Faculty of Mathematics and Physics, Comenius University

Mathematics of Economy and Finance

ECONOMETRIC MODELS OF PRODUCER BEHAVIOUR

Master Thesis

Supervisor: Univ.Doz.Dr.Bernhard Böhm
Master student: Katarína Zigová

Bratislava 2000

" I declare this thesis has been written by myself with the help of my supervisor and the referred literature "

“ I would like to express special thanks to God for the possibilities he gave me; to my parents who have created suitable conditions for my personal growth and to my supervisor for a great amount of worthy comments. ”   K.Z.

Contents

1 Introduction **3**

2 Microeconomic Theory **5**
2.1 Technology and Production Function 5
2.2 Cost Function 7
2.3 Duality 9
2.4 Functional Forms 10
2.4.1 Survey of the Linear-in-Parameters Forms 10
2.4.2 Deriving the Translog Cost Function 11

3 Econometric Methods **15**
3.1 Estimation Methods 15
3.1.1 Ordinary and Generalized Least Squares 16
3.1.2 Maximum Likelihood Estimation 19
3.1.3 Seemingly Unrelated Regression 21
3.1.4 Estimation under Restrictions 23
3.2 Statistical Inference 24
3.2.1 Test of Linear Restrictions 24
3.2.2 Likelihood Ratio, Wald and Lagrange Multiplier Test Statistics 27
3.2.3 Durbin-Watson test 28
3.3 Problem of Vector Autocorrelation 30

4 Applications **33**
4.1 Translog Cost Function for the Paper Industry 33
4.1.1 Methods 35
4.1.2 Results 36
4.1.3 Estimating the Cost Function 40
4.2 Testing 41
4.2.1 Testing the Constraints 41
4.2.2 Recursive Estimates 41
4.2.3 One-Step Ahead Prediction Errors 42
4.2.4 Chow Test for Constancy of Parameters 43

4.3 Estimating with Vector Autocorrelation 44
4.3.1 Methods 44
4.3.2 Results 47
4.3.3 Comparing with the Basic Model 49

5 Summary **51**
5.1 Challenges for Future Research 51
5.2 Concluding Remarks 52

6 Resume (in Slovak) **55**

Bibliography **57**

A Appendix **59**

Chapter 1

Introduction

The purpose of this work is to provide an exposition of econometric methods for modelling producer behaviour. The objective of econometric modeling is to determine the nature of substitution among inputs, the character of differences in technology and the role of economies of scale.

The empirical analysis of input demands and input substitution patterns provides an example of the strong links between economic theory and econometric implementation. In my example, while the underlying economic theory emphasizes the joint nature of energy-inputs demand decisions for paper production, econometric implementations of this interdependence involve simultaneous estimation of parameters in systems of factor demand equations having cross-equation constraints. I also consider alternative procedures for obtaining statistical inference on the empirical validity of hypotheses involving cross-equation parameter restriction, the measurements of goodness of fit in equation system and special properties of singular equation system.

Important innovations in specifying econometric models have arisen from the dual formulation of the theory of production. The dual formulation of production theory has made it possible to overcome the limitations of the traditional approach to econometric modelling. The key features of the dual formulation are, first to characterize the production function by means of a dual representation such as a price or cost function and, second to generate explicit demand and supply function as derivatives of the price or cost function. The dual approach has a crucial advantage in the development of econometric methodology. Demands and supplies can be generated as explicit functions of relative prices without imposing the arbitrary constraints on production patterns required by the classic methodology. The econometric modelling of producer behaviour requires parametric forms for demand and supply functions. These functions can be parametrized by treating measures of substitution, technical change and economies of scale as unknown parameters to be estimated on the basis of empirical data.

Econometric implementations of cost and production functions differ in their assumptions concerning exogeneity. In the production function regression equation, output is endogenous and input quantities are exogenous. By contrast, in the dual cost function, production cost and input quantities are endogenous, while input prices and the level of output are exogenous. It follows that in our case input prices are regressors and input prices and the level of output can be moreover assumed as exogenous. Thus there exist two good reasons to prefer cost function rather than production one. First, that we can characterize the production function by means of a dual representation and, second, to generate explicit demand and supply functions as derivatives of the price or cost fuction.

What will be showed later.

Econometric models of producer behaviour take the form of system of demand and supply functions. The variables may enter these functions in a nonlinear manner as it is in case of translog demand function which I have implemented in this work. Translog formula belongs to linear-in-parameters forms. This work connects to some previous one done in 1996 by Florian Haider [5], an Austrian student. My main aim is to provide some extensions, or improvements of his model. There are many ways which are offered by theory but as you will see reality is a very special case of the theory.

