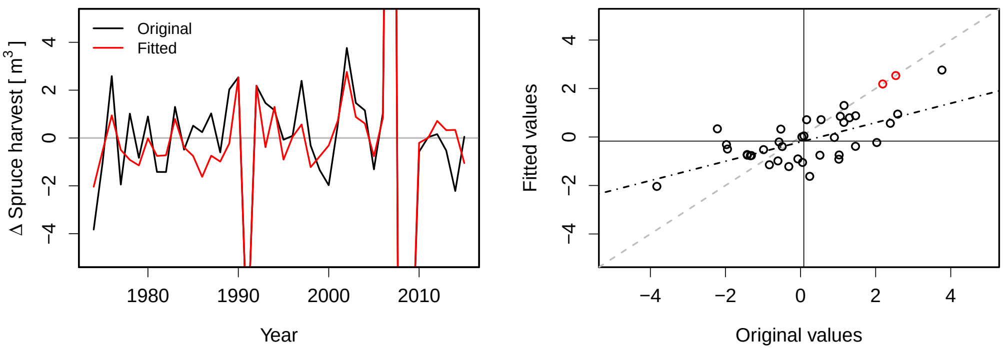


Figure 4.8: Fitted and original values for ring one in 4.2. Values are plotted together over time (left) and against each other (right). The dashed grey line has intercept zero and slope one, the dash-dotted line is a linear model of the displayed values (without the storm years), the thin black lines indicate the means of the respective values. Note that the storm years are not visible in the figure for the spruce, since it would impair visibility of the other data and the fit is essentially perfect.

	Ring 1	**Ring 2**	**Ring 3**
Portmanteau (adj.)	0.343	0.675	0.054
Edgerton-Shukur	0.874	0.961	0.512

Table 4.12: Results of tests for autocorrelated residuals (p-values) for the models estimated in 4.2.

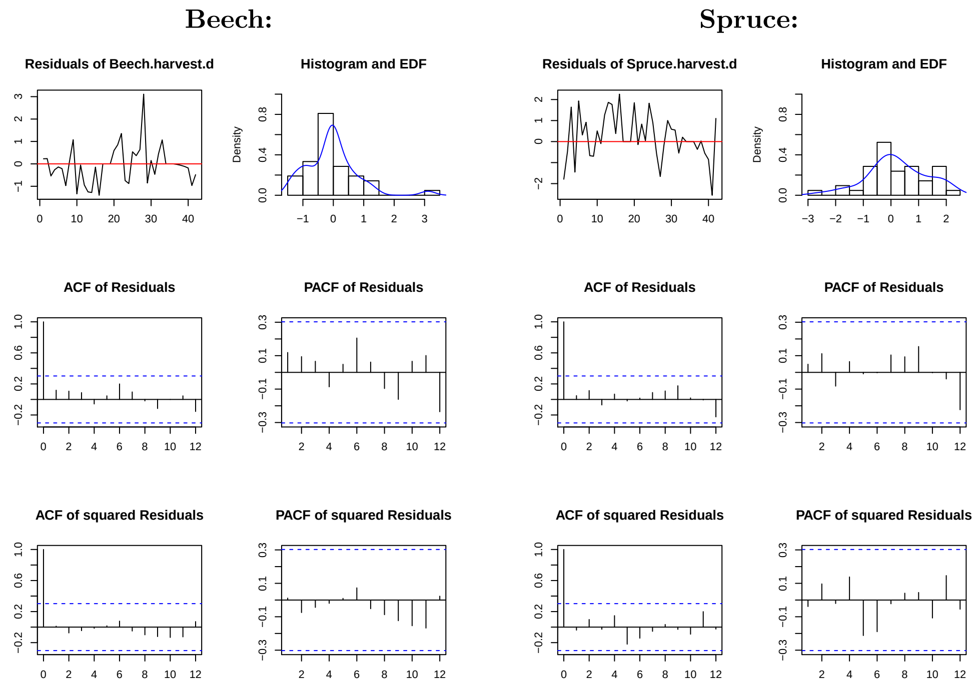


Figure 4.9: `plot()` function used on `serial.test()` for ring one in 4.2. Graphics for the Beech are in the two columns to the left, for the Spruce in the two columns to the right.

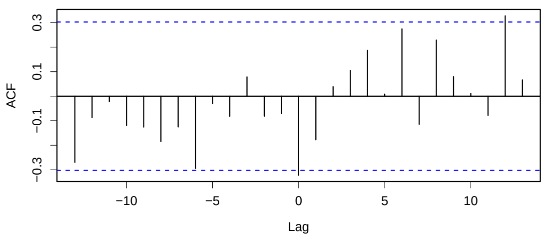


Figure 4.10: Cross-correlation function for the residuals of Ring 1 in 4.2.

	Beech *(univariate)*	**Spruce** *(univariate)*	**Both** *(multivariate)*
Ring 1	0.664	0.914	0.993
Ring 2	0.661	0.997	0.222
Ring 3	0.648	0.980	0.426

Table 4.13: ARCH test results (p-values) for the residuals of the models estimated in 4.2.

option `multivariate.only` was set to `FALSE`. The test found strong evidence against normally distributed data in every ring.

However, when looking at fig. 4.9, an outlier (≥ 3) can be seen in the histogram for the residuals of the beech in ring one, which might have influenced the test result. Because the other values seemed to behave quite well, a univariate Jarque-Bera test (`jarque.bera.test()` from the package `tseries` by Trapletti and Hornik (2018)) was also conducted on a dataset without the outlier. Same goes for ring 2, where an outlier $\geq$ 2.5 can be seen in fig. 8.4. The corresponding p-values are displayed in tab. 4.14.

The results clearly indicate normality for both beech series without the outliers. Therefore the other, multivariate results of the Jarque-Bera test were also assumed to be influenced by these outliers, hence the assumption of normally distributed residuals was seen to be fulfilled for the beech in all rings.

For the spruce there are also outliers visible, $<$-5 in ring two (fig. 8.4) and $\geq$5 in ring 3 (fig. 8.5). However, in this case the exclusion of the outliers did not bring the desired result, in the second ring the null hypothesis was deemed even less likely. However, the graphical display for ring two looked quite well and the first p-value of 0.098 was not too critical in the first place. Thus the assumption was seen to be fulfilled.

For the spruce in the third ring the p-value stayed considerably small, possibly due to some irregularities in the values $<$-2 and $>$4. However, because the histogram in fig. 8.5 otherwise indicates normally distributed data, apart from the outliers, the decision was made to see the assumption of normality as fulfilled.

```
normality.test(pt.var.r1, multivariate.only = FALSE)
```

	Ring 1	**Ring 2**	**Ring 3**
Beech	<0.001	<0.001	0.277
(tseries)	(0.870)	(0.457)	
Spruce	0.888	0.099	<0.001
(tseries)		(0.039)	(<0.001)
Both	<0.001	0.001	<0.001
Skewness	0.002	0.009	0.168
Kurtosis	<0.001	0.007	<0.001

Table 4.14: Results of the Jarque-Bera normality test (p-values) for the models in 4.2. The value in parenthesis was calculated without the outliers visible in figures 4.9, 8.4 and 8.5, respectively.

4.2.4 Results and Discussion

The objective of this case study was, to examine whether the forest enterprises shift logging activity due to price changes (cp. $\mathcal{Q}_2$). The data used was the harvest of beech and spruce in [m^3 / ha] for each economical year (fig. 4.7), calculated from the database of the BVGL separately for each consulting ring (see 3.1). The separation by ring was chosen, to make different adaption strategies visible, if they are present.

Furthermore the trunk wood price index per economical year of the national forest enterprises obtained from DESTATIS was included (fig. 4.6). The BVGL also contained a total revenue per tree species, yet it was only available since 1985 which would have made the sample even smaller, hence it was not used here. Since a change in logging activities in accordance with a change in the price was of interest, the data was differenced.

The price was used as an exogenous variable, since only the changes in logging activity were of primary interest for $\mathcal{Q}_2$. The exogenous dataset was expanded further by a matrix containing dummy variables for the storm years 1990 to 1992 and 2007 to 2009 (cp. 4.1.1).

Because the data was determined to be *I(0)* and centered around zero, a VAR without deterministic terms was chosen as a model. A separate model was fitted for each ring. The respective lags were determined in accordance with information criteria (see tab. 4.10) resulting in lags of three, two and three, respectively, for rings one to three. The assumptions were seen to be fulfilled for each model, relying mostly on graphical displays but also on statistical tests. The results can be seen in R Output 4.1 in this section as well as R Output 8.2 and 8.3 in the appendix.

Ring 1 (*Spruce*): When looking at R Output 4.1, it is first noticed that the changes in the harvest of beech wood only seem to depend on their own lagged values (ll. 18 - 23). Further the storm Kyrill had a noticeable impact with an increased harvest in year zero and quite a strong drop in year two (ll. 29 - 31). Such a drop after two years can also be noticed for Vivian / Wiebke, possibly indicating that all wood damaged by the storm has been removed and the stands are now being regenerated. A change due to differing prices, however, could not be noticed. The respective estimated values were close to zero and moreover insignificant (ll. 24 - 25).

For the spruce the parameter estimates drew quite a different picture: When looking at the autoregressive part (ll. 46 - 51), it seems that the harvest is not so much influenced by its own past values, but rather by the ones of the beech harvest. Because the respective parameters were strictly positive, an aligned adjustment of the changes in the harvest could be assumed.

Further the expected influences of the two storms are visible with an increased harvest in the first, and distinctly reduced ones in the following years (ll. 54 - 59). However, the parameters also show that the impact of Vivian / Wiebke was not as severe as the one of Kyrill.

When looking at the parameters for the price changes (ll. 52 - 53), the negative value for `Beech.price.d`, -0.158, implies, that the spruce harvest seems to be reduced when the price for beech wood is climbing. Therefore an adjustment of logging activity according to the price can be seen in the parameters. However, it has to be recalled, that the price is only given as an index. Therefore an exact connection between Euro and m^3 can not be made. The spruce price itself did not seem to be influential, the estimate of 0.071 was rather small.

Ring 2 (*Deciduous wood*): In this ring the changes in the beech harvest only seem to be influenced by the previous year, all other autoregressive parameters were close to zero and further rather insignificant (R Output 8.2, ll. 18 - 21). Same goes for the price changes (ll. 22 - 23). The estimated parameters for the storm years were all negative (ll. 24 - 29), indicating a possible concentration of activities on the conifers, which are more prone to be damaged by storms. However, their influence does not seem to be too big.

The spruce harvest in this ring seems to depend more on itself than before (ll. 44 - 47). The adaption to price changes, on the other hand, is quite similar to one before, the values, -0.164 for the beech price and 0.074 for the spruce, are quite identical to the estimates in ring one, though their significance was lower.

When looking at the storms, the impact of Vivian / Wiebke seems comparable (*l0*: 7.785, *l1*: -8.940, *l2*: 0.093), whereas the damage done by Kyrill seems to be distinctly smaller (ll. 50 - 55) (*l0*: 19.652, *l1*: -12.996, *l2*: 5.817).

Ring 3 (*Pine*): In the last ring the Beech harvest seems to not only be influenced by itself, but also by the spruce harvest (R Output 8.3, ll. 18 - 23). Again, the prices (ll. 24 - 25) and the storms (ll. 26 - 31) did not have a huge influence on the beech harvest. Albeit the first parameter for Vivian / Wiebke achieved significance, it is still rather small in absolute value (-1.738).

The spruce harvest in this ring, however, seemed to be influenced mainly by its own previous value in this ring, the parameter was estimated to be -0.485 (ll. 46 - 51). All other autoregressive parameters were comparably small. The parameters for the price again showed the same tendencies as before, a reduced harvest with an increase in the beech wood price and no distinct effect of the spruce price (ll. 52 - 53). However, the impact of the storms seems to be small compared to the other rings (ll. 54 - 59), as it could already be seen in fig. 4.7.

However, the results showed above should be interpreted carefully. Since the fit of the models did not seem satisfactory, as explained in 4.2.3, the explanatory power of the used variables does not seem to be very high (except for the storm dummies, obviously). Thus the estimated parameters can not be seen to reflect the truth, albeit their possible significance.

In conclusion, no clear answer can be given to $\mathcal{Q}_2$ with the data used here. Although the parameters showed a reduction in logging activities for the spruce when the beech price increased, therefore affirming $\mathcal{Q}_2$, interpretation can not be too strict at this point, as mentioned before. A broader study of the topic would be needed to confirm or refute the indications found here.

4.2.5 Further applications

When looking for applied work in forestry with VARs, not as many scientific articles as for the VECM (see 4.1.5) have been found in the scope of this work. This is probably due to the fact, that cointegration relations are present frequently, thus making the VAR inadequate, or that the researchers chose more advanced versions of a VAR, such as an SVAR (see 6), as a modeling approach. However, (Hetemäki and Mikkola, 2005) used VARs with and without exogenous data to forecast Germany's paper imports. Further application can be found in Hetemäki et al. (2004), where the Finnish lumber export is forecasted. Another example from Canada is the VAR analysis by Jennings et al. (2011) of the Canadian lumber industry and the macroeconomy.

4.3 Price interdependencies (IRF)

In this last case study, it should be investigated, whether there are interdependencies between the prices for the different tree species considered in 4.2 (oak, beech, spruce and pine), and what the effect of a price change for one species had on all species in the last decades. The question to answer will be:

$\mathcal{Q}_3$: What influence does a shock in the wood price of one tree species have on itself and the other species?

