DIPLOMARBEIT

Cost and Time Series Analysis of the Austrian Mechanical and Plant Engineering Industry

ausgeführt zum Zwecke der Erlangung des Akademischen Grades eines
DIPLOM-INGENIEURS

unter der Leitung von
Univ.Prof. Mag.rer.soc.oec. Dr.rer.soc.oec. Walter S.A. SCHWAIGER, MBA
E330 Institut für Managementwissenschaften

eingereicht an der Technischen Universität Wien
Fakultät für Maschinenwesen und Betriebswissenschaften

von
Matthias ONDRA, BSc BSc
0925005

Einverständniserklärung

Ich habe zur Kenntnis genommen, dass ich zur Drucklegung meiner Arbeit unter der Bezeichnung

Diplomarbeit

nur mit Bewilligung der Prüfungskommission berechtigt bin.

Ich erkläre weiters an Eides statt, dass ich meine Diplomarbeit nach den anerkannten Grundsätzen für wissenschaftliche Abhandlungen selbstständig ausgeführt habe und alle verwendeten Hilfsmittel, insbesondere die zugrunde gelegte Literatur genannt habe.

Weiters erkläre ich, dass ich dieses Diplomarbeitsthema bisher weder im In- noch im Ausland (einer Beurteilerin/ einem Beurteiler zur Begutachtung) in irgendeiner Form als Prüfungsarbeit vorgelegt habe und dass diese Arbeit mit der vom Begutachter beurteilten Arbeit übereinstimmt

January 23, 2016 Matthias Ondra

Danksagung

An dieser Stelle möchte ich mich bei all jenen bedanken, die mich bei der Fertigstellung meiner Diplomarbeit unterstützt haben.

Mein Dank gilt vor allem meinem Betreuer Prof.Dr.Walter Schwaiger, welcher durch seine hervorragenden Vorlesungen mein Interesse an dem Thema geweckt hat. Prof.Dr. Schwaiger hatte jederzeit ein offenes Ohr, um etwaige Probleme zu besprechen. Weiters möchte ich mich für die Freiheit diverse Themenstellungen zu bearbeiten, bedanken.

Mein besonderer Dank gilt meinen Eltern Andrea und Rudolf Ondra, die mir mein Studium erst ermöglicht haben und mich jederzeit in all meinen Entscheidungen unterstützt haben.

Besonders Danken möchte ich ebenso MMag.Roland Vögel von der WKO, für die sehr gute Zusammenarbeit in diesem Projekt.

Matthias Ondra

Contents

List of Figures

List of Tables

Abstract

Abstract. This thesis deals with the problem of analysing the structure of the austrian mechanical and plant engineering industry, especially with the problem of calibrating a cost function and setting up time series models for an industry analysis. The first part is based on the previous work of [Ivaz, 2014] where a cost function, based on a Leontief production function was calibrated. By considering a CES-production function, a more general approach is given. It is proven that other common production functions as for example the Leontief, Cobb-Douglas or linear production functions are special cases of the CES-production function.

Based on this CES-production function the analytic expression of the cost function $\mathcal{C}$ is derived, by minimizing the aggregate costs subject to a certain amount of outcome, modelled by the CES-production function. Given the data, the parameters of the cost function are estimated by fitting the cost function using an ordinary least squares model.

Since this method is very unstable, from a numerical point of view, another more application oriented approach based on [Kmenta, 1967] is given. More precisely, a Taylor approximation of second order around an initial point is performed and then fitted to the data points in order to estimate these parameters.

The second part of this thesis deals with the problem of performing an industry analysis. Therefore time series models of the different data are used, in order to give a prediction about the future. The statistics program `R` is used to give a compact method of selecting and estimating ARIMA(p,d,q) and VAR(p) models, on which the prediction of the future is based. Moreover this vector autoregressive model is used to investigate the dynamic of the system, using an impulse response functions. This theory allows us to model different price shocks in one variable and simulating the model to study the influence on the other variables.

Zusammenfassung

ZUSAMMENFASSUNG. Die vorliegende Diplomarbeit befasst sich mit dem Thema, den Sektor der Maschinenbau und Anlagenbranche zu untersuchen und analysieren um eine Kostenfunktion aufzustellen. Weiters wird eine Zeitreihenanalyse verschiedener Komponenten durchgeführt. Der erste Teil basiert auf der Arbeit von [Ivaz, 2014], in welcher eine lineare Kostenfunktion basierend auf einer Leontief Produktionsfunktion kalibriert wurde. In dieser Arbeit, wird ein allgemeinerer Ansatz untersucht, indem die Kostenfunktion basierend auf einer CES Produktionsfunktion kalibriert wird. Weiters wird gezeigt, dass einige bekannte Produktionsfunktionen, wie die Lenotief, Cobb-Douglas oder die lineare Produktionsfunktion als Spezialfälle der CES Produktionsfunktion gesehen werden können.

Basierend auf dieser CES Produktionsfunktion, wird der analytische Ausdruck der Kostenfunktion $\mathcal{C}$ über das Problem die aggregierten Kosten unter der Nebenbedingung einer bestimmten Ausbringungsmenge, welche über die CES Produktionsfunktion modelliert wird, zu minimieren, berechnet. Die Parameter dieser bestimmten Funktion werden dann über die Methode der Kleinsten-Quadrate-Schätzer bestimmt.

Aufgrund der Tatsache, dass diese Methode, von einem numerischen Standpunkt gesehen, sehr instabil ist, wird eine weiterer, anwendungsorientierter Ansatz beschriebn, der auf die Arbeit von [Kmenta, 1967] basiert. Dabei wird die Kostenfunktion durch ein Taylorpolynom zweiter

Ordnung um einem bestimmten Punkt angenährt. Unter Verwendung dieses Funktionstypen werden dann die Parameter, basierend auf den historischen Daten bestimmt.

Der zweite Teil dieser Arbeit befasst sich mit dem Problem einer Branchenanalyse des Maschinenbausektors. Dazu wurden verschiedne Zeitreihenmodelle relevanter Größen aufgestellt. Um eine kompakte Methode darzulegen, wird die Statistik Software `R` verwendet. Hierbei wird ein ARIMA(p,d,q) oder VAR(p) Model ausgewählt und deren Parameter geschätzt. Aufbauend auf diesem Modell werden dann die Zukünftigen Werte geschätzt. Weiters wird diese Vektor autoregressives Modell verwendet, um die Dynamik des Systems zu untersuchen. Dabei wird die Theorie der Impulse Response functions benützt um unterschiedliche Preisschocks in einer Variable zu modellieren. Ab diesem Zeitpunkt wird das System dann simuliert, um die Auswirkung dieser Schocks, im Vergleich zum ungeschockten Modell, zu untersuchen.

Part I.

Constructing the cost function

CHAPTER 1

Fundamentals of production theory

In the first chapter we will give an introduction to the theoretical background of production functions, in the context of an input and output oriented point of view. Therefore different terms and methods are defined, where the concept of elasticity will play a major role. A general class of production functions is generated by the property of a constant elasticity of substitution, for all input vectors. We will proof that this class of functions include other common production functions as for example the Cobb-Douglas, Leontief or linear production functions and is therefore very useful in the theory of flexible modelling of production functions.

1.1. Introduction

To describe the technical boundaries of the production of certain goods in a manufacture or a certain sector, we will make use of the commonly known production function. To do this, we will first of all define production in a general way, as the process of transforming inputs into outputs [Geoffrey et al., 2001, p.126]. We will therefore choose a process oriented approach, where the production function maps the amount of input factors of the manufacture to its output [Wied-Nebbeling and Schott, 2006, p.104].

According the notation, the amount of the i-th input factors will be denoted with x_i. Of course we want the input factors, as well as the output factors, to be positive real numbers, which restricts the domain, as well as the possible values of the production function.

1.1 Definition. [Geoffrey et al., 2001, p.127] Let $(x_i, \ldots, x_n) = \mathbf{x}$ be the amount of input factors. The function $f : \mathbb{R}^n_+ \to \mathbb{R}_+, (x_1, \ldots, x_n) \mapsto f(x_1, \ldots, x_n) = f(\mathbf{x})$, mapping the amount of input factors to the amount of output that can be produced, is called a production function.

With this definition given above, we are now able to describe and classify production functions, according to their mathematical properties. Doing so, we additionally require further characteristics.

1.2 Assumption. [Geoffrey et al., 2001, p.127] Let $f : \mathbb{R}^n_+ \to \mathbb{R}_+$ be a production function. Then f is continuous, strictly increasing, strictly quasiconcave and $f(\mathbf{0}) = 0$ holds .

Continuity of f ensures that small changes in the amount of input factors result in a small change of the amount of output. We don't want that small changes of the amount of input factors result in a 'dramatic' change of the amount of output, as it is in case of discontinuity at a certain point. According to the mathematical definition of continuity it is true that $\lim_{\mathbf{x} \to \mathbf{x}_0} f(\mathbf{x}) = f(\mathbf{x}_0)$ [Geoffrey et al., 2001, p.127].

Because of the logic argument that more amount of input can't effect a smaller amount of output we furthermore require f to generate strictly more output when strictly increasing the amount of input, [Geoffrey et al., 2001, p.127]. Thus it is always true, that for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n_+$ with $\mathbf{x} > \mathbf{y}$, which means that $x_i > y_i$ for all $i = 1, \ldots, n$, it follows that also $f(\mathbf{x}) > f(\mathbf{y})$ holds.

Considering the assumption of a strict quasiconcave production function, we will first take a look at the definition. A function $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ is strictly quasiconcave if and only if, for all $\mathbf{x}, \mathbf{y} \in D$, $\mathbf{x} \neq \mathbf{y}$: $f(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) > \min\{f(\mathbf{x}), f(\mathbf{y})\}$, for all $t \in (0, 1)$ [Geoffrey et al., 2001, p.541]. To economically interpret this situation, consider the special case of only two input factors, e.g. capital and work and two input combinations $\mathbf{x}, \mathbf{y}$. This input combinations are chosen, such that $\mathbf{x}$ has an extremely high amount of work and an extremely low amount of capital. The other input combination $\mathbf{y}$ is chosen such that it has an extremely high amount of capital and an extremely low amount of work, respectively. In this situation, one is always able to form an input factor $\mathbf{z} = \lambda\mathbf{x} + (1 - \lambda)\mathbf{y}$, for some $t_1 \in (0, 1)$. Choosing $t_1 = 1/2$ for example, results in an equally weighted combination of $\mathbf{x}$ and $\mathbf{y}$. In case of a strictly quasiconcave function we therefore find that choosing the input vector $\mathbf{z}$, differing from the extreme case, yields a higher output than one of the extreme cases $f(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) > \min\{f(\mathbf{x}), f(\mathbf{y})\}$ [Geoffrey et al., 2001, p.127]. Summing up, this assumption demands some sort of economic rationality.

The assumption $f(\mathbf{0}) = 0$ is just stated to make clear, that outputs can't be created using no inputs, which is obvious anyway.

Additionally, the production function is sometimes also considered to be differentiable. Thus we can compute its partial derivative with respect to a certain variable x_i and interpret this number as the rate at which output changes per additional unit of input i employed, which also motivates the following definition.

1.3 Definition. [Geoffrey et al., 2001, p.127], [Wied-Nebbeling and Schott, 2006, p.108] Let $f : \mathbb{R}^n_+ \to \mathbb{R}_+$ be a production function. The marginal product of the i-th input is defined as

$$\mathrm{MP}_i(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial x_i} \tag{1.1}$$

and the average productivity of the i-th input factor as

$$\mathrm{AP}_i(\mathbf{x}) = \frac{f(\mathbf{x})}{x_i}. \tag{1.2}$$

Now consider any fixed level of output represented by a positive real number, $c \in \mathbb{R}_+$. Obviously, the case might occur, that there are many different combinations of inputs, creating the same amount of output. This is the case, when one can easily substitute some of the input factors by others available. All possible combinations of input vectors leading to this amount of outcome, are given by the set $\{\mathbf{x} \in \mathbb{R}^n_+ : f(\mathbf{x}) = c\}$ [Schmidt, 2013, p.369]. In most cases this set is going to be a curve, defined in the space $\mathbb{R}^n_+$, such that we can state that this set can be represented by a curve, showing all possible combinations of inputs that yield the same output

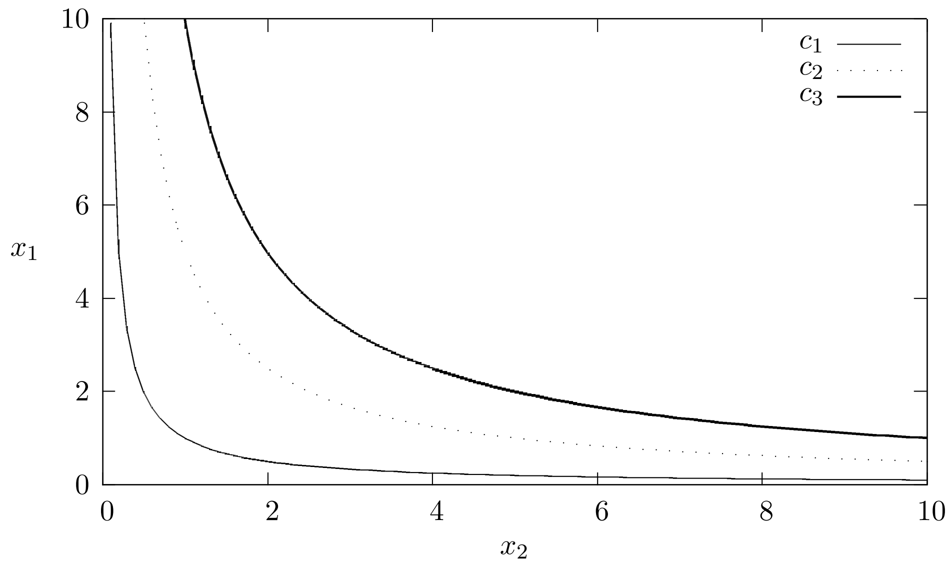


FIGURE 1.1.: Different c-level-isoquants for the production function $f(x_1, x_2) = \sqrt{x_1}\sqrt{x_2}$ with $c_1 = 1$, $c_2 = \sqrt{5}$ and $c_3 = \sqrt{10}$.

[Pindyck and Rubinfeld, 2013, p.216].

1.4 DEFINITION. Let $f : \mathbb{R}^n_+ \to \mathbb{R}_+$ be a production function and $c \in \mathbb{R}_+$. The c-level-isoquant is defined as the set $\{\mathbf{x} \in \mathbb{R}_+ : f(\mathbf{x}) = c\}$.

An illustrative example of different level-isoquants is given in Figure 1.1, where we can see that one is free to choose the input combination, to get a certain amount of output.

1.5 DEFINITION. [Uebe, 2013, p.21] Let $f : \mathbb{R}^n_+ \to \mathbb{R}_+$ be a differentiable production function and $c \in \mathbb{R}_+$. Consider a certain c-level-isoquant, choose two input factors $i, j = 1, \ldots, n$ and fix every component x_k of the input vector, where $k = 1, \ldots, n$, $k \neq i$ and $k \neq j$ except for the two x_i and x_j. The marginal rate of substitution of the factors i and j is defined as

$$\mathrm{MRS}_{ij}(\mathbf{x}) = \frac{\frac{\partial f(\mathbf{x})}{\partial x_i}}{\frac{\partial f(\mathbf{x})}{\partial x_j}}.$$

Trying to give meaning to the definition stated above, we first note that we only consider a certain isoquant and only two variable factors x_i and x_j. Thus we find that $f(x_1, \dots, x_n) = f(x_i, x_j) = c$, since these are the only variables left. Taking the total differential of both sides of this equation yields

$$\mathrm{d}f = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k} \mathrm{d}x_k = \frac{\partial f}{\partial x_i} \mathrm{d}x_i + \frac{\partial f}{\partial x_j} \mathrm{d}x_j = 0. \tag{1.3}$$

Reformulating (1.3) yields

$$\mathrm{MRS}_{ij} = \frac{\frac{\partial f(\mathbf{x})}{\partial x_i}}{\frac{\partial f(\mathbf{x})}{\partial x_j}} = -\frac{\mathrm{d}x_j}{\mathrm{d}x_i},$$

which can be precisely explained with the help of the theorem of implicit functions and simply represents the negative slope of the c-level-isoquant. This is due to the fact that one can find an expression of the form $x_j(x_i)$ based on the equation $f(x_1, x_2) = c$, under certain conditions.

A further remark may be added to the fact, that the slope of the c-level-isoquant does not depend on the chosen constant c itself, as it can be seen from (1.3) and Figure 1.1.

1.2. Homogeneous production functions

1.6 Definition. [Schmidt, 2013, p.367] A function $f : \mathbb{R}_+^n \to \mathbb{R}_+$ is called homogeneous of degree $\alpha \in \mathbb{R}$ if for all $\lambda \in \mathbb{R}_+ \setminus \{0\}$ and $\mathbf{x} \in \mathbb{R}_+^n : f(\lambda\mathbf{x}) = \lambda^\alpha f(\mathbf{x})$ holds.

Homogeneous production functions show a very pleasant characteristic, commonly used to describe some economic features of a production function in a very elegant way. Increasing the amount of input by a factor λ yields an increase of the amount of output by a factor λ^α. Thus, knowing the degree of homogenity makes it possible to determine the change of output, only with the help of the change of the input.

1.7 Example. [Schmidt, 2013, p.368] Consider production functions of the form $f(\mathbf{x}) = \left(\sum_{i=1}^n \alpha_i x_i^{-\rho}\right)^{-\frac{\nu}{\rho}}$ and $g(\mathbf{x}) = \prod_{i=1}^n x_i^{\alpha_i}$. Since

$$f(\lambda\mathbf{x}) = \left(\sum_{i=1}^{n} \alpha_i (\lambda x_i)^{-\rho}\right)^{-\frac{\nu}{\rho}} = \lambda^\nu \sum_{i=1}^{n} \left(\alpha_i x_i^{-\rho}\right)^{-\frac{\nu}{\rho}} = \lambda^\nu f(\mathbf{x})$$

and

$$g(\lambda\mathbf{x}) = \prod_{i=1}^{m}(\lambda x_i)^{\alpha_i} = \prod_{i=1}^{m}\lambda^{\alpha_i}\prod_{i=1}^{m}x_i^{\alpha_i} = \lambda^{\sum_{i=1}^{m}\alpha_i}\prod_{i=1}^{m}x_i^{\alpha_i} = \lambda^{\sum_{i=1}^{m}\alpha_i}\prod_{i=1}^{m}g(\mathbf{x})$$

holds, f is homogeneous of degree ν and g is homogeneous of degree $\sum_{i=1}^{m}\alpha_i$.

1.8 Definition. [Schmidt, 2013, p.368] Let $f : \mathbb{R}_+^n \to \mathbb{R}_+$ be a production function and $\gamma \in (1, \infty)$. If the following condition holds for all $\mathbf{x} \in \mathbb{R}_+^n$ we say that f has

(i) increasing returns of scale if $f(\gamma\mathbf{x}) > \gamma f(\mathbf{x})$,

(ii) constant returns of scale if $f(\gamma\mathbf{x}) = \gamma f(\mathbf{x})$ and

(iii) decreasing returns of scale if $f(\gamma\mathbf{x}) < \gamma f(\mathbf{x})$.

To describe the concept of returns of scale using homogeneous production functions, we want f to be a homogeneous production function of degree α. According to definition 1.6, f has increasing returns of scale, if $\alpha > 1$, constant returns of scale, if $\alpha = 1$ and decreasing returns of scale, if $\alpha < 1$.

1.3. The concept of elasticity

Following the standard characteristics to describe productionfunctions thoroughly, we now come to the point of considering different elasticities in a general setup, as well as in the context of production functions. To introduce a general concept, assume that we have two economic quantities u and w that are related to each other, $w = w(u)$. Later on the concept will be expanded for general production functions.

1.9 Definition. [Uebe, 2013, p.24] Let u, w be two economic quantities. The elasticity between u and w is defined as

$$\varepsilon = \frac{u}{w}\frac{\mathrm{d}w}{\mathrm{d}u} = \frac{\mathrm{d}w/w}{\mathrm{d}u/u} \approx \frac{\Delta w/w}{\Delta u/u}.$$

It is scarcely necessary to point out that, no matter what dimension u and w have, its according elasticity is in any case a dimensionless number. Regarding the definition above, we can give an interpretation. Changing the amount of u by 1%, which means that $\frac{\Delta u}{u} = 1\%$, is going to lead to a ε%-change of w. This general way of considering elasticities can be assigned to various type of variables u and w, as it will be done concerning production functions.

A further remark may be added, that the elasticity may also be defined as the logarithmic derivative $\varepsilon = \frac{\mathrm{d} \ln w}{\mathrm{d} \ln u}$. Using the chain rule yields

$$\begin{aligned}\varepsilon = \frac{\mathrm{d} \ln w}{\mathrm{d} \ln u} &= \frac{\mathrm{d} \ln w}{\mathrm{d} w} \frac{\mathrm{d} w}{\mathrm{d} u} \frac{\mathrm{d} u}{\mathrm{d} \ln u} \\ &= \frac{\mathrm{d} \ln w}{\mathrm{d} w} \frac{\mathrm{d} w}{\mathrm{d} u} \frac{1}{\mathrm{d} \ln u / \mathrm{d} u} \\ &= \frac{1}{w} \frac{\mathrm{d} w}{\mathrm{d} u} \frac{1}{1/u} \\ &= \frac{u}{w} \frac{\mathrm{d} w}{\mathrm{d} u},\end{aligned}$$

showing that both ways to define elasticity are equivalent [Uebe, 2013, p.24].

1.10 Example. Consider an economic relation via the formula $w(u) = \alpha u^{\beta}$, where $\alpha, \beta \in \mathbb{R}$. Applying definition 1.14 yields

$$\varepsilon = \frac{u}{w} \frac{\mathrm{d} w}{\mathrm{d} u} = \frac{1}{\alpha u^{\beta-1}} \alpha \beta u^{\beta-1} = \beta.$$

It may also be added, that we could also use the logarithmic definition of the elasticity, since both are equivalent. According to the logarithmic definition, we find

$$\varepsilon = \frac{\mathrm{d} \ln w}{\mathrm{d} \ln u} = \frac{\mathrm{d}}{\mathrm{d} \ln u} \beta \left(\ln u + \ln \alpha \right) = \beta,$$

resulting in the same conclusion as seen above.

Consequently, the next step is to extend the concept of elasticity into the setting of production functions. Doing so, we will define three different terms, namely the elasticity of production, the elasticity of scales and the elasticity of substitution.

1.3.1. Elasticitiy of production

1.11 Definition. [Schmidt, 2013, p.374] Let $f : \mathbb{R}_+^n \to \mathbb{R}_+$ be a differentiable production function. The partial elasticity of the i-th input factor of f is defined as

$$\varepsilon_i(\mathbf{x}) = \frac{x_i \frac{\partial f(\mathbf{x})}{\partial x_i}}{f(\mathbf{x})} = \frac{\frac{\partial f(\mathbf{x})}{\partial x_i}}{\frac{f(\mathbf{x})}{x_i}}. \tag{1.4}$$

At a very first sight, this definition seems rather complicated but can easily made plausible by considering definition 1.3. Hence we conclude that the elasticity of the production function is equal to

$$\varepsilon_i(\mathbf{x}) = \frac{\mathrm{MP}_i(\mathbf{x})}{\mathrm{AP}_i(\mathbf{x})}.$$

So the elasticity gives information about the ratio of marginal product and the marginal productivity.

Analysing the elasticity of production, we may also conclude that a 1%-change of the amount of the i-th input factor yields a ε_i%-change of the expected amount of outcome [Uebe, 2013, p.25].

Again, we can connect the concept of the elasticity of a production function with the concept of homogeneous functions. This yields a quite remarkable result which is also known as the 'Eulersche Homogenitätsrelation'.

1.12 Theorem. [Schmidt, 2013, p.378] Let $f : \mathbb{R}_+^n \to \mathbb{R}_+$ be a differentiable production function and homogeneous of degree $\alpha \in (0, \infty)$. Then

$$\alpha = \frac{\langle \mathbf{x} | \mathrm{grad} f(\mathbf{x}) \rangle}{f(\mathbf{x})} = \sum_{i=1}^{n} \varepsilon_i(\mathbf{x})$$

holds true for any arbitrary fixed input vector $\mathbf{x} \in \mathbb{R}_+^n$.

Proof. Let $\mathbf{x}$ be a fixed input vector in $\mathbb{R}_+^n$ and define the functions $\xi : \mathbb{R}_+ \to \mathbb{R}_+$, $\xi(c) = f(c\mathbf{x})$ and $\eta : \mathbb{R}_+ \to \mathbb{R}_+^n$, $\eta_i(c) = cx_i$ for all $i = 1, \ldots, n$. It can easily be seen, that we can write the function ξ as the composition of both functions f and η, which yields $\xi(c) = f(\eta(c)) = (f \circ \eta)(c)$. Since both functions f and η are differentiable, we can also differentiate its composition with respect to its variable c. Applying the chain rule yields

$$\frac{\mathrm{d}\xi}{\mathrm{d}c} = \xi'(c) = (f \circ \eta)'(c) = \sum_{i=1}^{n} \frac{\partial f}{\partial \eta_i} (\eta_1(c), \ldots, \eta_n(c)) \frac{\partial \eta_i}{\partial c}$$

$$= \sum_{i=1}^{n} \frac{\partial f}{\partial \eta_i} (cx_1, \ldots, cx_n)\, x_i \tag{1.5}$$

On the other side, we required the production function f to be homogeneous of degree α. This yields that

$$\xi(c) = f(c\mathbf{x}) = c^{\alpha} f(\mathbf{x}). \tag{1.6}$$

Differentiating (1.6)with respect to c therefore yields

$$\xi'(c) = \alpha c^{\alpha-1} f(\mathbf{x}). \tag{1.7}$$

According to both representations of the derivative of ξ in(1.8) and(1.7), which have both to be true for all c, we find for $c = 1$

$$\alpha f(\mathbf{x}) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} (x_1, \ldots, x_n)\, x_i.$$

Thus it follows that

$$\alpha = \sum_{i=1}^{n} \frac{x_i \frac{\partial f(\mathbf{x})}{\partial x_i}}{f(\mathbf{x})} = \frac{\langle \mathrm{grad} f(\mathbf{x}) | \mathbf{x} \rangle}{f(\mathbf{x})} = \sum_{i=1}^{n} \varepsilon_i(\mathbf{x}).$$

□

For the sake of an example, we now want this illustrate this theorem using the production functions of example 1.7.

1.13 Example. [Schmidt, 2013, p.379] Consider two production functions of the form $f(\mathbf{x}) = \left(\sum_{i=1}^{n} \alpha_i x_i^{-\rho}\right)^{-\frac{\nu}{\rho}}$ and $g(\mathbf{x}) = \prod_{i=1}^{n} x_i^{\alpha_i}$. First of all we consider the production function g. According to the definition, we have to compute its partial elasticity for an arbitrary $i \in \{1, \ldots, n\}$

$$\begin{aligned} \varepsilon_{g,i}(\mathbf{x}) = \frac{x_i \partial f(\mathbf{x}) / \partial x_i}{f(\mathbf{x})} &= \frac{x_i (x_1^{\alpha_1} \cdot \ldots \cdot \alpha_i x_i^{\alpha_i - 1} \cdot \ldots \cdot x_n^{\alpha_n})}{x_1^{\alpha_1} \cdot \ldots \cdot x_n^{\alpha_n}} \\ &= \frac{\alpha_i \prod_{i=1}^{n} x_i^{\alpha_i}}{\prod_{i=1}^{n} x_i^{\alpha_i}} = \alpha_i. \end{aligned} \tag{1.8}$$

Considering theorem 1.12, we have tu sum up all partial elasticities for $i = 1, \ldots, n$.

Thus (1.8) yields

$$\sum_{i=1}^{n} \varepsilon_{g,i}(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i,$$

which is precisely the degree of homogenity for this class of production functions, we considered in example 1.7, so the theorem holds.

Now we want to compute the sum of partial elasticities of the production function f in an analogous way, according to the procedure shown above. Thus we again calculate its partial elasticity for an arbitrary $i \in \{1, \dots, n\}$

$$\begin{aligned}\varepsilon_{f,i}(\mathbf{x}) = \frac{x_i \frac{\partial f(\mathbf{x})}{\partial x_i}}{f(\mathbf{x})} &= \frac{x_i}{(\sum_{i=1}^{n} \alpha_i x_i^{-\rho})^{-\frac{\nu}{\rho}}} \left(-\frac{\nu}{\rho} \left(\alpha_1 x_1^{-\rho} + \ldots + \alpha_n x_n^{-\rho} \right)^{-\frac{\nu}{\rho}-1} (-\rho)\alpha_i x_i^{-\rho-1} \right) \\ &= \nu \frac{\alpha_i x_i^{-\rho}}{(\sum_{i=1}^{n} \alpha_i x_i^{-\rho})} \end{aligned} \tag{1.9}$$

Again, summing up all partial elasticities in (1.9) for $i = 1, \dots, n$ and concidering the fact that the sum in the denominator is a constant and can therefore be multiplied with the sum, yields

$$\sum_{i=1}^{n} \varepsilon_{f,i}(\mathbf{x}) = \frac{\nu}{\sum_{j=1}^{n} \alpha_j x_j^{-\rho}} \left(\sum_{i=1}^{n} \alpha_i x_i^{-\rho} \right) = \nu,$$

which again equals the degree of homogenity of the production function in example 1.7, according to theorem 1.12.