Perhaps nowhere else has the increased sophistication of statistical software made a greater mark on econometric practise than in microeconomics applications. For modeling of my example of paper production behaviour I have used two software packages Eviews and Limdep7.0 as well.

This work begins with two chapters where is short overview of the theory, underlying the applications which are presented in the 4th chapter. In the last chapter I conclude the work by discussing the challenges for future research and outlining frontiers which gives theory by modeling reality.

Chapter 2

Microeconomic Theory

In this chapter some elements of the microeconomic theory associated with my thesis will be explained. Namely, the theory about production, cost function and relationship between them. There are two main purposes of the following chapter. First, it is necessary to understand the reasons for the existence of some special properties of the theoretical model which are then transformed into restrictions imposed on the estimation process. Second, some remarks on elasticities and measures of scale arising as unknown parameters in regression will be made.

2.1 Technology and Production Function

The most common way to describe the technology of a firm is the **production function**. A firm produces output y from various combination of inputs $\mathbf{x}^T = (x_1, x_2, \ldots, x_m)$. For us are important only outputs which are possible to produce from some given amount. So equivalently we can say that firm is described by a set of feasible combinations of inputs and outputs. All feasible combinations $(y, \mathbf{x}) \in R_+^{m+1}$ form a set Y which is known as **production possibility set**:

$$Y \subset R_+^{m+1}.$$

Set Y has to fullfil following axioms to be production possibility set.

$$\begin{aligned}
&\mathrm{A1}: \quad (0, \mathbf{x}) \in Y \qquad \forall \mathbf{x} \geq \mathbf{0} \\
&\mathrm{A2}: \quad (y, \mathbf{0}) \in Y \Rightarrow y = 0 \\
&\mathrm{A3}: \quad \forall \mathbf{x} \geq \mathbf{0} \quad \text{exists} \quad y \geq 0 \qquad \text{so that} \qquad (y, \mathbf{x}) \notin Y \\
&\mathrm{A4}: \quad (y, \mathbf{x}) \in Y \wedge x' \geq x \Rightarrow (y, \mathbf{x}') \in Y \\
&\mathrm{A5}: \quad Y \quad \text{is convex and closed}
\end{aligned}$$

where $\mathbf{x}^T = (x_1, \ldots, x_m)$, $\mathbf{x}^{T'} = (x'_1, \ldots, x'_m)$ are two different input vectors and $\mathbf{x} \geq \mathbf{x}' \Leftrightarrow x_i \geq x'_i \quad \forall i$; $\mathbf{0}$ is the $m \times 1$ vector of zeroes. Production function f is then defined as $f : R_+^m \to R$

$$f(x) = \max\{y : (y, \mathbf{x}) \in Y\} \tag{2.1}$$

Simply said it is the maximal output produced by some amount of input vector $\mathbf{x}$. The following theorem describes the production function.

Theorem: If production function f is defined on R_+^m and has following properties

$$\begin{aligned} f &\quad \text{is concave} \\ \forall \mathbf{x}' \geq \mathbf{x} &\quad \text{is} \quad f(\mathbf{x}') \geq f(\mathbf{x}) \\ f(\mathbf{0}) = 0; f(\mathbf{x}) \geq 0 &\quad \text{for} \quad \mathbf{x} \geq \mathbf{0} \end{aligned} \tag{2.2}$$

then $Y = \{(y, \mathbf{x}) : y \leq f(\mathbf{x})\}$ is production possibility test. For proof see [10].

All input bundles which gives the same amount of output y determines **isoquant** curve:

$$y = f(\mathbf{x}) = \text{const} \tag{2.3}$$

Isoquant gives all input bundles that produce exactly const units of output. All such bundles create a set

$$I_c : \{\mathbf{x} : f(\mathbf{x}) = \text{const}\}$$

Then $f(\mathbf{x}) - c = 0$ is implicit given function. Assume now that we are producing at a particular point $\hat{\mathbf{x}}$:

$$\text{const} = f(\hat{x}_1, \ldots, \hat{x}_m) \tag{2.4}$$

Suppose that we want to increase the amount of input i and decrease of input j. So as to maintain a constant level of output.

$$f(\hat{x}_1, \ldots, \hat{x}_i + dx_i, \hat{x}_{i+1}, \ldots, \hat{x}_j + dx_j, \hat{x}_{j+1}, \ldots, \hat{x}_m) = \text{const}$$

There exist a measure, which shows how one of the input must adjust in order to keep output constant when another input changes. Such measure between ith and jth factor is known as **technical rate of substitution** (TRS). It is determined by totally differentiating the identity (2.4) with respect to $x_i(x_j)$:

$$\begin{aligned} \tfrac{\partial f}{\partial x_i} dx_i + \tfrac{\partial f}{\partial x_j} dx_j &= 0 \\ \Rightarrow \quad \tfrac{dx_j}{dx_i} &= -\tfrac{\partial f / \partial x_i}{\partial f / \partial x_j} \end{aligned} \tag{2.5}$$

TRS$= -dx_j/dx_i > 0$ because of concavity of the production function. The expression (2.5) could be rewritten as

$$\frac{dx_j}{dx_i} = -\frac{MP_i}{MP_j} \qquad \text{where} \qquad MP_i = \frac{\partial y}{\partial x_i}$$

MP_i is known as marginal product and therefore dx_j/dx_i is often called **marginal rate of technical substitution**.