In this case study the influences of past prices on the current ones won't be of primary interest, but the influence of a price change in season t on seasons $t+1, \dots, t+h$.

4.3.1 Used data

For once, no data from the BVGL will be used, mainly due to the small sample sizes which are available there. The data used instead is again the trunk wood price index for the national forest enterprises of Germany obtained from DESTATIS. However, this time monthly data, available from 1968 to 2017, will be used from Table 61231-0002 of the GENESIS database, displayed in fig. 4.11. The decision was made to only use the last 30 years (1988 to 2017) for the analysis, as it was seen to be a good compromise between having a sufficiently large dataset and working with values which are still relevant. Furthermore, the variance of the beech was considerably smaller in the years before 1988 than after it, thus indicating heteroscedasticity. The data and its differences were stored in objects called `pi.data` and `pi.data.d`, respectively. It should be noted, that the data was differenced before the values before 1988 were excluded, leading to the same dimensions for both objects.

Further, the possible effects of the usual storm years (1990 to 1992 and 2007 to 2009) were considered by including a dummy matrix. However, since monthly data was being dealt with, the matrix had to be expanded to dimensions (360×6), containing a column of twelve ones for each month of the respective storm years and zeros elsewhere, such that one parameter per storm year is estimated. Said matrix was stored as an object called `pi.storm`.

4.3.2 VAR modeling approach

When looking for an answer to $\mathcal{Q}_3$, an Impulse Response Function (IRF, see 2.3) was deemed the most adequate choice of the methods and models introduced in part I. At this point, one should recall, that the IRF is based on the moving average representation of a VAR (see 2.3), or more precisely on $\mathbf{\Phi}_i$ (see 2.1.9). Thus a VAR is needed either way, whether by direct estimation or by converting a VECM to its VAR representation (see (2.16)).

When looking at the data in 4.11, the assumption of non-stationarity does not seem to be too far fetched, especially for the spruce and the pine. Hence the usual tests were computed (see tab. 4.15), all indicating non-stationary data for oak, spruce and pine. For the beech the evidence was not as strong, hence it may be stationary. When testing the first differences, all results indicated stationarity. Thus all series, with a possible exception of the beech, are assumed to be *I(1)*.

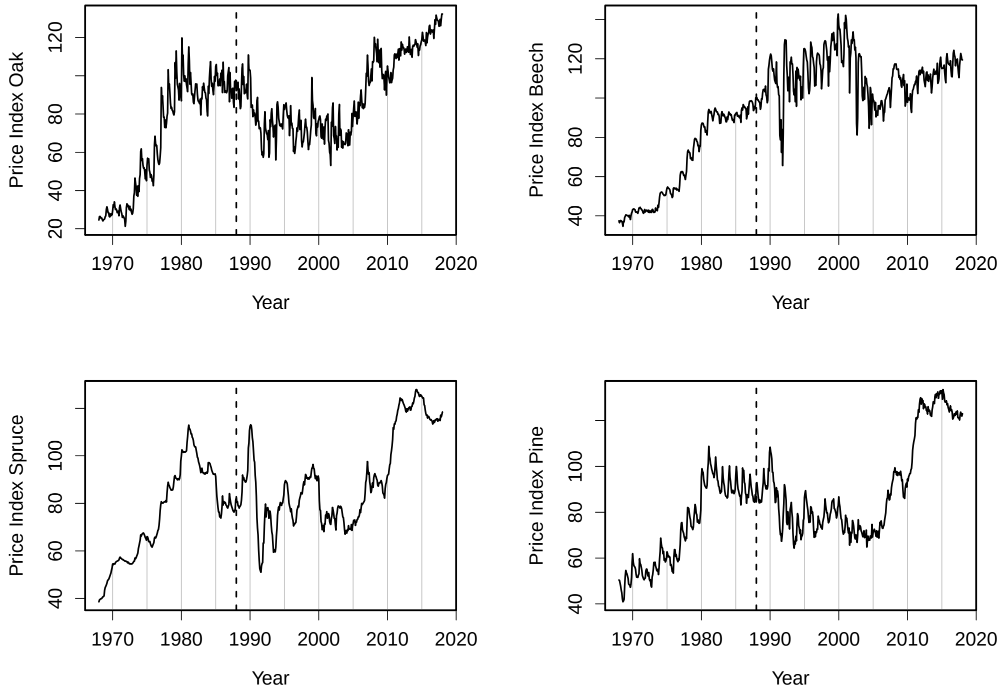


Figure 4.11: Wood price index from the national forest enterprises of Germany for the trunk wood of the tree species Oak (top left), Beech (top right), Spruce (bottom left) and Pine (bottom right). The dashed line indicates the year 1988, from which on the data was used.

	Oak	**Beech**	**Spruce**	**Pine**
ADF	0.531	0.117	0.338	0.768
KPSS	<0.010	0.065	<0.010	<0.010
	ΔOak	**ΔBeech**	**ΔSpruce**	**ΔPine**
ADF	<0.010	<0.010	<0.010	<0.010
KPSS	>0.100	>0.100	>0.100	>0.100

Table 4.15: Results for stationarity tests (p-values) for the data used in 4.3.

Because (some of) the data is possibly integrated, it might be cointegrated as well. Figure 4.11 indicates, that spruce, pine and possibly oak may follow a common trend. For the beech, no direct connection is visible. To gather further information, the Johansen test for cointegration was computed. The lag order, which is needed for the test, was again selected by consulting `VARselect()` from the `vars` package. The maximum lag was set to three years, or 36 months. The argument `season = 12` is used to incorporate twelve seasonal dummies, one for each month, into the model to account for seasonal effects.[25] Further a constant and the storm dummies were included, the result can be seen in tab. 4.10.

```
VARselect(pi.data,
          lag.max = 36,
          season = 12,
          type = "const",
          exogen = pi.storm)
```

	AIC(n)	HQ(n)	SC(n)	FPE(n)
1	9.795	10.204	**10.821**	17950.82
	⋮			⋮
3	9.497	**10.056**	10.898	13354.093
	⋮			⋮
12	9.38	10.61	12.461	**12127.996**
	⋮			⋮
30	**9.292**	11.863	15.733	13676.207
	⋮			⋮

Table 4.16: Partial output of `VARselect()` for the data used in 4.3. Lines without the respective minimum have been omitted, bold font indicates said minimum.

Due to the possibly (co-)integrated data, the AIC and FPE may be inconsistent (see 2.1.8). However, a lag of twelve was seen to be adequate since monthly data was being dealt with. Furthermore, when models were estimated with a lag smaller than twelve, they showed high autocorrelations for the residuals at lag twelve. Hence in the final model the lag was set to twelve. The Johansen test was then computed with the same settings as above. The results can be found in tab. 4.17.

```
pi.vecm <- ca.jo(pi.data,
                 K = 12,
                 season = 12,
                 ecdet = "const",
                 dumvar = pi.storm)
summary(pi.vecm)
```

The test indicated the presence of one cointegration relationship at all significance levels. However, the value for `r <= 1` was close to the critical value at $\alpha = 10\,\%$. If the beech series was truly stationary, a cointegration rank of two would be more adequate to account for the

[25] The necessity of and reasoning behind seasonal dummies will not be discussed here. The interested reader is referred to Zwirglmaier (2012) (German) for a detailed analysis based on the same dataset as used here.

	test	10pct	5pct	1pct
r <= 3	1.73	7.52	9.24	12.97
r <= 2	5.70	13.75	15.67	20.20
r <= 1	19.69	19.77	22.00	26.81
r = 0	39.00	25.56	28.14	33.24

Table 4.17: Results for the cointegration test conducted with `ca.jo()` for the data used in 4.3.

already stationary data. Hence neither the order of integration, nor the number of cointegration relations present can be deduced clearly here.

For such circumstances, Gospodinov et al. (2013) proposed the use of an unrestricted VAR for the non-differenced data in combination with an orthogonal IRF, as described in 2.3.1, as the most reliable strategy (when an IRF is the goal) (Gospodinov et al., 2013). Hence an unrestricted VAR for the non-differenced data was chosen as a basis and estimated with the parameters set in unison with the previous findings. The head of the VAR estimation output can be seen in R Output 4.2, the complete output is not displayed in this work due to its length, but is available on the enclosed CD and on Github: https://github.com/jan-schick/Masters_Thesis. However, the covariance and correlation matrices can be found in tab. 4.18, the autoregressive coefficients are displayed in fig. 4.12, the ones for the seasonal and storm dummies in fig. 4.13.

```
pi.var <- VAR(pi.data,
              p = 12,
              season = 12,
              type = "const",
              exogen = pi.storm)
summary(pi.var)
```

`R Output 4.2:` Partial VAR output for section 4.3.

```
VAR Estimation Results:
=========================
Endogenous variables: Oak, Beech, Spruce, Pine
Deterministic variables: const
Sample size: 348
Log Likelihood: -3366.596
Roots of the characteristic polynomial:
0.9919 0.9919 0.9551 0.9551 0.9434 0.9434 0.9131 0.9131 0.9105 0.9105 0.8985 0.8985 0.8877
    0.8877 0.8833 0.8833 0.8754 0.8754 0.8679 0.8679 0.8598 0.8598 0.8582 0.8582 0.857 0.857
    0.856 0.856 0.8552 0.8552 0.8382 0.8382 0.8331 0.8331 0.8235 0.8235 0.8157 0.8157 0.777
    0.777 0.7759 0.7759 0.7168 0.7168 0.6583 0.6583 0.5762 0.4136
Call:
VAR(y = pi.data, p = 12, type = "const", season = 12L, exogen = pi.storm)
```

Covariance:

	Oak	**Beech**	**Spruce**	**Pine**
Oak	22.042	2.737	1.24	1.791
Beech	2.737	26.518	-0.218	3.364
Spruce	1.24	-0.218	3.129	1.128
Pine	1.791	3.364	1.128	4.809

Correlation:

	Oak	**Beech**	**Spruce**	**Pine**
Oak	1.000	0.113	0.149	0.174
Beech	0.113	1.000	-0.024	0.298
Spruce	0.149	-0.024	1.000	0.291
Pine	0.174	0.298	0.291	1.000

Table 4.18: Variance / Covariance and Correlation matrix for the residuals of the VAR estimated in 4.3.

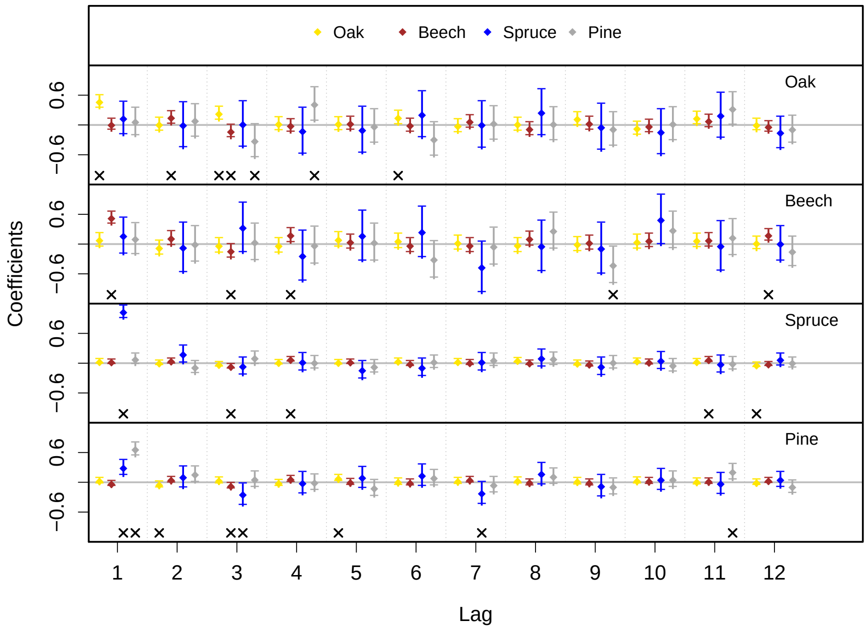


Figure 4.12: Estimated coefficients (diamonds) for the $\boldsymbol{A}_i$ matrices ($i = 1, \ldots, 12$, "columns" of the graphic) for each species ("rows" of the graphic). Lines and horizontal bars indicate $\pm$ two times the standard deviation. If the zero is not contained in this interval, a $\times$ is displayed below.

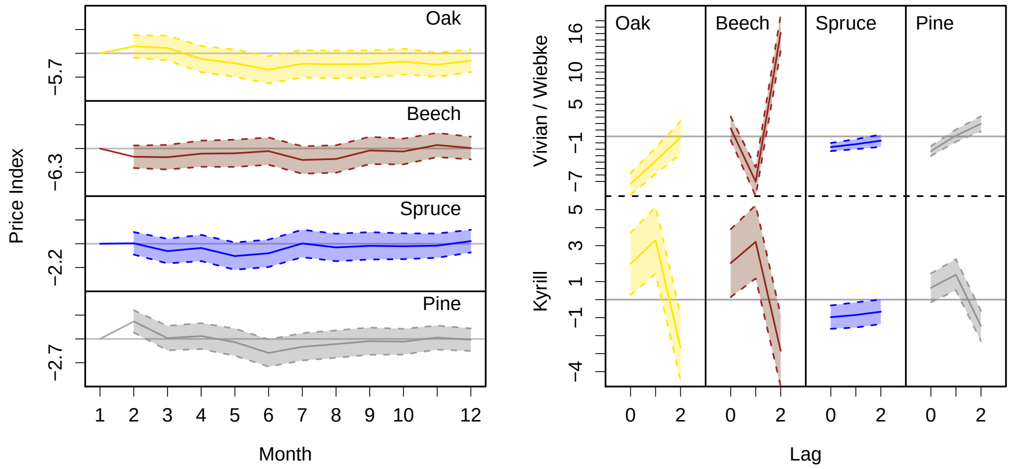


Figure 4.13: Estimated coefficients for the seasonal dummies (left) and the storm dummies (right), dashed lines indicate ± two standard deviations. For the seasonal dummies, January was set to zero as a reference point. On the right, the y-axes give the change in the price index for each storm. Horizontal gray lines indicate zero.