1.3.2. Elasticity of scale

1.14 Definition. [Uebe, 2013, p.25] Let $f : \mathbb{R}_+^n \to \mathbb{R}_+$ be a production function. The number

$$\zeta(\mathbf{x}) = \lim_{\lambda \to 1} \frac{\partial f(\lambda \mathbf{x})}{\partial \lambda} \frac{\lambda}{f(\lambda \mathbf{x})}$$

is called the elasticity of scale.

This definition can be seen as an extension of the definition of the elasticity of production. The analogue interpretation would be, that a proportional change of every input factor of 1% leads to a change of amount of the outcome of ζ%. Naturally the question arouses, if we can find a connection between the elasticity of scale and the elasticity of production [Uebe, 2013, p.24]. This connection is given by the theorem of Wicksell- Johnsonn.

1.15 THEOREM. [Uebe, 2013, p.24] Let $f : \mathbb{R}^n_+ \to \mathbb{R}_+$ be a continuously differentiable production function, and let ζ, ε_i be its corresponding elasticity of scale and elasticity of production for the i-th input factor. Then

$$\zeta(\mathbf{x}) = \sum_{i=1}^{n} \varepsilon_i(\mathbf{x})$$

holds for every input combination $\mathbf{x} \in \mathbb{R}^n_+$.

PROOF. Let $\mathbf{x} \in \mathbb{R}^n_+$ be a fixed input vector. By definition we know, that $f(\lambda\mathbf{x}) = f(\lambda x_1, \ldots, \lambda x_n)$. Thus we can again consider its derivation with respect to λ. Define the functions $\xi : \mathbb{R}_+ \to \mathbb{R}_+$, $\xi(\lambda) = f(\lambda\mathbf{x})$ and $\eta : \mathbb{R}_+ \to \mathbb{R}^n_+$, $\eta_i(\lambda) = \lambda x_i$ for all $i = 1, \ldots, n$. Thus we can rewrite $\xi(\lambda) = f(\eta(\lambda)) = (f \circ \eta)(\lambda)$. Using the chain rule yields

$$\begin{aligned} \frac{\mathrm{d}\xi}{\mathrm{d}\lambda} = \xi'(\lambda) = (f \circ \eta)'(\lambda) &= \sum_{i=1}^{n} \frac{\partial f}{\partial \eta_i}(\eta_1(\lambda), \ldots, \eta_n(\lambda)) \frac{\partial \eta_i}{\partial \lambda} \\ &= \sum_{i=1}^{n} \frac{\partial f}{\partial \eta_i}(\lambda x_1, \ldots, \lambda x_2) x_i. \end{aligned} \tag{1.10}$$

According to the definition of the elasticity of scale and (1.10), we compute

$$\frac{\partial f(\lambda\mathbf{x})}{\partial \lambda} \frac{\lambda}{f(\lambda\mathbf{x})} = \sum_{i=1}^{n} \frac{\partial f}{\partial \eta_i}(\lambda x_1, \ldots, \lambda x_2) \frac{\lambda x_i}{f(\lambda\mathbf{x})}.$$

Taking the limit $\lambda \to 1$ results in the elasticity of scale:

$$\zeta(\mathbf{x}) = \lim_{\lambda \to 1} \frac{\partial f(\lambda\mathbf{x})}{\partial \lambda} \frac{\lambda}{f(\lambda\mathbf{x})} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x_1, \ldots, x_2) \frac{x_i}{f(\mathbf{x})} = \sum_{i=1}^{n} \varepsilon_i(\mathbf{x}),$$

which completes the proof. $\square$

1.3.3. Elasticity of substitution

1.16 Definition. [Uebe, 2013, p.26] Let $f : \mathbb{R}_+^n \to \mathbb{R}_+$ be a differentiable production function and $c \in \mathbb{R}_+$. Consider a certain c-level-isoquant, choose two input factors $i, j = 1, \ldots, n$ and fix every component x_k of the input vector, where $k = 1, \ldots, n$, $k \neq i$ and $k \neq j$ except for the two x_i and x_j. Thus x_i and x_j are the only variables left. The elasticity of substitution between the input factor x_i and x_j is defined as

$$\sigma_{ij} = \frac{\mathrm{d}\ln\left(\frac{x_i}{x_j}\right)}{\mathrm{d}\ln(\mathrm{MRS}_{ji})}.$$

Thus we can state, that the definition of σ_{ij} is only true, given $\frac{\partial f}{\partial x_i}\mathrm{d}x_i + \frac{\partial f}{\partial x_j}\mathrm{d}x_j = 0$, which is the mathematical way to describe that the input combinations lie on the same isoquant.

Due to the fact that this formula turns out to be much more easier, an equivalent definition, reformulated in terms of logarithmic derivations would also be possible:

$$\sigma_{ij} = \frac{\mathrm{d}\left(\frac{x_i}{x_j}\right)}{\mathrm{d}(\mathrm{MRS}_{ji})} \frac{\mathrm{MRS}_{ji}}{\frac{x_i}{x_j}} = \left(\frac{\mathrm{d}(\mathrm{MRS}_{ji})}{\mathrm{d}\left(\frac{x_i}{x_j}\right)}\right)^{-1} \frac{\mathrm{MRS}_{ji}}{\frac{x_i}{x_j}}.$$

Thus the elasticity of substitution describes the change of the marginal rate of substitution along an isoquant while changing the ratio x_i/x_j.

1.4. The Constant elasticity of substitution function

In this section we want to introduce a special production function, which is in some sense more general, than the well known Leontief, Cobb-Douglas or linear production function. The most remarkable feature of this function is, that it's elasticity of substitution between any two arbitrary input factors is constant.

1.17 Definition. [Sydsæter et al., 2008, p.72] Let $\beta \in [0, \infty)$, $\rho \in [-1, \infty) \setminus \{0\}$, $\nu \in (0, \infty)$ and $\delta_i > 0$ for all $i = 1, \ldots, n$ with $\sum_{i=1}^n \delta_i = 1$. A production function of the form $f : \mathbb{R}^n_+ \to \mathbb{R}_+$,

$$f(\mathbf{x}) = \beta \left(\sum_{i=1}^{n} \delta_i x_i^{-\rho} \right)^{-\frac{\nu}{\rho}}$$

is calld a CES-function.

For a production function of this form, we can interpret the four different types of coefficients as follows

(i) The coefficient β determines the productivity of the whole process,

(ii) the parameters $\delta_1, \dots, \delta_n$ can be seen as some sort of share factor since they add up to 100 % and are weighted with the input factors.

(iii) The parameter ρ is closely connected with the elasticity of substitution, via the relation $\sigma_{ij} = \frac{1}{1+\rho}$, for all $i, j = 1, \dots, n$, $i \neq j$.

(iv) The last parameter ν simply represents the degree of homogenity and thus also equals the elasticity of scale.

It is a straightforward calculation to compute all the different characteristic numbers for production functions, introduced in this chapter. The result is listed in Table 1.1. At this point it may be added, that a general CES-production function has also a constant elasticity of scale.

In [Yasui, 1965], it was proven for the special case of $n = 2$, that every function that has a constant elasticity of substitution has to be of the form

$$f(x_1, x_2) = F\left[\left(\delta x_1^{-\rho} + (1-\delta) x_2^{-\rho} \right)^{-\frac{1}{\rho}} \right],$$

where $F : \mathbb{R} \to \mathbb{R}$ is an arbitrary differentiable function. We get the general form of the CES-production function we defined in definition 1.17, when we additionally require the production function to be homogeneous of degree ν, using Theorem 1.12 [Yasui, 1965].

As already stated above, the CES-function is a more general way to describe production functions. This is due to the fact, that a special choice of certain parameters of the CES-function involves other common production functions. This circumstance is discussed and explained thoroughly in theorem 1.18.

TABLE 1.1.: Characteristic numbers of the CES-production function, where λ is the degree of homogenity, MP_i the marginal product of the i-th input factor, MRS_{ij} the marginal rate of substitution of the factors i and j, σ_{ij} the elasticity of substitution of the factors i and j, as well as the elasticity of scale ζ.

λ	MP_i	MRS_{ij}	σ_{ij}	ζ
ν	$\beta\left(\nu\delta_i x_i^{-\rho-1}\sum_{i=1}^n \alpha_i x_i^{-\rho}\right)$	$\frac{\alpha_i}{\alpha_j}\left(\frac{x_i}{x_j}\right)^{-\rho-1}$	$\frac{1}{1+\rho}$	ν

1.18 THEOREM. [Böhm, 2013, p.38] Let $f : \mathbb{R}_+^n \to \mathbb{R}_+$, $f(\mathbf{x}) = \beta\left(\sum_{i=1}^n \delta_i x_i^{-\rho}\right)^{-\frac{\nu}{\rho}}$ be a CES-production function. Then the following holds true:

(i) Taking the Limit $\rho \to 0$ results in a general Cobb-Douglas function, with the degree of homogenity of ν.

(ii) Taking the limit $\rho \to \infty$ results in a general Leontief function, with a degree of homogenity of ν.

(iii) Taking the limit $\rho \to -1$, with $\nu = 1$ results in a linear production function.

PROOF. First if all consider (i). As it follows from the statement of the theorem we calculate the limit of the CES-function straightforwardly. Therefore we use a little trick, taking the exponential function of the logarithmic of the function itself. This yields

$$\begin{aligned}
\lim_{\rho\to 0} f(\mathbf{x}) &= \lim_{\rho\to 0} \exp\ln\left(\beta\left(\sum_{i=1}^n \delta_i x_i^{-\rho}\right)^{-\frac{\nu}{\rho}}\right) \\
&= \lim_{\rho\to 0} \beta\exp\left(-\frac{\nu}{\rho}\ln\left(\sum_{i=1}^n \delta_i x_i^{-\rho}\right)\right) \\
&= \beta\exp\left(\lim_{\rho\to 0} -\frac{\nu}{\rho}\ln\left(\sum_{i=1}^n \delta_i x_i^{-\rho}\right)\right).
\end{aligned}$$

Thus we only need to compute the argument of the exponential function, which can be analysed by using L'Hospitals rule, since we find an expression of the form $0/0$ as $\rho \to 0$. Using the rule of derivation with respect to ρ: $\frac{\mathrm{d}x^{-\rho}}{\mathrm{d}\rho} = -x^{-\rho}\ln x$, differentiating both nominator and denominator yields,

$$\begin{aligned}\lim_{\rho\to 0} -\frac{\nu}{\rho}\ln\left(\sum_{i=1}^{n}\delta_i x_i^{-\rho}\right) &= \lim_{\rho\to 0}\frac{\nu\sum_{i=1}^{n}\delta_i\ln(x_i)x_i^{-\rho}}{\sum_{i=1}^{n}\delta_i x_i^{-\rho}}\\ &= \nu\sum_{i=1}^{n}\delta_i\ln x_i,\end{aligned}$$

since all the share factors δ_i add up to one, according to the definition of the CES-production function. Summing up we found, that the limit

$$\begin{aligned}\lim_{\rho\to 0} f(\mathbf{x}) &= \beta\exp\left(\nu\sum_{i=1}^{n}\delta_i\ln x_i\right)\\ &= \beta\prod_{i=1}^{n} x_i^{\nu\delta_i},\end{aligned}$$

which represents a general Cobb-Douglas function with degree of homogenity of ν, as it is the case of the CES-production function. In fact this result is also reasonable, since choosing the coefficient $\rho = 0$ yields an elasticity of substitution of one, as it is the case concerning the Cobb-Douglas production function.

Secondly, to proof (ii) we would first of all guess, that $\rho \to \infty$ would probably result in a Leontief production function, since the elasticity of substitution converges to zero. To see this we take a closer look at

$$\begin{aligned}\lim_{\rho\to\infty}\beta\left(\sum_{i=1}^{n}\delta_i x_i^{-\rho}\right)^{-\frac{\nu}{\rho}} &= \lim_{\rho\to\infty}\beta\left(\delta_1 x_1^{-\rho} + \ldots + \delta_n x_n^{-\rho}\right)^{-\frac{\nu}{\rho}}\\ &= \beta\lim_{\rho\to\infty}\left(\frac{1}{\sqrt[\rho]{\delta_1 x_1^{-\rho} + \ldots + \delta_n x_n^{-\rho}}}\right)^{\nu}.\end{aligned}$$

Since the input parameters $x_1, \ldots, x_n$ are fixed, we can find the smallest of them, denoted by $x_M = \min_{i=1,\ldots,n} x_i$ and its associated share parameter δ_M, respectively. This yields

$$\lim_{\rho\to\infty}\beta\left(\sum_{i=1}^{n}\delta_i x_i^{-\rho}\right)^{-\frac{\nu}{\rho}} = \beta\lim_{\rho\to\infty}\left(\frac{1}{\sqrt[\rho]{\left(\frac{1}{x_M}\right)^{\rho}}\sqrt[\rho]{\delta_1(\frac{x_M}{x_1})^{\rho} + \ldots + \delta_M + \ldots + \delta_n(\frac{x_M}{x_n})^{\rho}}}\right)^{\nu}$$

$$= \beta \lim_{\rho\to\infty} \left(\frac{x_M}{\sqrt[\rho]{\delta_1(\frac{x_M}{x_1})^\rho + \ldots + \delta_M + \ldots + \delta_n(\frac{x_M}{x_n})^\rho}} \right)^\nu .$$

Since we chosen x_M to be the minimum of all the input factors, each and every ratio x_M/x_i is smaller or even one, for $i = 1, \ldots, n$. Therefore the root in the denominator converges to one. Summing up, we found that

$$\lim_{\rho\to\infty} \beta \left(\sum_{i=1}^{n} \delta_i x_i^{-\rho} \right)^{-\frac{\nu}{\rho}} = \beta \left(\min_{i=1,\ldots,n} x_i \right)^\nu ,$$

which represents a general Leontief function, again with the degree of homogenity equal to ν, as the CES- production function.

Concerning (iii), we can simply calculate the limit, setting $\nu = 1$

$$\lim_{\rho\to -1} \beta \left(\sum_{i=1}^{n} \delta_i x_i^{-\rho} \right)^{-\frac{\nu}{\rho}} = \beta \left(\delta_1 x_1 + \ldots + \delta_n x_n \right),$$

resulting in a linear production function. □

If we define the sets

(i) $\mathcal{COB} = \{f(\mathbf{x}) : \mathbb{R}^n_+ \to \mathbb{R}_+ : f(\mathbf{x}) = \beta \prod_{i=1}^n x_i^{\alpha_i} : \beta > 0, \alpha_i > 0 \quad \text{for all} \quad i = 1, \ldots, n\}$

(ii) $\mathcal{LEON} = \{f(\mathbf{x}) : \mathbb{R}^n_+ \to \mathbb{R}_+ : f(\mathbf{x}) = \beta(\min_{i=1,\ldots,n} x_i) : \beta > 0\}$

(iii) $\mathcal{LIN} = \{f(\mathbf{x}) : \mathbb{R}^n_+ \to \mathbb{R}_+ : f(\mathbf{x}) = \sum_{i=1}^n \alpha_i x_i : \alpha_i > 0 \quad \text{for all} \quad i = 1, \ldots, n\}$ and

(iv) $\mathcal{CES} = \{f(\mathbf{x}) : \mathbb{R}^n_+ \to \mathbb{R}_+ : f(\mathbf{x}) = \beta \left(\sum_{i=1}^n \delta_i x_i^{-\rho}\right)^{-\frac{\nu}{\rho}} : \beta > 0, \rho \in [0,\infty) \setminus \{0\}, \nu > 0 \quad \text{and} \quad \delta_i > 0 \quad \text{with} \quad \sum_{i=1}^n \delta_i = 1 \quad \text{for all} \quad i = 1, \ldots, n\} \cup \{\hat{f}_1(\mathbf{x}, \hat{f}_2(\mathbf{x})\}$,

where $\hat{f}_1(\mathbf{x}) = \lim_{\rho\to\infty} \beta \left(\sum_{i=1}^n \delta_i x_i^{-\rho}\right)^{-\frac{\nu}{\rho}}$ and $\hat{f}_2(\mathbf{x}) = \lim_{\rho\to 0} \beta \left(\sum_{i=1}^n \delta_i x_i^{-\rho}\right)^{-\frac{\nu}{\rho}}$. We immediately conclude from theorem 1.18 that $\mathcal{COB} \subseteq \mathcal{CES}$, $\mathcal{LEON} \subseteq \mathcal{CES}$ and $\mathcal{LIN} \subseteq \mathcal{CES}$ holds true.

Thus we can choose an arbitrary element of the set $\mathcal{COB}$, $\mathcal{LEON}$ or $\mathcal{LIN}$, which represents a Cobb- Douglas, Leontief or linear production function and see that it is also included in the set $\mathcal{CES}$. Therefore each of these production function can be written using special parameters of the CES production function.

Having this result in mind, we will model the cost function of a certain sector based on a CES-function, since this is obviously the most general way to describe certain properties. This is due to the fact that all other common production functions are only special cases of the more general CES-function.

CHAPTER 2

Basic results of previous works

In this chapter we shortly summarize and give an overview about the previous work done in this field, based on [Ivaz, 2014]. Following this ideas, the producer price index is introduced and a first approximation of the cost function, based on separate linear regression models, for different input variables, is given.

2.1. The producer price index

To get rid of the effect of inflation, we renormalise all data by using the producer price index (PPI)of the mechanical engineering branch published by statistics austria. According to [Ivaz, 2014, p.44] the PPI is calculated, choosing the basis year 2011, because it was the latest year providing the data. Based on this data, a factor p_t, where $t = 2008, \ldots, 2011$ denotes the year, has to be construed to normalize given values of different years, such that multiplying this factor results in inflation-adjusted values that can be compared.

In Table 2.1 the characteristic numbers to compute the price adjustment are given.

TABLE 2.1.: The values v_t of the years 2008-2011, on which the calculation of the correction factor p_t is based, as well as the annual change rate Δ_t.

Year	2008	2009	2010	2011
Annual change Δ_t	-0,753 %	3,507%	1,500%	/
Value with basis year 2011 v_t	95,9057	95,1840	98,5222	100
Correction factor p_t	1,0427	1,0506	1,0150	1

Given the values v_t, we can compute the annual change rate via the formula

$$\Delta_t = \frac{v_{t+1} - v_t}{v_t}$$

for the years $2008, 2009, 2010$.

Having this annual change rate Δ_t, we can now calculate the values v_t for $t < 2011$ with the basis year of 2011, given the annual change rate. This can be done by using the following formula

$$v_t = \frac{v_{2011}}{\prod_{k=t}^{2010}(1+\Delta_k)} = \frac{100}{\prod_{k=t}^{2010}(1+\Delta_k)}.$$

The correction factor for a certain year p_t can then be calculated by

$$p_t = \frac{v_{2011}}{v_t} = \frac{100}{v_t} = \prod_{k=t}^{2010}(1+\Delta_k).$$

According to the statistics of the WKO, we find the data for the whole expenditures E_t for the years 2008,2009,2010 and 2011, given in Table 2.2. This data will then be multiplied with the factor p_t according to the considerations stated above, to find the re-assessed expenditures $\tilde{E}_t = E_t p_t$ for each year. The result of this approach is also given in Table 2.2.

TABLE 2.2.: This table shows the accumulated entire expenditures E_t for the years 2008-2010, as well as the re-assessed accumulated entire expenditures $\tilde{E}_t$ and the working hours x.

Year	2008	2009	2010	2011
Entire expenditures E_t	15.977.799.000	13.186.539.000	12.844.611.000	15.485.238.000
Re-assessed Entire expenditures $\tilde{E}_t$	16.660.051.017	13.853.777.873	13.037.280.165	15.485.238.000
Working hours x	98558670	91024723	85718242	94131837

2.2. Computation of the Cost function as a function of working hours

According to the given data of the WKO, we are able to use the data points to construct a linear cost function $\mathcal{C}(x)$ with the working hours x as the variable. According to the considerations above we will use the expenditures E_t as well as the re-assessed expenditures $\tilde{E}_t$. This values are given in Table 2.2. The plot of the cost function using the normal expenditures is given in Figure 2.1 and the plot using the re-assessed expenditures is given in Figure 2.2. Concerning the accuracy of the linear regression model we used, we find that $R^2 = 0,69441$ for the normal expenditures. Another linear regression with the re-assessed data $\tilde{E}_t$ yields a better accuracy with $R^2 = 0,7557$.

At this point the idea of modelling the cost function, with the working hours as the depended variable is rejected, since the working hours does not show the sufficient flexibility to regulate certain scenarios in an accurate manner.

2.3. Computation of the Cost function as a function of the production value

Since the accuracy of the fit wasn't satisfactory, the next idea was to create an own variable, corresponding to the production value x_{PV} of the branch. According to the production statistics of Statistic Austria this calculated production values x_{PV}, as well as the re-assessed ones $\tilde{x}_{PV}$, using the method of the PPI, are given in Table

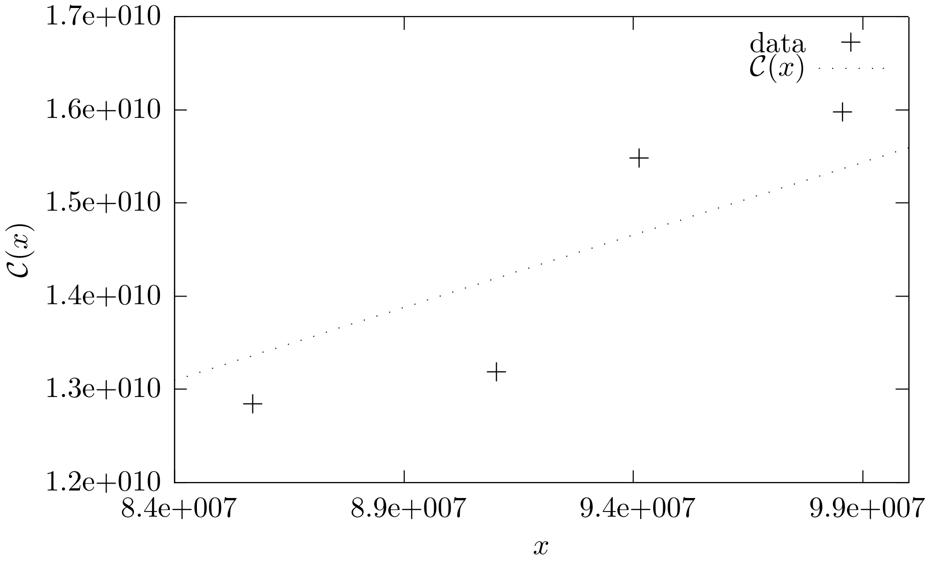


FIGURE 2.1.: The linear approximated cost function $\mathcal{C}(x)$ as a function of the working hours x, using the normal expenditures for the years 2008-2011.

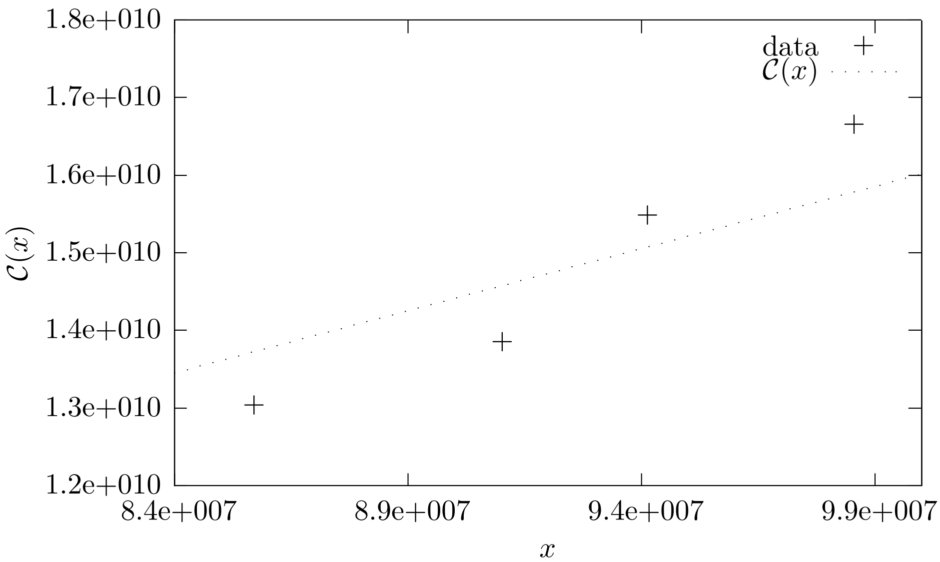


FIGURE 2.2.: The linear approximated cost function $\mathcal{C}(x)$ as a function of the working hours x, using the re-assessed expenditures for the years 2008-2011.

Table 2.3.: The production value PV, the re-assesed production value $\tilde{PV}$, the variable x_{PV} and the re-assessed variable $\tilde{x}_{PV}$ used for the computation of the cost function dependend of the variable of the production value.

Year	2008	2009	2010	2011
PV	15.891.633.000	12.980.208.000	12.920.163.000	15.842.013.000
x_{PV}	158.916.330	129.802.080	129.201.630	158.420.130
$\tilde{PV}$	16.570.205.729	13.637.006.525	13.113.965.445	15.842.013.000
$\tilde{x}_{PV}$	165.702.057	136.370.065	131.139.654	158.420.130

2.3. Given the data of Statistics Austria we compute the variable x_{PV} based on the given production values via the formula

$$x_{PV} = \frac{PV}{100},$$

such that this variable has the unit of monetary units divided by a factor 100 for the sake of convenience. Again we can compute a linear cost function $\mathcal{C}(\tilde{x}_{PV})$ but this time with the self created variable $\tilde{x}_{PV}$, which should lead to more accurate results. A linear regression model yields an accuracy of $R^2 = 0,97854$ and is shown in Figure 2.3. Thus we found a much more accurate model but since the fixed costs of this model are about 4% of the whole expenditures,this approach was also rejected.

2.4. Construction of an output orientated cost function

At this point, the expenditures are separated into costs of materials, costs of labour and costs of technology. A further remark may be added, that we add any costs, that are not possible to allocate to any of this three cost categories, to the third one. Based on a Leontief production function we assume the corresponding cost function $\mathcal{C}(\tilde{x}_{PV}, x)$ to be linear, whereas we want to consider the two variables: re-assessed production value $\tilde{x}_{PV}$ and the working hours x. Having this in mind, we split the cost function in three parts,

$$\begin{aligned} \mathcal{C}(x_{PV}, x) &= \mathcal{C}_M + \mathcal{C}_L + \mathcal{C}_T \\ &= c_{v_M}\tilde{x}_{PV} + C_{f_M} + c_{v_L}x + C_{f_L} + c_{v_T}\tilde{x}_{PV} + C_{f_T}, \end{aligned} \tag{2.1}$$

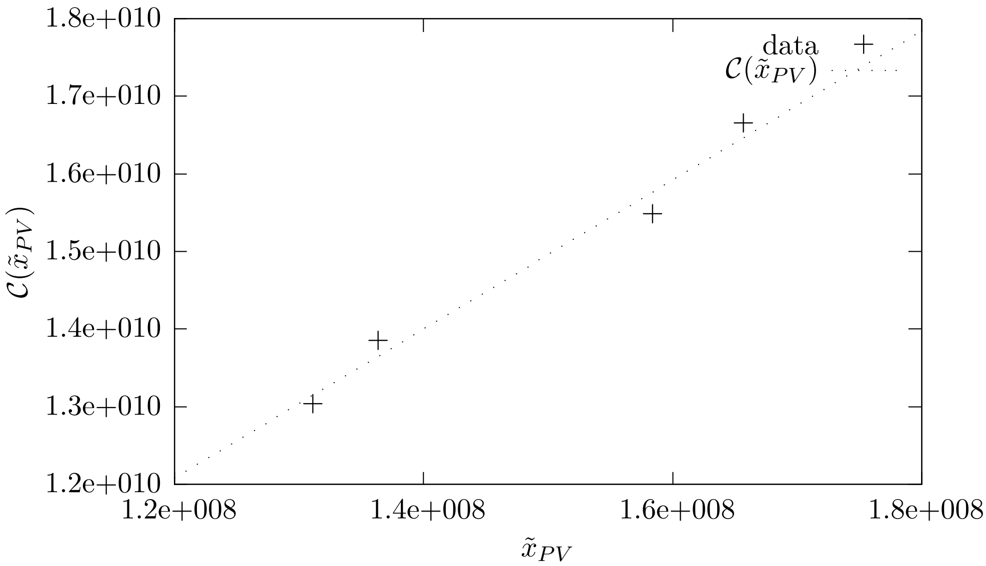


FIGURE 2.3.: This plot shows the linear approximated cost function $\mathcal{C}(\tilde{x}_{PV})$ as a function of the re-assessed production value $\tilde{x}_{PV}$ using the re-assessed expenditures for the years 2008-2011.

where the index M stands for material, L for labour and T for technology, according to the three different types of cost categories. In the next step, we have to determine the variable costs for all three different cost categories. Again this is done by a linear regression. Analysing the linear regression model yields the variable as well as the fixed costs for every cost category listed above. Thus we find $c_{v_M} = 36,75761, c_{v_L} = 28,06162, c_{v_T} = 47,25623$ as well as $C_{f_M} = 1.027.323.254, C_{f_L} = 641.949.365, C_{f_T} = -1.908.392.250$. According to the fact that the fixed costs of the technology category is negative this basic approach is replaced by a polynom of second order with pre defined fixed costs of $C_{f_T} = 3.500.000.000$, which is a plausible value. Thus we find for the cost function of technology

$$\mathcal{C}_T(x_{PV}) = 0,0000002635x_{PV}^2 + 28,63538x_{PV} + 3.500.000.000.$$

Summing up all single parts according to (2.1) with the estimated variable and fixed costs, yields the total cost function

$$\mathcal{C}(x_{PV}, x) = 0,0000002635x_{PV}^2 + 65,39299x_{PV} + 28,06162x + 5.169.272.619,$$

TABLE 2.4.: The costs of material MC, the costs of labour LC, the costs of technology TC, the working hours x as well as the production value $\tilde{x}_{PV}$ for the years 2008-2011.