Another point of view offers **elasticity of substitution** measure. It gives the percentage change in the factor ratio divided by the percentage change in the TRS. It is expressed as

$$\sigma = \frac{\frac{\Delta(x_i/x_j)}{x_i/x_j}}{\frac{\Delta TRS}{TRS}} \tag{2.6}$$

where $\Delta(x_i/x_j)$ is the change in the factor ratio and ΔTRS is the change in the technical rate of substitution. In practise we think of the percent change as being very small and we take limit of (2.6) as $\Delta \to 0$. Thus σ becomes

$$\sigma = -\frac{TRS}{x_j/x_i} \frac{d(x_j/x_i)}{dTRS}$$

It is often convenient to calculate σ using the logaritmic derivative:

$$\sigma = -\frac{dx_j/x_j}{dx_i/x_i} = -\frac{dx_j}{dx_i}\frac{x_i}{x_j} = -\frac{d(\ln x_i)}{d(\ln x_j)} \tag{2.7}$$

Graphically in x_i, x_j plane TRS (2.5) measures the slope of an isoquant and elasticity of substitution (2.7) measures the curvature of an isoquant (2.3).

In the cases where we wanted only to scale output up by some amount we use the concept of **returns to scale** (RTS). RTS reflect the degree to which a proportional increase in all inputs, increases output. For const a which indicate the proportional increase we say, that technology exhibits

$$\begin{aligned} &\text{constant RTS if } f(a\mathbf{x}) = af(\mathbf{x}) \wedge \epsilon = 1 \wedge a > 0 \\ &\text{increasing RTS if } f(a\mathbf{x}) > af(\mathbf{x}) \wedge \epsilon > 1 \wedge a > 1 \\ &\text{decreasing RTS if } f(a\mathbf{x}) < af(\mathbf{x}) \wedge \epsilon < 1 \wedge a > 1 \end{aligned}$$

where $\epsilon = \sum_{i=1}^{m} \frac{\partial \ln y}{\partial \ln x_i}$ is the sum of partial elasticities of production.

Last remark is about **homogenous** and **homothetic** technologies. A function $f(\mathbf{x})$ is homogenous of degree k if $f(t\mathbf{x}) = t^k f(\mathbf{x})$ for all $t > 0$. From economic point of view are two important degrees. A zero-degree homogenous function is one for which $f(t\mathbf{x}) = f(\mathbf{x})$ and a first-degree homogenous function is one for which $f(t\mathbf{x}) = tf(\mathbf{x})$. Comparing the previous with concept of RTS, production function has constant RTS if and only if is homogenous of degree one. A homothetic function is a monotonic transformation of a function that is homogenous of degree one. Simply said $f(\mathbf{x})$ is homothetic if and only if it can be written as $f(\mathbf{x}) = g(h(\mathbf{x}))$, where $h(\cdot)$ is homogenous of degree one and $g(\cdot)$ is a monotonic function [10].

2.2 Cost Function

The rational behaviour of the firm can also be described, instead of maximizing of production function (2.1), by cost minimizing.

Let us consider the problem of finding a cost minimizing input vector $\hat{\mathbf{x}}$ to produce a given level of output. The total cost of the firm is determined by vector of input prices $\mathbf{p}^T = (p_1, \ldots, p_m)$ multipied by values of input factors vector $\mathbf{x}$.

$$\begin{aligned} &\min_{\mathbf{x}} \mathbf{p}^T\mathbf{x} \\ &\text{s.t} \quad f(\mathbf{x}) = y. \end{aligned} \tag{2.8}$$

By using the method of Lagrange multipliers

$$L(\lambda, \mathbf{x}) = \mathbf{p}^T\mathbf{x} - \lambda(f(\mathbf{x}) - y)$$

and differentiate it with respect to each x_i and the Lagrange multiplier, λ. The first-order conditions are

$$\begin{aligned} p_i - \lambda\frac{\partial f(\hat{\mathbf{x}})}{\partial x_i} = 0 \quad \text{for} \quad i = 1, \ldots, m \\ f(\hat{\mathbf{x}}) = y \end{aligned} \tag{2.9}$$