4.3.3 VAR model adequacy

First, the roots in line eight of R Output 4.2 are checked. They are very close to one, however, since they are still smaller than one no unstable process will be generated by the VAR. The next step is a comparison of the fitted over the real values, displayed in fig. 4.15. Further their correlation as well as the intercept and slope of a corresponding linear model are reported in tab. 4.19. The intercept, however, should not be interpreted too strictly as the data is not centered around zero. Also, in this case study the storm years were not excluded from the calculations (as in 4.1.3 and 4.2.3), because there were twelve values per storm year and one corresponding parameter, hence the fit is not necessarily perfect.

It can be seen in the graphics, that the fit was pretty good for all tree species. Correlation and slope are also close to one. Given the scale, even the intercepts are reasonably close to zero. The estimated slope and intercept for the beech were a little off, but still seen as satisfying. Hence the VAR seemed to be able to grasp the data generating process quite well.

	Oak	**Beech**	**Spruce**	**Pine**
Correlation	0.979	0.923	0.997	0.996
Intercept	3.762	16.565	0.605	0.777
Slope	0.958	0.852	0.993	0.992

Table 4.19: Correlation and the parameters of a linear model for the fitted and original values in 4.3. The storm years (1990 to 1992 and 2007 to 2009) were excluded from the calculations.

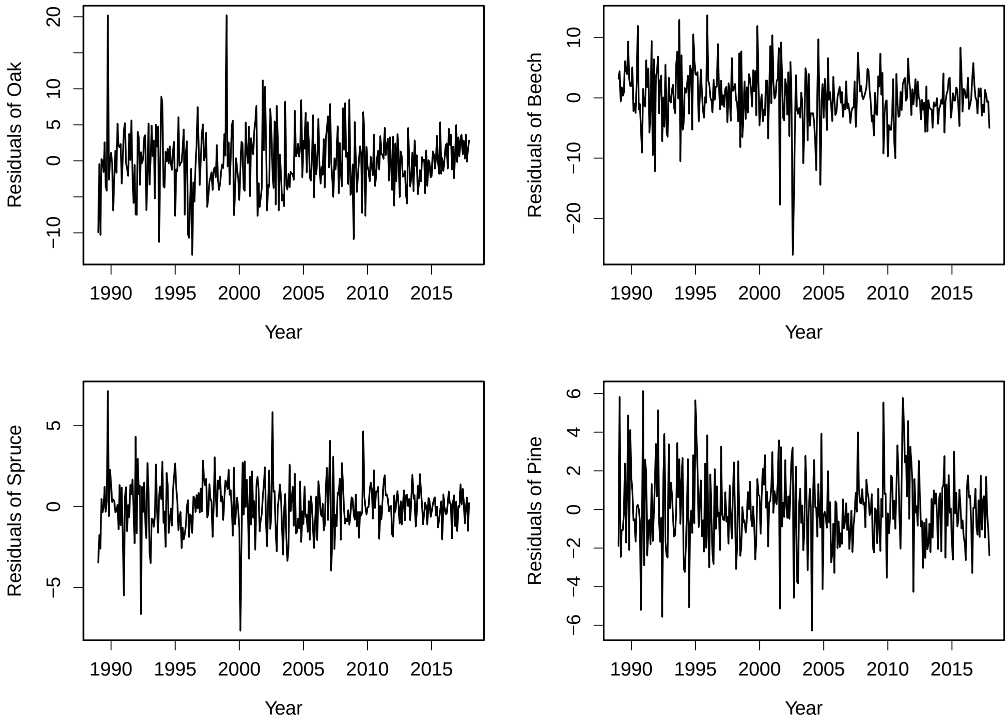


Figure 4.14: Residuals of the VAR in 4.3 plotted over time for oak (top left), beech (top right), spruce (bottom left) and pine (bottom right).

The assumptions for the residuals (no autocorrelation, homoscedasticity, normal distribution) were examined with the usual tests (`serial.test()`, `arch.test()` and `normality.test()` from the `vars` package (Pfaff, 2008b)). Where possible, the lags were set to twelve to account for eventual effects of the values recorded one year ago. Further, the multi- and univariate tests where computed, when available. Results can be found in tab. 4.20.

In all tests, at least one strong violation of the assumptions was indicated. Hence the graphical analysis in form of plots and histograms as well as auto- and cross-correlation functions was consulted to review the test results (figures 4.14, 4.16 and 8.7).

```
serial.test(pi.var,
            lags.pt = 12)
arch.test(pi.var,
          lags.single = 12,
          lags.multi = 12,
          multivariate.only = F)
normality.test(pi.var,
               multivariate.only = F)
```

	Oak	Beech	Spruce	Pine	All
Portmanteau					<0.001
ARCH	0.964	<0.001	0.966	0.33	<0.001
Jarque-Bera	<0.001	<0.001	<0.001	0.002	<0.001

Table 4.20: Test results (p-value) for the different assumptions (no autocorrelation, heteroscedasticity or non-normality) of the VAR estimated in 4.3, univariate results for each species (where available) are in the first four columns, the last one displays the multivariate results.

The least critical test was the ARCH test for heteroscedasticity, which only found irregularities for the Beech. When looking at the residuals over time (fig. 4.14), the years 2002 and 2003 seemed to have an increased fluctuation which may have distorted the test. Because the series behaved quite well otherwise, the assumption of homoscedasticity was seen to be fulfilled for all series.

To check the assumption of normality, which was strongly denied by the Jarque-Bera test, histograms were plotted for all series in fig. 4.16. Except for some outliers, the graphics clearly indicate normally distributed data. Especially the first three series beautifully show an almost perfect distribution. Therefore this assumption also seems to hold.

The last test to check is the one of no autocorrelation. The ACFs in fig. 4.16 showed no critical behavior of the data. Further fig. 8.7 in the appendix displays the CCFs, from which no irregular properties can be deduced either. Thus all assumptions were seen as fulfilled.

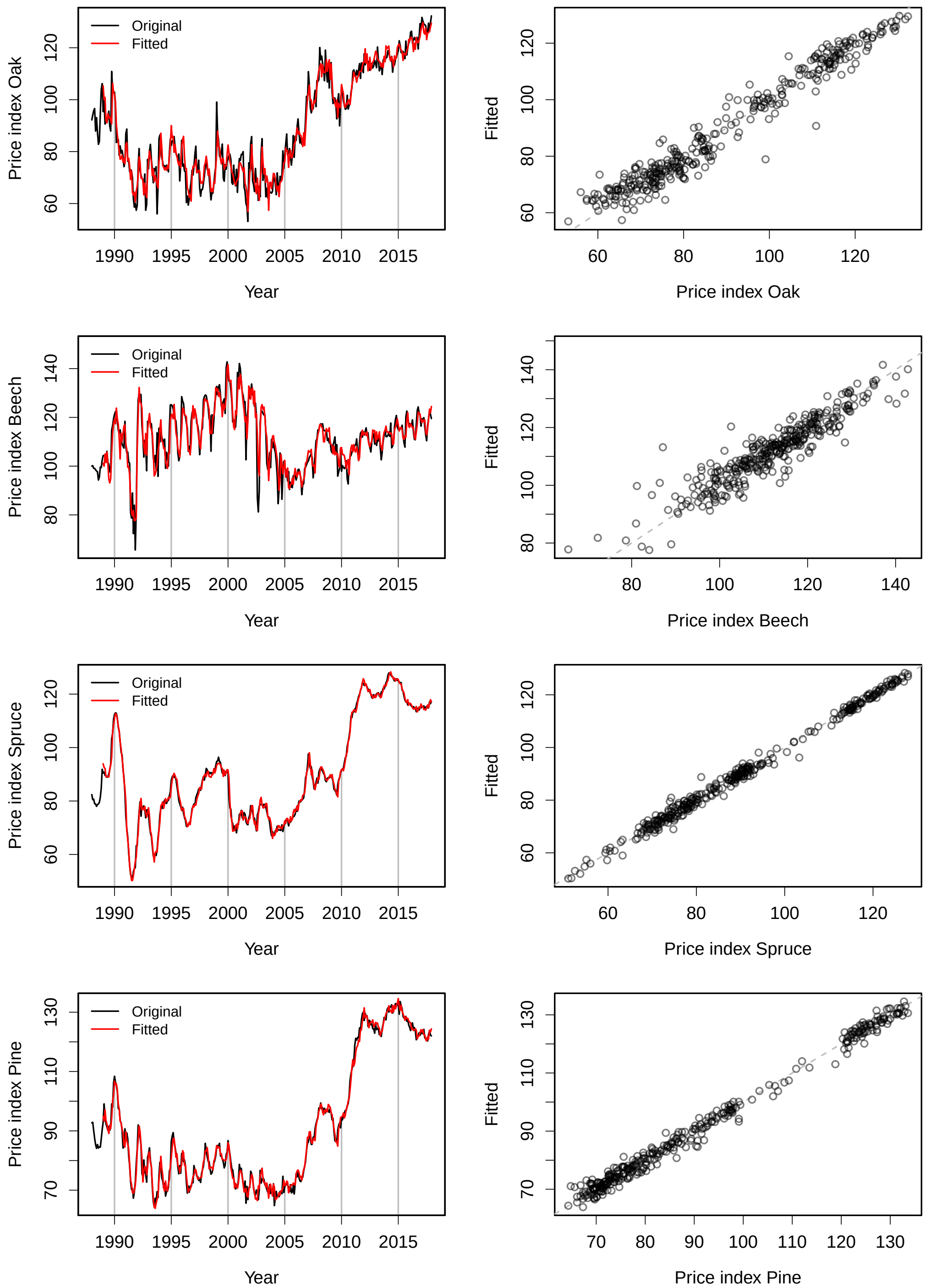


Figure 4.15: Fitted and original values of the model used in 4.3, plotted over time (left column) and against each other (right column) for oak, beech, spruce and pine (in that order from top to bottom). The color in the right plot is a semitransparent gray, such that overlapping values appear to be darker.

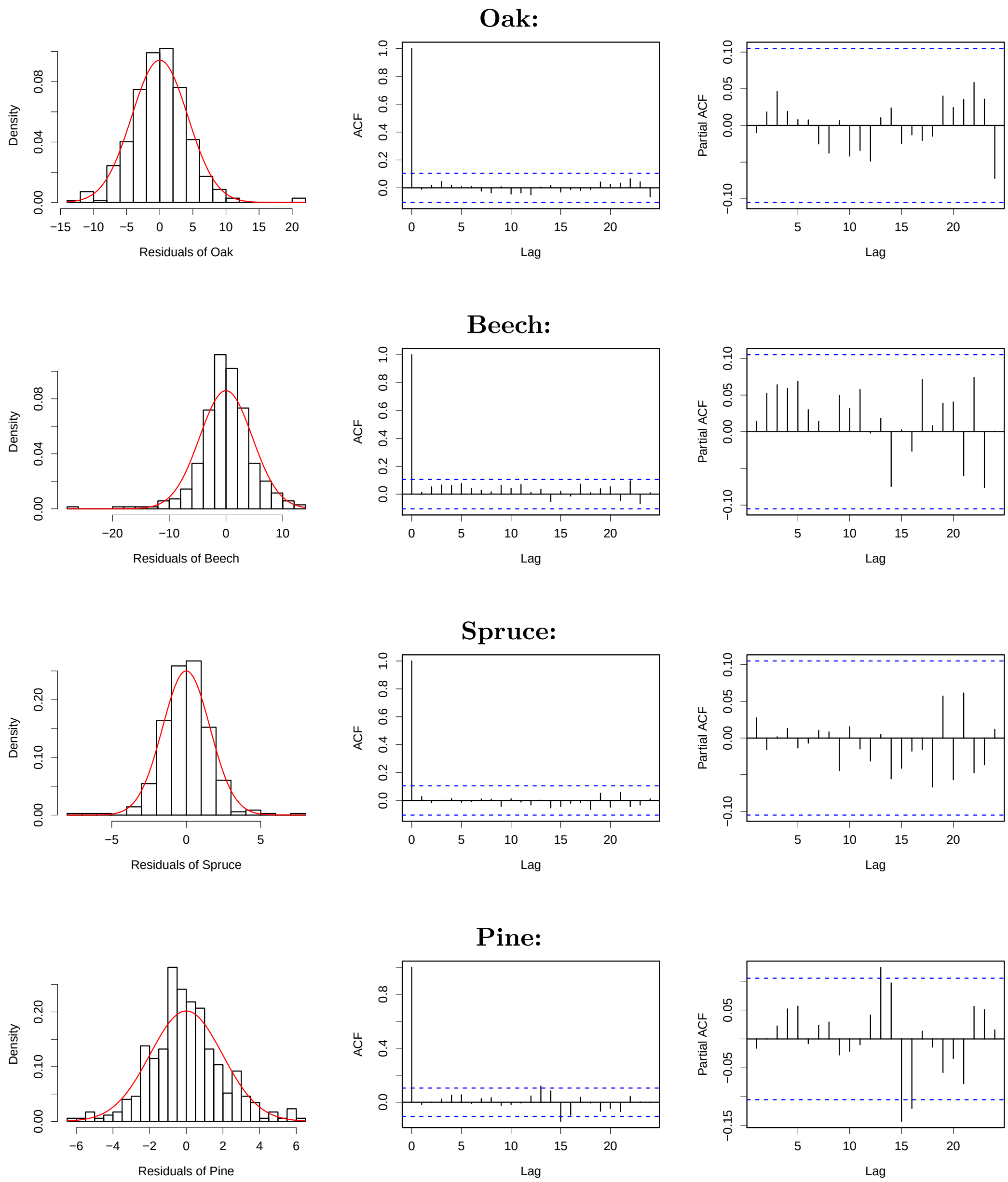


Figure 4.16: Graphical display of residual properties for the VAR estimated in 4.3: Histogram (left column) including a fitted normal distribution (red), autocorrelation function (center column) and partial autocorrelation function (right column) for oak (first row), beech (second row), spruce (third row) and pine (fourth row).