Year	2008	2009	2010	2011
MC	6.980.529.000	6.061.653.000	5.520.656.000	6.731.340.000
$\tilde{LC}$	3.370.819.706	3.197.549.428	3.027.476.025	3.338.854.000
TC	5.764.490.000	4.081.340.000	4.341.220.000	5.415.044.000
x_{PV}	158.916.330	129.802.080	129.201.630	158.420.130
x	98.558.670	91.024.723	85.718.242	94.131.837

which is plotted in Figure 2.4.

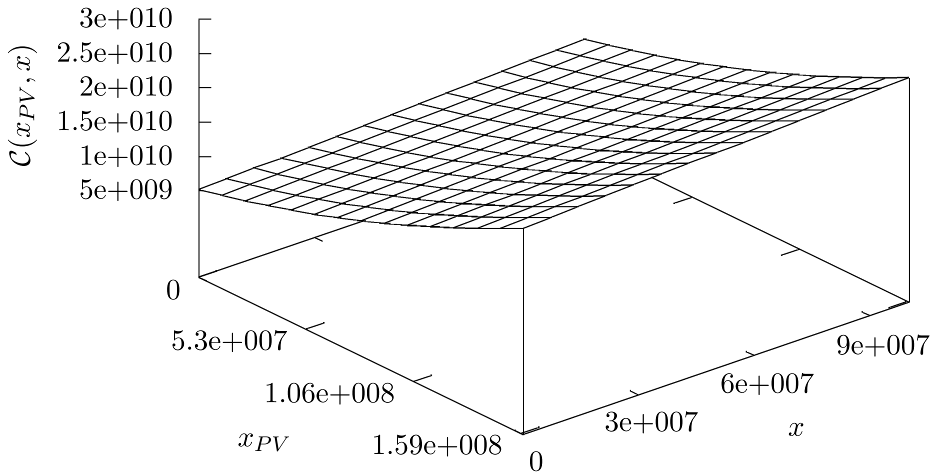


FIGURE 2.4.: The cost function $\mathcal{C}(x_{PV}, x)$ as a function of the production value x_{PV} and the working hours x.

CHAPTER 3

Constructing the Cost function on the basis of a CES production function

In this chapter we will derive the mathematical form of the cost function based on a general CES-production function, by solving a minimization problem using the method of Lagrange multipliers. Further more the factor prices and the fixed costs associated with the cost function are estimated, again based on the work [Ivaz, 2014]. In the last part of this chapter, another method of estimating the parameter of the CES-production function, following the idea of [Kmenta, 1967], is introduced and analysed, showing the advantages and disadvantages of this procedure.

3.1. Deriving the cost function

Since we find ourselves in the situation of having only information about the cost structure available, we have to determine the cost function, based on its production function. With the determination of the production function, we are able to discuss the technical possibilities of a certain sector, whereas the cost function characterises its economic features [Wied-Nebbeling and Schott, 2006, p.132]. Doing so, we are confronted with the problem of computing the cost function, given a certain production function.

3.1.1. The general case of n input factors

Lets assume that we have given an arbitrary production function $f : \mathbb{R}^n_+ \to \mathbb{R}_+$ mapping n input variables to a certain amount of outcome. The task is to find the minimal cost combination of the amount of input factors $x_1, \ldots, x_n$, knowing the factor prices q_i for one unit of amount of x_i, for all $i = 1, \ldots, n$ and calculate its costs $\mathcal{C}(\mathbf{x}) = \sum_{i=1}^n q_i x_i + \mathrm{FC}$, where FC denotes the fixed costs. The cost function is restrained by a certain amount of outcome due to economic reasons. Summing up, we have to solve the minimisation problem

$$\min_{\substack{x_i \in \mathbb{R}_+, \\ i=1,\ldots,n}} (q_1 x_1 + \ldots + q_n x_n + \mathrm{FC}) = \min_{\substack{x_i \in \mathbb{R}_+, \\ i=1,\ldots,n}} \mathcal{C}(\mathbf{x}),$$

given the constraint of a certain production level $\bar{c}$, with $\bar{c} \in \mathbb{R}_+$

$$f(\mathbf{x}) = \bar{c}.$$

Note, that we can ignore the fixed costs in the minimization problem, since it represents a constant factor and does not change the position of the minimum, but only the value itself. Thus we ignore the fixed costs, solve the minimization problem and add them to the cost function afterwards.

One can interpret this problem as follows: Assume we have given a certain $\bar{c}$-level-isoquant. This means that we have a set of input variables available, namely all input combinations, referring to that $\bar{c}$-level-isoquant. We then have to look for the combination of input factors of this set, having the characteristic feature of being the input combination, referring to minimal costs.

This is a classical problem to be solved with the method of Lagrange multipliers [Uebe, 2013, p.38]. Thus we set up the Lagrangian function $\mathcal{L}(\mathbf{x}, \lambda) = \mathcal{C}(\mathbf{x}) + \lambda(\bar{c} - f(\mathbf{x}))$ and set its partial derivatives with respect to x_i, for $i = 1, \ldots, n$ and λ to zero, to find the minimum of the cost function. Doing so, we generally get $n+1$ non linear equations of the form

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial x_1} &= \frac{\partial \mathcal{C}}{\partial x_1} - \lambda \frac{\partial f}{\partial x_1} = 0, \\
&\vdots \\
\frac{\partial \mathcal{L}}{\partial x_n} &= \frac{\partial \mathcal{C}}{\partial x_n} - \lambda \frac{\partial f}{\partial x_n} = 0, \\
\frac{\partial \mathcal{L}}{\partial \lambda} &= \bar{c} - f(\mathbf{x}) = 0.
\end{aligned} \tag{3.1}$$

Since we know that $\frac{\partial \mathcal{C}}{\partial x_i} = q_i$ for all $i = 1, \ldots, n$ and additionally $\frac{\partial f}{\partial x_i} = \mathrm{MP}_i(\mathbf{x})$ (see definition 1.3) holds, we can reformulate (3.1) to a system of equations of the form

$$\begin{aligned}
q_1 &= \lambda \mathrm{MP}_1(\mathbf{x}) \\
&\vdots \\
q_n &= \lambda \mathrm{MP}_n(\mathbf{x}) \\
\bar{c} &= f(\mathbf{x}).
\end{aligned} \tag{3.2}$$

Thus we have to look for the combinations of amount of input factors where

$$\frac{q_i}{q_j} = \frac{\mathrm{MP}_i(\mathbf{x})}{\mathrm{MP}_j(\mathbf{x})} = \mathrm{MRS}_{ij}(\mathbf{x}) \tag{3.3}$$

holds, for all $i, j = 1, \ldots, n$, $i \neq j$, given a certain $\bar{c}$-level-isoquant $\bar{c} = f(\mathbf{x})$ [Uebe, 2013, p.39]. Remarkably we find the result, that the optimal combination of amount of input factors is to be found, when the ratio of factor prices equals the marginal rate of substitution. At this point we are not able to simplify the result further on, since the marginal rate of substitution depends on the given production function.

3.1.2. The special case of two input factors for the CES function

In many situations we find ourself confronted with the problematic of two input factors, commonly labour and capital. This motivates the special case to compute the cost function only for two input factors x_1 and x_2 with given factor prices q_1, q_2 and a certain amount of outcome $\bar{c} = f(x_1, x_2)$. Due to the fact, that the CES production function is more general, we want to consider a function of the form $f(x_1, x_2) : \mathbb{R}^2_+ \to \mathbb{R}_+$, $f(x_1, x_2) = \beta \left(\delta_1 x_1^{-\rho} + \delta_2 x_2^{-\rho}\right)^{-\frac{\nu}{\rho}}$, where $\delta_1 + \delta_2 = 1$, with only two input factors. From (3.3) and Table 1.1, we know that

$$\frac{q_1 x_1}{q_2 x_2} = \frac{\delta_1}{\delta_2} \left(\frac{x_1}{x_2}\right)^{-\rho}$$

holds. Adding one to both sides of the equations yields

$$\frac{q_1 x_1 + q_2 x_2}{q_2 x_2} = \frac{\delta_1 x_1^{-\rho} + \delta_2 x_2^{-\rho}}{\delta_2 x_2^{-\rho}} = \frac{\mathcal{C}(x_1, x_2)}{q_2 x_2}. \tag{3.4}$$

Since we consider a $\bar{c}$- isoquant we can transform $\bar{c} = f(x_1, x_2) = \beta \left(\delta_1 x_1^{-\rho} + \delta_2 x_2^{-\rho}\right)^{-\frac{\nu}{\rho}}$ to get

$$\delta_1 x_1^{-\rho} + \delta_2 x_2^{-\rho} = \left(\frac{\bar{c}}{\beta}\right)^{-\frac{\rho}{\nu}}. \tag{3.5}$$

Thus we can combine (3.4) and (3.5) resulting in

$$\frac{\mathcal{C}(x_1, x_2)}{q_2 x_2} = \frac{\left(\frac{\bar{c}}{\beta}\right)^{-\frac{\rho}{\nu}}}{\delta_2 x_2^{-\rho}},$$

or equivalently

$$x_2^{\rho+1} = \frac{\mathcal{C}(x_1, x_2) \left(\frac{\bar{c}}{\beta}\right)^{\frac{\rho}{\nu}} \delta_2}{q_2}, \tag{3.6}$$

$$q_2 x_2 = \mathcal{C}(x_1, x_2)^{\frac{1}{1+\rho}} \left(\frac{\bar{c}}{\beta}\right)^{\frac{\rho}{\nu(\rho+1)}} \delta_2^{\frac{1}{1+\rho}} q_2^{\frac{\rho}{\rho+1}}. \tag{3.7}$$

A similar argument, in fact on has only to change $2 \mapsto 1$, yields

$$q_1 x_1 = \mathcal{C}(x_1, x_2)^{\frac{1}{1+\rho}} \left(\frac{\bar{c}}{\beta}\right)^{\frac{\rho}{\nu(\rho+1)}} \delta_1^{\frac{1}{1+\rho}} q_1^{\frac{\rho}{\rho+1}}. \tag{3.8}$$

Since adding up (3.6) and (3.8) yields the total costs $\mathcal{C}(x_1, x_2)$ we get

$$q_1 x_1 + q_2 x_2 = \mathcal{C}(x_1, x_2) = \mathcal{C}(x_1, x_2)^{\frac{1}{1+\rho}} \left(\frac{\bar{c}}{\beta}\right)^{\frac{\rho}{\nu(\rho+1)}} \left(\delta_1^{\frac{1}{1+\rho}} q_1^{\frac{\rho}{\rho+1}} + \delta_2^{\frac{1}{1+\rho}} q_2^{\frac{\rho}{\rho+1}}\right).$$

From this it follows that

$$\mathcal{C}(x_1, x_2) = \left(\frac{\bar{c}}{\beta}\right)^{\frac{1}{\nu}} \left(\delta_1^{\frac{1}{1+\rho}} q_1^{\frac{\rho}{1+\rho}} + \delta_2^{\frac{1}{1+\rho}} q_2^{\frac{\rho}{1+\rho}}\right)^{\frac{1+\rho}{\rho}}. \tag{3.9}$$

A further remark may be added, that setting $\sigma = \frac{1}{1+\rho}$, which is the elasticity of substitution for a general CES-production function according to Table 1.1 yields

$$\mathcal{C}(q_1, q_2) = \left(\frac{\bar{c}}{\beta}\right)^{\frac{1}{\nu}} \left(\delta_1^{\sigma} q_1^{1-\sigma} + \delta_2^{\sigma} q_2^{1-\sigma}\right)^{\frac{1}{1-\sigma}}. \tag{3.10}$$

Note, that (3.9) has a very similar mathematical structure to the generalised CES-production function. Rewriting (3.10) in an input oriented form, by using $\left(\frac{\bar{c}}{\beta}\right)^{\frac{1}{\nu}} = \left(\delta_1 x_1^{-\rho} + \delta_2 x_2^{-\rho}\right)^{-\frac{1}{\rho}}$ from (3.5) yields

$$\mathcal{C}(x_1, x_2) = \left(\delta_1 x_1^{-\rho} + \delta_2 x_2^{-\rho}\right)^{-\frac{1}{\rho}} \left(\delta_1^{\frac{1}{1+\rho}} q_1^{\frac{\rho}{1+\rho}} + \delta_2^{\frac{1}{1+\rho}} q_2^{\frac{\rho}{1+\rho}}\right)^{\frac{1+\rho}{\rho}}.$$

Considering the fixed costs we simply add them to the cost function, which we ignored in the beginning of the problem. Thus we find

$$\mathcal{C}(x_1, x_2) = \left(\delta_1 x_1^{-\rho} + \delta_2 x_2^{-\rho}\right)^{-\frac{1}{\rho}} \left(\delta_1^{\frac{1}{1+\rho}} q_1^{\frac{\rho}{1+\rho}} + \delta_2^{\frac{1}{1+\rho}} q_2^{\frac{\rho}{1+\rho}}\right)^{\frac{1+\rho}{\rho}} + \mathrm{FC}.$$

In the last step, we use the fact that both share parameters δ_1 and δ_2 should add up to one, according to the definition of a general CES-production function. With $\delta_1 = \delta$ and $\delta_2 = 1 - \delta$ we therefore find

$$\mathcal{C}(x_1, x_2) = \left(\delta x_1^{-\rho} + (1-\delta) x_2^{-\rho}\right)^{-\frac{1}{\rho}} \left(\delta^{\frac{1}{1+\rho}} q_1^{\frac{\rho}{1+\rho}} + (1-\delta)^{\frac{1}{1+\rho}} q_2^{\frac{\rho}{1+\rho}}\right)^{\frac{1+\rho}{\rho}} + \mathrm{FC}, \tag{3.11}$$

which is the final form of the cost function we will use further on. What we can see from (3.11), is that the first part of the cost function is in fact a CES production function with a special choice of parameters $\nu = \beta = 1$. Defining $\left(\delta x_1^{-\rho} + (1-\delta) x_2^{-\rho}\right)^{-\frac{1}{\rho}} = f^*_{x_1,x_2}(\delta, \rho)$ and $\Psi(\delta, \rho) = \left(\delta^{\frac{1}{1+\rho}} q_1^{\frac{\rho}{1+\rho}} + (1-\delta)^{\frac{1}{1+\rho}} q_2^{\frac{\rho}{1+\rho}}\right)$

allows us to rewrite the cost function in a more compact form

$$\mathcal{C}(x_1, x_2) = f^*_{x_1,x_2}(\delta, \rho)\Psi(\delta, \rho)^{\frac{1+\rho}{\rho}} + \mathrm{FC}. \tag{3.12}$$

The only parameters left, are the share parameter δ and the parameter ρ, referring to the elasticity of substitution. To calibrate the cost function based on a CES-production function, we therefore have to estimate only two parameters that refer to the occurring costs.

3.2. Calibration of the Cost function for a general CES-production-function

3.2.1. Estimation of the factor prices

Before starting the calibration of the parameter of the cost function $\mathcal{C}(x_1, x_2)$ we have to estimate the factor prices q_1, q_2 of both of the input variables x_1 and x_2, which are chosen the following way. The input variable x_2 represents the amount of working hours and the input variable $x_1 = x_{RMC}$ refers to the remaining manufacturing costs, computed via the formula $x_{RMC} = \frac{1}{100}(100x_{PV} - x_2q_2)$. Again this input variable is given in units of 100 to get smaller numbers.

The factor prices are computed by dividing the given accumulated variable costs through the amount of input according to the three different cost categories costs of material, costs of labour and costs of technology over the years 2008-2011 and then this numbers are averaged over the years. The fixed costs are estimated by the approximation of linear cost functions, see [Ivaz, 2014]. The result is given in Table 3.1. Finally we can compute the average factor prices over the years, which yields $q'_1 = 36,757$, $q_2 = 27,874$ and $q'_3 = 9,332$ [Ivaz, 2014].

Concerning the depended variables, we have only chosen $x_1 = x_{RMC}$ and $x_2 = x$. Thus we can sum up $q'_1 + q'_3 = q_1$ to get the variable costs for the input variable x_1, representing the remaining manufacturing costs of the sector. This yields the variable costs of $q_1 = 56,139$ and $q_2 = 27,874$, respectively.

Yet another transformation has to be performed, namely dividing all costs by a factor of 10^8. This is done due to the fact that we have to fit the function to the given data points in the best way, by using a nonlinear least squares fit. This is a numeric

TABLE 3.1.: The total costs of material MC, the total re-assesed costs of labour $\tilde{LC}$, the total costs of technology TC as well as its total variable costs,with its associated fixed costs $\mathcal{C}_{f_i}$, where $i = \mathrm{M, L, T}$. Additionally the input variables $x_{RMC} = x_1$, $x = x_2$ (as well as x_{PV}), its estimated factor prices $\bar{q_1}'$, $\bar{q_2}$ and $\bar{q_3}'$ are given, for every year and cost category where the index 1 refers to costs of material, 2 to re-assessed costs of labour and 3 to costs of technology, respectively.

Year	2008	2009	2010	2011
MC	6.980.529.000	6.061.653.000	5.520.656.000	6.731.340.000
$\tilde{LC}$	3.370.819.706	3.197.549.428	3.027.476.025	3.338.854.000
TC	5.764.490.000	4.081.340.000	4.341.220.000	5.415.044.000
$\mathrm{MC} - \mathcal{C}_{f_M}$	5.953.205.746	5.034.329.746	4.493.332.746	5.704.016.746
$\tilde{\mathrm{LC}} - \mathcal{C}_{f_L}$	2.701.459.103	2.523.117.425	2.375.897.420	2.696.904.635
$\mathrm{TC} - \mathcal{C}_{f_T}$	2.264.490.000	581.340.000	841.220.000	1.915.044.000
x_{PV}	158.916.330	129.802.080	129.201.630	158.420.130
$x_1 = x_{RMC}$	131.901.739	104.570.906	105.442.656	131.451.084
$x_2 = x$	98.558.670	91.024.723	85.718.242	94.131.837
$\bar{q_1}'$	45,13	48,14	42,61	43,39
$\bar{q_2}$	27,41	27,719	27,718	28,65
$\bar{q_3}'$	17,17	5,56	7,98	14,57

TABLE 3.2.: The result of rescaling the total variable costs $\mathrm{MC} - \mathcal{C}_{f_M}, \tilde{\mathrm{LC}} - \mathcal{C}_{f_L}$ and $\mathrm{TC} - \mathcal{C}_{f_T}$, respectively. Additionally the rescaled input variables $x_{RMC} = x_1$, $x = x_2$ are given.

Year	2008	2009	2010	2011
$\mathrm{MC} - \mathcal{C}_{f_M}$	59,53205746	50,34329746	44,93332746	57,04016746
$\tilde{\mathrm{LC}} - \mathcal{C}_{f_L}$	27,01459103	25,23117425	23,75897420	26,96904635
$\mathrm{TC} - \mathcal{C}_{f_T}$	22,64490000	5,81340000	8,41220000	19,15044000
$x_1 = x_{RMC}$	1,31901739	1,04570906	1,05442656	1,31451084
$x_2 = x$	0,98558670	0,91024723	0,85718242	0,94131837

method and since this method can fail, when using high numbers, dividing by a factor 10^8 makes the process much more stable. Thus we divide the total variable costs and its variables x_1 and x_2 by this rescaling factor. The result is given in Table 3.2.

Note that the variable costs remain the same, only the interpretation of the variable has changed. The values given in Table 3.2 are used for fitting the parameters of the cost function.

Given these factor prices q_1, q_2, the cost function $\mathcal{C}(x_1, x_2)$ according to (3.11) has the form

$$\mathcal{C}(x_1,x_2) = \left(\delta x_1^{-\rho} + (1-\delta)x_2^{-\rho}\right)^{-\frac{1}{\rho}} \left(\delta^{\frac{1}{1+\rho}} 56,139^{\frac{\rho}{1+\rho}} + (1-\delta)^{\frac{1}{1+\rho}} 27,874^{\frac{\rho}{1+\rho}}\right)^{\frac{1+\rho}{\rho}} + 51,69272 \tag{3.13}$$

Having this important results for the factor prices as well as the fixed costs, we can now start calibrating the cost function.

3.2.2. Calibration of the CES-based cost function based on the given data

At this point we are confronted with the problem of fitting the cost function based on the given data points by a nonlinear least square fit. This procedure is done using the program Mathematica. The function used is FindFit, which approximates the parameter of a defined model which is in our case the cost function, according to given data points. The code for this procedure is given in listing B.1.

As a result we get a list with the parameters fitting the function the best. We find for the values of the parameters, that $\delta = 0,738711$ and $\rho = 0,100836$. A further remark may be added, that is just the best approximation the program did find. Furthermore it is also very sensitive to changes in the model. Thus a very small change of given model parameters can end up in a significant different result. According to the calibrated results, the cost function in (3.13) has the form

$$\mathcal{C}(x_1, x_2) = 51,69272 + \frac{83,1225}{\left(\frac{0,738711}{x_1^{0,100836}} + \frac{0,261289}{x_2^{0,100836}}\right)^{9,91714}}. \tag{3.14}$$

3.2.3. Interpretation of the result

One of the most important aspect in analysing the result of the cost function may be the interpretation of the parameter ρ, since we know that it is closely related to the elasticity of substitution σ via the formula $\sigma = \frac{1}{1+\rho}$. This yields that the elasticity of substitution is $\sigma = 0,908401$. According to theorem 1.18 a ρ value close to zero results in a Cobb-Douglas production function, which has a linear cost function. A plot of the cost function in (3.14) is given in Figure 3.1. One can clearly see, that this function is in an approximation linear. Thus assuming a linear function produces an error.

Concerning the quality of the fit, we can compute the relative error of the data points and the calibrated cost function, which gives information about the existing deviations. By averaging all absolute values of the relative errors over the years, we find an average relative error of approximately $\frac{1}{4}\sum_{i=1}^{4} \frac{\mathcal{C}(x_{1_{t_i}}, x_{2_{t_i}}) - \tilde{E}_{t_i}}{\tilde{E}_{t_i}} \approx 4\%$, which is, compared to the average relative error of approximately 60% made in the firs attempt [Ivaz, 2014], a very good result.

To sum up it may be said that this procedure has both advantages and disadvantages. On the one hand using a numerical nonlinear least square fit, is a very unstable procedure which may also fail. This is obviously the worst case scenario in using this model, making it very unreliable. On the other hand, as we just concluded the model is much more accurate.

Consequently, the next step is to give an approximation of the cost function, such that the nonlinear least square fit, using this special type of function, isn't necessary any more. This would result in a less accurate fit, but a more stable procedure,

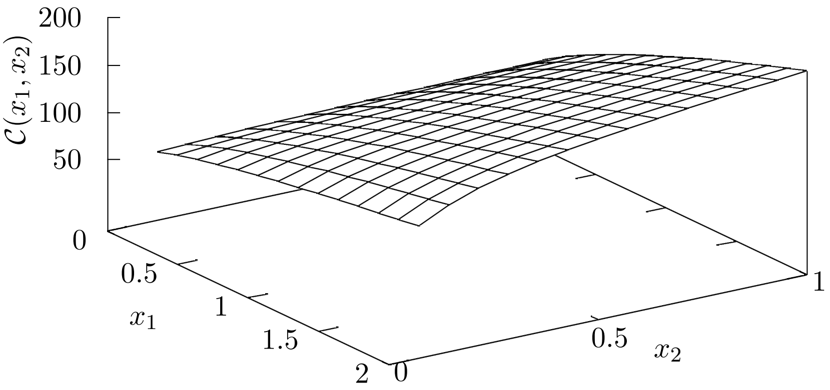


FIGURE 3.1.: The calibrated cost function $\mathcal{C}(x_1, x_2)$ as a function of the production value x_1 and the working hours x_2.

which is needed, to achieve a much more application-oriented method.

3.3. Approximation of the Cost function

In order to find a more stable procedure, we now calibrate the parameters of the cost function, based on an approximative method, using a Taylor expansion. We can, under certain circumstances (see [Königsberger, 2013, p.66]), approximate a function $f : \mathbb{R}^n \to \mathbb{R}$ around a point $\mathbf{a} \in \mathbb{R}^n$ via the formula

$$f(\mathbf{x}) \approx f(\mathbf{a}) + \sum_{i=1}^{n} \frac{\partial f(\mathbf{a})}{\partial x_i}(x_i - a_i) + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial f^2}{\partial x_i \partial x_j}(x_i - a_i)(x_j - a_j),$$

or equivalently

$$f(\mathbf{x}) \approx f(\mathbf{a}) + \mathrm{grad} f(\mathbf{a})(\mathbf{x} - \mathbf{a}) + \frac{1}{2}(\mathbf{x} - \mathbf{a})^{\intercal} H_f(\mathbf{a})(\mathbf{x} - \mathbf{a}), \qquad (3.15)$$

where $(H_f)_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$ denotes the hessian matrix of the function f [Königsberger, 2013, p.60].

Having this in mind, we can fix the two input variables x_1, x_2 and consider the cost function $\mathcal{C}_{x_1,x_2}(\delta, \rho)$ as a function of its parameters δ and ρ, $\mathcal{C}[0,1] \times [-1, \infty) \setminus \{0\} \to \mathbb{R}_+$, $(\delta, \rho) \mapsto \mathcal{C}_{x_1,x_2}(\delta, \rho)$. Thus we can choose a point (δ_m, ρ_m) and use Taylors expansion to approximate the Cost function, using a constant term $\mathcal{C}_{x_1,x_2}(\delta_m, \rho_m)$, a linear function $\mathfrak{L}_{x_1,x_2}(\delta, \rho)$ and a quadratic deviation $\mathfrak{D}_{x_1,x_2}(\delta, \rho)$, such that

$$\mathcal{C}_{x_1,x_2}(\delta, \rho) \approx \mathcal{C}_{x_1,x_2}(\delta_m, \rho_m) + \mathfrak{L}_{x_1,x_2}(\delta, \rho) + \mathfrak{D}_{x_1,x_2}(\delta, \rho)$$

holds. According to Talyor's expansion, we find

$$\mathfrak{L}_{x_1,x_2}(\delta, \rho) = \frac{\partial \mathcal{C}_{x_1,x_2}(\delta_m, \rho_m)}{\partial \delta}(\delta - \delta_m) + \frac{\partial \mathcal{C}_{x_1,x_2}(\delta_m, \rho_m)}{\partial \rho}(\rho - \rho_m) \tag{3.16}$$

$$\begin{aligned}\mathfrak{D}_{x_1,x_2}(\delta, \rho) = \frac{1}{2}\Bigg(&\frac{\partial^2 \mathcal{C}_{x_1,x_2}(\delta_m, \rho_m)}{\partial \delta^2}(\delta - \delta_m)^2 + 2\frac{\partial^2 \mathcal{C}_{x_1,x_2}(\delta_m, \rho_m)}{\partial \delta \partial \rho}(\delta - \delta_m)(\rho - \rho_m) \\ &+ \frac{\partial^2 \mathcal{C}_{x_1,x_2}(\delta_m, \rho_m)}{\partial \rho^2}(\rho - \rho_m)^2\Bigg)\end{aligned} \tag{3.17}$$

At this point, it may also be stated, that this idea follows closely [Kmenta, 1967], where a Taylor approximation of the production function with the initial value $\rho = 0$ is performed, also known as Kmenta's approximation. From this linearised form, the parameters are amenable to estimation by a linear regression analysis, which are then the parameters of the CES-production function. Due to its simplicity, the Kmenta approximation has received a wide acceptance [Mishra, 2006, p.2].