4.3.4 IRF and FEVD modeling approach

Since the VAR was deemed valid and the fit was satisfying, the next step was to estimate the IRFs. As it can be seen in tab. 4.18, there is correlation between the residuals, such that a shock in one variable will not likely happen in isolation. Thus orthogonal IRFs (see 2.3.1) had to be used. Further, as mentioned in 4.3.2, this approach should be the most reliable (see Gospodinov et al. (2013)).

Since one of the key elements of the orthogonal IRFs is a lower triangular matrix, i.e. the Choleski decomposition of the error covariance $\boldsymbol{\Sigma}_u = \boldsymbol{PP}'$, the ordering of the variables in the VAR is important: $y_{i,t}$ can only directly affect $y_{j,t}$ with $i > j$ in the IRF (cp. 2.3.1). For the given dataset, however, ordering the variables is not trivial, because the prices will influence each other and no price is an obvious result of another one. Further, no study with the same objective could be found in the scope of this work, hence no "reference order" exists.

First and last position, however, could be chosen rather easily: The spruce was set as the first variable because it is commonly seen as a leading price in forestry and will therefore have the largest influence on all other prices. Oak was chosen as the last variable, since its price is known to develop independently from external influences, due to its special uses, such as veneer, parquet or railroad ties. Positions two and three were not so clear. Finally, pine was set as the second variable, as its price was seen to be more stable (cp. fig. 4.11) and less influenced by trends, hence it should exert greater influence in the long run.

The VAR was estimated again with the same settings as in 4.3.2, but with ordered variables.[26] This model was used to calculate the IRFs with the command `irf()` from the `vars` package by Pfaff (2008b). The impulse responses were calculated for the next three years, or 36 months (`n.ahead = 36`), to also get an impression of the long-term effects. The argument `ortho = T` was used to calculate the orthogonal impulse responses, `runs = 1000` specifies the number of runs for the bootstrap confidence intervals. The resulting functions can be seen in fig. 4.18. To further get an impression of how the different variables affect each others fluctuation, the FEVD was also computed for the same time span of 36 months, displayed in fig. 4.17.

To assure the stability of the impulse responses, they were also calculated for 600 months. All functions approached zero in the long run, hence the process is indeed stable. To asses information about the sensitivity of the IRFs to the ordering of the variables, IRFs for all permutations, ergo all possibilities of the variable ordering, were computed, and are shown in fig. 4.18 (cp. Sims (1981)).

[26] Note, that the order of the variables has no effect on the VAR parameter estimation. Thus the VAR could already have been estimated with the chosen variable ordering in 4.3.2. This was not done here solely due to didactical reasons.

```
pi.var.sort <- VAR(pi.data[, c("Spruce", "Pine", "Beech", "Oak")],
                   p = 12,
                   season = 12,
                   type = "const",
                   exogen = pi.storm)
pi.irf <- irf(pi.var.sort,
              n.ahead = 36,
              ortho = T,
              runs = 1000)
pi.fevd <- fevd(pi.var.sort,
                n.ahead = 36)
```

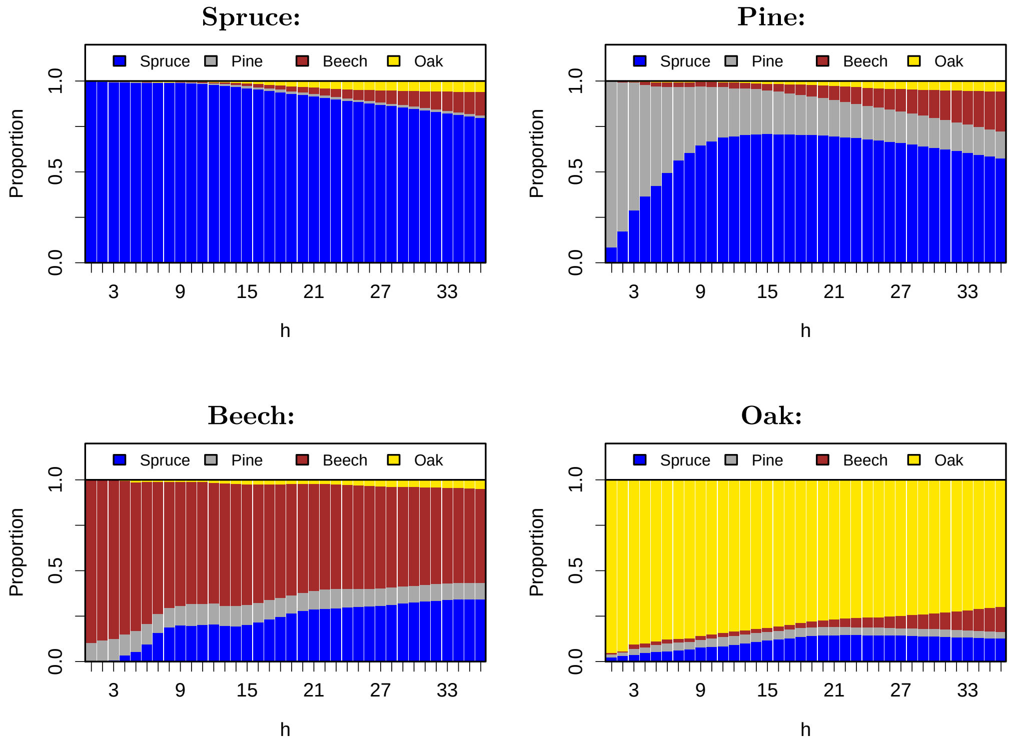


Figure 4.17: Forecast error variance decomposition for the VAR estimated in 4.3.

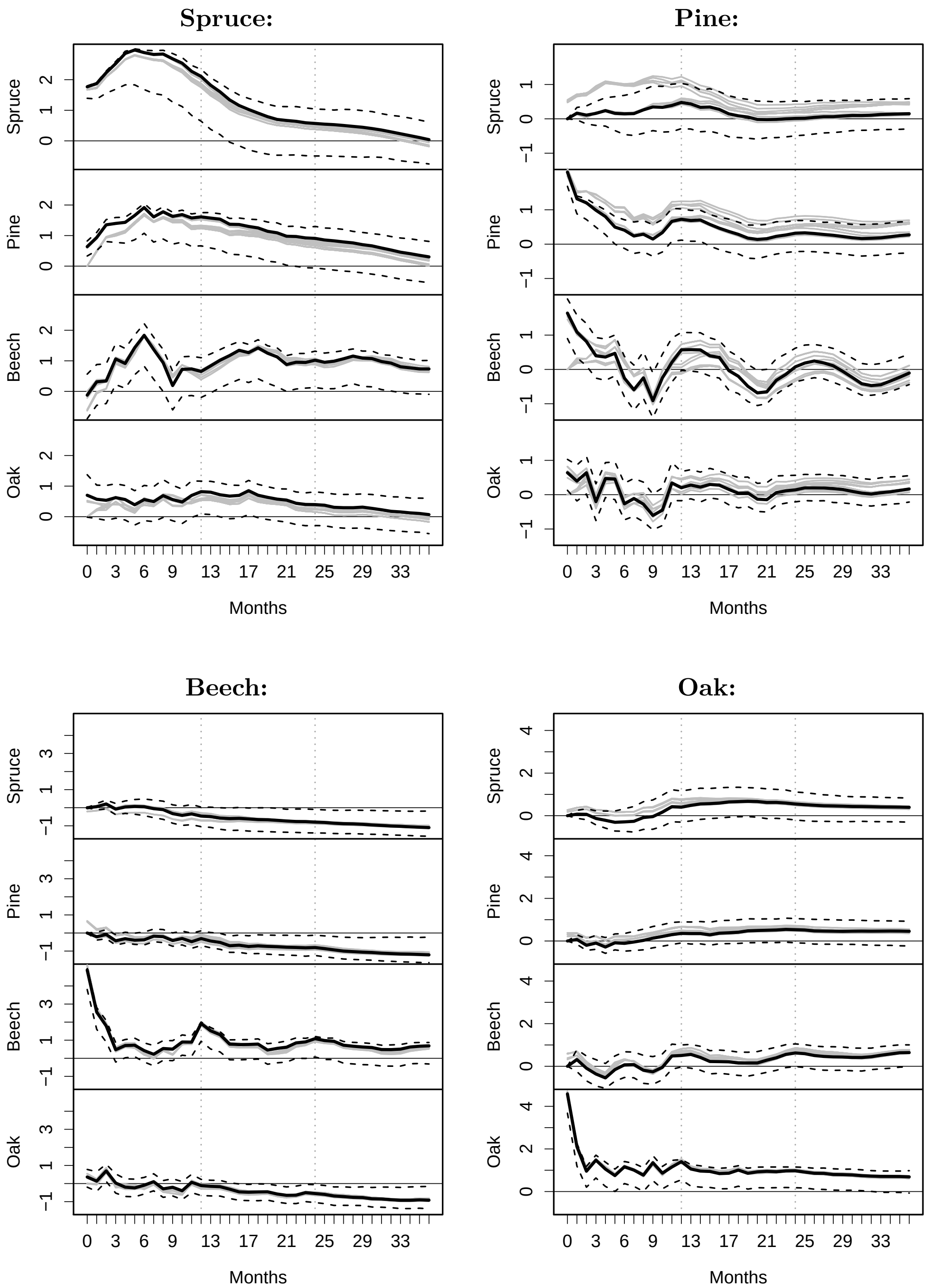


Figure 4.18: Orthogonal impulse responses (solid black lines) for the ordered VAR estimated in 4.3, including 95 % bootstrap confidence bands (dashed black lines) obtained from 1000 runs. Grey solid lines indicate IRFs from all permutations of the variable ordering (see 4.3.4). The vertical dotted gray lines indicate 12 and 24 months after the shock. The variable in which the shock occured is written above each plot.

4.3.5 Results and discussion

In this last case study, the underlying question was, what the effect of a price change in one tree species would be on itself as well as on the other species (cp. $\mathcal{Q}_3$). To increase the sample in comparison to the second case study (see 4.2), only data from DESTATIS was used. The values chosen were the monthly trunk wood price indexes for oak, beech, spruce and pine (see 4.3.1), whose yearly values had already been used in 4.2. All indexes were used as endogenous data.

When searching for an answer to $\mathcal{Q}_3$, an impulse response was seen to be adequate. Due to uncertainties concerning the order of integration and the cointegration rank, the reasoning of Gospodinov et al. (2013) was followed and an unrestricted VAR in levels was estimated. The lag order was set to twelve, as the FPE in tab. 4.16 suggested, to account for annually recurring effects. Dummies were used for each month, to include the month-specific effects, as well as one coefficient (via a column of twelve ones, respectively) for each storm year. As the data was not centered around zero, a constant was also included.

As it can be seen in fig. 4.15 and tab. 4.19, the fit of the model was quite satisfying. The assumptions of the VAR were checked and seen to be fulfilled, mostly relying on graphical displays (figures 4.14, 4.16 and 8.7). The model was rated as valid. Finally, the VAR was re-estimated with differently ordered variables, and an orthogonal IRF and a FEVD were calculated for the next three years or 36 months, displayed in figures 4.18 and 4.17, respectively.

First it should be noted, that the influence of the ordering of the variables did not seem to be too large in general. The first value of the IRFs at $t = 0$ naturally changes with the ordering (cp. 2.3.1), but the curves approached each other rather quickly. Noticable differences, however, were found when looking at the effects of a shock in the pine, especially on the spruce. Particularly in the first year the effect was distinctly larger when the spruce was ordered below the pine (ergo non-zero values for $t = 0$). Also the effect on itself seemed to be larger in some of the cases. As mentioned, there is no statistical procedure to determine which order is correct (cp. Lütkepohl (2007)), and no reference literature is available. Hence this delivers no rational reason to change the order selected earlier, yet it should be kept in mind when interpreting the results for the pine. For all other species, however, the ordering had only small to marginal effects, such that the results would not substantially differ.

The IRFs and FEVDs will be discussed below, separately for each tree species in which the shock occured. It should be noted again, that no literature with a comparable analysis or a suitable theoretical basis for the results obtained here could be found in the scope of this work. Hence all interpretation, and the results themselves, need further validation.