3.3.1. Calculation of the linear term $\mathfrak{L}_{x_1,x_2}(\delta, \rho)$ of Taylors expansion for the cost function

We now have to compute the partial derivatives of $\mathcal{C}_{x_1,x_2}(\delta, \rho)$ with respect to δ and ρ. Doing this we first of all define $\mathcal{C}_{x_1,x_2}(\delta, \rho) - \mathrm{FC} = \mathfrak{C}_{x_1,x_2}(\delta, \rho)$ and find that both partial derivatives are the same, since

$$\frac{\partial \mathfrak{C}_{x_1,x_2}(\delta, \rho)}{\partial \rho} = \frac{\partial(\mathcal{C}_{x_1,x_2}(\delta, \rho) - \mathrm{FC})}{\partial \rho} = \frac{\partial \mathcal{C}_{x_1,x_2}(\delta, \rho)}{\partial \rho}$$

holds true. Due to the fact, that the calculation gets more clearly laid out we rewrite the derivation with the help of $\mathfrak{C}_{x_1,x_2}(\delta,\rho)$. Considering the chain rule yields

$$\begin{aligned}
\frac{\partial \mathfrak{C}_{x_1,x_2}(\delta,\rho)}{\partial \rho} &= \frac{\partial \mathfrak{C}_{x_1,x_2}(\delta,\rho)}{\partial \ln \mathfrak{C}_{x_1,x_2}(\delta,\rho)} \frac{\partial \ln \mathfrak{C}_{x_1,x_2}(\delta,\rho)}{\partial \rho} \\
&= \frac{1}{\frac{\partial \ln \mathfrak{C}_{x_1,x_2}(\delta,\rho)}{\partial \mathfrak{C}_{x_1,x_2}(\delta,\rho)}} \frac{\partial \ln \mathfrak{C}_{x_1,x_2}(\delta,\rho)}{\partial \rho} = \mathfrak{C}_{x_1,x_2}(\delta,\rho) \frac{\partial \ln \mathfrak{C}_{x_1,x_2}(\delta,\rho)}{\partial \rho}
\end{aligned} \tag{3.18}$$

and

$$\frac{\partial \mathfrak{C}_{x_1,x_2}(\delta,\rho)}{\partial \delta} = \mathfrak{C}_{x_1,x_2}(\delta,\rho) \frac{\partial \ln \mathfrak{C}_{x_1,x_2}(\delta,\rho)}{\partial \delta}, \tag{3.19}$$

respectively. With the form of the cost function, given in (3.12), we find

$$\ln \mathfrak{C}_{x_1,x_2}(\delta,\rho) = \ln f^*(\delta,\rho) + \frac{1+\rho}{\rho} \ln \Psi(\delta,\rho).$$

With (3.18), we find

$$\frac{\partial \mathfrak{C}_{x_1,x_2}(\delta,\rho)}{\partial \rho} = \mathfrak{C}_{x_1,x_2}(\delta,\rho) \frac{\partial}{\partial \rho} \left(\ln f^*(\delta,\rho) + \frac{1+\rho}{\rho} \ln \Psi(\delta,\rho) \right).$$

Thus we conclude

$$\begin{aligned}
\frac{\partial \mathfrak{C}_{x_1,x_2}(\delta,\rho)}{\partial \rho} &= \mathfrak{C}_{x_1,x_2}(\delta,\rho) \left(\frac{1}{f^*(\delta,\rho)} \frac{\partial f^*(\delta,\rho)}{\partial \rho} + \ln \Psi(\delta,\rho) \frac{1}{\rho^2} + \frac{(1+\rho)}{\rho} \frac{\partial \ln \Psi(\delta,\rho)}{\partial \rho} \right) \\
&= \mathfrak{C}_{x_1,x_2}(\delta,\rho) \left(\frac{1}{f^*(\delta,\rho)} \frac{\partial f^*(\delta,\rho)}{\partial \rho} + \ln \Psi(\delta,\rho) \frac{1}{\rho^2} + \frac{(1+\rho)}{\rho} \frac{1}{\Psi(\delta,\rho)} \frac{\partial \Psi(\delta,\rho)}{\partial \rho} \right),
\end{aligned}$$

as well as

$$\frac{\partial \mathfrak{C}_{x_1,x_2}(\delta,\rho)}{\partial \delta} = \mathfrak{C}_{x_1,x_2}(\delta,\rho) \left(\frac{1}{f^*(\delta,\rho)} \frac{\partial f^*(\delta,\rho)}{\partial \delta} + \frac{(1+\rho)}{\rho} \frac{1}{\Psi(\delta,\rho)} \frac{\partial \Psi(\delta,\rho)}{\partial \delta} \right).$$

All in all, we find for the first partial derivatives, with respect to δ and ρ

$$
\begin{aligned}
\frac{\partial \mathfrak{C}_{x_1,x_2}(\delta,\rho)}{\partial \delta} &= \frac{\left(\delta^{\frac{1}{1+\rho}} q_1^{\frac{\rho}{1+\rho}} + (1-\delta)^{\frac{1}{1+\rho}} q_2^{\frac{\rho}{1+\rho}}\right)^{\frac{1}{\rho}} \left(\delta x_1^{-\rho} + (1-\delta) x_2^{-\rho}\right)^{-\frac{1}{\rho}}}{(\delta-1)\delta\rho\left((\delta-1)x_1^{\rho} - \delta x_2^{\rho}\right)} \\
&\quad \frac{\left((\delta-1)\delta^{\frac{1}{1+\rho}} q_1^{\frac{\rho}{1+\rho}} x_1^{\rho} + (1-\delta)^{\frac{1}{1+\rho}} \delta q_2^{\frac{\rho}{1+\rho}} x_2^{\rho}\right)}{(\delta-1)\delta\rho\left((\delta-1)x_1^{\rho} - \delta x_2^{\rho}\right)}, \\
\frac{\partial \mathfrak{C}_{x_1,x_2}(\delta,\rho)}{\partial \delta} &= \frac{1}{\rho^2} \left(\delta^{\frac{1}{1+\rho}} q_1^{\frac{\rho}{1+\rho}} + (1-\delta)^{\frac{1}{1+\rho}} q_2^{\frac{\rho}{1+\rho}}\right)^{1+\frac{1}{\rho}} \left(\delta x_1^{-\rho} + (1-\delta) x_2^{-\rho}\right)^{-\frac{1}{\rho}} \\
&\quad \left[\frac{\rho}{1+\rho} \left(-\ln(1-\delta) + \ln q_2 + \frac{\delta^{\frac{1}{1+\rho}} q_1^{\frac{\rho}{1+\rho}} (2\arctan(1-2\delta) + \ln q_1 - \ln q_2)}{\left(\delta^{\frac{1}{1+\rho}} q_1^{\frac{\rho}{1+\rho}} + (1-\delta)^{\frac{1}{1+\rho}} q_2^{\frac{\rho}{1+\rho}}\right)} \right)\right. \\
&\quad - \ln\left(\delta^{\frac{1}{1+\rho}} q_1^{\frac{\rho}{1+\rho}} + (1-\delta)^{\frac{1}{1+\rho}} q_2^{\frac{\rho}{1+\rho}}\right) + \frac{\rho\left(-\delta x_2^{\rho} \ln x_1 + (\delta-1) x_1^{\rho} \ln x_2\right)}{(\delta-1)x_1^{\rho} - \delta x_2^{\rho}} \\
&\quad \left. + \ln\left(\delta x_1^{-\rho} + (1-\delta) x_2^{-\rho}\right)\right].
\end{aligned}
\tag{3.20}
$$

3.3.2. Calculation of the quadratic Deviation $\mathfrak{D}_{x_1,x_2}(\delta,\rho)$ of Taylors expansion for the cost function

Consequently, for computing the term $\mathfrak{D}_{x_1,x_2}(\delta,\rho)$ describing the quadratic deviation according to the Taylor expansion in (3.17), we also need the partial derivatives of second order. Since this calculation is very complex, only the final results of the partial derivatives via the program Mathematica are given in the appendix in listing A.1, A.2 and A.3.

3.3.3. Computation of the initial value for Taylors expansion

Now it comes to choosing the right initial values of (δ_m, ρ_m) for the Taylor expansion, which is estimated, by using the given data points of the form $(x_{1_j}, x_{2_j}, \tilde{E}_j)$, where j denotes the year. Doing this we find a certain x_1, x_2 and cost level, for every given year. It does make sense, to choose the initial value as close as possible to the real values, due to the fact that the Taylor approximation is more accurate.

Consider the case where we have n different data points of the form $(x_{1_j}, x_{2_j}, \tilde{E}_j)$, where $j = t_1, \dots, t_n$ denotes the points of time. So we are able to plot this informations in a $\delta - \rho$ diagram, for every data set j, by a contour plot of this cost level, given by $\mathcal{C}_{x_{1_j}, x_{2_j}}(\delta, \rho) = \tilde{E}_j$.

Consequently we can do this for every set of data points, and have to find a certain (δ_m, ρ_m) value, fitting the best in all contour plots.

Doing this, we have to analyse every data set on its own. Therefore we consider every data set i, where $i \in \{t_1, \dots, t_n\}$. Naturally we can distinguish between three different cases:

(i) $\max_{(\delta,\rho)\in[0,1]\times[-1,\infty)\setminus\{0\}}\{\mathcal{C}_{x_{1_i}, x_{2_i}}(\delta, \rho)\} < \tilde{E}_i$,

(ii) $\max_{(\delta,\rho)\in[0,1]\times[-1,\infty)\setminus\{0\}}\{\mathcal{C}_{x_{1_i}, x_{2_i}}(\delta, \rho)\} = \tilde{E}_i$, and

(iii) $\max_{(\delta,\rho)\in[0,1]\times[-1,\infty)\setminus\{0\}}\{\mathcal{C}_{x_{1_i}, x_{2_i}}(\delta, \rho)\} > \tilde{E}_i$, respectively.

This means that the given re-assessed expenditures, computed on the basis of historical data by using the producer price index, can be greater, less or equal to the maximum of the cost function. The next step in analysing the data, is to order them according to the cases (i)-(iii), given above. We define the set of indices $\mathscr{F} = \{t_m, \dots, t_k\}$ to be representatives of the first case (i), $\mathscr{S} = \{t_{k+1}, \dots, t_l\}$ of the second case (ii) and $\mathscr{T} = \{t_{l+1}, \dots, t_n\}$ of the third case (iii), respectively.

Concerning (i) and (ii), we will obviously choose

$$(\delta_{m_j}, \rho_{m_j}) = \underset{(\delta,\rho)\in[0,1]\times[-1,\infty)\setminus\{0\}}{\operatorname{argmax}} \{\mathcal{C}_{x_{1_j}, x_{2_j}}(\delta, \rho)\},$$

referring to the maximum itself as the best estimation, for all $j \in \mathscr{F}$ and $j \in \mathscr{S}$. Averaging them over the associated years yields a first approximation

$$(\bar{\delta}_{m_I}, \bar{\rho}_{m_I}) = \frac{1}{\#\mathscr{F}} \sum_{j\in\mathscr{F}} \underset{(\delta,\rho)\in[0,1]\times[-1,\infty)\setminus\{0\}}{\operatorname{argmax}} \{\mathcal{C}_{x_{1_j}, x_{2_j}}(\delta, \rho)\}, \tag{3.21}$$

where $\#\mathscr{F}$ is called the cardinality of the set and denotes the number of elements included. In the same way, we find for the second case

$$(\bar{\delta}_{m_{II}}, \bar{\rho}_{m_{II}}) = \frac{1}{\#\mathscr{S}} \sum_{j\in\mathscr{S}} \underset{(\delta,\rho)\in[0,1]\times[-1,\infty)\setminus\{0\}}{\operatorname{argmax}} \{\mathcal{C}_{x_{1_j}, x_{2_j}}(\delta, \rho)\},$$

TABLE 3.3.: The maximum of the cost function $\mathcal{C}_{x_1,x_2}(\delta,\rho)$, with the restrictions $0 \leq \delta \leq 1$ and $-1 \leq \rho$, $\rho \neq 0$, as well as its position.

Year	2008	2009	2010	2011
$\max_{(\delta,\rho)\in[0,1]\times[-1,\infty)\setminus\{0\}} \mathcal{C}_{x_1,x_2}(\delta,\rho)$	153,213	135,77	134,78	151,726
$\operatorname{argmax}_{(\delta,\rho)\in[0,1]\times[-1,\infty)\setminus\{0\}} \mathcal{C}_{x_1,x_2}(\delta,\rho)$	$\begin{pmatrix}0,740706\\0,199386\end{pmatrix}^T$	$\begin{pmatrix}0,70766\\0,325618\end{pmatrix}$	$\begin{pmatrix}0,727565\\0,362457\end{pmatrix}$	$\begin{pmatrix}0,761577\\0,381116\end{pmatrix}^T$

Ad (iii) the contour plot of a certain data set $i \in \{t_{l+1},\ldots,t_n\}$ will be a curve in the $\delta-\rho$ plane. Consequently, we can do this for every data set referring to case (iii). Thus we find $n-l$ different curves, which might or might not intersect with each other and we choose the point $(\bar{\delta}_{m_{III}}, \bar{\rho}_{m_{III}})$, fitting the best.

Averaging these values with equal weights refers to the final initial value for the Taylor approximation (δ_m,ρ_m).

3.3.4. Computing the Taylor expansion based on the given data

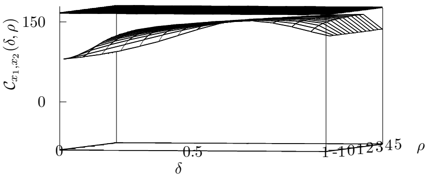
Having this theoretical foundation, we can now start computing the Taylor approximation according to the given data used in the previous chapter. Here the index i can take values from the set $\{t_1,t_2,t_3,t_4\} = \{2008, 2009, 2010, 2011\}$, corresponding to the years of the data set. First of all we have to allocate the years to the different cases occurring, described in chapter 3.3.3. To verify this, a plot of the corresponding costs for the concerning years is given in Figure 3.2.

Therefore we find that $\{2008, 2009, 2011\} = \{t_1,t_2,t_4\}$ refer to case (i) and $\{2010\} = \{t_3\}$ refers to case (iii). A numerical analysis of the cost function according to its maximum yields the maximum value and its position, given in Table 3.3. Additionally the code for finding these maximums using Mathematica, is given in listing B.2.

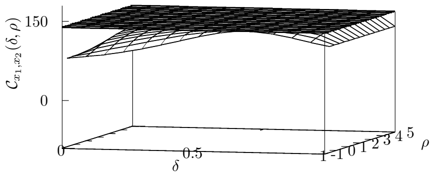
According to (3.21), we find

$$(\bar{\delta}_{m_I}, \bar{\rho}_{m_I}) = (0,736648, 0,30204).$$

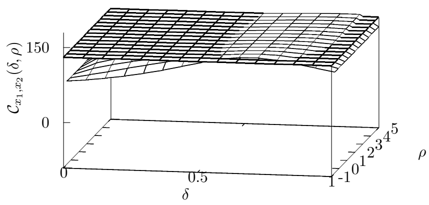
Analysing the year 2010, we find that it refers to case (iii). To illustrate this situation, we make a contour plot of $\mathcal{C}_{x1_{t_3},x2_{t_3}}(\delta,\rho)$ for the contour level of $\tilde{E}_{t_3}$, which



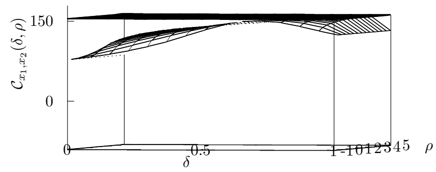
(A) $t_1 = 2008$, case (i)



(B) $t_2 = 2009$, case (i)



(C) $t_3 = 2010$, case (iii)



(D) $t_4 = 2011$, case (i)

FIGURE 3.2.: The cost function $\mathcal{C}_{x_1,x_2}(\delta, \rho)$ for the different years 2008-2011 in (A)-(D). Additionally the assignment to its associated cases is given.

is given in Figure 3.3. All points on the dotted line refer to the same costs. Since we have have only one candidate refering to case (iii), we just round the primary estimation such that it fits the values of the dotted curve better. Thus we just round $(\bar{\delta}_{m_I}, \bar{\rho}_{m_I})$, to get

$$(\delta_m, \rho_m) = (0.7, 0.3),$$

which is the final initial value for the Taylor approximation.

We now want to fit the model of the form $\mathcal{C}_I(\delta, \rho) = \mathcal{C}(\bar{\delta}_m, \bar{\rho}_m) + \mathfrak{L}(\delta, \rho) + \mathfrak{D}(\delta, \rho)$ using the given data. According to the partial derivatives given in (3.20) and the appendix, we find the best approximation for the parameters to be

$$(\delta, \rho) = (0,714658, -0,999998),$$

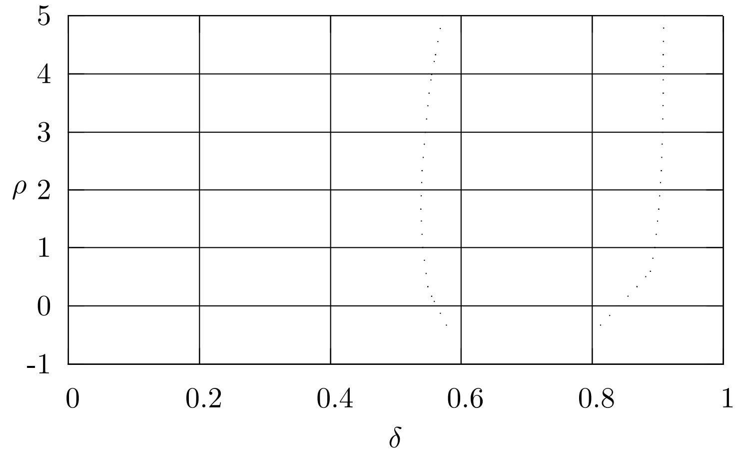


FIGURE 3.3.: Contour plot of the cost function $\mathcal{C}_{x1_{t_3},x2_{t_3}}(\delta,\rho) = \tilde{E}_{t_3}$ in the year 2010 in the $\delta - \rho$ plane.

which is again done by using the Mathematica implemented method FindFit, given in listing B.3.

We find that the best approximation is $\rho \approx -1$. Thus, we get

$$\mathcal{C}_I(x_1,x_2) = 51,6927 + 78,5537(0,714658x_1^{0,999998} + 0,285342x_2^{0.999998})^{1,0000018688167596}.$$

3.3.5. Interpretation of the result

Comparing this result $\mathcal{C}_I(x_1,x_2)$ with the cost function $\mathcal{C}(x_1,x_2)$ we can take a closer look at the associated deviations $\Sigma(x_1,x_2) = \mathcal{C}_I(x_1,x_2) - \mathcal{C}(x_1,x_2)$. Both plots of $\mathcal{C}_I(x_1,x_2)$ and $\Sigma(x_1,x_2)$ are given in Figure 3.4(A) and Figure 3.4(B), respectively.

In the region, where almost all data points $x1_{t_i}$ and $x2_{t_i}$ are located, this is around its mean value $\bar{x}_1 = \sum_{j=t_1}^{t_4} x1_j = 1,18342$ and $\bar{x}_2 = \sum_{j=t_1}^{t_4} x2_j = 0,923584$, the deviations become smaller, as it can be seen from Figure 3.5. In fact, the deviation at the center of the data points, which is the mean value, is $\Sigma(\bar{x}_1,\bar{x}_2) \approx -5$.

All in all we can state, that we found a way to give a good estimation of the parameters δ and ρ for the CES cost function $\mathcal{C}_I(x_1,x_2)$. The main advantage concerning this method is, the reduced complexity of the model, due to a Taylor expansion

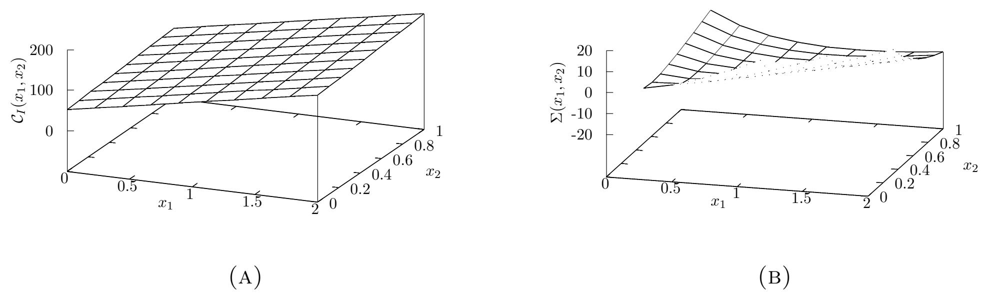
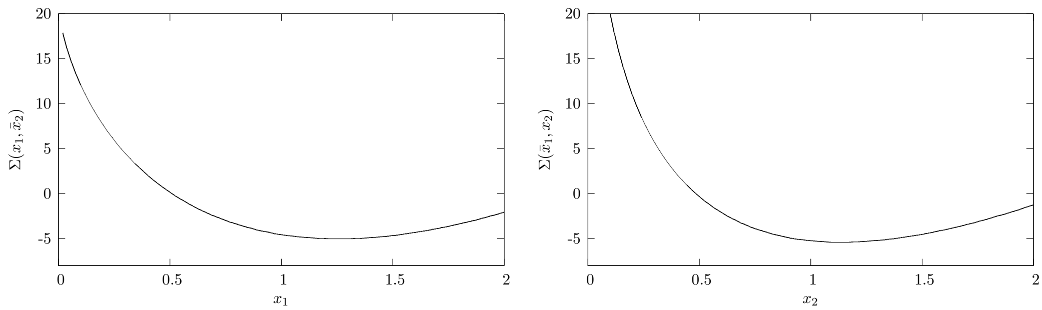


FIGURE 3.4.: (A) Cost function $\mathcal{C}_I(x_1, x_2)$ based on the estimation of the parameters δ and ρ from a regression model. (B) Deviations $\Sigma(x_1, x_2)$ of the cost function $\mathcal{C}_I(x_1, x_2)$ and the fitted cost function $\mathcal{C}(x_1, x_2)$

of second order. Hence fitting the function to the according data points becomes much more stable than performing the fit with the original form of the cost function $\mathcal{C}(x_1, x_2)$. It is also necessary to point out that the quality of the fit, speaking in terms of deviations to the given data points, reduces in order to make the model more application-oriented.

One should also be aware of the fact, that we have only four data points available, to preform the fit of the function $\mathcal{C}(x_1, x_2)$. But this method will certainly work better in the future, when the analyst has more data points available.



(A) $x_2 = \bar{x}_2$ (B) $x_1 = \bar{x}_1$

FIGURE 3.5.: Figure (A) shows the deviations $\Sigma(x_1, \bar{x}_2)$ for a constant x_2-level, namely the mean value $\bar{x}_2$ of all data points. Figure (B) shows the deviations $\Sigma(\bar{x}_1, x_2)$ for a constant x_1-level, which is the mean value $\bar{x}_1$ of all data points.

Part II.

Time series analysis

CHAPTER 4

Theoretical foundations of Time Series Analysis

This chapter introduces the method of time series analysis to predict certain values in the future based on an ARIMA(p,d,q) model. Therefore, all necessary basic ideas and definitions are given and explained. Further more a step by step instruction in setting up a general ARIMA(p,d,q) model including model selection and parameter estimation based on a maximum likelihood estimation, is given.

4.1. Introduction

The first step in performing a time series analysis is to select a certain model for the given data. Considering the unpredictable nature of future, it is reasonable to construct this models with the help of random variables X_t, where t denotes the point of time, with a certain realisation (also denoted with X_t) which represents the actual data points [Brockwell and Davis, 2013, p.8].

4.1 DEFINITION. [Neusser, 2009, p.8] A stochastic process is a sequence of random variables $\{\mathrm{X}_t\}_{t\in\mathscr{T}}$, where $\mathscr{T} \subseteq \mathbb{Z}$.

Generally, a random variable describes the realisation of a certain outcome whereas the set $\mathscr{T}$ is used to indicate the time period in which this outcome was measured.

Since this point of times are discrete, we want the set to be a subset of Integers. From here on we will always choose $\mathscr{T} = \mathbb{Z}$, since it allows a more elegant way to work out different concepts. It is straightforward to define every method used, on a certain subset of $\mathbb{Z}$. A time series is then simply a certain realisation $\{\mathrm{X}_t\}_{t\in\mathbb{Z}}$ of this stochastic process [Neusser, 2009, p.9].

It is necessary to point out, that we use time series analysis to investigate the connection and influence of the realisation of this random numbers in different points of time. It is for example reasonable to assume that the production costs of a certain year somehow influences the production costs of the next period of time.

4.2 Definition. [Neusser, 2009, p.10] Let $t, s \in \mathbb{Z}$ and $\{\mathrm{X}_t\}_{t\in\mathbb{Z}}$ be a stochastic process, with $V[X_t] < \infty$, for all $t \in \mathbb{Z}$, where $V[.]$ denotes the variance and $E[.]$ denotes the expected value. The function $\gamma_{\mathrm{X}} : \mathbb{Z} \times \mathbb{Z} \to \mathbb{R}$, defined as

$$\gamma_{\mathrm{X}}(t, s) = \mathrm{cov}(\mathrm{X}_t, \mathrm{X}_s) = E[\mathrm{X}_s\mathrm{X}_t] - E[\mathrm{X}_s]E[\mathrm{X}_t],$$

is called autocovariancefunction of X_t.

For practical reasons further simplifications have to be assumed. We will therefore define a special class of stochastic processes, called stationary stochastic processes.

4.3 Definition. [Neusser, 2009, p.11] A stochastic process $\{\mathrm{X}_t\}_{t\in\mathbb{Z}}$ is called stationary, if for all $t, s, r \in \mathbb{Z}$

(i) $E[\mathrm{X}_t] = \mu$,

(ii) $V[\mathrm{X}_t] < \infty$ and

(iii) $\gamma_{\mathrm{X}}(t, s) = \gamma_{\mathrm{X}}(t + r, s + r)$

holds true.

Roughly speaking a stationary process has the property of the same mean, variance and autocovariance structure. It is, for example, possible to translate all points of time, by a number $r \in \mathbb{Z}$ and the autocovariancefunction does not change at all, since property (iii) holds.

Considering a stationary stochastic process we find, by choosing $r = -s$, that $\gamma_{\mathrm{X}}(t, s) = \gamma_{\mathrm{X}}(t - s, 0)$, again because property (iii) holds. Thus the autocovariancefunction does not depend on the absolute values of time t and s itself but on its difference $h = t - s$. Thus we can ignore the second argument, since it is always zero and define the autocovariancefunction via $\gamma_X(h) := \gamma_X(h, 0)$. It may also be pointed out, that since $\gamma_X(t, s) = \gamma_X(s, t)$, also $\gamma_X(h) = \gamma_X(-h)$ holds true. Therefore it makes sense, to define the autocovariancefunction only on the set of positive integers of $\mathbb{Z}$ [Neusser, 2009, p.12].

Now consider a finite time series $\{X_t\}_{t\in[1,T]}$ and its associated covariances. Writing them into a matrix of the form

$$\Gamma_T = \begin{pmatrix} \mathrm{cov}(X_1, X_1) & \mathrm{cov}(X_1, X_2) & \cdots & \mathrm{cov}(X_1, X_T) \\ \mathrm{cov}(X_2, X_1) & \mathrm{cov}(X_2, X_2) & \cdots & \mathrm{cov}(X_2, X_T) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(X_T, X_1) & \cdots & & \mathrm{cov}(X_T, X_T) \end{pmatrix} \tag{4.1}$$

defines the covariance matrix [Arens et al., 2008, p.1310]. Note that for stationary processes the covariance matrix takes the form

$$\Gamma_T = \begin{pmatrix} \gamma_X(0) & \gamma_X(1) & \cdots & \gamma_X(1-T) \\ \gamma_X(1) & \gamma_X(0) & \cdots & \gamma_X(2-T) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_X(1-T) & \cdots & & \gamma_X(0) \end{pmatrix}. \tag{4.2}$$

4.4 Definition. [Neusser, 2009, p.13] Let $h \in \mathbb{Z}$. The autocorrelationfunction (ACF) of $\{X_t\}_{t\in\mathbb{Z}}$ is defined as

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}.$$

It may also be added that the ACF of stationary time series represents just the normal correlation coefficient, defined by

$$\varrho(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sqrt{V[X]}\sqrt{V[Y]}},$$

between the corresponding random variables. Hence all characteristics of the correlation coefficient can be transferred to the ACF.

The concept of time series analysis is to describe processes with the most easiest modules available. One of these easiest modules is the white noise process.

4.5 Definition. [Neusser, 2009, p.14] The process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is called white noise, if $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is stationary and

(i) $E[\varepsilon_t] = 0$, as well as

(ii) $\gamma_\varepsilon(h) = \begin{cases} \sigma^2 & \text{for } h = 0 \\ 0 & \text{for } h \neq 0, \end{cases}$

holds true. This is denoted with $\varepsilon_t \sim \mathrm{WN}(0, \sigma^2)$.

One of the easiest time series that can be created using white noise, is the so called moving average process.

4.6 Definition. [Neusser, 2009, p.14] Let $\theta \in \mathbb{R}$ be a parameter and $\varepsilon_t \sim \mathrm{WN}(0, \sigma^2)$. The process, defined by

$$\mathrm{X}_t = \varepsilon_t + \theta\varepsilon_{t-1},$$

is called a moving average process, denoted with MA(1).

4.7 Example. [Brockwell and Davis, 2013, p.13] For the sake of an example, we will now compute the autocovariancefunction as well as the ACF of a MA(1)-process. Let us start with

$$\begin{aligned}
\gamma_{\mathrm{X}}(t+h,t) &= \mathrm{cov}(\mathrm{X}_{t+h}, \mathrm{X}_t) = \mathrm{cov}(\varepsilon_{t+h} + \theta\varepsilon_{t+h-1}, \varepsilon_t + \theta\varepsilon_{t-1}) \\
&= E\left[(\varepsilon_{t+h} + \theta\varepsilon_{t+h-1})(\varepsilon_t + \theta\varepsilon_{t-1})\right] - E[\varepsilon_t + \theta\varepsilon_{t-1}]E[\varepsilon_{t+h} + \theta\varepsilon_{t+h-1}] \\
&= E[\varepsilon_{t+h}\varepsilon_t + \theta\varepsilon_{t+h}\varepsilon_{t-1} + \theta\varepsilon_{t+h-1}\varepsilon_t + \theta^2\varepsilon_{t+h-1}\varepsilon_{t-1}] \\
&\quad - (E[\varepsilon_t]E[\varepsilon_{t+h}] + \theta E[\varepsilon_t]E[\varepsilon_{t+h-1}] + \theta^2 E[\varepsilon_{t-1}]E[\varepsilon_{t+h-1}]).
\end{aligned}$$

Using a case-by-case analysis we find by using the linearity of the expected value and with $E[\varepsilon_t^2] = \sigma^2, E[\varepsilon_t] = 0$ as well as $E[(\varepsilon_{t+r}\varepsilon_{t+s})] = 0,\ r, s \in \mathbb{Z}, r \neq s$

$$\gamma_{\mathrm{X}}(t+h,t) = \begin{cases} \sigma^2(1+\theta^2), & \text{for } h = 0 \\ \theta\sigma^2, & \text{for } h = \pm 1 \\ 0, & \text{else.} \end{cases} \tag{4.3}$$

By using, that $\gamma_{\mathrm{X}}(0) = \gamma_{\mathrm{X}}(t,t) = \sigma^2(1+\theta^2)$, we can also compute the ACF of a MA(1) process:

$$\rho_{\mathrm{X}}(h) = \begin{cases} 1, & \text{for } h = 0 \\ \frac{\theta}{(1+\theta^2)}, & \text{for } h = \pm 1 \\ 0, & \text{else.} \end{cases}$$

4.2. Models for stationary time series analysis

One of the most important models, with which a wide class of problems can be described is the autoregressive moving average model, denoted with ARMA(p,q).

4.8 Definition. [Brockwell and Davis, 2013, p.78] Let $\{\mathrm{X}_t\}_{t\in\mathbb{Z}}$ be a stochastic process. The process $\{\mathrm{X}_t\}_{t\in\mathbb{Z}}$ is said to be an ARMA(p,q) process, if $\{\mathrm{X}_t\}_{t\in\mathbb{Z}}$ is stationary and if for every $t \in \mathbb{Z}$

$$\mathrm{X}_t - \phi_1 \mathrm{X}_{t-1} - \ldots - \phi_p \mathrm{X}_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}, \tag{4.4}$$

with $\varepsilon_t \sim \mathrm{WN}(0,\sigma^2)$ holds true.

It may also be added, that sometimes a constant $c \in \mathbb{R}$ is added to the model given in (4.4). For the sake of a compact notation, we further more define a certain operator, called the lag operator, to rewrite (4.4) in a more convenient form.

4.9 Definition. [Neusser, 2009, p.21] The lag operator, or back shift operator $\mathbb{L}$ has the following properties:

(i) for $c \in \mathbb{R}$, using the lag operator yields $\mathbb{L}(c) = c$,

(ii) for a stochastic process $\{\mathrm{X}_t\}_{t\in\mathbb{Z}}$, using the lag operator yields $\mathbb{L}(\mathrm{X}_t) = \mathrm{X}_{t-1}$,

(iii) using the lag operator n-times yields $(\mathbb{L} \circ \ldots \circ \mathbb{L})(\mathrm{X}_t) = \mathbb{L}^n(X_t) = \mathrm{X}_{t-n}$,

(iv) for $m, n \in \mathbb{Z}$, using the lag operator yields $(\mathbb{L}^m \circ \mathbb{L}^n)(\mathrm{X}_t) = \mathbb{L}^{m+n}(\mathrm{X}_t) = \mathrm{X}_{t-m-n}$,

(v) for $a, b \in \mathbb{R}$ and another stochastic process $\{Y_t\}_{t\in\mathbb{Z}}$, using the lag operator yields $\mathbb{L}(a\mathrm{X}_t + b\,Y_t) = a\mathrm{X}_{t-1} + b\,Y_{t-1}$.

To rewrite (4.4) in a more compact form, we define the polynomials $\Phi(\mathbb{L}) = 1 - \sum_{i=1}^{p} \phi_i \mathbb{L}^i$ and $\Theta(\mathbb{L}) = 1 + \sum_{i=1}^{q} \theta_i \mathbb{L}^i$. Hence we can write a general ARMA(p,q) process in a more compact notation:

$$\Phi(\mathbb{L})(\mathrm{X}_t) = \Theta(\mathbb{L})(\varepsilon_t).$$

Therefore it is possible to determine the properties of a model for a time series analysis, using this polynomials given above.

Some very useful special cases which can be construct out of a general ARMA(p,q) process are given by a certain choice of p and q, respectively. Choosing $p = 0$, yields a moving average process of order q, denoted with MA(q)=ARMA(0,q). On the other hand, setting $q = 0$ yields a so called autoregressive process of order p, denoted with AR(p)=ARMA(p,0) [Neusser, 2009, p.23].

It turns out, that many economic time series aren't stationary. Thus we have to apply a transformation first, such that the time series becomes stationary. In many applications differencing the values, by a certain degree d, or taking the logarithm, yields the necessary properties. Hence the transformation has the form

$$Y_t = (1 - \mathbb{L})^d \mathrm{X}_t.$$

With this new time series $\{Y_t\}_{t \in \mathbb{Z}}$ we can set up the ARMA(p,q) model as described above. This model is then called an ARIMA(p,d,q) model.

4.10 Definition. [Neusser, 2009, p.83] Let $\{\mathrm{X}_t\}_{t \in \mathbb{Z}}$ be a stochastic process. The process $\{\mathrm{X}_t\}_{t \in \mathbb{Z}}$ is said to be an ARIMA(p,d,q) process, if $\{(1 - \mathbb{L})^d \mathrm{X}_t\}_{t \in \mathbb{Z}}$ is stationary and if for every $t \in \mathbb{Z}$

$$\Phi(\mathbb{L})((1 - \mathbb{L})^d \mathrm{X}_t) = \Theta(\mathbb{L})(\varepsilon_t)$$

holds true.

4.3. Estimation of the autocovariancefunction

From this point on all methods described will make use of the autocovariancefunction $\gamma_X(h)$ of stationary processes. Since we have only T data points of the form

$\{X_t\}_{t\in[1,T]}$ available and have no information according the probability distributions, we have to estimate the autocovariance function based on this data points. With $\bar{X}_T = \frac{1}{T}\sum_{j=1}^{T} X_j$ the estimation for the autocavariance function we will use is

$$\hat{\gamma}_X(h) = \frac{1}{T}\sum_{j=1}^{T-h}(X_{t+h} - \bar{X}_T)(X_t - \bar{X}_T), \tag{4.5}$$

where $\hat{\gamma}_X(h) = 0$ for $h \geqslant T$, see [Kreiß and Neuhaus, 2006, p.32].

4.4. Estimating time series models

In practice, we are often confronted with the problem of estimating the parameters of a certain model, given a set of observables $\{X\}_{t\in[1,T]}$, which are already corrected by its mean value, by applying the transformation $\{X_t\}_{t\in[1,T]} \mapsto \{X_t - \bar{X}_T\}_{t\in[1,T]}$. In fact there exist very good methods to determine the parameters $\phi_1, \ldots, \phi_p$ and $\theta_1, \ldots, \theta_q$ of special ARMA(p,q) models, namely the Yule-Walker estimation and an OLS estimation. Both methods turn out to work really well considering a general AR(p) process, which is a special choice of an ARMA(p,q) model, but fail, when it comes to estimate the parameters of a general ARMA(p,q) model. Then in most cases, a maximum-likelihood estimation is performed [Neusser, 2009, p.71].

4.4.1. The maximum Likelihood method

Since we are using the maximum likelihood method to estimate the parameters, a short introduction of this technique is given. This section follows closely the book of [Fahrmeir et al., 2013].

Given the case of $\{X_i\}_{i\in[1,n]}$ identical and independent replications of an experiment and assuming a certain probability distribution $f_i(X_i|\theta_i)$ for each X_i, where θ_i describes the parameters of the distribution, the probability density function of all n experiments is given by

$$f(X_1, \ldots, X_n|\theta) = \prod_{j=1}^{n} f_j(X_j|\theta_j). \tag{4.6}$$

Considering the actual realisation of those n experiments it is obvious, that the parameters θ_i in (4.6) are the only variables left. We will denote $L(\theta) = f(\mathrm{X}_1, \ldots, \mathrm{X}_n|\theta)$ as the likelihood function, and interpret it as the probability density function of the parameters θ_i of the probability distributions, where $i = 1, \ldots, n$.

The maximum likelihood principle implies that the best estimation for the parameter θ, which is a vector containing all θ_i, denoted by $\hat{\theta}$ is found, where the likelihood function has its maximum,

$$L(\hat{\theta}) = \max_{\theta} L(\theta).$$

For practical reasons, working with the ln-likelihood function is easier, since the position of the optimal value $\hat{\theta}$ does not change using a ln-transformation.

4.11 Example. In many theoretical and practical settings, we find ourselves in the situation of each X_j being normally distributed with mean μ and variance σ^2, where $j = 1, \ldots, n$. This is denoted with $\mathrm{X}_j \sim \mathcal{N}(\mu, \sigma^2)$. With $\theta = (\mu, \sigma)$ we find for the likelihood function

$$L(\mu, \sigma) = \frac{1}{(\sqrt{2\pi\sigma})^n} \exp\left(\frac{(\mathrm{X}_1 - \mu)^2}{2\sigma^2}\right) \cdot \ldots \cdot \exp\left(\frac{(\mathrm{X}_n - \mu)^2}{2\sigma^2}\right). \tag{4.7}$$

As already stated above, using a ln-transformation yields

$$\ln L(\mu, \sigma) = \sum_{i=1}^{n} \left(- \ln(\sqrt{2\pi}) - \ln \sigma - \frac{(\mathrm{X}_i - \mu)^2}{2\sigma^2} \right).$$

We are now looking for a maximum, thus we have to solve the equations

$$\left.\frac{\partial \ln L(\hat{\mu}, \hat{\sigma})}{\partial \mu}\right|_{(\hat{\mu}_{ML}, \hat{\sigma}_{ML})} = \sum_{i=1}^{n} \frac{(\mathrm{X}_i - \hat{\mu}_{ML})}{\hat{\sigma}^2} = 0$$

$$\frac{\partial \ln L(\hat{\mu}, \hat{\sigma})}{\partial \sigma} = \sum_{i=1}^{n} \left(\frac{1}{\hat{\sigma}_{ML}} + \frac{(\mathrm{X}_i - \hat{\mu}_{ML})^2}{\hat{\sigma}_{ML}^3} \right) = 0.$$

The solution of this system yields the maximum likelihood estimations of both parameters, $\hat{\mu}_{ML} = \frac{1}{n} \sum_{i=1}^{n} \mathrm{X}_i = \bar{\mathrm{X}}$, which is simply the mean value of the data set and $\hat{\sigma}_{ML} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\mathrm{X}_i - \bar{\mathrm{X}})}$.

4.4.2. Estimation of a ARMA(p,q) process based on a maximum likelihood estimation

Considering a general ARMA(p,q) process, given by $\Phi(\mathbb{L})(\mathrm{X}_t) = \Theta(\mathbb{L})(\varepsilon_t)$, estimating the parameters ϕ_i, θ_j, $i = 1, \ldots, p$ and $j = 1, \ldots, q$ of the model becomes more difficult. We now make the assumption that the white noise $\{\varepsilon_t\}_{t\in\mathbb{Z}} \sim \mathrm{WN}(0, \sigma^2)$ is additionally independently and identically distributed and the polynomials $\Phi(z)$ and $\Theta(z)$ don't have the same zeros, for theoretical reasons. Summing up, the parameters of the model are given by

$$\beta = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)^T \quad \text{and} \quad \sigma^2,$$

respectively. Other theoretical considerations [Neusser, 2009, p.77] require that possible values for β are elements of the set $\mathscr{P}$

$$\mathscr{P} = \{\beta \in \mathbb{R}^{p+q} : \Phi(z)\Theta(z) \neq 0 \text{ for } |z| \leqslant 1,\ \phi_p\theta_q \neq 0 \\ \text{and } \Phi(z), \Theta(z) \text{ don't have same zeros}\}.$$

According to the maximum likelihood method the parameters β and σ are chosen, such that the probability that the actual data points $\mathrm{X}_T = (\mathrm{X}_1, \ldots, \mathrm{X}_T)$ occur is maximized under the assumption of a joint distribution of the random variables $\mathrm{X}_1, \ldots, \mathrm{X}_T$. The most important case is represented by the random variables X_i being multivariate normally distributed with zero mean. Setting $G_T(\beta)\sigma^2 = \Gamma_T$, where Γ_T denotes the covariance matrix (4.1), the likelihood function is

$$L(\beta, \sigma^2|\mathrm{X}) = \frac{1}{(2\pi\sigma^2)^{\frac{T}{2}}} \frac{1}{\sqrt{\det G_T(\beta)}} \exp\left(-\frac{1}{2\sigma^2}\mathrm{X}_T^{\intercal} G_T^{-1}(\beta)\mathrm{X}_T\right), \tag{4.8}$$

see [Neusser, 2009, p.77]. Again, it is easier to work with the log likelihood function. Hence taking the logarithm of (4.8) yields

$$\ln L(\beta, \sigma^2|\mathrm{X}) = -\frac{T}{2}\ln(2\pi\sigma^2) - \frac{1}{2}\ln(\det G_T(\beta)) - \frac{1}{2\sigma^2}\mathrm{X}_T^{\intercal} G_T^{-1}(\beta)\mathrm{X}_T. \tag{4.9}$$

According to the maximum likelihood method we now have to differentiate (4.9) with respect to σ and set the term to zero. Thus we find

$$\frac{\partial \ln L_t(\beta, \sigma^2|\mathrm{X})}{\partial \sigma}\Big|_{(\hat{\sigma}^2_{ML}, \hat{\beta}_{ML})} = -\frac{T}{\hat{\sigma}_{ML}} + \frac{1}{\hat{\sigma}^3_{ML}} \mathrm{X}_T^\intercal G_T^{-1}(\hat{\beta}_{ML})\mathrm{X}_T = 0,$$

concluding that

$$\hat{\sigma}^2_{ML} = \frac{1}{T}\mathrm{X}_T^\intercal G_T^{-1}(\hat{\beta}_{ML})\mathrm{X}_T. \tag{4.10}$$

Since we now found an expression for the maximum value of the maximum likelihood estimation of σ^2, we can simply maximize (4.9) using relation (4.10). Thus we have to solve the optimisation problem

$$\max_{\beta \in \mathscr{P}} \left[-\ln(2\pi) - \frac{T}{2} \ln\left(\frac{1}{T}\mathrm{X}_T^\intercal G_T(\beta)^{-1}\mathrm{X}_T\right) - \frac{1}{2}\ln\det G_T(\beta) - \frac{T}{2}\right],$$

which is equivalent to minimise the same function multiplied with a factor of minus one and leaving the constant terms out. Hence we have to minimise

$$\min_{\beta \in \mathscr{P}} \left[\frac{T}{2} \ln\left(\frac{1}{T}\mathrm{X}_T^\intercal G_T(\beta)^{-1}\mathrm{X}_T\right) + \frac{1}{2}\ln G_T(\beta)\right], \tag{4.11}$$

which can be done numerically and refers to the maximum likelihood estimator $\hat{\beta}_{ML}$. As a matter of fact there exist many different ways of estimating the parameters but in theory the most common method is the maximum likelihood estimation. In practical applications however, the minimisation problem is transferred via a certain algorithm (see [Neusser, 2009, p.78]) to another minimisation problem, being more stable from a numerical point of view.

4.4.3. Estimating the order (p,q) of an ARMA(p,q) model

The first step in estimating the parameters of an ARMA(p, q) model, based on a set of observables $\{\mathrm{X}_t\}_{t\in[1,T]}$, is to find the orders p and q of the model. Concerning this procedure, generally two possible mistakes may occur. The first one is called over fitting, when both parameters p and q are too high than the actual value. On the other hand either p or q can be chosen too small, which is called under fitting [Neusser, 2009, p.80].

All in all one can estimate the order of the parameters by analysing the autocorrelation function, but nowadays automatic selection processes are performed, where a certain function, called the information criterion, is minimised. The most important minimisation problems of this information criterion have the form

$$\min_{(p,q)\in\mathbb{N}\times\mathbb{N}} \left[\ln \hat{\sigma}^2_{p,q} + (p+q)\frac{C(T)}{T}\right], \tag{4.12}$$

where $\hat{\sigma}^2_{p,q}$ denotes the estimated variance of the residuals, given by the estimation of σ of the maximum of the likelihood function [Hannan, 1980]. In practice three different forms of the function $C(T)$ in (4.12) are used:

(i) $C(T) = 2$ is called the Akaike information criterion AIC,

(ii) $C(T) = \ln(T)$ is called the Bayesian information criterion BIC and

(iii) $C(T) = 2\ln(\ln T)$ is called the Hannan-Quinn information criterion,

see [Neusser, 2009, p.82].

It may also be stated, that in most cases the AIC number is used, where AIC is a numerical value, by which to rank competing models in terms of information loss in approximating the unknowable truth. It derives meaning from comparison with the AIC values of other models with the model having the lowest AIC value representing the best approximating model [Symonds and Moussalli, 2011, p.14].

Concerning different approaches in the literature [Kreiß and Neuhaus, 2006, p.288], the AIC function may also be defined with the help of the likelihood function itself as

$$AIC(p,q) = -\frac{2}{n}L_T\left(\hat{\beta}_{ML}, \hat{\sigma}^2_{ML}|\mathrm{X}\right) + 2\frac{(1+p+q)}{T},$$

or with the help of a linear prediction using $\hat{\sigma}^2_{p,q} = \frac{\mathrm{RSS}}{T-p-q}$,

$$AIC(p,q) = T\ln\frac{\mathrm{RSS}}{T-p-q} + 2(p+q), \tag{4.13}$$

see [Maddala and Lahiri, 1992, p.540], where RSS denotes the regression sum of squares. This is given by the formula

$$\mathrm{RSS} = \sum_{i=1}^{T} \left(\mathrm{X}_i - (a + bt_i)\right)^2 , \tag{4.14}$$

where the coefficients a and b are given by

$$a = \bar{\mathrm{X}} - b\bar{t} \quad \text{and} \quad b = \frac{\sum_{j=1}^{T}(\mathrm{X}_i - \bar{\mathrm{X}})(t_i - \bar{t})}{\sum_{k=1}^{T}(\mathrm{X}_k - \bar{\mathrm{X}})}, \tag{4.15}$$

respectively.

4.4.4. Summary of setting up an ARMA(p,q) model

Having all individual steps in setting up an ARMA(p,q) model, we are now ready to combine them to give a short guidance of performing a time series analysis, which is illustrated in Figure 4.1.

We begin by an observation of the given data, denoted by $\{\mathrm{X}_t\}_{t\in[1,T]}$. Assuming that this time series is stationary, we now have to estimate the model based on this data. We can rearrange the data, by subtracting its mean value $\mu = \frac{1}{T}\sum_{i=1}^{T}\mathrm{X}_i$ from all data points. Having this centered data, which we will also denote with X_i form here on, we need to estimate the autocovariance function $\gamma_{\mathrm{X}}(h)$ and its covariance matrix Γ_T, respectively. This is done by using the formulas (4.5) and (4.2). Form this point on, we have two ways of preceding, based on different ways of estimating the order of the model.

The first one is illustrated on the left side of Figure 4.1. In this case we have to choose $P, Q \in \mathbb{N}$ and set up the likelihood function $L(\beta, \sigma^2|\mathrm{X})$ given in (4.8) in order to be able to solve the maximisation problem

$$\max_{(\beta,\sigma^2)\in\mathscr{P}\times\mathbb{R}_+} \left[L_T(\beta, \sigma^2|\mathrm{X})\right] ,$$

for every $p \in [1, P]$ and $q \in [1, Q]$. This maximum value of the likelihood function refers to the optimal choice $(\hat{p}, \hat{q})$ of the order of the model, by using the AIC number discussed in section 4.4.3.

Another, different way to find the order of the model is based on a different definition of the AIC number using the regression sum of squares, given in (4.13), (4.14) and (4.15).

Since we now know the order of the model we come to the point of estimating its parameters, which is done by the maximum likelihood method, given in (4.11) and (4.10).

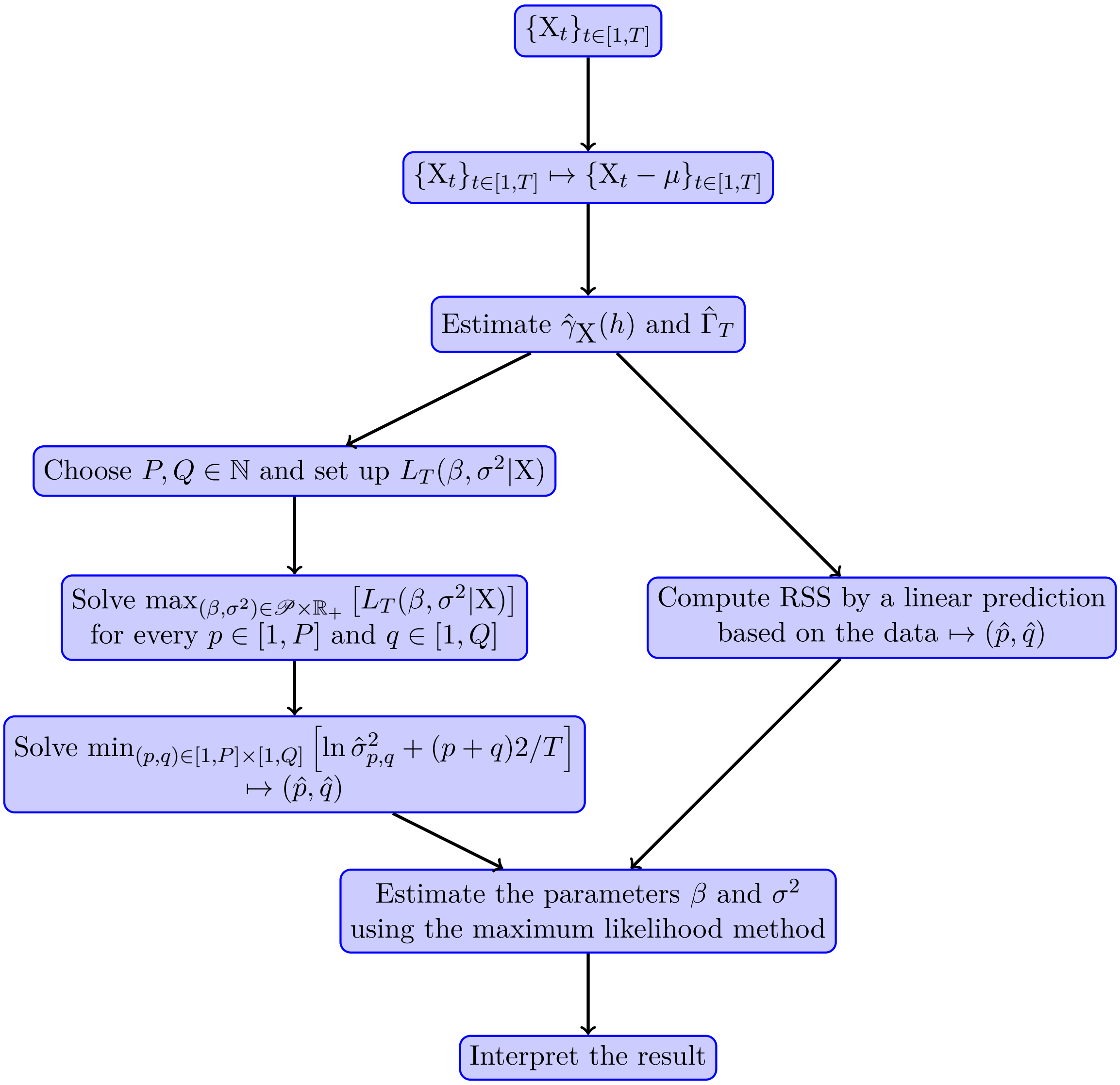


FIGURE 4.1.: Actions to be taken in setting up an ARMA(p,q) model based on historic data.

CHAPTER 5

Multivariate time series analysis

This chapter expands the concept of univariate to simple multivariate time series. Therefore we introduce a concept to describe higher dimensional time series. Further more, an introduction to impulse response functions, describing the dynamic properties of a model is given and illustrated in an example.

5.1. Introduction

In many problems concerning time series analysis, the data may be available on several variables of interest. Sometimes this data show a certain dynamic relationship among each other. It is for example possible that one series leads another one, or there may be a feedback relationship. This is a reason to analyse such multivariate time series jointly [Peña et al., 2011, p.365].

5.1 DEFINITION. [Neusser, 2009, p.165] A n-dimensional stochastic process is a family of random variables $\{\mathbf{X}_t\}_{t\in\mathbb{Z}}$, where $\mathbf{X}_t \in \mathbb{R}^n$.

This definition is very similar to the definition according to the univariate case. In fact, we expand the whole concept of univariate time series to a vector model.

Therefore we now consider the case, where the time series is represented by a vector of dimension n, where every component represents a univariate time series. We therefore adopt the notation

$$\mathbf{X}_t = \begin{pmatrix} X_{1t} \\ X_{2t} \\ \vdots \\ X_{nt} \end{pmatrix},$$

and interpret each component X_{it} as a one-dimensional time series with the possibility of each component being connected to several others.

Again, we characterise this multivariate time series by its mean value $\mu_{it} = E[X_{it}]$ and autocovariance function $\gamma_{ij}(t,s) = \mathrm{Cov}(X_{it}, X_{js})$ for $i,j = 1,\dots,n$ and a point of time $t,s \in \mathbb{Z}$. For the sake of a convenient notation we can write this quantities in matrix form and get

$$E[\mathbf{X}_t] = \begin{pmatrix} E[X_{1t}] \\ E[X_{2t}] \\ \vdots \\ E[X_{nt}] \end{pmatrix}, \quad \mathbf{\Gamma}(t,s) = \begin{pmatrix} \gamma_{11}(t,s) & \gamma_{12}(t,s) & \dots & \gamma_{1n1}(t,s) \\ \gamma_{21}(t,s) & \dots & & \vdots \\ \vdots & & \ddots & \vdots \\ \gamma_{n1}(t,s) & \dots & & \gamma_{nn}(t,s) \end{pmatrix}.$$

A n dimensional stochastic process is called stationary, if definition 4.3 holds for every component. It is because of this equivalent definition in terms of components of a multivariate time series, that all properties of the autocovariance function transfer to the matrix $\mathbf{\Gamma}(t,s)$. Thus we can characterise this matrix just by the time difference $h = t - s$, such that the only variable left, is the time difference h.

To give information about the interdependencies between the n different time series, we define the correlationfunction

$$\rho_{ij}(h) = \frac{\gamma_{ij}(h)}{\sqrt{\gamma_{ii}(0)\gamma_{jj}(0)}},$$

for $i,j = 1,\dots,n$, see [Neusser, 2009, p.166]. For $i \neq j$, this function investigates the cross correlation between different time series. Again, we can construct a higher dimensional white noise process using the univariate definition.

5.2 Definition. [Neusser, 2009, p.167] The process $\{\boldsymbol{\varepsilon}_t\}_{t\in\mathbb{Z}}$ is called white noise, if $\{\boldsymbol{\varepsilon}_t\}_{t\in\mathbb{Z}}$ is stationary and

(i) $E[\boldsymbol{\varepsilon}_t] = 0$, as well as

(ii) $\boldsymbol{\Gamma}(h) = \begin{cases} \Sigma & \text{for } h = 0 \\ 0 & \text{for } h \neq 0, \end{cases}$

holds true. This is denoted with $\boldsymbol{\varepsilon}_t \sim \mathrm{WN}(0, \Sigma)$.