Spruce: First, it can be seen that, generally speaking, a positive shock in the spruce price resulted in higher prices for all other species as well, which speaks for the theory of it being the leading price in forestry. Moreover, the effect of an increased spruce price appeared to be quite persistent in time and rather detached from seasonal behavior, as, except for the beech, no annually recurring effects could clearly be seen. Moreover, the effect on the spruce price itself appeared to be self-enhancing for the first six to eight months.

When looking at the FEVD in fig. 4.17, it could clearly be seen that the variation of the spruce price is almost entirely its own and not influenced by any other species, especially in the first twelve months. Even at the end of the considered time span, the summed influence of all other species were only slightly larger than 20 %.

Pine: The impact of a shock in the pine price did not seem to be as long-living when compared to the spruce. Annually recurring effects, on the other hand, were visible for the beech, the pine itself and possibly the oak. However, the overall effects on the spruce and the oak were rather marginal, thus indicating a small influence of the pine price.

The FEVD on the other hand showed, that, in the long run, the variance of the pine price is heavily influenced by the spruce price. After about one and a half years, the beech also gets an increasing influence.

Beech: The effect of a shock in the beech price was quite contrary to the results of the conifers. First, the increased price quickly moved back towards zero and stayed rather low, but above zero, in the next years. With the exception of an increase after twelve months. However, the beech price stayed on a higher level, whereas the effect was negative for all other species in the long run.

When looking at the influences on the variance of the beech price in fig. 4.17, the analysis suggested, that the variation was mainly its own, even in the long run. However, after three months the influence of the spruce increased. The proportion of the pine, on the other hand, stayed rather constant.

Oak: A shock in the oak price also seemed to die out rather quickly within two months. However, the price did stay higher for a long time and was not drastically influenced by seasonal effects. For all other species no large effect could be seen in the first year, however, with the second year the prices increased slightly.

The variance showed a similar picture as the spruce, the fluctuation in the oak price was mainly its own, even after three years the cumulated influence of all other species was only at about 30 %.

Generally speaking, annually recurring patterns were not strongly present in the IRFs of the coniferous tree species, except for the reaction of the pine price to a shock in itself. Especially for the spruce this result is congruent with the original values in fig. 4.11, where no seasonal behavior can clearly be deduced either. In combination with the seasonal dummies, which were comparably small and unsignificant (fig. 4.13), the spruce price seemed to be quite stable and unaffected by seasonal trends, compared to the other prices.

For the oak slight seasonal behavior could be anticipated in the original data (fig. 4.11). Their absence in most IRFs in fig. 4.18 may be due to the fact, that the price has a specific effect for each month of the year, which was accounted for by the seasonal dummies (fig. 4.13). Hence no further annually recurring effects may be relevant for the oak price. The beech on the other hand showed no clear month-specific coefficients (fig. 4.13), although annually recurring effects were present in the impulse responses, especially for a shock in the pine price. Thus for the beech the price achieved twelve months ago may be more important than the month in which it is sold.

As a side note, another striking aspect of fig. 4.13 is the effect of the storms on the price indexes. For the coniferous wood, almost no effect is present and for the deciduous species a slight increase during the first 24 months after Kyrill was indicated. For Vivian / Wiebke the effects were also not as negative as one might expect.

For the latter one this might be due to a higher demand for construction wood in the eastern part of Germany after the reunification, such that the wood could still be sold. (Ripberger, 1994) Concerning both storms, it should further be remembered, that the data used here is a national price index, whereas the storms had their biggest impact in the western and southern parts of Germany. (cp. e.g. Ripberger (1994)) Hence at least some damping of the effects by the influence of the other federal states has to be expected. Moreover, some wood may have been sold by the terms of previously negotiated contracts where no influence on the price could have been detected. Finally, a general decrease in the prices after the storms, as visible in 4.11, especially for Vivian / Wiebke, may already have been captured by the autoregressive parameters. Hence another approach than the one chosen here may be more suitable when one is actually interested in the effect of storm induced-price changes.

Concerning the FEVD (fig. 4.17), the spruce price seemed to exert noticeable influence on the fluctuation of the pine especially, but of the beech as well. This can be seen as further evidence for it being a leading price in forestry. The influence of the pine, however, was comparably small, even concerning its own variance. After six months, more than half of the influences came from other species. The beech on the other hand showed behavior comparable to the IRFs: Only small influences on all other species, but a substantial impact on its own variance. Finally, the oaks fluctuation was also mainly its own and had almost no influence on other species. Thus giving further affirmation of the assumption, that it has no large effects on other prices, possibly due to its special uses.

In conclusion, the analysis showed strong interdependencies with mostly positive effects between price changes induced by the coniferous wood, especially the spruce. A shock in the beech price, however, seemed to have negative effects on all other species in the long run. The oak on the other hand had a positive influence on the remaining species. Thus giving a first answer to $\mathcal{Q}_3$. Further, no strong evidence against the variable order selected in 4.3.4 could be found, if anything it was affirmed.

However, as mentioned before, the analysis crucially depends on the order of the variables. As no reference work or theoretical basis could be found in the scope of this work, the results still need further validation.

4.3.6 Further applications

As already noted in 4.2.5, an “ordinary” VAR approach in combination with orthogonal impulse responses seems to be rarely used in recent research. Other, more advanced models like structural VARs or VECMs (see 6), are more common, because the IRFs derived from said models can be advantageous, as they do for example not depend on variable ordering (Lütkepohl, 2007, Pfaff, 2008a). Moreover, other calculation methods for the IRF have been proposed, like the generalized IRF (GIRF) (Koop et al., 1996). However, the interpretation of the (G)IRFs is still similar to the ones used earlier.

Application can for example be found in the study by Alavalapati et al. (1997), which was already mentioned in 4.1.5. Here the VAR representation of a VECM is used to generate impulse responses for a reaction of the Canadian pulp wood price to a shock in the U.S. / Canada currency exchange rate as well as one in the U.S. pulp price. Shahi and Kant (2009) used GIRFs to study the effect of a shock in different wood prices on several types of timber. Another approach can be found in Hansen et al. (1999), where an IRF is computed from a VECM to analyze the demand for roundwood imports in Denmark.

Further examples can be found outside of forestry. For example, Herwartz and Plödt (2016) and Kilian (2009) used an SVAR (see 6) as a basis for the IRF to investigate properties of the oil market and price. The same approach was used by Carstensen et al. (2013) to examine the influence of the oil market on the German economy as well as connections to the 2009 recession.

5 Concluding remarks

The aim of this work was, to give the reader an impression of the possibilities of multivariate time series analysis, such that one can develop own ideas or see how existing questions can be answered, respectively. The theory necessary was introduced in part I. The application of said theory followed in part II in the form of three case studies, which used preferably easy datasets.

At this point it should be stressed, that the first two case studies were only possibly because the BVGL provides continuous data on numerous variables since 1969. A time series gathered over decades obviously has its difficulties and drawbacks (see 3.1). However, when dealing with forest economics, data in sufficient quantity and quality, and especially without data gaps or changing variables (cp. Nellen (2012)), is hard to come by. This was also faulted by Toppinen and Kuuluvainen (2010) as well as Kolo and Tzanova (2017). The wood price index from DESTATIS also was rather the exception than the rule.

Although not all variables in the BVGL have been collected since 1969, and therefore their sample size may still be too small for the models introduced in this work (cp. 4.2), their sample will naturally increase with the years such that they will be available for future analysis. Thus continuing the expansion of this unique dataset is crucial.

When applying the theory presented in this work for future research, the reader should once more be reminded that the models presented here are merely the basic ones, especially the VAR. They do have their uses, as shown in the case studies, but can be inferior to the more complex structural models, like SVARs or SVECMs (see 6), especially when IRFs are the objective. Those models were not dealt with in part I due to their complexity, the interested reader is again referred to Lütkepohl (2007), section 9, or Kilian and Lütkepohl (2017).

Hopefully it could still be shown in part II, that there is great potential in multivariate time series analysis such that future research is inspired.

6 Further Topics

In the course of this work only the basic models have been used. To allow a broader view, this section will provide a very brief overview of some more advanced models, such that the interested reader may be able to judge their usefulness for specific analysis.

Structural Models: Structural Models, like SVARs and SVECMs, have been mentioned a few times in this work. The basic idea behind an SVAR is, to find a matrix, here called $\boldsymbol{B}$, by which the original VAR equation, as given in (2.3), without the intercept, can be multiplied such that the error term is instantaneously uncorrelated. $\boldsymbol{B}$ is usually based on the covariance matrix of the error term, $\boldsymbol{\Sigma}_u$. Finding a suitable matrix, however, is not trivial as there are several methods. Some identification methods for $\boldsymbol{B}$ can, for example, be found in the package `svars` by Lange et al. (2018). (Lütkepohl, 2007)

Though if $\boldsymbol{B}$ is chosen correctly, the Impulse Responses obtained from an SVAR have the main advantage that they are unique and do no longer depend on the variable ordering, as it would be the case with orthogonal impulses. Structural forms of a VECM can also be calculated, however, they are based on the Beveridge-Nelson MA representation of a VECM, which was not discussed in this work. For more details see Lütkepohl (2007), section 9, or Kilian and Lütkepohl (2017). (Lütkepohl, 2007)

VARMA: A Vector Autoregressive Moving Average (VARMA) model is also based on the standard VAR equation (2.3), yet the error term, here denoted $\boldsymbol{\epsilon}_t$, does not have to be white noise but can still contain autocorrelation, defined by the MA representation $\boldsymbol{\epsilon}_t = \boldsymbol{u}_t + \boldsymbol{M}_1\boldsymbol{u}_{t-1} + \cdots + \boldsymbol{M}_q\boldsymbol{u}_{t-q}$. (Lütkepohl, 2007)

ARCH: Autoregressive Conditional Heteroscedasticity (ARCH) models are used on processes with time-invariant variances. They model the current variance of the error term $\boldsymbol{u}_t$ as a function of the variance of the past q error terms. (Lütkepohl, 2007)

This list is by no means concluding, further topics can be found e.g. in Lütkepohl (2007), Pfaff (2008a), or, for univariate series only, in Shumway and Stoffer (2006).

7 Acknowledgements

At this point thanks should be expressed towards the numerous people who made this thesis possible. First of all, the examiners

Prof. Dr. Bernhard Möhring
and
Prof. Dr. Helmut Herwartz,

who kindly agreed on grading this work, are thanked for their expertise and advice. Further gratitude is expressed towards

Dr. Kai Husmann,

who had the initial idea for this work and closely supervised it with constructive criticism and great patience. Additional consulting and advise came from

Alexander Lange,

whose expertise helped greatly in understanding some more complex aspects of the topic. Moreover, thanks are expressed towards Dr. Markus Dög, Lennart Empting and Dr. Bernhard Pfaff. Further gratitude belongs to family and friends for their support during this work, especially Lea Schmeil, Ingo Schick and Birgit Schick. Finally, Hartley R. Alley should be thanked for keeping the authors back clear.

8 Appendix

R Output 8.1: Direct Output from the `VAR()` function used in 2.1.2.

```
VAR Estimation Results:
=======================

Estimated coefficients for equation do:
=======================================
Call:
do = do.l1 + dw.l1

do.l1       dw.l1
1.01954489 -0.09189592


Estimated coefficients for equation dw:
=======================================
Call:
dw = do.l1 + dw.l1

do.l1     dw.l1
0.5025056 0.7981194
```

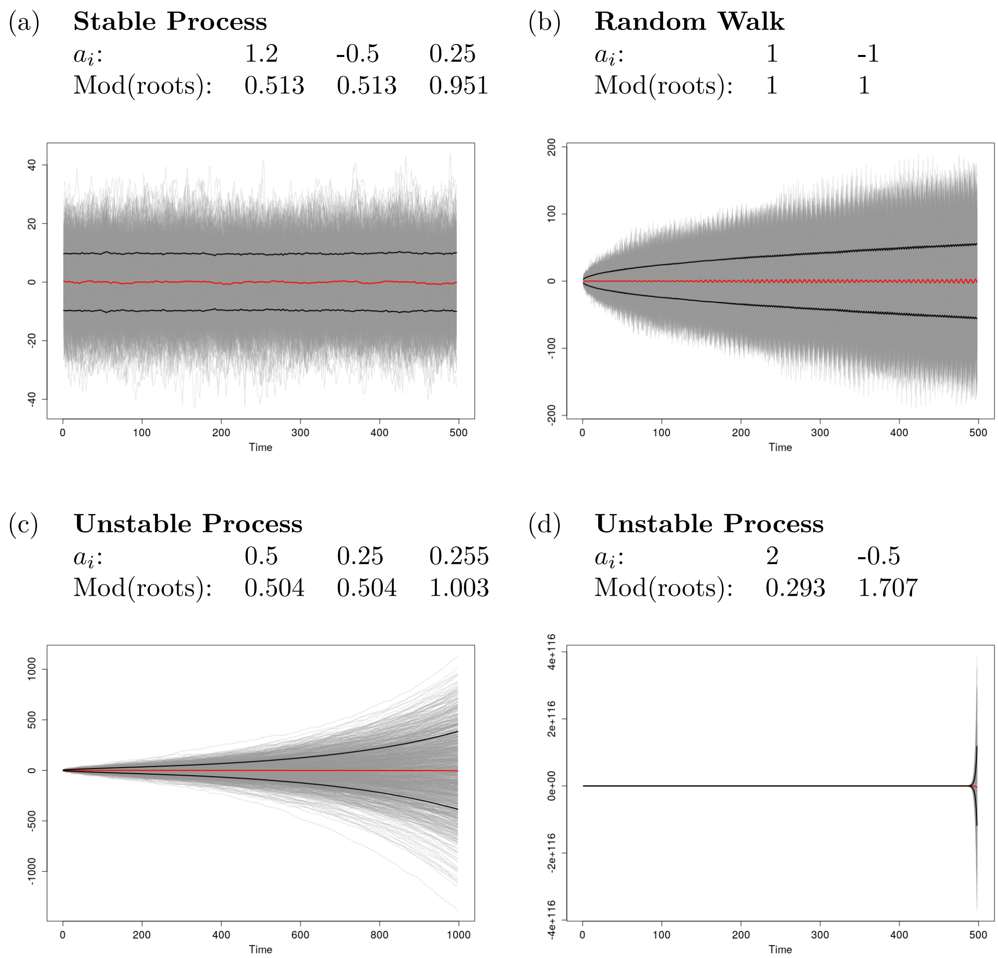


Figure 8.1: Different *AR(p)* processes with a $wn(0,\ 3)$ error term. For each combination 1000 samples were created and plotted (grey), along with the empirical mean (red) and standard deviation (black) over time.