5.1.1. Estimating the mean and the autocavariances

In many applications, we are confronted with the problem of estimating the mean and the autocavariance function by using a certain realisation $\{\mathbf{X}_t\}_{t\in[1,T]}$ of the time series. This estimations are given by

$$\hat{\mathbf{S}}_T = \bar{\mathbf{S}}_T = \frac{1}{T}\sum_{i=1}^{T} \mathbf{X}_i,$$

$$\hat{\boldsymbol{\Gamma}}(h) = \begin{cases} \sum_{i=1}^{T-h} (\mathbf{X}_{t+h} - \bar{\mathbf{X}}_T)(\mathbf{X}_t - \bar{\mathbf{X}}_T)^T, & 0 \leqslant h \leqslant T-1 \\ \hat{\boldsymbol{\Gamma}}^T(-h) & -T+1 \leqslant h < 0, \end{cases}$$

see [Neusser, 2009, p.172].

5.2. Vector Autoregressive moving average processes

5.3 Definition. [Neusser, 2009, p.179] Let $\{\mathbf{X}_t\}_{t\in\mathbb{Z}}$ be a stochastic process. The process $\{\mathbf{X}_t\}_{t\in\mathbb{Z}}$ is said to be an VARMA(p,q) process, if $\{\mathbf{X}_t\}_{t\in\mathbb{Z}}$ is stationary and if for every $t \in \mathbb{Z}$

$$\mathbf{X}_t - \boldsymbol{\phi}_1\mathbf{X}_{t-1} - \ldots - \boldsymbol{\phi}_p\mathbf{X}_{t-p} = \boldsymbol{\varepsilon}_t + \boldsymbol{\theta}_1\boldsymbol{\varepsilon}_{t-1} + \ldots + \boldsymbol{\theta}_q\boldsymbol{\varepsilon}_{t-q}, \tag{5.1}$$

with $\boldsymbol{\varepsilon}_t \sim \mathrm{WN}(0, \Sigma)$ holds true.

In the definition given above, the parameters $\boldsymbol{\phi}_i$ and $\boldsymbol{\theta}_j$ are $n \times n$ matrices whose entries determine the characteristics of the model. Using the Lag operator who acts on the components of a time series and has the same properties as in definition 4.9, makes it able to rewrite (5.1) in a more convenient form

$$\mathbf{\Phi}(\mathbb{L})(\mathbf{X}_t) = \mathbf{\Theta}(\mathbb{L})(\mathbf{X}_t),$$

with $\mathbf{\Phi}(\mathbb{L}) = \mathbb{I} - \sum_{i=1}^{p} \boldsymbol{\phi}_i \mathbb{L}^i$ and $\mathbf{\Theta}(\mathbb{L}) = \mathbb{I} + \sum_{i=1}^{q} \boldsymbol{\theta}_i \mathbb{L}^i$, respectively.

5.3. Vector Autoregressive Processes

It is due to its relative simple estimation of the parameters of a VAR(p) model, that it can be used in various applications. As in the case of univariate time models, we well define a vector autoregressive process of order p using the equation

$$\mathbf{X}_t = \boldsymbol{\phi}_1 \mathbf{X}_{t-1} + \ldots + \boldsymbol{\phi}_p \mathbf{X}_{t-p} + \boldsymbol{\varepsilon}_t, \tag{5.2}$$

where $\mathbf{W}_t \sim \mathrm{WN}(0, \Sigma)$ holds true. Sometimes a constant $\mathbf{c} \in \mathbb{R}^n$ is also included in the model.

If we denote $\phi_{ij}^{(k)}$ as the (i,j) the element of the matrix $\boldsymbol{\phi}_k$, we can rewrite (5.2)using the matrix notation

$$\begin{pmatrix} X_{1t} \\ X_{2t} \\ \vdots \\ X_{nt} \end{pmatrix} = \begin{pmatrix} \phi_{11}^{(1)} & \phi_{12}^{(1)} & \cdots & \phi_{1n}^{(1)} \\ \phi_{21}^{(1)} & \cdots & & \vdots \\ \vdots & \ddots & & \vdots \\ \phi_{n1}^{(1)} & \cdots & & \phi_{nn}^{(1)} \end{pmatrix} \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \\ \vdots \\ X_{n,t-1} \end{pmatrix} + \begin{pmatrix} \phi_{11}^{(p)} & \phi_{12}^{(p)} & \cdots & \phi_{1n}^{(p)} \\ \phi_{21}^{(p)} & \cdots & & \vdots \\ \vdots & \ddots & & \vdots \\ \phi_{n1}^{(p)} & \cdots & & \phi_{nn}^{(p)} \end{pmatrix} \begin{pmatrix} X_{1,t-p} \\ X_{2,t-p} \\ \vdots \\ X_{n,t-p} \end{pmatrix} + \ldots$$
$$+ \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \vdots \\ \varepsilon_{nt} \end{pmatrix}.$$

Interpreting the components of this matrix equation, we find that this is equivalent to the system of n different equations, given by

$$\begin{aligned} X_{1t} = {} & \phi_{11}^{(1)} X_{1,t-1} + \phi_{12}^{(1)} X_{2,t-1} + \ldots + \phi_{1n}^{(1)} X_{n,t-1} \\ & + \phi_{11}^{(2)} X_{1,t-1} + \phi_{12}^{(2)} X_{2,t-1} + \ldots + \phi_{1n}^{(2)} X_{n,t-1} \\ & + \ldots \end{aligned}$$

$$
\begin{aligned}
&+ \phi_{11}^{(p)} X_{1,t-1} + \phi_{12}^{(p)} X_{2,t-1} + \ldots + \phi_{1n}^{(p)} X_{n,t-1} \\
&+ W_{1t} \\
&\vdots \\
X_{nt} &= \phi_{n1}^{(1)} X_{1,t-1} + \phi_{n2}^{(1)} X_{2,t-1} + \ldots + \phi_{nn}^{(1)} X_{n,t-1} \\
&+ \phi_{n1}^{(2)} X_{1,t-1} + \phi_{n2}^{(2)} X_{2,t-1} + \ldots + \phi_{nn}^{(2)} X_{n,t-1} \\
&+ \ldots \\
&+ \phi_{n1}^{(p)} X_{1,t-1} + \phi_{n2}^{(p)} X_{2,t-1} + \ldots + \phi_{nn}^{(p)} X_{n,t-1} \\
&+ W_{nt}. \qquad (5.3)
\end{aligned}
$$

We can see from (5.3) that each component of the n dimensional time series influences the future values. Hence this model also includes the interdependencies of the different time series. Concerning the number of variables of the model, we have to estimate $n^2 p$ different parameters. These parameters are then estimated using an ordinary least squares approximation.

5.3.1. OLS estimation of the coefficients of a VAR(p) model

We now have to estimate the coefficients of the VAR(p) model, using the information of $T + p$ observations, where the points of time these observations have been recognized, range from $t = -p + 1, -p + 2, \ldots, 0, 1, \ldots, T$. We can then use the points of time $t = 1, \ldots, T$ and write all regression variables from (5.3) into a $(T \times np)$ matrix, denoted with $\tilde{\mathbf{X}}$ and obtain

$$
\tilde{\mathbf{X}} = \begin{pmatrix} X_{1,0} & \cdots & X_{n,0} & X_{1,1} & \cdots & X_{n,1} & \cdots & X_{1,-p+1} & \cdots & X_{n,-p+1} \\ \vdots & \cdots & & & & & & & & \vdots \\ X_{1,T-1} & \cdots & X_{n,T-1} & X_{1,T} & \cdots & X_{n,T} & \cdots & X_{1,-p+T} & \cdots & X_{n,-p+T} \end{pmatrix}.
$$

According to [Neusser, 2009, p.193], we find for the OLS-estimation for the parameters of the VAR(p) model, give by the vector $\boldsymbol{\beta} =$ (Coefficients of the first equation, $\ldots$, Coefficients of the n-th equation)$^\intercal$, to be

$$\hat{\boldsymbol{\beta}}_{OLS} = \begin{pmatrix} (\tilde{\mathbf{X}}^\intercal\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^\intercal & 0 & \cdots & 0 \\ 0 & (\tilde{\mathbf{X}}^\intercal\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^\intercal & \cdots & 0 \\ \vdots & & \ddots & \\ 0 & \cdots & & (\tilde{\mathbf{X}}^\intercal\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^\intercal \end{pmatrix} \mathbf{Y},$$

where $\mathbf{Y} = (X_{1,1}, X_{1,2}, \ldots, X_{1,T}, X_{2,1}, \ldots, X_{2,T}, \ldots, X_{n,1}, \ldots, X_{n,T})^\intercal$ is a vector of observations.

5.3.2. Impulse Response function

Since a general VAR(p) involves many different parameters, it is in many cases very difficult to investigate the dynamic interactions between every variable. A possible way to gain information about the dynamic of the model, is provided by the method of impulse response functions. The impulse response function gives information about the dynamic effects of a structural interference, called a shock, of a certain variable [Neusser, 2009, p.206].

5.4 Example. [1] Given a VAR(1) model $\mathbf{X}_t = \boldsymbol{\phi}_1\mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t$ we want to compute the change of the time series at a point of time $t+h$, $\Delta\mathbf{X}_{t+h}$ when the j-th component is shocked by $\Delta\varepsilon_{j,t} = 1$. Thus we find

$$\Delta\mathbf{X}_{t+h} = \frac{\Delta\mathbf{X}_{t+h}}{\Delta\varepsilon_{j,t}} = \frac{\partial\mathbf{X}_{t+h}}{\partial\varepsilon_{j,t}}. \tag{5.4}$$

To compute (5.4), we make multiple use of the recursion of the VAR(1) process and observe that $\mathbf{X}_{t+h} = \boldsymbol{\phi}_1 X_{t+h-1} + \boldsymbol{\varepsilon}_{t+h}$ holds true. Thus we find

$$\frac{\partial}{\partial\varepsilon_{j,t}}\left(\mathbf{X}_{t+h}\right) = \frac{\partial}{\partial\varepsilon_{j,t}}\left(\boldsymbol{\phi}_1\mathbf{X}_{t+h-1} + \boldsymbol{\varepsilon}_{t+h}\right). \tag{5.5}$$

Again, using the recursion, we find $\mathbf{X}_{t+h-1} = \boldsymbol{\phi}_1\mathbf{X}_{t+h-2} + \boldsymbol{\varepsilon}_{t+h-1}$. Hence we can

[1] see http://stats.stackexchange.com/questions/143214/how-to-calculate-the-impulse-response-function-of-a-var1-with-example, accessed:2015-12-08

insert this information into (5.5) and get

$$\begin{aligned}\frac{\partial}{\partial \varepsilon_{j,t}}\left(\boldsymbol{\phi}_1 \mathbf{X}_{t+h-1}+\boldsymbol{\varepsilon}_{t+h}\right) &= \frac{\partial}{\partial \varepsilon_{j,t}}\left(\boldsymbol{\phi}_1\left(\boldsymbol{\phi}_1 \mathbf{X}_{t+h-2}+\boldsymbol{\varepsilon}_{t+h-1}\right)+\boldsymbol{\varepsilon}_{t+h}\right)\\ &= \boldsymbol{\phi}_1^2 \mathbf{X}_{t+h-2}+\boldsymbol{\phi}_1 \boldsymbol{\varepsilon}_{t+h-1}+\boldsymbol{\varepsilon}_{t+h}.\end{aligned}$$

By induction, we conclude that

$$\begin{aligned}\frac{\partial}{\partial \varepsilon_{j,t}}\left(\mathbf{X}_{t+h}\right) &= \frac{\partial}{\partial w_{j,t}}\left(\boldsymbol{\phi}_1^{h+1} \mathbf{X}_{t-1}+\sum_{i=0}^{h} \boldsymbol{\phi}_1^{i} \boldsymbol{\varepsilon}_{t+h-i}\right).\\ &= \frac{\partial}{\partial \varepsilon_{j,t}}\left(\boldsymbol{\phi}_1^{h} \boldsymbol{\varepsilon}_t\right)\\ &= \boldsymbol{\phi}_1^{h} \mathbf{e}_j,\end{aligned}$$

where $\mathbf{e}_j = (0, \ldots, 1, \ldots, 0)^\intercal$, with a 1 in the j-th component, denotes the j-th unit vector. By taking a closer look at the expression $\boldsymbol{\phi}_1^h \mathbf{e}_j$, we obtain, that this is simply the j-th column of the matrix $\boldsymbol{\phi}_1^h$. This recursive method can then be expanded to a general VAR(p) process.

In many applications the impulse response function is computed by simulating the VAR(p) process as follows: To implement the simulation, we have to set $\mathbf{X}_{t-1}, \ldots, \mathbf{X}_{t-p} = \mathbf{0}$. Further more we set $\varepsilon_{j,t} = 1$ and all other elements of $\boldsymbol{\varepsilon}_t$ to zero and compute the future values $t, t+1, \ldots$. The value of the vector $\mathbf{X}_{t+h}$ corresponds to the coefficients of the impulse response functions [Hamilton, 1994, p.319].

CHAPTER 6

Industry analysis of the engineering sector

This chapter deals with the problem of performing an industry analysis of the engineering industry by applying the theory of time series analysis using the data of the Austrian chambers of economics. Furthermore a prediction of all time series is given, as well as an impulse response analysis of the vector-autoregressive model of the cost components.

6.1. Introduction

A firm competing in a certain industry usually has a competitive strategy in order to be present at the market. This competitive strategy links the company to its associated environment. The relevant environment is naturally very broad but nevertheless the key aspect of the firm's environment is the industry in which it competes. The structure of this industry influences the choice of the different strategies taken [Porter, 1998, p.3]. It is therefore very important to analyse the industry thoroughly. The content of this industry analysis can differ in certain cases but according to the key aspects of the industry, it includes in many cases an analysis of

(i) the cost structure

(ii) the factor of success and

(iii) the growth of the industry,

see [Aaker, 2013, p.92]. According to [Aaker, 2013, p.101], analysing the cost structure includes, besides the investigation of the distribution of the main cost factors, also the study of the value added. In the case of the mechanical engineering industry the factor of success is chosen to be the sales volume of the whole industry. According to [Blaschke, 2008, p.29] the costs of investment play a major role of the development of the industry. Many companies react on different circumstances by investing a certain amount of money, which changes the position of power, as well as the barrier of entrance in a sector. Thus analysing the costs of investment of the mechanical engineering industry makes it possible to predict possible future trends.

6.2. Time series analysis with `R`

When performing a time series analysis using the statistics program `R`, we have to distinguish between the case of a univariate time series $\{X_t\}_{t\in\mathscr{T}}$ and a multivariate time series $\{\mathbf{X}_t\}_{t\in\mathscr{T}}$.

First of all, we want to consider the case of a univariate time series. Generally the code for estimating a univariate time series model is given in listing C.1. The first step is to load the package forecast and the data as a `.csv` file (denoted with A) into the program. This data will then be transformed into a time series such that the program can perform all methods described in the previous chapters. The name of the row therefore has to be inserted in B. Concerning the time structure, the frequency is denoted with C and the first year with D.

To start with the actual ARIMA(p,d,q) model we have to declare the maximal p and q values that are allowed in the model estimation and write these numbers into position E and F, respectively. The last information we have to give is the number of time periods for the forecast, represented by the number G.

A time series analysis of a multivariate model is performed in a similar way, using the package vars, see listing C.2. Thus we have to load this package, with the command library(vars) and follow the protocol in the same way as in the univariate case.

6.3. Time series analysis of the mechanical engineering industry

Concerning the structure of the data, the sales Volume, the costs of investment, the value added, the costs of material, the costs of labour, the costs of energy and the residual costs are given in the years $\mathscr{T} = \{1997, \ldots, 2013\}$ are given in Table 6.1. All this values are given in units if 1000 € throughout this chapter.

TABLE 6.1.: The values of the realisations of the time series.

Year	S_t [1000 €]	I_t [1000 €]	VA_t [1000 €]	M_t [1000 €]	L_t [1000 €]	E_t [1000 €]	R_t [1000 €]
1997	4157231,53	145646,90	1462495,44	2786607,41	1178423,14	43163,52	135770,73
1998	4655229,76	167015,04	1674943,93	3035432,22	1245053,89	41290,89	130722,66
1999	4714592,85	172322,77	1670641,70	3071602,94	1279798,84	41234,28	136791,35
2000	5198114,00	194975,00	1823325,00	3427617,00	1346445,00	42538,00	167973,00
2001	5630818,00	204352,00	1999148,00	3791271,00	1395484,00	42892,00	173414,00
2002	5795127,00	182087,00	1988010,00	3777899,00	1420317,00	41538,00	174498,00
2003	5902390,00	180009,00	2025265,00	4028743,00	1500512,00	43600,00	176273,00
2004	6533261,00	204691,00	2136658,00	4697114,00	1571650,00	47391,00	171467,00
2005	7371482,00	181218,00	2457459,00	5150754,00	1640265,00	52365,00	173010,00
2006	8301750,00	218813,00	2718800,00	6026979,00	1757768,00	62505,00	175566,00
2007	9720147,00	253768,00	3139504,00	7126837,00	1915223,00	68441,00	190060,00
2008	10888973,00	386989,00	3374732,00	7918153,00	2078720,00	75502,00	233617,00
2009	8958063,00	249536,00	2859751,00	6232248,00	2008390,00	71129,00	227087,00
2010	8953702,00	160765,00	2869962,00	6034471,00	2006814,00	75317,00	221280,00
2011	10192998,00	232582,00	3566400,00	7507848,00	2217971,00	79646,00	209558,00
2012	10957192,00	300316,00	3490774,00	8286481,00	2401151,00	81883,00	207635,00
2013	11294417,00	268705,00	3691630,00	8304521,00	2570634,00	85145,00	224949,00

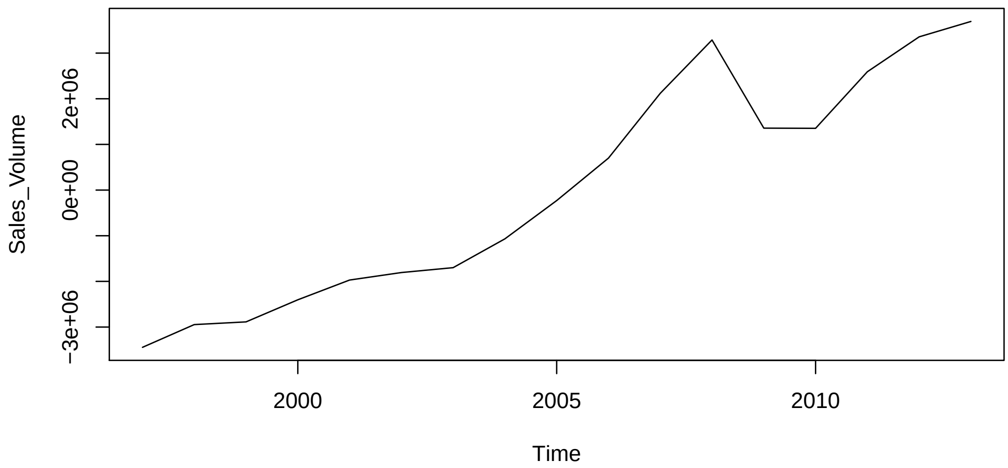


FIGURE 6.1.: The centred time series of the sales volume in the years 1997-2013.

Since we are interested in the cost structure of the mechanical engineering industry, we the relative percentage of the cost components are given in Table 6.2.

6.3.1. The sales volume

Concerning the analysis of the sales volume $\{S_t\}$, we obtain the centred time series in Figure 6.1. Note that the time series is centred around its mean value and therefore some values become negative. This is the case, when they are smaller than the mean value. Thus the time series is given by $\{\bar{S}_t\}_{t\in\mathscr{T}} = \{S_t - \mu_S\}_{t\in\mathscr{T}}$, where $\mu_S = \frac{1}{\#\mathscr{T}} \sum_{t\in\mathscr{T}} S_t \approx 7601499$. We obtain a small decrease in the sales volume around the time of the economic crisis, see Figure 6.1. Beside this years the mechanical engineering industry always increased the sales volume, which implies that this sector is growing. According to the data, the recent growth in the years 2010-2013 tends to get smaller. If we insert the data and realize the protocol given above, we find the best model fitting the data, to be an ARIMA(1,0,0) model, which is in fact a AR(1) model of the form

$$\bar{S}_t = \phi \bar{S}_{t-1} + \varepsilon_t, \tag{6.1}$$

Table 6.2.: The relative percentage of the cost data.

Year	L_t [%]	M_t [%]	E_t [%]	R_t [%]
1997	28,44	67,24	1,04	3,28
1998	27,96	68,17	0,93	2,94
1999	28,26	67,81	0,91	3,02
2000	27,01	68,76	0,85	3,37
2001	25,83	70,17	0,79	3,21
2002	26,23	69,78	0,77	3,22
2003	26,10	70,08	0,76	3,07
2004	24,23	72,40	0,73	2,64
2005	23,38	73,41	0,75	2,47
2006	21,91	75,12	0,78	2,19
2007	20,59	76,63	0,74	2,04
2008	20,17	76,83	0,73	2,27
2009	23,52	72,99	0,83	2,66
2010	24,07	72,37	0,90	2,65
2011	22,15	74,97	0,80	2,09
2012	21,87	75,49	0,75	1,89
2013	22,98	74,25	0,76	2,01

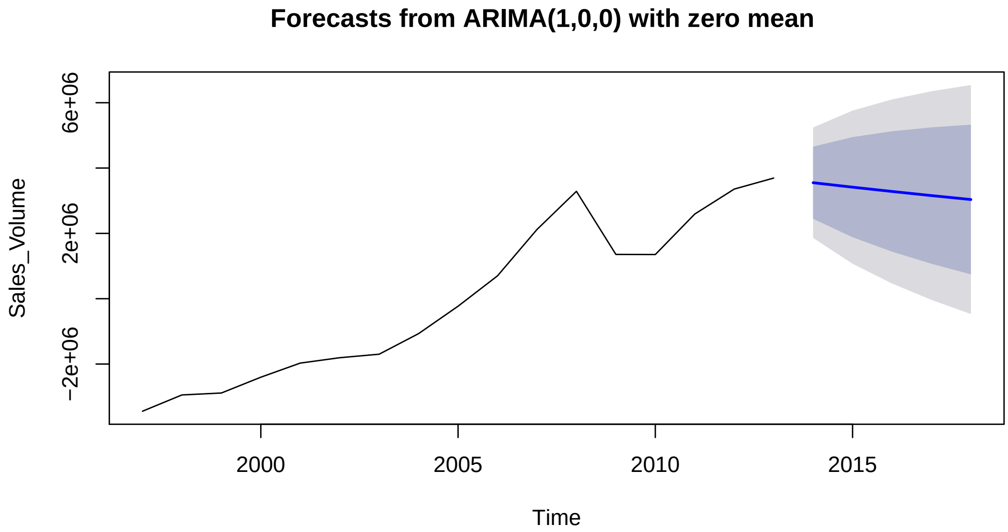


FIGURE 6.2.: The predicted values of the sales volume of the mechanical engineering industry for the years 2014-2018 with the 80 % an 90 % confidence interval.

where $\varepsilon_t \sim \mathrm{WN}(0, \sigma^2)$ is a white noise process. Due to the maximum likelihood estimation, we find for the parameters $\phi = 0,9616 \pm 0,0490$ and $\sigma = 8,6 \cdot 10^5$, such that (6.1) takes the form

$$\bar{S}_t = 0,9616 \bar{S}_{t-1} + \varepsilon_t. \tag{6.2}$$

A forecast of the sales volume, based on the calibrated model in (6.2), with the standard confidence interval, is given in Figure 6.2. According to this model, the sales volume is going to decrease in the years 2014-2018. The actual values at this points of time, are

$$\bar{S}_{2014} = 3551032$$
$$\bar{S}_{2015} = 3414597$$
$$\bar{S}_{2016} = 3283404$$
$$\bar{S}_{2017} = 3157252$$

$$\bar{S}_{2018} = 3035947.$$

It is scarcely necessary, to point out, that these transformed values are subtracted by the mean value of the data series. In order to get the right values, one has to add this mean value. Thus the final values are

$$\begin{aligned} S_{2014} &= 11152531 \\ S_{2015} &= 11016096 \\ S_{2016} &= 10884903 \\ S_{2017} &= 10758751 \\ S_{2018} &= 10637446. \end{aligned}$$

Despite the fact, that the sales volume was increasing most of the time, the predicted values in the future decrease, due to the structure break in the time of the economic crisis which acts negatively on the prediction in the future. Here the question arises if this structure break should be eliminated in the data set in order to get an uninfluenced forecast. Another way of dealing with the problem is, to split the data set into two separate sets, one before the economic crisis and the other one afterwards, respectively. Due to the lack of data points and the most realistic description of the past, it was choose not to ignore this structure break of the economic crisis, resulting in decreasing values of the sales volume.

6.3.2. The costs of investment

Again we centred the data of the costs of investment around the mean value $\mu_I = \frac{1}{\#\mathscr{T}} \sum_{t \in \mathscr{T}} I_t \approx 217870$. The graph of the centred time series of the costs of investment $\{\bar{I}_t\}_{t \in \mathscr{T}} = \{I_t - \mu_I\}_{t \in \mathscr{T}}$ is given in Figure 6.3. Concerning the structure of the data, we obtain a small increase of the costs of investment in the years 1997-2005, meaning that companies in the mechanical engineering industry were willing to invest in the future. In the years right before the economic crisis these investments increased rapidly by a factor of 2,1 from 181218 in 2005 up to 386989 in 2007. Naturally, in the time of the economic crises these investments decreased quickly and dropped to a level in 2010 comparable to the level in 1998. After the time of the crisis the investments increased again, showing that the mechanical engineering industry is

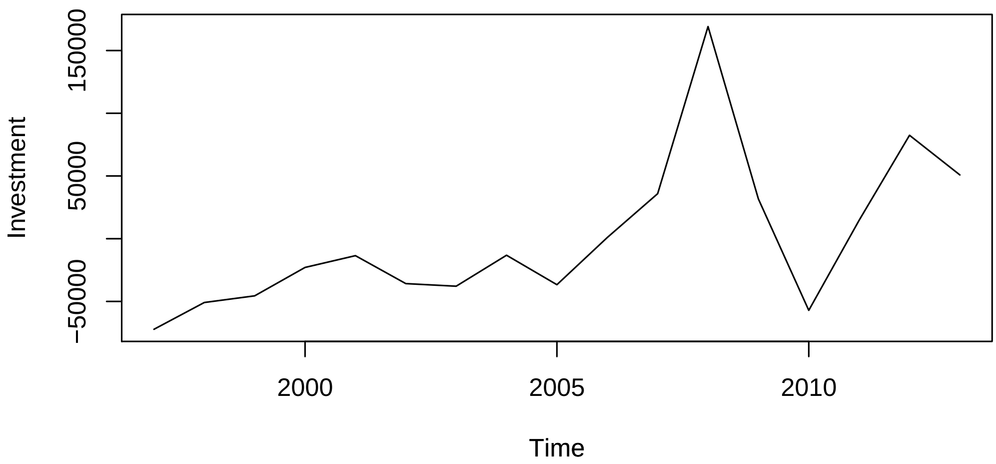


FIGURE 6.3.: The centred time series of the costs of investment in the years 1997-2013.

growing again, except the last year the data have been recorder, where we can notice a slight decrease again.

It is because of this unstable structure of increasing and decreasing costs of investment, that we may conclude that the best model, describing the development of the costs of investment is a white noise process. In fact, analysing the data using the AIC-criterion confirms this assumption and we find for the time series model of the centred costs of investment

$$\bar{I}_t = \varepsilon_t, \tag{6.3}$$

which is a ARIMA(0,0,0) process, where $\varepsilon_t \sim \mathrm{WN}(0, \sigma^2)$ is a white noise process with $\sigma = 5,8 \cdot 10^4$. A forecast of the sales volume, based on the calibrated model in (6.3), with the standard confidence interval, is given in Figure 6.4. Due to the easy structure of the model given in (6.3), also the forecast is very easy for the white noise process. Since $E[\varepsilon_t] = 0$, the best prediction for the future values $\{\bar{I}_t\}$ itself is zero. We have to keep in mind, that we have to transform this data, by adding the

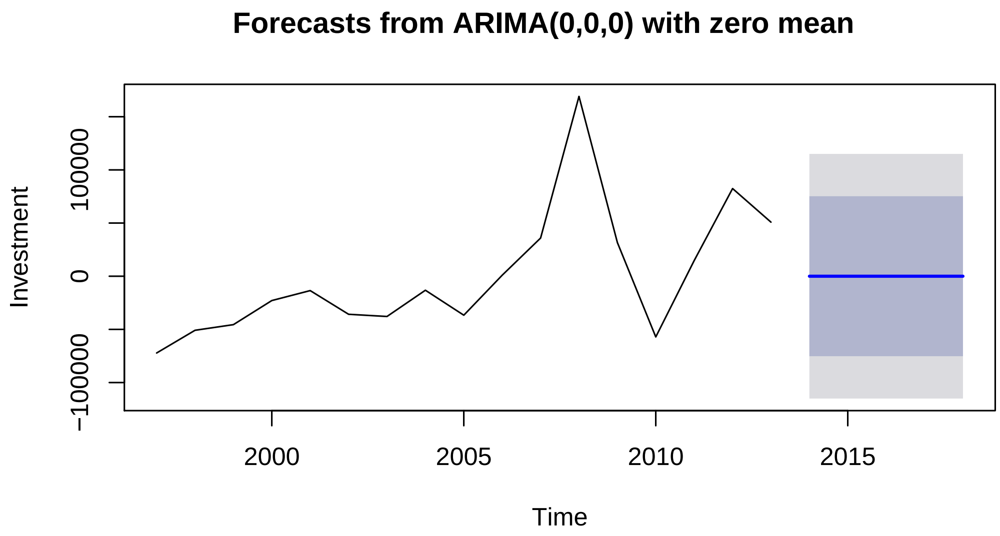


FIGURE 6.4.: The predicted values of the costs of investment of the mechanical engineering industry for the years 2014-2018 with the 80 % an 90 % confidence interval.

mean value μ_I to get the real values for the prediction of $\{I_t\}$, which gives us

$$I_{2015} = \ldots = I_{2018} = \mu_I.$$

Finally we can conclude that the best prediction based on the given data is given by the mean value of the data set in the years the data have been recorded.