R Output 8.2: Var results for ring two from section 4.2.

```
VAR Estimation Results:
=========================
Endogenous variables: Beech.harvest.d, Spruce.harvest.d
Deterministic variables: none
Sample size: 43
Log Likelihood: -148.351
Roots of the characteristic polynomial:
0.6479 0.4074 0.3896 0.06846
Call:
VAR(y = pt.data.r2.d, p = 2, type = "none", exogen = pt.exo.d)


Estimation results for equation Beech.harvest.d:
===============================================
Beech.harvest.d = Beech.harvest.d.l1 + Spruce.harvest.d.l1 + Beech.harvest.d.l2 +
     Spruce.harvest.d.l2 + Beech.price.d + Spruce.price.d + Vivian.Wiebke + Vivian.Wiebke.l1 +
     Vivian.Wiebke.l2 + Kyrill + Kyrill.l1 + Kyrill.l2

                  Estimate Std. Error t value Pr(>|t|)
Beech.harvest.d.l1 -0.45780  0.17488 -2.618  0.0136 *
Spruce.harvest.d.l1 0.03974  0.05980  0.665  0.5112
Beech.harvest.d.l2 0.02131   0.20011  0.107  0.9159
Spruce.harvest.d.l2 0.03877  0.03910  0.991  0.3292
Beech.price.d      0.01368   0.03486  0.392  0.6974
Spruce.price.d     0.01545   0.02227  0.694  0.4930
Vivian.Wiebke     -0.43494   1.06514 -0.408  0.6858
Vivian.Wiebke.l1  -0.32023   1.38912 -0.231  0.8192
Vivian.Wiebke.l2  -0.83077   1.25590 -0.661  0.5132
Kyrill            -1.09140   1.00287 -1.088  0.2849
Kyrill.l1         -1.90431   1.63734 -1.163  0.2537
Kyrill.l2         -1.59215   1.58990 -1.001  0.3244
---
Signif. codes: 0 ‘***‘ 0.001 ‘**‘ 0.01 ‘*‘ 0.05 ‘.‘ 0.1 ‘ ‘ 1


Residual standard error: 0.9551 on 31 degrees of freedom
Multiple R-Squared: 0.3167, Adjusted R-squared: 0.0522
F-statistic: 1.197 on 12 and 31 DF, p-value: 0.3278


Estimation results for equation Spruce.harvest.d:
================================================
Spruce.harvest.d = Beech.harvest.d.l1 + Spruce.harvest.d.l1 + Beech.harvest.d.l2 +
     Spruce.harvest.d.l2 + Beech.price.d + Spruce.price.d + Vivian.Wiebke + Vivian.Wiebke.l1 +
     Vivian.Wiebke.l2 + Kyrill + Kyrill.l1 + Kyrill.l2

                   Estimate Std. Error t value Pr(>|t|)
Beech.harvest.d.l1 -0.07463   0.49294 -0.151 0.88065
```

```
Spruce.harvest.d.l1 -0.24076  0.16857 -1.428 0.16322
Beech.harvest.d.l2  0.30207   0.56407  0.536 0.59612
Spruce.harvest.d.l2 0.21904   0.11022  1.987 0.05578 .
Beech.price.d      -0.16395   0.09826 -1.669 0.10527
Spruce.price.d      0.07404   0.06276  1.180 0.24708
Vivian.Wiebke       7.88688   3.00239  2.627 0.01327 *
Vivian.Wiebke.l1   -8.89548   3.91561 -2.272 0.03019 *
Vivian.Wiebke.l2    0.07158   3.54008  0.020 0.98400
Kyrill             19.66525   2.82686  6.957 8.35e-08 ***
Kyrill.l1         -12.96602   4.61526 -2.809 0.00852 **
Kyrill.l2          -5.81134   4.48155 -1.297 0.20430
---
Signif. codes: 0 ‘***‘ 0.001 ‘**‘ 0.01 ‘*‘ 0.05 ‘.‘ 0.1 ‘ ‘ 1


Residual standard error: 2.692 on 31 degrees of freedom
Multiple R-Squared: 0.819, Adjusted R-squared: 0.749
F-statistic: 11.69 on 12 and 31 DF, p-value: 2.186e-08



Covariance matrix of residuals:
               Beech.harvest.d Spruce.harvest.d
Beech.harvest.d         0.8832          -0.3021
Spruce.harvest.d       -0.3021           7.1733

Correlation matrix of residuals:
               Beech.harvest.d Spruce.harvest.d
Beech.harvest.d           1.00            -0.12
Spruce.harvest.d         -0.12             1.00
```

R Output 8.3: Var results for ring three from section 4.2.

```
VAR Estimation Results:
=========================
Endogenous variables: Beech.harvest.d, Spruce.harvest.d
Deterministic variables: none
Sample size: 42
Log Likelihood: -109.765
Roots of the characteristic polynomial:
0.7417 0.7417 0.6552 0.6552 0.223 0.223
Call:
VAR(y = pt.data.r3.d, p = 3, type = "none", exogen = pt.exo.d)


Estimation results for equation Beech.harvest.d:
==============================================
Beech.harvest.d = Beech.harvest.d.l1 + Spruce.harvest.d.l1 + Beech.harvest.d.l2 +
    Spruce.harvest.d.l2 + Beech.harvest.d.l3 + Spruce.harvest.d.l3 + Beech.price.d +
    Spruce.price.d + Vivian.Wiebke + Vivian.Wiebke.l1 + Vivian.Wiebke.l2 + Kyrill + Kyrill.l1
    + Kyrill.l2

                   Estimate Std. Error t value Pr(>|t|)
Beech.harvest.d.l1 -0.337588 0.143134 -2.359 0.025561 *
Spruce.harvest.d.l1 -0.170908 0.059158 -2.889 0.007379 **
Beech.harvest.d.l2 -0.529979 0.138946 -3.814 0.000690 ***
Spruce.harvest.d.l2 -0.091534 0.062882 -1.456 0.156612
Beech.harvest.d.l3 -0.190371 0.128854 -1.477 0.150724
Spruce.harvest.d.l3 -0.231791 0.060093 -3.857 0.000615 ***
Beech.price.d     -0.011315  0.025677 -0.441 0.662821
Spruce.price.d     0.009177  0.017413  0.527 0.602316
Vivian.Wiebke     -1.738032  0.794055 -2.189 0.037112 *
Vivian.Wiebke.l1  -0.461162  1.044677 -0.441 0.662285
Vivian.Wiebke.l2  -0.057033  0.917738 -0.062 0.950888
Kyrill             1.236447  0.821933  1.504 0.143698
Kyrill.l1          0.186962  0.887752  0.211 0.834723
Kyrill.l2          0.783533  0.872440  0.898 0.376794
---
Signif. codes: 0 ‘***‘ 0.001 ‘**‘ 0.01 ‘*‘ 0.05 ‘.‘ 0.1 ‘ ‘ 1


Residual standard error: 0.6978 on 28 degrees of freedom
Multiple R-Squared: 0.6277, Adjusted R-squared: 0.4415
F-statistic: 3.372 on 14 and 28 DF, p-value: 0.003002


Estimation results for equation Spruce.harvest.d:
===============================================
Spruce.harvest.d = Beech.harvest.d.l1 + Spruce.harvest.d.l1 + Beech.harvest.d.l2 +
    Spruce.harvest.d.l2 + Beech.harvest.d.l3 + Spruce.harvest.d.l3 + Beech.price.d +
```

```
    Spruce.price.d + Vivian.Wiebke + Vivian.Wiebke.l1 + Vivian.Wiebke.l2 + Kyrill + Kyrill.l1
    + Kyrill.l2

                  Estimate Std. Error t value Pr(>|t|)
Beech.harvest.d.l1 0.01545   0.35378  0.044 0.96549
Spruce.harvest.d.l1 -0.48541 0.14622 -3.320 0.00251 **
Beech.harvest.d.l2 -0.05517  0.34343 -0.161 0.87354
Spruce.harvest.d.l2 -0.17293 0.15542 -1.113 0.27532
Beech.harvest.d.l3 0.13360   0.31849  0.419 0.67807
Spruce.harvest.d.l3 0.10097  0.14853  0.680 0.50223
Beech.price.d     -0.10495   0.06346 -1.654 0.10935
Spruce.price.d     0.02761   0.04304  0.642 0.52635
Vivian.Wiebke      3.44782   1.96265  1.757 0.08990 .
Vivian.Wiebke.l1  -1.29677   2.58211 -0.502 0.61944
Vivian.Wiebke.l2   0.36851   2.26835  0.162 0.87211
Kyrill             6.15497   2.03156  3.030 0.00522 **
Kyrill.l1         -4.31072   2.19424 -1.965 0.05946 .
Kyrill.l2         -2.06062   2.15639 -0.956 0.34746
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 1.725 on 28 degrees of freedom
Multiple R-Squared: 0.6538, Adjusted R-squared: 0.4806
F-statistic: 3.776 on 14 and 28 DF, p-value: 0.001348



Covariance matrix of residuals:
               Beech.harvest.d Spruce.harvest.d
Beech.harvest.d        0.4866          0.1119
Spruce.harvest.d       0.1119          2.9710

Correlation matrix of residuals:
               Beech.harvest.d Spruce.harvest.d
Beech.harvest.d       1.00000         0.09309
Spruce.harvest.d      0.09309         1.00000
```

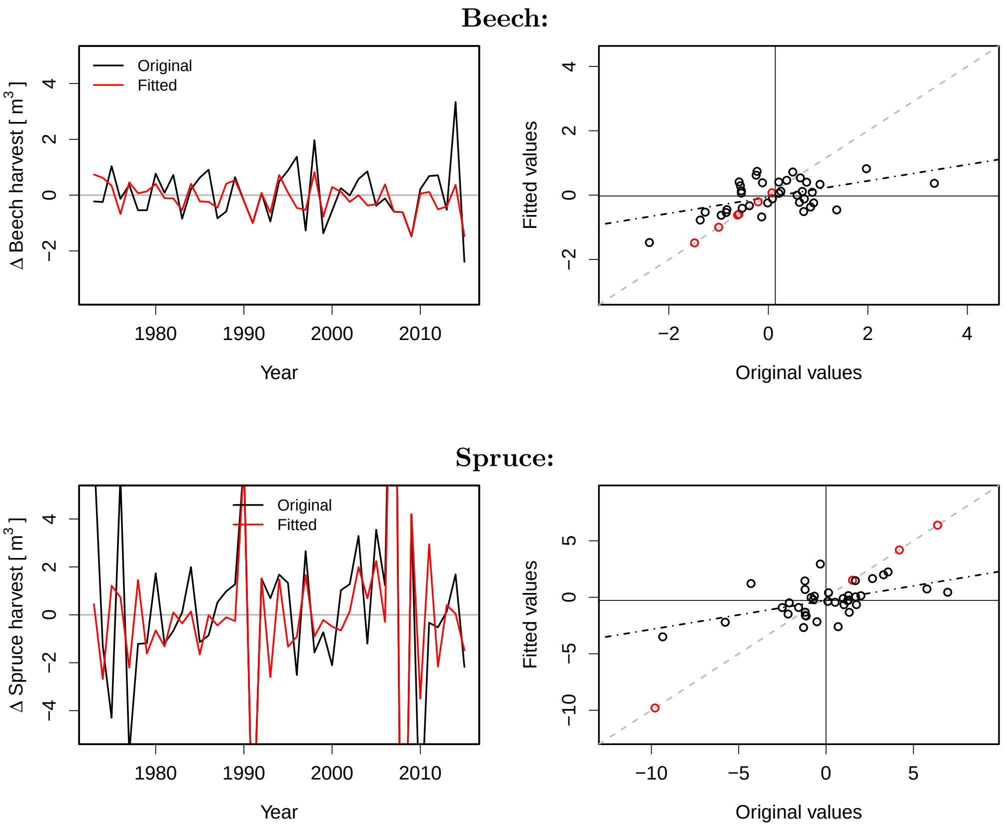


Figure 8.2: Fitted and original values for ring two in 4.2. Values are plotted together over time (left) and against each other (right). The dashed grey line has intercept zero and slope one, the dash-dotted line is a linear model of the displayed values, the thin black lines indicate the means of the respective values. In the right graphics, red dots indicate storm years.