6.3.3. The value added

In chapter one, we described the production process as the process of transforming inputs into a certain amount of output. Due to the transformation, this created output has a higher value than the sum of the values of the input factors. This is also the definition of the value added according to Müller-Stewens and Lechner, where the value added is described as the process of creating an additional value by using a certain treatment [Wulfsberg and Redlich, 2011, p.21].

There are many different definitions of the value added, summarized in [Wulfsberg and Redlich, 2011, p.22] but the most general definition of the value added is given by

$$VA = \text{Output} - \text{Input}.$$

A further remark may be added, that in the data the influence of taxes as well as the influence of subventions have been eliminated.

Again, we take a closer look at the centred time series of the value added, given by $\{\bar{VA}_t\}_{t\in\mathscr{T}} = \{VA_t - \mu_{VA}\}_{t\in\mathscr{T}}$, where $\mu_{VA} = \frac{1}{\#\mathscr{T}} \sum_{t\in\mathscr{T}} VA_t \approx 2526441$. A plot of the resulting centred time series is given in Figure 6.5. We obtain the same structural behaviour as in the case of the time series of the sales volume, given in Figure 6.1. The value added is increasing in the years of before the economic crisis, where we can divide this increasing structure into two different stages: In the years 1997-2001 this increase can be approximated by a linear function, whereas the increase in the years 2002-2008 shows an exponential increase as it was in the case concerning the sales volume. In the time of the economic crisis this numbers naturally dropped but only by a factor of 0,8 from 3374732 in 2008 to 2859751 in 2009. After this decrease, the value added stagnated in the next year and then increased again in the years afterwards, except for one year.

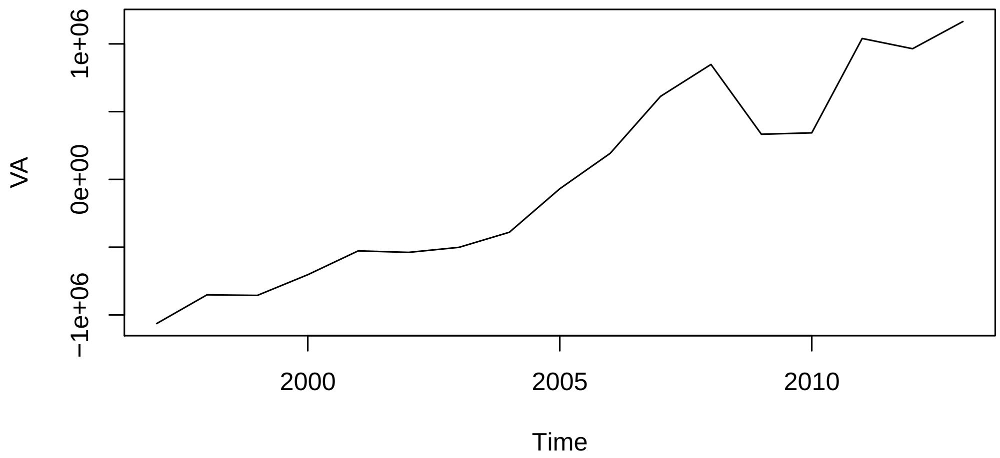


FIGURE 6.5.: The centred time series of the value added in the years 1997-2013.

The best model, describing this time series is given by an ARIMA(1,0,0) model

$$\bar{VA}_t = \phi \bar{VA}_{t-1} + \varepsilon_t, \tag{6.4}$$

where we find for the maximum likelihood estimation of the parameter $\phi = 0,9546 \pm 0.0570$ and $\sigma = 28 \cdot 10^4$, such that (6.4) takes the form

$$\bar{VA}_t = 0,9546 \bar{VA}_{t-1} + \varepsilon_t. \tag{6.5}$$

A forecast based on the model given in (6.5) with the standard confidence interval, is given in Figure 6.6. Again we find the same structural behaviour as in the case of the forecast of the sales volume, showing decreasing values in the future. This future values of the centred time series are given by

$$\bar{VA}_{2014} = 1112347,3$$
$$\bar{VA}_{2015} = 1061902$$
$$\bar{VA}_{2016} = 1013744,5$$

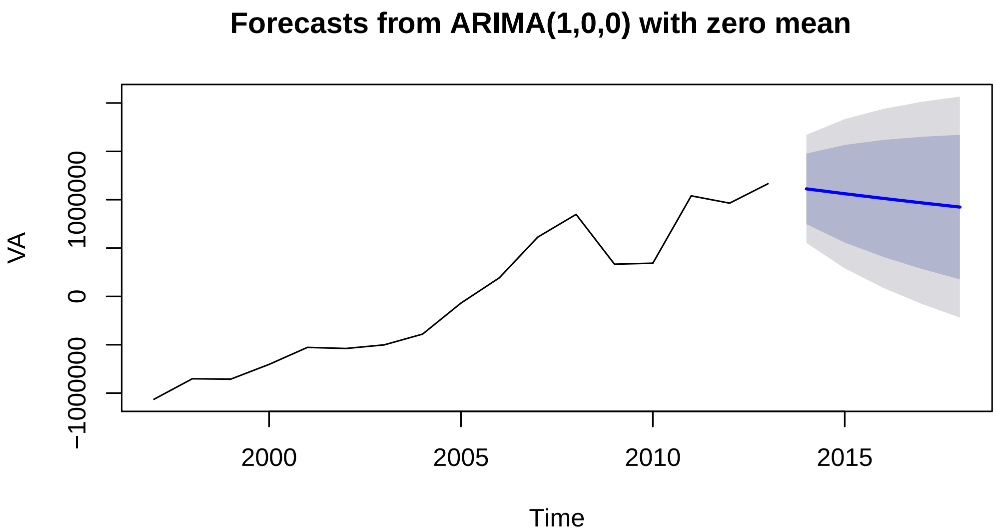


FIGURE 6.6.: The predicted values of the value added of the mechanical engineering industry for the years 2014-2018 with the 80 % an 90 % confidence interval.

$$\bar{VA}_{2017} = 967770,9$$
$$\bar{VA}_{2018} = 923882,2.$$

Again, we have to consider the fact, that this are just the values of the centred time series. To get the real values, we have to add the mean value to them and obtain

$$VA_{2014} = 3638788,3$$
$$VA_{2015} = 3588343$$
$$VA_{2016} = 3540185,5$$
$$VA_{2017} = 3494211,9$$
$$VA_{2018} = 3450323,2.$$

6.3.4. The cost components

To analyse the structure if the cost components of the mechanical engineering industry, we take a closer look at the shift of the relative percentage of the costs of material $\{M_t\}_{t\in\mathscr{T}}$, costs of labour $\{L_t\}_{t\in\mathscr{T}}$, costs of energy $\{E_t\}_{t\in\mathscr{T}}$ and the residual costs $\{R_t\}_{t\in\mathscr{T}}$. To get a stationary time series we first of all transform the data, given in Table 6.2 by taking the logarithm. These transformed time series are denoted with $\{\bar{M}_t\}_{t\in\mathscr{T}}$, $\{\bar{L}_t\}_{t\in\mathscr{T}}$, $\{\bar{E}_t\}_{t\in\mathscr{T}}$ and $\{\bar{R}_t\}_{t\in\mathscr{T}}$, respectively. Again we have to keep in mind to re transform the data by using the exponential function. A plot of the transformed time series is given in Figure 6.7. Concerning the trends in this plot, we observe a break in the structure of the data at the years 2008-2009, due to the economic crisis.

Concerning the costs of labour, we recognize a decrease until the point of the economic crisis. In the year 1997 companies in the engineering industry spent $28,4\%$ of the whole expenditures for personnel issues. This number drops to the minimum of $20,1\%$ in the year 2008. After that there is a tendency to increase the costs of labour by a small amount. This decrease of costs of labour can be explained by the fact, that nowadays more and more places of employment are replaced by automatic machines.

Both time series $\{L_t\}_{t\in\mathscr{T}}$ and $\{E_t\}_{t\in\mathscr{T}}$, representing the costs of labour and the costs of energy, show very similar characteristics, in terms of the observed trend at a first sight. The costs of energy dropped from a maximum of 1% to a minimum of

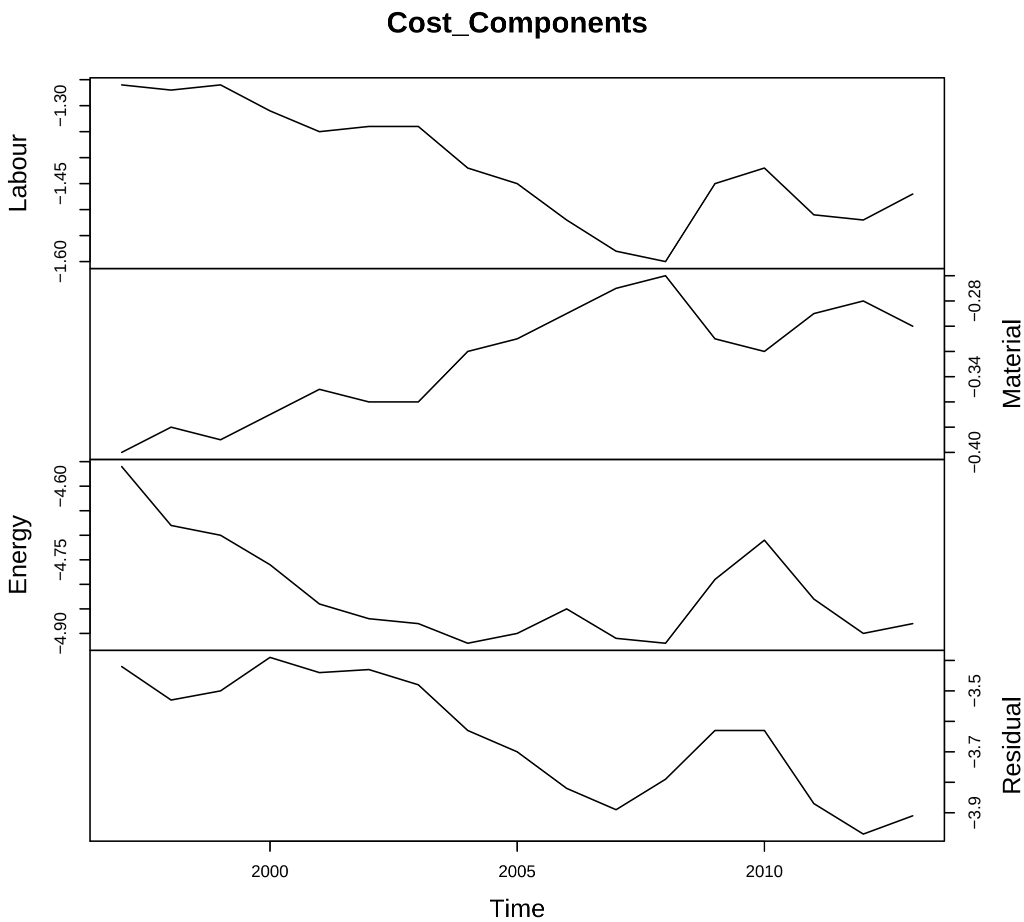


Figure 6.7.: The time series of the cost components.

$0,73\%$ such that we can conclude, that the decrease was much smaller. Surprisingly, we obtain the result, that the costs of energy almost stay the same. Considering the effect of inflation, we notice that the costs of energy were lowered, approximately by this factor such that the costs of energy almost stay the same. This could be argued by the fact, that more companies tend to outsource a part of their manufacturing, resulting in a lower energy consumption but nevertheless this effect is, compared to others relative small in the mechanical engineering industry.

Concerning the costs of material, where also costs of services have been included, we find the complete opposite behaviour. While the costs of labour and energy were decreasing over the last years, the costs of material and service increased form a minimum of 67% to a maximum of 77%. Again, we could interpret the higher costs of services with the effect of outsourcing the production. On the other hand a higher percentage of the costs of material could also indicate the growing of the mechanical engineering industry, as it can also be seen from the growing volume of sales.

The best vector autoregressive time series model, describing the observed data, is a VAR(2) model, which has the form

$$\begin{pmatrix} \bar{L}_t \\ \bar{M}_t \\ \bar{E}_t \\ \bar{R}_t \end{pmatrix} = \begin{pmatrix} \phi_{11}^{(1)} & \phi_{12}^{(1)} & \phi_{13}^{(1)} & \phi_{14}^{(1)} \\ \phi_{21}^{(1)} & \phi_{22}^{(1)} & \phi_{23}^{(1)} & \phi_{24}^{(1)} \\ \phi_{31}^{(1)} & \phi_{32}^{(1)} & \phi_{33}^{(1)} & \phi_{34}^{(1)} \\ \phi_{41}^{(1)} & \phi_{42}^{(1)} & \phi_{43}^{(1)} & \phi_{44}^{(1)} \end{pmatrix} \begin{pmatrix} \bar{L}_{t-1} \\ \bar{M}_{t-1} \\ \bar{E}_{t-1} \\ \bar{R}_{t-1} \end{pmatrix} + \begin{pmatrix} \phi_{11}^{(2)} & \phi_{12}^{(2)} & \phi_{13}^{(2)} & \phi_{14}^{(2)} \\ \phi_{21}^{(2)} & \phi_{22}^{(2)} & \phi_{23}^{(2)} & \phi_{24}^{(2)} \\ \phi_{31}^{(2)} & \phi_{32}^{(2)} & \phi_{33}^{(2)} & \phi_{34}^{(2)} \\ \phi_{41}^{(2)} & \phi_{42}^{(2)} & \phi_{43}^{(2)} & \phi_{44}^{(2)} \end{pmatrix} \begin{pmatrix} \bar{L}_{t-2} \\ \bar{M}_{t-2} \\ \bar{E}_{t-2} \\ \bar{R}_{t-2} \end{pmatrix}$$
$$+ \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \\ \varepsilon_{4t} \end{pmatrix} + \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix}.$$

An OLS-estimation of the matrix-coefficients yields the result

$$\boldsymbol{\phi}_1 = \begin{pmatrix} \phi_{11}^{(1)} & \phi_{12}^{(1)} & \phi_{13}^{(1)} & \phi_{14}^{(1)} \\ \phi_{21}^{(1)} & \phi_{22}^{(1)} & \phi_{23}^{(1)} & \phi_{24}^{(1)} \\ \phi_{31}^{(1)} & \phi_{32}^{(1)} & \phi_{33}^{(1)} & \phi_{34}^{(1)} \\ \phi_{41}^{(1)} & \phi_{42}^{(1)} & \phi_{43}^{(1)} & \phi_{44}^{(1)} \end{pmatrix} = \begin{pmatrix} 1.8201 & 2.1850 & -0.7979 & 0.3413 \\ -0.54301 & -0.47997 & 0.27073 & -0.12417 \\ 1.40045 & 2.54331 & 0.35015 & 0.11529 \\ -1.4830 & -6.0618 & -1.1081 & 1.1358 \end{pmatrix},$$

$$\boldsymbol{\phi}_2 = \begin{pmatrix} \phi_{11}^{(2)} & \phi_{12}^{(2)} & \phi_{13}^{(2)} & \phi_{14}^{(2)} \\ \phi_{21}^{(2)} & \phi_{22}^{(2)} & \phi_{23}^{(2)} & \phi_{24}^{(2)} \\ \phi_{31}^{(2)} & \phi_{32}^{(2)} & \phi_{33}^{(2)} & \phi_{34}^{(2)} \\ \phi_{41}^{(2)} & \phi_{42}^{(2)} & \phi_{43}^{(2)} & \phi_{44}^{(2)} \end{pmatrix} = \begin{pmatrix} -1.2844 & -2.1207 & 0.7468 & -0.2830 \\ 0.47284 & 0.83585 & -0.26067 & 0.10098 \\ -2.02538 & -4.47359 & 0.07783 & -0.26420 \\ -2.1683 & -5.6194 & 0.7886 & -0.8908 \end{pmatrix}$$

and

$$\begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix} = \begin{pmatrix} -0.6887 \\ -0.33851 \\ -4.82683 \\ -13.2939 \end{pmatrix}.$$

Thus we found a vector autoregressive model, which describes the time series of the cost components in an adequate way. Further more it is possible to make a prediction about future values.

The dynamic of the model is described by the matrices $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$, respectively. If we want to interpret the coefficients of these matrices, we observe that the first component of the model is

$$\begin{aligned} \ln L_t = {} & \phi_{11}^{(1)} \ln L_{t-1} + \phi_{12}^{(1)} \ln M_{t-1} + \phi_{13}^{(1)} \ln E_{t-1} + \phi_{14}^{(1)} \ln R_{t-1} \\ & + \phi_{11}^{(2)} \ln L_{t-2} + \phi_{12}^{(2)} \ln M_{t-2} + \phi_{13}^{(2)} \ln E_{t-2} + \phi_{14}^{(2)} \ln R_{t-2} + c_1 + \varepsilon_{1t}. \end{aligned}$$

In order to get the re-transformed values, we apply the exponential function to get

$$L_t = L_{t-1}^{\phi_{11}^{(1)}} M_{t-1}^{\phi_{12}^{(1)}} E_{t-1}^{\phi_{13}^{(1)}} R_{t-1}^{\phi_{14}^{(1)}} \cdot L_{t-2}^{\phi_{11}^{(2)}} M_{t-2}^{\phi_{12}^{(2)}} E_{t-2}^{\phi_{13}^{(2)}} R_{t-2}^{\phi_{14}^{(2)}} \cdot e^{c_1} e^{\varepsilon_{1t}}.$$

According to the prediction, given in Figure 6.8, we find a tendency of the evolution of the next periods. The predicted values for the next three periods are

$$\begin{pmatrix} \bar{L}_{2014} \\ \bar{M}_{2014} \\ \bar{E}_{2014} \\ \bar{R}_{2014} \end{pmatrix} = \begin{pmatrix} -1.450036 \\ -0.3082658 \\ -4.809315 \\ -3.787143 \end{pmatrix}, \quad \begin{pmatrix} \bar{L}_{2015} \\ \bar{M}_{2015} \\ \bar{E}_{2015} \\ \bar{R}_{2015} \end{pmatrix} = \begin{pmatrix} -1.470081 \\ -0.3034879 \\ -4.789557 \\ -3.739135 \end{pmatrix}, \quad \begin{pmatrix} \bar{L}_{2016} \\ \bar{M}_{2016} \\ \bar{E}_{2016} \\ \bar{R}_{2016} \end{pmatrix} = \begin{pmatrix} -1.485602 \\ -0.2989982 \\ -4.823437 \\ -3.756276 \end{pmatrix}$$

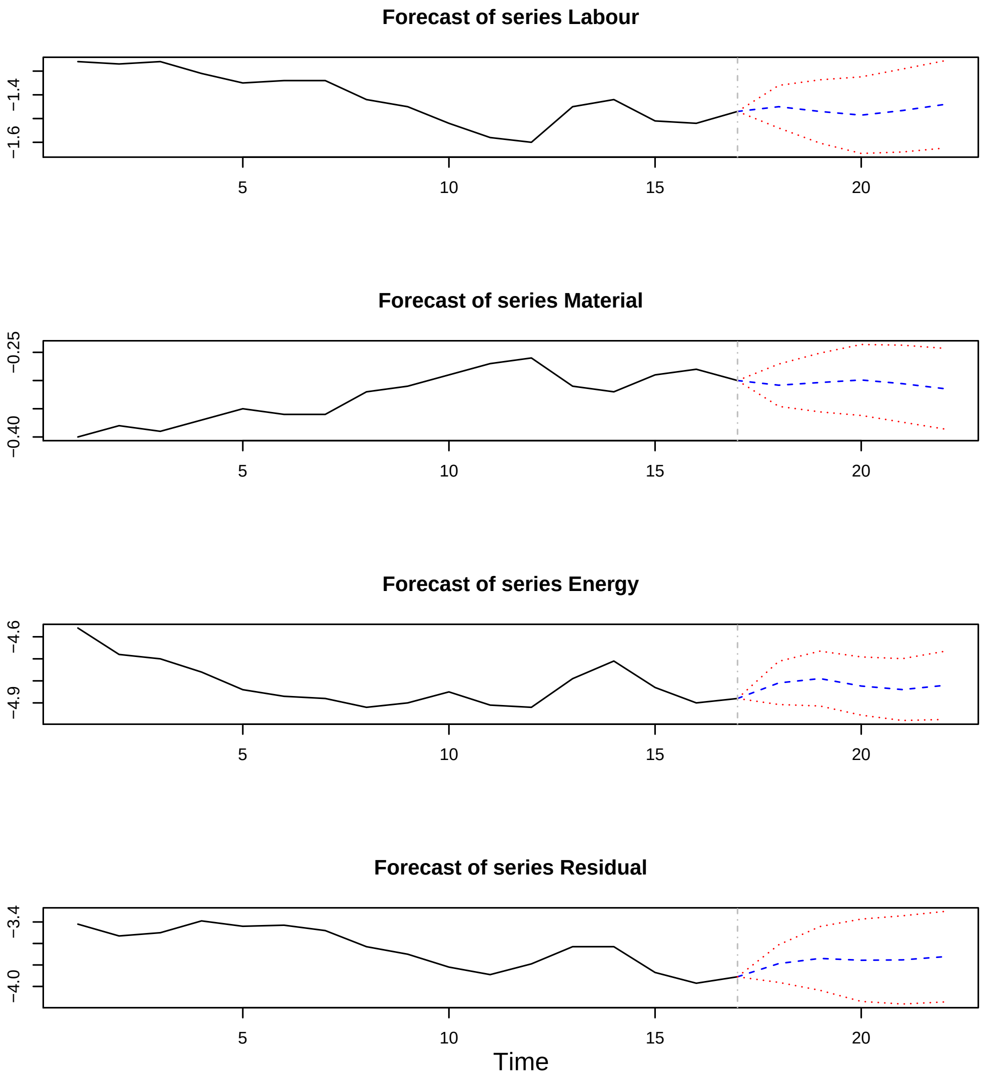


FIGURE 6.8.: The prediction of the cost components with a 95% confidence interval.

For the rounded re-transformed values we find

$$\begin{pmatrix} L_{2014} \\ M_{2014} \\ E_{2014} \\ R_{2014} \end{pmatrix} = \begin{pmatrix} 0,23456 \\ 0,73472 \\ 0,00815 \\ 0,02267 \end{pmatrix}, \quad \begin{pmatrix} L_{2015} \\ M_{2015} \\ E_{2015} \\ R_{2015} \end{pmatrix} = \begin{pmatrix} 0,22991 \\ 0,73824 \\ 0,00832 \\ 0,02377 \end{pmatrix}, \quad \begin{pmatrix} L_{2016} \\ M_{2016} \\ E_{2016} \\ R_{2016} \end{pmatrix} = \begin{pmatrix} 0,22637 \\ 0,74156 \\ 0,00804 \\ 0,02337 \end{pmatrix}$$

This values have to be normalised, such that the sum of all components add up to 100%. Thus we construct a normalising factor $\mathcal{N}_j$ for every year $j = 2014, 2015, 2016$, where $\mathcal{N}_j = \frac{1}{\sum_{i=1}^{4} x_i^j}$ and x_i denotes the component of the j-th vector. The final result for the prediction is given by the multiplying these normalising factors:

$$\begin{pmatrix} L_{2014} \\ M_{2014} \\ E_{2014} \\ R_{2014} \end{pmatrix} = \begin{pmatrix} 0,23454 \\ 0,73465 \\ 0,00815 \\ 0,02267 \end{pmatrix}, \quad \begin{pmatrix} L_{2015} \\ M_{2015} \\ E_{2015} \\ R_{2015} \end{pmatrix} = \begin{pmatrix} 0,22984 \\ 0,73802 \\ 0,00832 \\ 0,02376 \end{pmatrix}, \quad \begin{pmatrix} L_{2016} \\ M_{2016} \\ E_{2016} \\ R_{2016} \end{pmatrix} = \begin{pmatrix} 0,22652 \\ 0,74205 \\ 0,00805 \\ 0,02339 \end{pmatrix}.$$

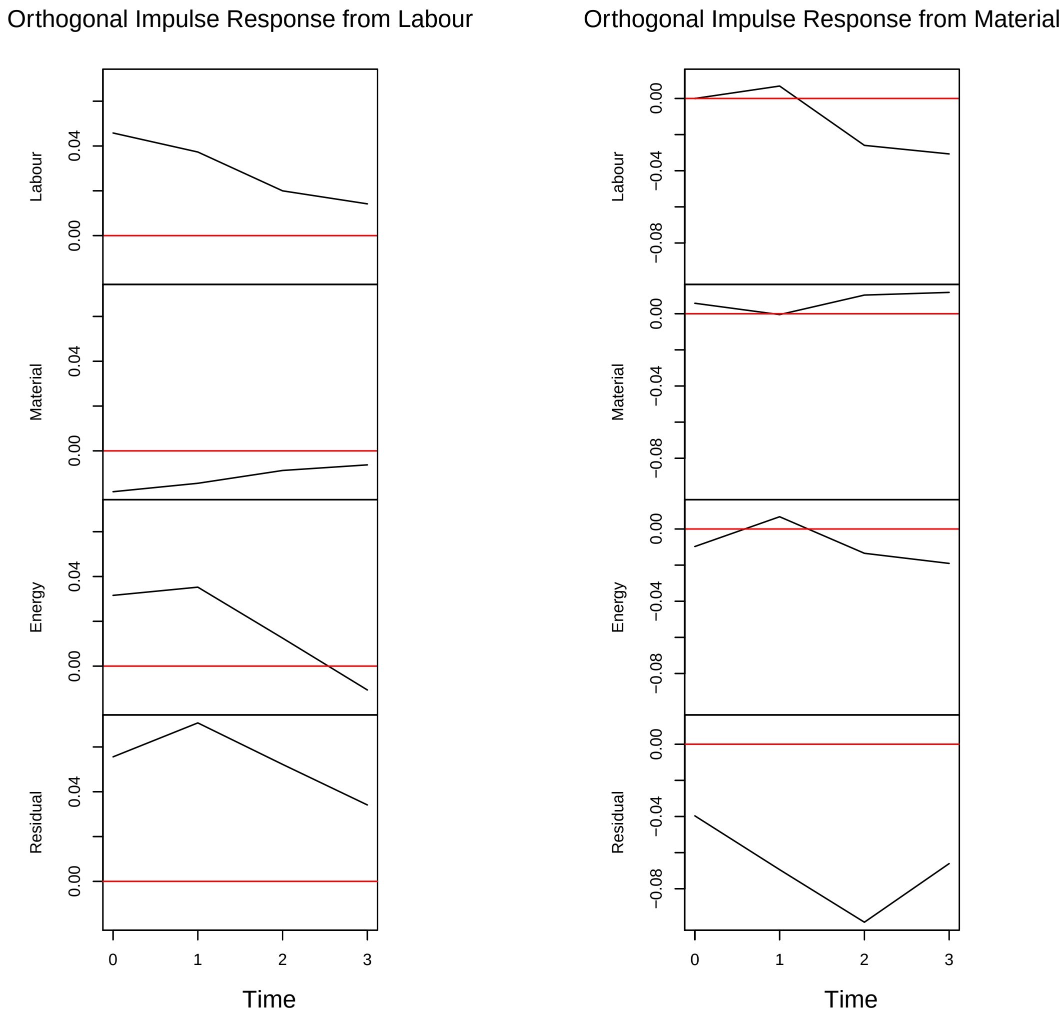
Having this result we now analyse the dynamic of the model. Here we ask ourselves the question: What happens if one component is shocked at a certain point of time? We therefore want to investigate how the system responds to higher values of a certain variable. This general treatment of the problem allows us to work out different economic problems, such as: What is going to happen if the price for a certain amount of energy is increasing? Obviously, this will become noticeable in increasing costs of energy in the engineering branch. But how will the other cost components react, according to the model? The theory of impulse response functions gives an answer to this question and simulates the VAR(2) model under a certain shock of a variable.

The impulse response functions are given in Figure 6.9 (A),(B) and Figure 6.10, respectively.

If the costs of labour are shocked in such a way that suddenly a higher percentage is used for the costs of labour, then the system tends to decrease the costs of labour and energy, while the costs of material show an increasing tendency, but remain below the base level in the next three periods.

Concerning a shock in the costs of material, both percentages of the costs of labour and the costs of energy are decreasing, since more money is needed for the costs of material.

Concerning a shock in the costs of energy, we observe that the costs of material and the costs of labour show a contrary behaviour. While the costs of material are increasing, the costs of labour are decreasing. After both cost components stagnate, the costs of material are decreasing, while the costs of labour are increasing.



(A) Costs of labour

(B) Costs of material

FIGURE 6.9.: Figure (A) shows the reaction of the cost components due to a shock of the costs of labour. Figure (B) shows the reaction of the cost components due to a shock of the costs of material.

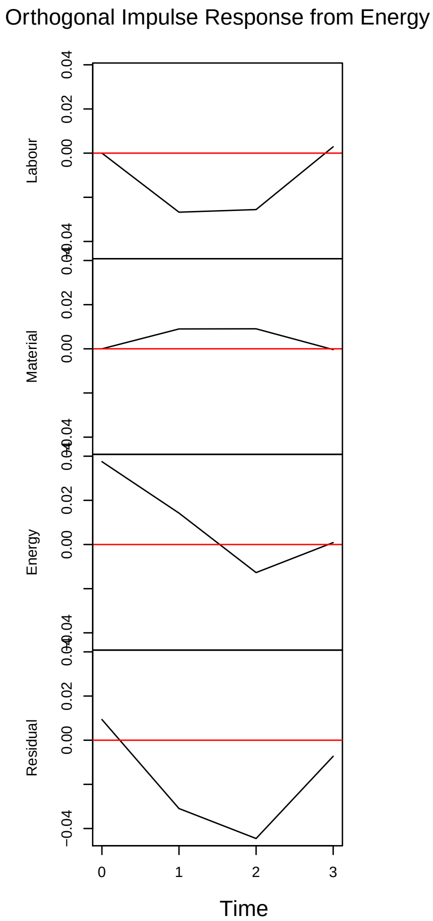


FIGURE 6.10.: The reaction of the cost components due to a shock of the costs of energy.

CHAPTER 7

Conclusion

In the FIRST CHAPTER of this thesis we introduced the production function and defined all necessary concepts in order to give a compact introduction in production theory. Therefore we defined the elasticity of a production function and its related concepts. We also introduced a special class of production functions called the CES production function which refers to a constant elasticity of substitution and proved that other common production functions, as for example the Leontief, Cobb-Douglas or linear production function can be derived a special choice of parameters of the CES production function. In that sense we want to refer the CES production function as a more general production function.

The SECOND CHAPTER is a summary of the previous work of [Ivaz, 2014], which was done concerning this topic. Basically a linear cost function, based on a Leontief production function was calibrated by a linear regression model, using the data of the aggregated costs of the engineering sector.

In the THIRD CHAPTER the approach of [Ivaz, 2014] was extended to a more general concept. We dropped the assumption of a Leontief production function and assumed a more general CES production function to get a better result. Further more the input variable was changed from the production value x_{PV} to the remaining manufacturing costs x_{RMC} in order to interpret the cost function in an input/output oriented context. In this input/output oriented setting, we were able to derive the cost function $\mathcal{C}(x_1, x_2)$ by a minimization problem subject to a certain amount of outcome $f(x_1, x_2) = \bar{c}$. In the next step the cost function was fitted to the given data points, in order to determine the parameters δ and ρ of the production function. Analysing this procedure, we have to be aware of the fact, that we have only four data points available. Therefore the result may not be very significant. But with this method, we found a useful way to calibrate the cost function for future applications, since then there will be more data points available. Further more, we found that fitting the complex form of the cost function to the data points is very unstable and may also fail but is more accurate than the first attempt of a linear cost function. Thus we also introduced a different approach, based on the idea of [Kmenta, 1967]. Following this ideas we approximated the cost function by a Taylor polynomial of second order and then fitted the function, which is more easy to handle, to the given data points in order to determine the parameters. We found, that for the parameter $\rho \approx -1$ holds true. Thus we can state, that using this method yields that a good approximation for the cost function is a linear cost function. It is, however scarcely necessary to point out, that we didn't assumed this cost function to be linear, but proved that an estimation using this technique refers to a linear cost function.

Concerning the FOURTH CHAPTER, we introduced the basic concepts and methods to give an introduction into univariate time series models. It was shown, how to set up and calibrate an ARIMA(p,d,q) model in order to be able to give a prediction about future values of a time series. All this calculations are only based on the given historical data in form of a realisation $\{\mathrm{X}_t\}_{t\in\mathscr{T}}$ of a time series. We provided all methods, including the choice of the orders of an ARIMA(p,d,q) model, to be able to understand how different implemented methods of time series analysis in `R` work.

In the FIFTH CHAPTER, we discussed multivariate time series in order to calibrate higher dimensional time series models. Furthermore we introduced the most common way to perform a multivariate time series analysis, by using vector autoregressive processes of order q. To investigate the dynamics of the model the impulse response analysis was discussed and computed for a simple VAR(1) model.

In the SIXTH CHAPTER we made use of all this techniques and performed an industry analysis of the engineering sector by setting up an univariate time series model for the sales volume, the value added, the costs of investment and a multivariate time series models for the four cost components: costs of labour, costs of energy, costs of material and residual costs of the engineering sector. A predication of all time series was computed and discussed, as well as an impulse response analysis concerning the costs components.

APPENDIX A

Second partial derivatives for the Taylor approximation

$$\frac{\partial^2 \mathcal{C}_{x_1,x_2}}{\partial \delta^2} =$$

```
(1/(rho^2))(delta^(1/(1 + rho)) q1^(rho/(1 + rho)) + (1 - delta)^(1/(
1 + rho)) q2^(rho/(1 + rho)))^(1/rho) (delta x1^-rho - (-1 + delta)
x2^-rho)^(-1/rho) ((delta^(-1 + 1/(1 + rho)) q1^(rho/(1 + rho)) -
(1 - delta)^(-1 + 1/(1 + rho)) q2^(rho/(1 + rho)))^2/((delta^(1/(1 + rho))
q1^(rho/(1 + rho)) + (1 - delta)^(1/(1 + rho)) q2^(rho/(1 + rho)))
(1 +rho)) + (delta^(-2 + 1/(1 + rho)) q1^(rho/(1 + rho)) +
(1 - delta)^(-2 + 1/(1 + rho)) q2^(rho/(1 + rho))) rho (-1 + 1/(1 + rho)) -
(2 (delta^(-1 + 1/(1 + rho)) q1^(rho/(1 + rho)) - (1 - delta)^
(-1 + 1/(1 + rho)) q2^(rho/(1 + rho))) (x1^-rho - x2^-rho))/
(delta x1^-rho - (-1 + delta) x2^-rho) + ((delta^(1/(1 + rho)) q1^
(rho/(1 + rho)) + (1 - delta)^(1/(1 + rho)) q2^(rho/(1 + rho)))
(1 + rho) (x1^rho - x2^rho)^2)/((-1 + delta) x1^ rho - delta x2^rho)^2)
```

Listing A.1: The second partial derivative with respect to δ.

$$\frac{\partial^2 \mathcal{C}_{x_1,x_2}}{\partial\delta\partial\rho} =$$

```
(1/(rho^2))(delta^(1/(1 + rho)) q1^(rho/(1 + rho))+
(1 - delta)^(1/(1 + rho)) q2^(rho/(1 + rho)))^(1/rho) (delta x1^-rho
- (-1 + delta) x2^-rho)^(-1/rho) (-delta^(-1 + 1/(1 + rho))
q1^(rho/(1 + rho))+ (1 - delta)^(-1 + 1/(1 + rho)) q2^(rho/( 1 + rho))
+ ((delta^(-1 + 1/(1 + rho)) q1^(rho/(1 + rho)) -
(1 - delta)^(-1 + 1/(1 + rho)) q2^(rho/(1 + rho)))rho)/(1 + rho) +
((delta^(1/(1 + rho)) q1^(rho/(1 + rho)) + (1 - delta)^(1/(1 + rho))
q2^(rho/(1 + rho))) (x1^rho - x2^rho))/((-1 + delta)x1^rho -
delta x2^rho) + (rho (-delta^(-1 + 1/(1 + rho)) q1^(rho/(1 + rho))
(1 + rho) + (1 - delta)^(-1 + 1/(1 + rho)) q2^(rho/(1 + rho)) (1 + rho)
+ (1 - delta)^(-1 + 1/(1 + rho)) q2^(rho/(1 + rho)) Log[1 - delta] -
delta^(-1 + 1/(1 + rho)) q1^(rho/(1 + rho)) Log[delta] + delta^
(-1 + 1/(1 + rho)) q1^(rho/(1 + rho)) Log[q1] - (1 - delta)^
(-1 + 1/(1 + rho)) q2^(rho/(1 + rho)) Log[q2]))/(1 + rho)^2 +
((delta^(-1 + 1/(1 + rho)) q1^(rho/(1 + rho)) - (1 - delta)^
(-1 + 1/(1 + rho)) q2^(rho/(1 + rho))) ((rho (-(1 - delta)^
((1/(1 + rho))) q2^(rho/(1 + rho))Log[1 - delta] - delta^(1/(1 + rho))
q1^(rho/(1 + rho)) Log[delta] + delta^(1/(1 + rho)) q1^(rho/(1 + rho))
Log[q1] + (1 - delta)^(1/(1 + rho)) q2^(rho/(1 + rho))Log[q2]))/
((delta^(1/(1 + rho))q1^(rho/(1 + rho)) + (1 - delta)^(1/(1 + rho))
q2^(rho/(1 + rho))) (1 + rho)^2) - Log[delta^(1/(1 + rho))
q1^(rho/(1 + rho)) + (1 - delta)^(1/(1 + rho)) q2^(rho/(1 + rho))]))/
rho - ((delta^(1/(1 + rho)) q1^(rho/(1 + rho)) + (1 - delta)^(1/(1 + rho))
q2^(rho/(1 + rho)))(x1^-rho - x2^-rho) (( rho (-(1 - delta)^((1/(1 + rho)))
q2^(rho/(1 + rho)) Log[1 - delta] - delta^(1/(1 + rho))
q1^(rho/(1 + rho)) Log[delta] +delta^(1/(1 + rho)) q1^(rho/(1 + rho))
Log[q1] + (1 - delta)^(1/(1 + rho))q2^(rho/(1 + rho))Log[q2]))/
((delta^(1/(1 + rho)) q1^(rho/(1 + rho)) + (1 - delta)^(1/(1 + rho))
q2^(rho/(1 + rho))) (1 + rho)) - Log[delta^(1/(1 + rho)) q1^(rho/(1 + rho))
+ (1 - delta)^(1/(1 + rho))q2^(rho/(1 + rho))]))/(rho (delta x1^-rho -
(-1 + delta) x2^-rho)) +((delta^(1/(1 + rho))q1^(rho/(1 + rho)) +
(1 - delta)^(1/(1 + rho)) q2^(rho/(1 + rho))) rho (-x2^rho Log[x1] +
x1^rho Log[x2]))/((-1 + delta) x1^rho -delta x2^rho) +
((delta^(-1 + 1/(1 + rho)) q1^(rho/(1 + rho)) - (1 - delta)^(-1 + 1/(1 + rho))
q2^(rho/(1 + rho))) (-delta rho x2^rho Log[x1]+ (-1 + delta)
rho x1^ rho Log[x2] + ((-1 + delta) x1^rho - delta x2^rho)Log[delta x1^-rho
- (-1 + delta) x2^-rho]))/(rho ((-1 + delta) x1^rho -delta x2^rho)) -
((delta^(1/(1 + rho)) q1^(rho/(1 + rho)) + (1 - delta)^(1/(1 + rho))
q2^(rho/(1 + rho))) (x1^rho - x2^rho) (-delta rho (1 + rho) x2^rho Log[x1]
+ (-1 + delta) rho (1 + rho) x1^rho Log[x2] + ((-1 + delta) x1^rho - delta
x2^rho) Log[delta x1^-rho - (-1 + delta) x2^-rho]))/( rho ((-1 + delta)
```

```
       x1^rho - delta x2^rho)^2))
```

Listing A.2: The mixed partial derivative.

$$\frac{\partial^2 \mathcal{C}_{x_1,x_2}}{\partial \rho^2} =$$

```
(1/(rho^4))(delta^(1/(1 + rho)) q1^(rho/(1 + rho)) + (1 - delta)^(1/(
    1 + rho)) q2^(rho/(1 + rho)))^(
 1 + 1/rho) (delta x1^-rho - (-1 + delta) x2^-rho)^(-1/
  rho) ((rho^2 ((1 - delta)^(1/(1 + rho)) q2^(rho/(1 + rho))
        Log[1 - delta] +
       delta^(1/(1 + rho)) q1^(rho/(1 + rho)) Log[delta] -
       delta^(1/(1 + rho)) q1^(rho/(1 + rho)) Log[q1] - (1 - delta)^(
        1/(1 + rho)) q2^(rho/(1 + rho)) Log[q2]))/((delta^(1/(1 + rho))
         q1^(rho/(1 + rho)) + (1 - delta)^(1/(1 + rho)) q2^(rho/(
        1 + rho))) (1 + rho)^2) + (
   rho^2 ((1 - delta)^(1/(1 + rho)) q2^(rho/(1 + rho))
        Log[1 - delta] +
       delta^(1/(1 + rho)) q1^(rho/(1 + rho)) Log[delta] -
       delta^(1/(1 + rho)) q1^(rho/(1 + rho)) Log[q1] - (1 - delta)^(
        1/(1 + rho)) q2^(rho/(1 + rho)) Log[q2]))/((delta^(1/(1 + rho))
         q1^(rho/(1 + rho)) + (1 - delta)^(1/(1 + rho)) q2^(rho/(
        1 + rho))) (1 + rho)) - (
   rho^3 ((1 - delta)^(1/(1 + rho)) q2^(rho/(1 + rho))
        Log[1 - delta] +
       delta^(1/(1 + rho)) q1^(rho/(1 + rho)) Log[delta] -
       delta^(1/(1 + rho)) q1^(rho/(1 + rho)) Log[q1] - (1 - delta)^(
        1/(1 + rho)) q2^(rho/(1 + rho)) Log[q2])^2)/((delta^(1/(
        1 + rho)) q1^(rho/(1 + rho)) + (1 - delta)^(1/(1 + rho)) q2^(
        rho/(1 + rho)))^2 (1 + rho)^3) + (
   rho^3 (-(1 - delta)^((1/(1 + rho))) q2^(rho/(1 + rho))
        Log[1 - delta] -
       delta^(1/(1 + rho)) q1^(rho/(1 + rho)) Log[delta] +
       delta^(1/(1 + rho)) q1^(rho/(1 + rho)) Log[q1] + (1 - delta)^(
        1/(1 + rho)) q2^(rho/(1 + rho)) Log[q2]))/((delta^(1/(1 + rho))
         q1^(rho/(1 + rho)) + (1 - delta)^(1/(1 + rho)) q2^(rho/(
        1 + rho))) (1 + rho)^2) + (
   1/((delta^(1/(1 + rho)) q1^(rho/(1 + rho)) + (1 - delta)^(1/(
        1 + rho)) q2^(rho/(1 + rho))) (1 + rho)^3))
   rho^3 ((1 - delta)^(1/(1 + rho)) q2^(rho/(1 + rho))
        Log[1 - delta]^2 +
       delta^(1/(1 + rho)) q1^(rho/(1 + rho)) Log[delta]^2 +
       2 delta^(1/(1 + rho)) q1^(rho/(1 + rho))
        Log[delta] (1 + rho - Log[q1]) -
       2 delta^(1/(1 + rho)) q1^(rho/(1 + rho)) Log[q1] -
```

```
        2 delta^(1/(1 + rho)) q1^(rho/(1 + rho)) rho Log[q1] +
        delta^(1/(1 + rho)) q1^(rho/(1 + rho)) Log[q1]^2 +
        2 (1 - delta)^(1/(1 + rho)) q2^(rho/(1 + rho))
          Log[1 - delta] (1 + rho - Log[q2]) -
        2 (1 - delta)^(1/(1 + rho)) q2^(rho/(1 + rho)) Log[q2] -
        2 (1 - delta)^(1/(1 + rho)) q2^(rho/(1 + rho))
          rho Log[q2] + (1 - delta)^(1/(1 + rho)) q2^(rho/(1 + rho))
          Log[q2]^2) +
   2 rho Log[
     delta^(1/(1 + rho)) q1^(rho/(1 + rho)) + (1 - delta)^(1/(
       1 + rho)) q2^(rho/(1 + rho))] + ((
     rho ((1 - delta)^(1/(1 + rho)) q2^(rho/(1 + rho))
            Log[1 - delta] +
          delta^(1/(1 + rho)) q1^(rho/(1 + rho)) Log[delta] -
          delta^(1/(1 + rho)) q1^(rho/(1 + rho)) Log[q1] - (1 - delta)^(
           1/(1 + rho)) q2^(rho/(1 + rho)) Log[q2]))/((delta^(1/(
           1 + rho)) q1^(rho/(1 + rho)) + (1 - delta)^(1/(1 + rho)) q2^(
           rho/(1 + rho))) (1 + rho)) +
     Log[delta^(1/(1 + rho)) q1^(rho/(1 + rho)) + (1 - delta)^(1/(
          1 + rho)) q2^(rho/(1 + rho))])^2 + (
   2 rho^2 (delta x2^rho Log[x1] - (-1 + delta) x1^rho Log[x2]))/((-1
       + delta) x1^rho - delta x2^rho) + (
   rho^3 (delta x2^rho Log[x1] - (-1 + delta) x1^
       rho Log[x2])^2)/((-1 + delta) x1^rho - delta x2^rho)^2 + (
   rho^3 (delta x2^rho Log[x1]^2 - (-1 + delta) x1^
       rho Log[x2]^2))/((-1 + delta) x1^rho - delta x2^rho) -
   2 rho Log[delta x1^-rho - (-1 + delta) x2^-rho] + (
   1/((-1 + delta) x1^rho - delta x2^rho))
   2 ((rho (-(1 - delta)^((1/(1 + rho))) q2^(rho/(1 + rho))
             Log[1 - delta] -
           delta^(1/(1 + rho)) q1^(rho/(1 + rho)) Log[delta] +
           delta^(1/(1 + rho)) q1^(rho/(1 + rho))
             Log[q1] + (1 - delta)^(1/(1 + rho)) q2^(rho/(1 + rho))
             Log[q2]))/((delta^(1/(1 + rho)) q1^(rho/(
            1 + rho)) + (1 - delta)^(1/(1 + rho)) q2^(rho/(
            1 + rho))) (1 + rho)) -
      Log[delta^(1/(1 + rho)) q1^(rho/(1 + rho)) + (1 - delta)^(1/(
           1 + rho)) q2^(rho/(1 + rho))]) (-delta rho x2^
       rho Log[x1] + (-1 + delta) rho x1^
       rho Log[x2] + ((-1 + delta) x1^rho - delta x2^rho) Log[
         delta x1^-rho - (-1 + delta) x2^-rho]) + (-delta rho x2^
      rho Log[x1] + (-1 + delta) rho x1^
      rho Log[x2] + ((-1 + delta) x1^rho - delta x2^rho) Log[
       delta x1^-rho - (-1 + delta) x2^-rho])^2/((-1 + delta) x1^rho -
      delta x2^rho)^2)
```

Listing A.3: The second partial derivative with respect to ρ.

APPENDIX B

Mathematica code

```
q1= 56.139
q2= 27.874
FC= 51.69272
data = {{1.31901739, 0.98558670, 166.600510173}, {1.04570906,
          0.91024723, 138.53777873}, {1.05442656, 0.85718242,
          130.37280165}, {1.31451084, 0.94131837, 154.85238000}}

FindFit[data, {(delta*x1^(-rho) + (1 - delta)*x2^(-rho))^(-1/
                  rho)*(delta^(1/(rho + 1))*(q1)^(rho/(rho + 1)) + (1 -
                  delta)^(1/(rho + 1))*(q2)^(rho/(rho + 1)))^((rho + 1)/
                  rho) + FC,
                {0 <= delta <= 1, -1 <= rho}}, {delta, rho}, {x1,x2}]
```

LISTING B.1: Mathematica code for finding the parameters of the model based on a nonlinear least square fit.

```
q1= 56.139
q2= 27.874
FC= 51.69272
Cost[x1_, x2_] := (delta*x1^(-rho) + (1 - delta)*x2^(-rho))^(-1/rho)*
                 (delta^(1/(rho + 1))*(q1)^(rho/(rho + 1)) + (1 -delta)^
                 (1/(rho + 1))*(q2)^(rho/(rho + 1)))^((rho + 1)/rho) + FC
NMaximize[{Cost[1.31901739, 0.98558670],
  0 <= delta <= 1, -1 <= rho}, {delta, rho}]
NMaximize[{Cost[1.04570906, 0.91024723],
```

```
  0 <= delta <= 1, -1 <= rho}, {delta, rho}]
NMaximize[{Cost[1.05442656, 0.85718242],
  0 <= delta <= 1, -1 <= rho}, {delta, rho}]
NMaximize[{Cost[1.31451084, 0.94131837],
  0 <= delta <= 1, -1 <= rho}, {delta, rho}]
```

LISTING B.2: Mathematica code for finding the maximum of the cost function.

```
q1= 56.139
q2= 27.874
FC= 51.69272
data = {{1.31901739, 0.98558670, 166.600510173}, {1.04570906,
    0.91024723, 138.53777873}, {1.05442656, 0.85718242,
    130.37280165}, {1.31451084, 0.94131837, 154.85238000}}
deltam = 0.7
rhom = 0.3
Cost1[x1_, x2_] := (delta*x1^(-rho) + (1 - delta)*x2^(-rho))^(-1/
      rho)*(delta^(1/(rho + 1))*(q1)^(rho/(rho + 1)) + (1 -
          delta)^(1/(rho + 1))*(q2)^(rho/(rho + 1)))^((rho + 1)/rho) +
   FC
Cost2[delta_, rho_] := (delta*x1^(-rho) + (1 - delta)*x2^(-rho))^(-1/
      rho)*(delta^(1/(rho + 1))*(q1)^(rho/(rho + 1)) + (1 -
          delta)^(1/(rho + 1))*(q2)^(rho/(rho + 1)))^((rho + 1)/rho) +
   FC
Taylorr[delta_, rho_] := Evaluate[D[Cost2[delta, rho], rho]]
Taylorrr[delta_, rho_] := Evaluate[D[Cost2[delta, rho], rho, rho]]
Taylord[delta_, rho_] := Evaluate[D[Cost2[delta, rho], delta]]
Taylordd[delta_, rho_] := Evaluate[D[Cost2[delta, rho], delta, delta]]
Taylorrd[delta_, rho_] := Evaluate[D[Cost2[delta, rho], rho, delta]]
App[x1_,x2_]:=Cost2[deltam, rhom] + Taylorr[deltam, rhom]*(rho - rhom) +
 Taylord[deltam, rhom]*(delta - deltam) +
 1/2*(Taylorrr[deltam, rhom]*(rho - rhom)^2 +
    2*Taylorrd[deltam, rhom]*(rho - rhom)*(delta - deltam) +
    Taylordd[deltam, rhom]*(delta - deltam)^2)
FindFit[data, {App[x1, x2], {0 <= delta <= 1, -1 <= rho}}, {delta,
  rho}, {x1, x2}]
```

LISTING B.3: Mathematica code for finding the parameters δ and ρ using a Taylor approximation.

APPENDIX C

R Code

```
library(forecast)
Daten = read.csv(file="A", dec=".", sep=",", header=TRUE)
DatenZeitreihe<-ts(Daten$B, freq= C, start = D)
plot(DatenZeitreihe, main="Title")
summary(DatenZeitreihe)
fit<-auto.arima(DatenZeitreihe, max.p = E, max.q = F)
summary(fit)
plot(forecast(fit,h=G))
forecast(fit,h=G)
```

LISTING C.1: Code to determine an ARIMA model.

```
library(vars)
Daten = read.csv(file="A", dec=",", sep=":", header=TRUE)
Cost_Components<-ts(Daten, freq=C, start = D)
VARselect(Cost_Components, lag.max = E, type = "const")
varmodel<-VAR(Cost_Components, p = P, type = "const")
summary(varmodel)
plot(predict(varmodel, n.ahead=5, ci=0.95))
```

LISTING C.2: Code to determine a VAR model.

Bibliography

[Aaker, 2013] Aaker, D. (2013). *Strategisches Markt-Management: Wettbewerbsvorteile Erkennen · Märkte Erschliessen · Strategien Entwickeln*. Gabler Verlag.

[Arens et al., 2008] Arens, T., Hettlich, F., Karpfinger, C., Kockelkorn, U., Lichtenegger, K., and Stachel, H. (2008). *Mathematik*. Spektrum Akademischer Verlag.

[Blaschke, 2008] Blaschke, C. (2008). *Das Marktpotential regenerativer Energien*. Diplom.de.

[Böhm, 2013] Böhm, V. (2013). *Arbeitsbuch zur Mikroökonomik I*. Heidelberger Taschenbücher. Springer Berlin Heidelberg.

[Brockwell and Davis, 2013] Brockwell, P. and Davis, R. (2013). *Time Series: Theory and Methods*. Springer Series in Statistics. Springer New York.

[Fahrmeir et al., 2013] Fahrmeir, L., Künstler, R., Pigeot, I., and Tutz, G. (2013). *Statistik: Der Weg zur Datenanalyse*. Springer-Verlag.

[Geoffrey et al., 2001] Geoffrey, A., Philip, J., and Reny, J. (2001). *Advanced microeconomic theory*. Boston, Addison-Wesley.

[Hamilton, 1994] Hamilton, J. (1994). *Time Series Analysis*. Princeton University Press.

[Hannan, 1980] Hannan, E. J. (1980). The estimation of the order of an arma process. *The Annals of Statistics*, pages 1071–1081.

[Ivaz, 2014] Ivaz, M. (2014). Konstruktion von kostenfunktionen für die maschinenbaubranche.

[Kmenta, 1967] Kmenta, J. (1967). On estimation of the ces production function. *International Economic Review*, 8(2):180–189.

[Königsberger, 2013] Königsberger, K. (2013). *Analysis 2*. Springer-Verlag.

[Kreiß and Neuhaus, 2006] Kreiß, J. and Neuhaus, G. (2006). *Einführung in Die Zeitreihenanalyse*. Statistik und ihre Anwendungen. Springer.

[Maddala and Lahiri, 1992] Maddala, G. S. and Lahiri, K. (1992). *Introduction to econometrics*, volume 2. Macmillan New York.

[Mishra, 2006] Mishra, S. K. (2006). A note on numerical estimation of sato's two-level ces production function. *Available at SSRN 947307*.

[Neusser, 2009] Neusser, K. (2009). *Zeitreihenanalyse in den Wirtschaftswissenschaften*. Studienbücher Wirtschaftsmathematik. Vieweg+Teubner Verlag.

[Peña et al., 2011] Peña, D., Tiao, G. C., and Tsay, R. S. (2011). *A course in time series analysis*, volume 322. John Wiley & Sons.

[Pindyck and Rubinfeld, 2013] Pindyck, R. S. and Rubinfeld, D. (2013). *Microeconomics (8th edn)*. Upper Saddle River, NJ: Pearson Prentice Hall.

[Porter, 1998] Porter, M. E. (1998). *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. Free Press, 1 edition.

[Schmidt, 2013] Schmidt, K. D. (2013). *Mathematik: Grundlagen für Wirtschaftswissenschaftler*. Springer-Verlag.

[Sydsæter et al., 2008] Sydsæter, K., Hammond, P., and Seierstad, A. (2008). *Further mathematics for economic analysis*. Pearson education.

[Symonds and Moussalli, 2011] Symonds, M. R. and Moussalli, A. (2011). A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using akaike's information criterion. *Behavioral Ecology and Sociobiology*, 65(1):13–21.

[Uebe, 2013] Uebe, G. (2013). *Produktionstheorie*, volume 114. Springer-Verlag.

[Wied-Nebbeling and Schott, 2006] Wied-Nebbeling, S. and Schott, H. (2006). *Grundlagen der Mikroökonomik*. Springer-Verlag.

[Wulfsberg and Redlich, 2011] Wulfsberg, J. and Redlich, T. (2011). *Wertschöpfung in der Bottom-up-Ökonomie*. VDI-Buch. Springer Berlin Heidelberg.

[Yasui, 1965] Yasui, T. (1965). The ces production function: A note. *Econometrica: Journal of the Econometric Society*, pages 646–648.