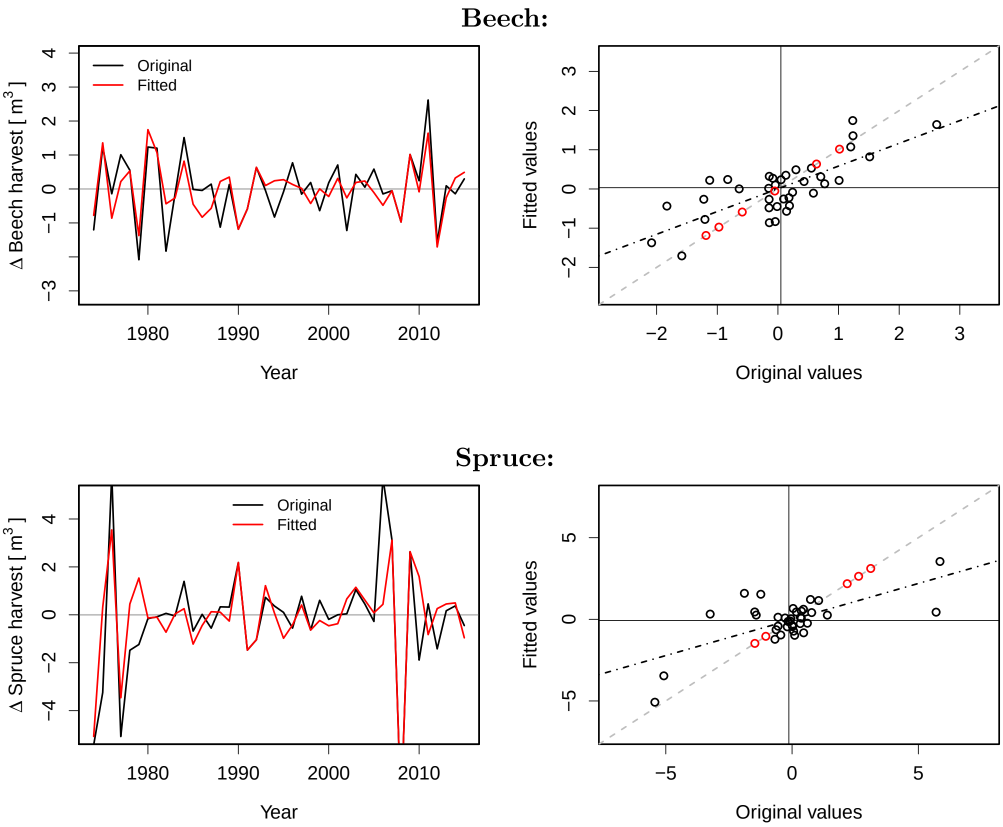


Figure 8.3: Fitted and original values for ring three in 4.2. Values are plotted together over time (left) and against each other (right). The dashed grey line has intercept zero and slope one, the dash-dotted line is a linear model of the displayed values, the thin black lines indicate the means of the respective values. In the right graphics, red dots indicate storm years.

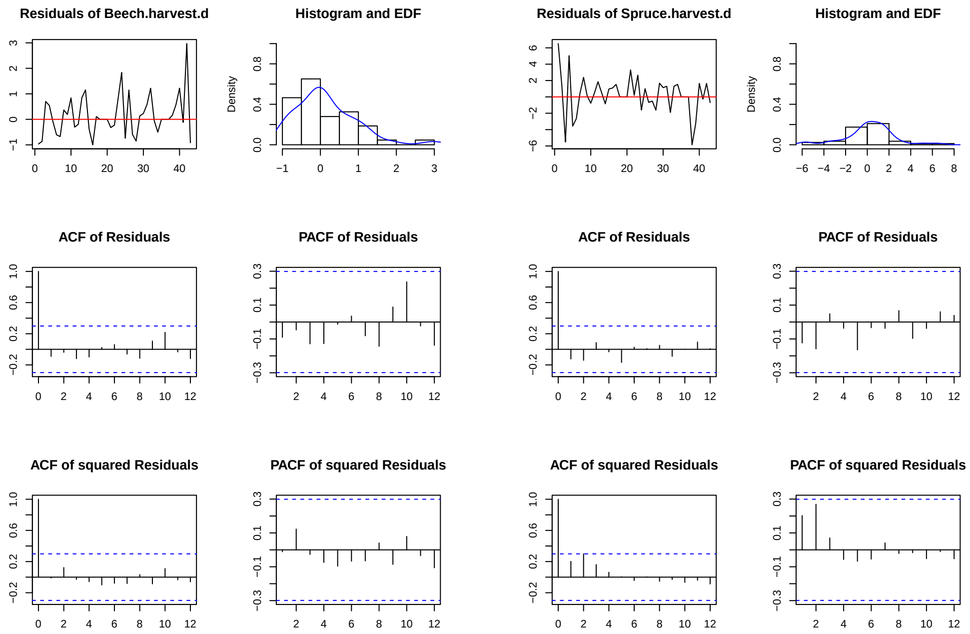


Figure 8.4: `plot()` function used on `serial.test()` for ring two in 4.2. Graphics for the Beech are in the two columns to the left, for the Spruce in the two columns to the right.

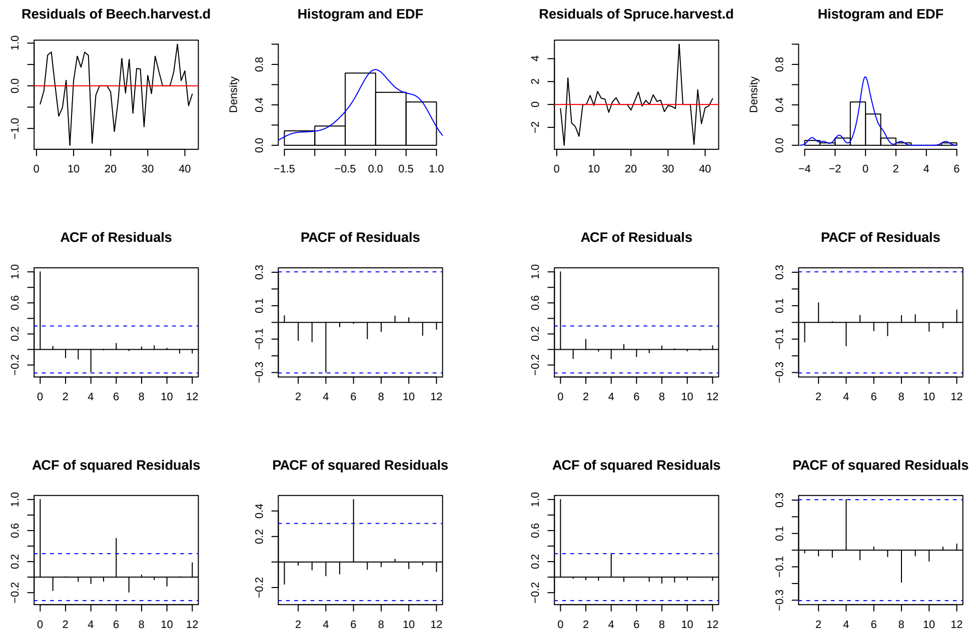


Figure 8.5: `plot()` function used on `serial.test()` for ring three in 4.2. Graphics for the Beech are in the two columns to the left, for the Spruce in the two columns to the right.

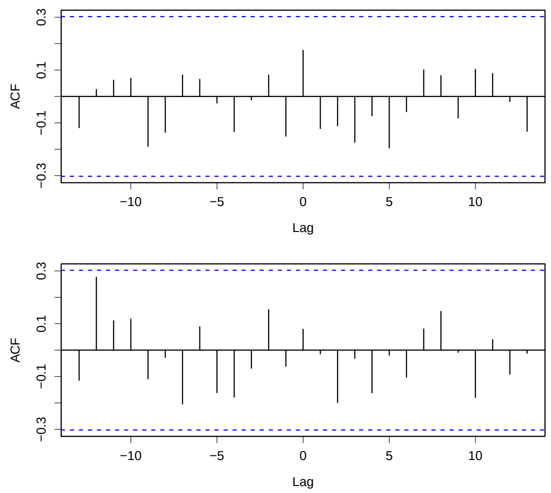


Figure 8.6: Cross-correlation function for the residuals of rings 2 and 3 in 4.2.

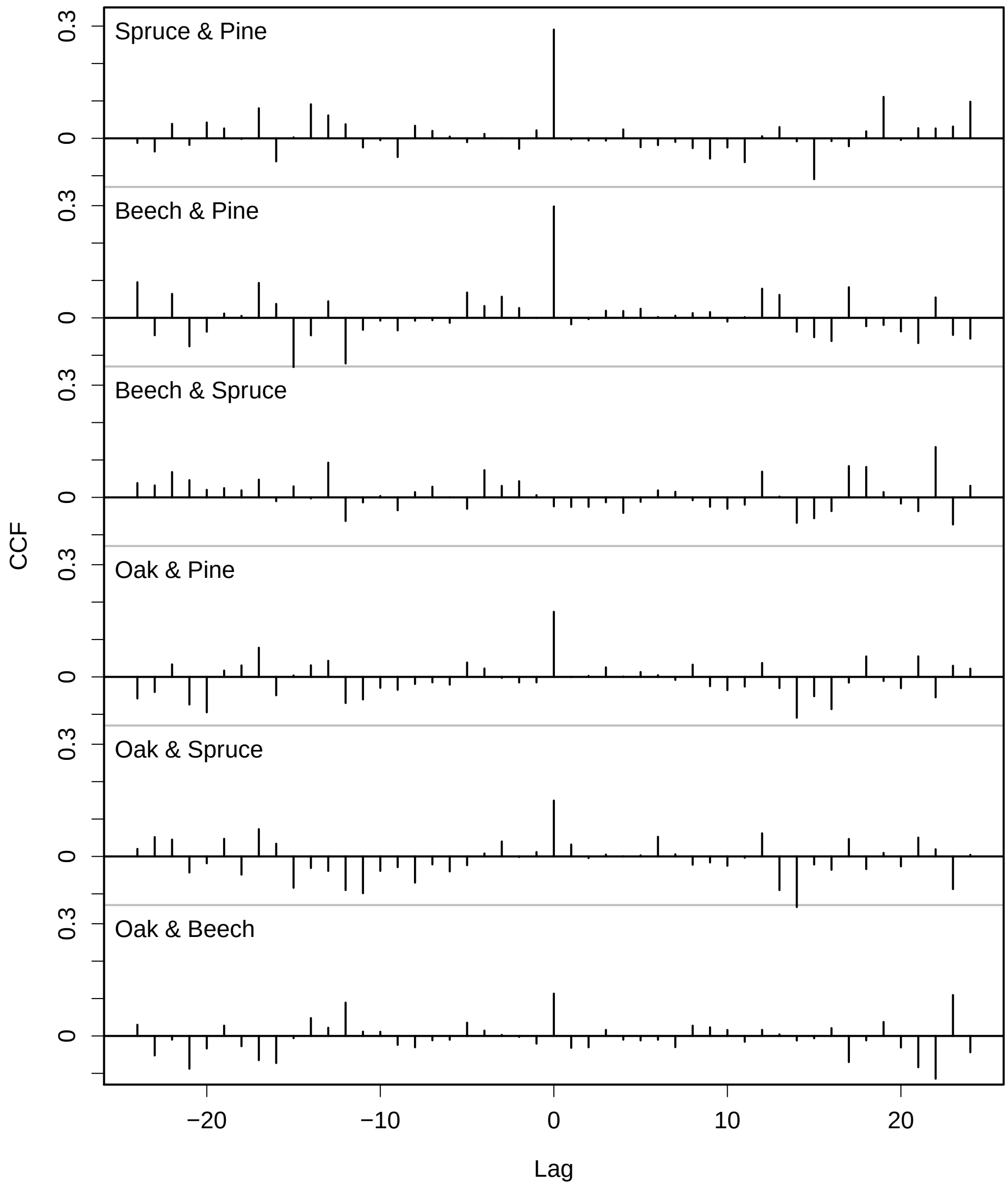


Figure 8.7: Cross-correlation functions for the residuals of the VAR estimated in 4.3.

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Meaning
ADF	Augmented Dickey-Fuller test
ACF	Autocorrelation function
AIC	Akaikes Information Criterion
AR	Autoregressive (Model)
ARCH	Autoregressive conditional heteroscedasticity
a.k.a.	Also known as
BIC	see SC
BVGL	German: “Betriebsvergleich Westfalen-Lippe”
	English: Business Comparison Westphalia-Lippe
cp.	Compare
CCF	Cross-correlation function
DESTATIS	German: Statistisches Bundesamt
	English: German Federal Statistical Office
e.g.	exempli gratia
eq.	Equation
FEVD	Forecast Error Variance Decomposition
fig.	Figure
FPE	Final Prediction Error
GDP	Gross Domestic Product
ha	hectare
HQ	Hannan and Quinn Information Criterion
i.e.	id est
IRF	Impulse Response Function
KPSS	Kwiatkowski-Phillips-Schmidt-Shin test
l.	line
ll.	lines
LS	Least squares (estimator)
MA	Moving Average representation
MSE	Mean Squared Error
NRW	German: Nordrhein-Westfalen
	English: North Rhine-Westphalia
p.	page
pp.	pages
SC	Schwarz Information Criterion
SVAR	Structural VAR
SVECM	Structural VECM

Continued on next page

Continued from previous page

tab.	Table
U.S.	United States (of America)
VAR	Vector Autoregressive Model
VECM	Vector Error Correction Model

Used Software

- **R**: Version 3.5.1 “Feather Spray”, 02.07.2018 (R Core Team, 2018)
- R-packages:
  - **gtools**: 3.8.1 (Warnes et al., 2018)
  - **tseries**: 0.10-44 (Trapletti and Hornik, 2018)
  - **urca**: 1.3-0 (Pfaff, 2008a)
  - **vars**: 1.5-2 (Pfaff, 2008b)

R Session Info

```
R version 3.5.1 (2018-07-02)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Debian GNU/Linux 9 (stretch)

Matrix products: default
BLAS: /usr/lib/libblas/libblas.so.3.7.0
LAPACK: /usr/lib/lapack/liblapack.so.3.7.0

locale:
[1] LC_CTYPE=de_DE.UTF-8    LC_NUMERIC=C             LC_TIME=de_DE.UTF-8
[4] LC_COLLATE=de_DE.UTF-8  LC_MONETARY=de_DE.UTF-8  LC_MESSAGES=de_DE.UTF-8
[7] LC_PAPER=de_DE.UTF-8    LC_NAME=C                LC_ADDRESS=C
[10] LC_TELEPHONE=C          LC_MEASUREMENT=de_DE.UTF-8 LC_IDENTIFICATION=C

attached base packages:
[1] stats     graphics grDevices utils     datasets methods   base

other attached packages:
[1] gtools_3.8.1     tseries_0.10-44 vars_1.5-3      lmtest_0.9-36
[5] strucchange_1.5-1 sandwich_2.4-0 zoo_1.8-1       MASS_7.3-50
[9] urca_1.3-0

loaded via a namespace (and not attached):
[1] quadprog_1.5-5 lattice_0.20-35 grid_3.5.1  nlme_3.1-137   quantmod_0.4-13
[6] curl_3.2       TTR_0.23-3      xts_0.10-2    tools_3.5.1    yaml_2.1.19
[11] compiler_3.5.1
```

References

Abildtrup, J., Helles, F., Holten-Andersen, P., Larsen, J. F., and Thorsen, B. J. (1999). *Modern Time Series Analysis in Forest Products Markets (Forestry Sciences)*. Kluwer Academic Publishers.

Akaike, H. (1981). Likelihood of a model and information criteria. *Journal of Econometrics*, 16(1):3 – 14.

Alavalapati, J. R. R., Adamowicz, W. L., and Luckert, M. K. (1997). A Cointegration Analysis of Canadian Wood Pulp Prices. *American Journal of Agricultural Economics*, 79(3):975 – 986.

Banerjee, A., Dolado, J. J., Galbraith, J. W., and Hendry, D. (1993). Co-integration, Error Correction, and the Econometric Analysis of Non-Stationary Data. *Oxford University Press*.

Baur, P., Bernath, K., Holthausen, N., and Roschewitz, A. (2003). *LOTHAR – Ökonomische Auswirkungen des Sturms Lothar im Schweizer Wald, Teil I. – Einkommens- und Vermögenswirkungen für die Waldwirtschaft und gesamtwirtschaftliche Beurteilung des Sturms*. Number 157 in Umwelt-Materialien. Bundesamt für Umwelt, Wald und Landschaft, Bern.

Beinhofer, B. (2010). *Zur Anwendung der Portfoliotheorie in der Forstwissenschaft : finanzielle Optimierungsansätze zur Bewertung von Diversifikationseffekten*. Number 208 in Schriftenreihe des Zentrum Wald-Forst-Holz Weihenstephan. Zentrum Wald-Forst-Holz Weihenstephan.

Box, G. E. P. and Jenkins, G. M. (1970). *Time series analysis: forecasting and control*. Holden-Day series in time series analysis. Holden-Day, San Francisco, California.

Breusch, T. S. (1978). Testing for autocorrelation in dynamic linear models. *Australian Economic Papers*, 17(31):334 – 355.

Buongiorno, J. (1996). Forest sector modeling: a synthesis of econometrics, mathematical programming, and system dynamics methods. *International Journal of Forecasting*, 12(3):329 – 343.

Carstensen, K., Elstner, S., and Paula, G. (2013). How Much Did Oil Market Developments Contribute to the 2009 Recession in Germany? *Scandinavian Journal of Economics*, 115(3):695 – 721.

Cheung, Y.-W. and Lai, K. S. (1993). Finite-sample sizes of Johansen's likelihood ratio test for cointegration. *Oxford Bulletin of Economics and Statistics*, 55(3):313 – 328.

Dög, M., Möhring, B., von Wrede, E., and Ellermann, W. (2017). Privatwald-Betriebsvergleich Westfalen-Lippe im Jahr 2016. *AFZ-DerWald*, 23:31 – 33.

Edgerton, D. and Shukur, G. (1999). Testing autocorrelation in a system perspective testing autocorrelation. *Econometric Reviews*, 4:343 – 386.

Engle, R. and Granger, C. (1987). Co-integration and error correction: Representation, estimation and testing. *Econometrica*, 55(2):251 – 276.

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50(4):987 – 1007.

Fahrmeir, L., Kneib, T., and Lang, S. (2009). *Regression: Modelle, Methoden und Anwendungen (Statistik und ihre Anwendungen)*. Springer.

Fahrmeir, L., Kneib, T., Lang, S., and Marx, B. (2013). *Regression: Models, Methods and Applications*. Springer.

Godfrey, L. G. (1978). Testing for Higher Order Serial Correlation in Regression Equations when the Regressors Include Lagged Dependent Variables. *Econometrica*, 46(6):1303 – 1310.

Gospodinov, N., Herrera, A. M., and Pesavento, E. (2013). Unit Roots, Cointegration, and Pretesting in Var Models. *Advances in Econometrics*, 32:81 – 115.

Hannan, E. J. and Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society. Series B (Methodological)*, 41(2):190 – 195.

Hansen, H., Larsen, J. F., Thorsen, B. J., and Abildtrup, J. (1999). Modelling the Danish Demand for Roundwood Imports — a Cointegration and Common Trends Analysis. In *Modern Time Series Analysis in Forest Products Markets*, pages 85–102. Kluwer Academic Publishers.

Herwartz, H. and Plödt, M. (2016). The macroeconomic effects of oil price shocks: Evidence from a statistical identification approach. *Journal of International Money and Finance*, 61:30–44. [Online; accessed 26. Feb. 2019].

Hetemäki, L., Hänninen, R., and Toppinen, A. (2004). Short-Term Forecasting Models for the Finnish Forest Sector: Lumber Exports and Sawlog Demand. *Forest Science*, 50(4):461 – 472.

Hetemäki, L. and Mikkola, J. (2005). Forecasting Germany's Printing and Writing Paper Imports. *Forest Science*, 51(5):483 – 497.

Jarque, C. M. and Bera, A. K. (1987). A Test for Normality of Observations and Regression Residuals. *International Statistical Review / Revue Internationale de Statistique*, 55(2):163 – 172.

Jenkins, S. P., Brandolini, A., Micklewright, J., and Nolan, B. (2013). *The Great Recession and the Distribution of Household Income*. OUP Oxford.

Jennings, S., Adamowicz, W., and Constantino, L. (2011). The Canadian lumber industry and the macroeconomy: a vector autoregression analysis. *Canadian Journal of Forest Research*.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12(2-3):231 – 254.

Johansen, S. (1991). Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica*, 59(6):1551 – 1580.

Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press.

Johansen, S. and Jusélius, K. (1990). Maximum likelihood estimation and inference on cointegration – with applications to the demand for money. *Oxford Bulletin of Economics & Statistics*, 52(2):169 – 210.

Kilian, L. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review*, 99(3):1053 – 1069.

Kilian, L. and Lütkepohl, H. (2017). *Structural Vector Autoregressive Analysis (Themes in Modern Econometrics)*. Cambridge University Press.

Kolo, H. and Tzanova, P. (2017). Forecasting the German forest products trade: A vector error correction model. *Journal of Forest Economics*, 26:30 – 45.

Koop, G., Pesaran, M. H., and Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics*, 74(1):119 – 147.

Kropp, J., Holsten, A., Lissner, T., Roithmeier, O., Hattermann, F., Huang, S., Rock, J., Wechsung, F., Lüttger, A., Pompe, S., Kühn, I.and Costa, L., Steinhäuser, M., Walther, C., Klaus, M., Ritchie, S., and Metzger, M. (2009). *Klimawandel in Nordrhein-Westfalen – Regionale Abschätzung der Anfälligkeit ausgewählter Sektoren (Abschlussbericht)*. Potsdam-Institut für Klimafolgenforschung (PIK), Potsdam.

Kuuluvainen, J., Korhonen, J., Xu, D., and Toppinen, A. (2018). Price integration for domestic and imported sawlogs and pulpwood in Finland: an update. *Scandinavian Journal of Forest Research*, 33(1):71 – 80.

Kwiatkowski, D., Phillips, P., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, 54(1-3):159 – 178.

Lange, A., Dalheimer, B., Herwartz, H., and Maxand, S. (2018). *svars: Data-Driven Identification of SVAR Models*. R package version 1.2.2.

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

Lütkepohl, H. (2007). *New introduction to multiple time series analysis*. Springer, Berlin, 1 edition.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1):77 – 91.

Markowitz, H. M. (2008). *Portfolio Selection: Effiziente Diversifikation von Anlagen*. FinanzBuch Verlag.

McKillop, W. (1967). Supply and demand for forest products—an econometric study. *Hilgardia*, 38(1):1 – 132.

Nellen, B. (2012). Preisentwicklung von 1919 bis 2010 - Nadel-Sägerundholz. *Wald und Holz*, (1):14 – 15.

Osterwald-Lenum, M. (1992). A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics. *Oxford Bulletin of Economics and Statistics*, 54(3):461 – 72.

Parajuli, R. and Zhang, D. (2016). Price Linkages between Spot and Futures Markets for Softwood Lumber. *Forest Science*, 62(5):482 – 489.

Paulsen, J. (1984). Order determination of multivariate autoregressive time series with unit roots. *Journal of Time Series Analysis*, 5(2):115 – 127.

Pfaff, B. (2008a). *Analysis of Integrated and Cointegrated Time Series with R*. Springer, New York, second edition.

Pfaff, B. (2008b). VAR, SVAR and SVEC Models: Implementation Within R Package vars. *Journal of Statistical Software*, 27(4):1 – 32.

Pringle, S. (1954). *An econometric analysis of the demand for newsprint in the United States*. PhD thesis, Syracuse University.

Quinn, B. G. (1980). Order determination for a multivariate autoregression. *Journal of the Royal Statistical Society. Series B (Methodological)*, 42(2):182 – 185.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ripberger, M., editor (1994). *Dokumentation der Sturmschäden 1990*. Number 75 in Schriftenreihe der Landesforstverwaltung Baden-Württemberg. Ministerium für Ländlichen Raum, Ernährung, Landwirtschaft und Forsten.

Schelhaas, M.-J., Nabuurs, G.-J., and Schuck, A. (2003). Natural disturbances in the European forests in the 19th and 20th centuries. *Global Change Biology*, 9(11):1620 – 1633.

Schmidt, M., Hanewinkel, M., Kändler, G., Kublin, E., and Kohnle, U. (2010). An inventory-based approach for modeling single-tree storm damage – experiences with the winter storm of 1999 in southwestern germany. *Canadian Journal of Forest Research*.

Schwarz, G. (1978). Estimating the Dimension of a Model. *Annals of Statistics*, 6(2):461 – 464.

Serepka, R. (2012). Analyzing and modelling exchange rate data using VAR framework. Master's thesis, KTH Royal institute of technology.

Shahi, C. K. and Kant, S. (2009). Cointegrating relationship and the degree of market integration among the North American softwood lumber product markets. *Canadian Journal of Forest Research*, 39(11):2129 – 2137.

Shumway, R. H. and Stoffer, D. S. (2006). *Time series analysis and its applications: with R examples*. Springer texts in statistics. Springer, New York, NY, 2 edition.

Sims, C. A. (1981). An autoregressive index model for the us 1948-1975. In Kmenta, J. and Ramsey, J., editors, *Large-Scale Macro-Econometric Models*, pages 283 – 327. Amsterdam: North-Holland.

Toppinen, A. (1998). Incorporating cointegration relations in a short-run model of the Finnish sawlog market. *Canadian Journal of Forest Research*, 28(2):291 – 298.

Toppinen, A. and Kuuluvainen, J. (2010). Forest sector modelling in Europe—the state of the art and future research directions. *Forest Policy and Economics*, 12(1):2 – 8.

Trapletti, A. and Hornik, K. (2018). *tseries: Time Series Analysis and Computational Finance*. R package version 0.10-44.

Warnes, G. R., Bolker, B., and Lumley, T. (2018). *gtools: Various R Programming Tools*. R package version 3.8.1.

Zwirglmaier, K. (2012). *Ursachen saisonaler Holzpreisschwankungen: Eine Analyse des deutschen Holzmarktes*. Gabler Verlag.

Hiermit versichere ich gemäß § 7 Abs. 5 der Master-Prüfungsordnung vom 23.09.2010, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Jan Schick

Göttingen, den 21.03.2019