The first equation can be rewritten into vector notation

$$\mathbf{p} = \lambda[Df(\hat{\mathbf{x}})] \tag{2.10}$$

where $Df(\mathbf{x})$ is the gradient vector of function f. There are also a second order conditions that must be satisfied at a cost-minimizing choice. Let us look on (2.9) and (2.10) from the implicit function theorem

$$\begin{pmatrix} D^2 f(\mathbf{x}) & [Df(\mathbf{x})]^T \\ Df(\mathbf{x}) & 0 \end{pmatrix}$$

is the matrix of the partial derivatives of the system (2.9) and (2.10) and $D^2 f(\mathbf{x})$ is the corresponding $m \times m$ Hessian matrix of production function. Determinant of this matrix $(Df)(D^2 f)(Df)^T \neq 0$. If $D^2 f$ is negative semidefinite then the second order conditions are fulfiled and there exists an unique solution of cost-minimizing problem (2.8). For each choice of $\mathbf{p}$ and y there will be some choice of $\hat{\mathbf{x}}$ that minimizes the cost of producing y units of output. We will call the function that gives us this optimal choice the **conditional factor demand function** and write it as $\mathbf{x}(\mathbf{p}, y)$. If the firm produces its output from m various inputs then we have whole system of factor demands. For each factor one demand function. So in multi-input case $\mathbf{x}(\mathbf{p}, y)$ is $m \times 1$ vector.

In general the cost function can always be expressed simply as the value of the conditional factor demands

$$C(\mathbf{p}, y) \equiv \mathbf{p}^T \mathbf{x}(\mathbf{p}, y) \tag{2.11}$$

and defines the minimum cost of producing particular output with given input prices. C indicates the cost function. Expression (2.11) is in fact definition of the costs $C = \mathbf{p}^T \mathbf{x}$, where conditional factor demands are substituted. Let us look deeper into cost function. Its properties will help to understand the restrictions which will be imposed by later investigation. A cost function has following properties [10]:

(i) Nondecreasing in $\mathbf{p}$: if $\mathbf{p}' \geq \mathbf{p}$, then $C(\mathbf{p}', y) \geq C(\mathbf{p}, y)$
(ii) Homogenous of degree one in $\mathbf{p}$: $C(t\mathbf{p}, y) = tC(\mathbf{p}, y)$; for $t > 0$
(iii) Concave in $\mathbf{p}$, convex in y

Proof:

(i) Let $\mathbf{x}$ and $\mathbf{x}'$ be cost-minimizing bundles associated with $\mathbf{p}$ and $\mathbf{p}'$. Then $\mathbf{p}^T\mathbf{x} \leq \mathbf{p}^T\mathbf{x}'$ by minimization and $\mathbf{p}^T\mathbf{x}' \leq \mathbf{p}^{T'}\mathbf{x}'$ since $\mathbf{p} \leq \mathbf{p}'$. Putting this two inequalities together gives $\mathbf{p}^T\mathbf{x} \leq \mathbf{p}^{T'}\mathbf{x}'$.
(ii) This property is automatically clear, because if $\mathbf{x}$ is the cost-minimizing bundle at prices $\mathbf{p}$, then $\mathbf{x}$ also minimizes costs at prices $t\mathbf{p}$.
(iii) Let $(\mathbf{p}, \mathbf{x})$ and $(\mathbf{p}', \mathbf{x}')$ be two cost-minimizing price-factor combinations and let $\mathbf{p}'' = \alpha\mathbf{p} + (1-\alpha)\mathbf{p}'$ for any $0 \leq \alpha \leq 1$. Now

$$C(\mathbf{p}'', y) = \mathbf{p}''^T\mathbf{x} = \alpha\mathbf{p}^T\mathbf{x}'' + (1-\alpha)\mathbf{p}'^T\mathbf{x}''.$$

Since $\mathbf{x}''$ is not necessarily the cheapest way to produce y at prices $\mathbf{p}'$ and $\mathbf{p}$ we have

$$C(\mathbf{p}'', y) \geq \alpha C(\mathbf{p}, y) + (1-\alpha)C(\mathbf{p}', y).$$

In case of y, y' and $(C(y), C(y'))$ as two different outputs and corresponding cost functions

$$\begin{aligned} \alpha C(y) + (1-\alpha)C(y') &= \alpha\mathbf{p}^T\mathbf{x}(y) + (1-\alpha)\mathbf{p}^T\mathbf{x}(y') = \\ &= \mathbf{p}^T(\alpha\mathbf{x}(y) + (1-\alpha)\mathbf{x}(y')) \end{aligned} \tag{2.12}$$

and from concavity of the production function follows

$$f(\alpha\mathbf{x}(y) + (1-\alpha)\mathbf{x}(y')) \geq \alpha f(\mathbf{x}(y)) + (1-\alpha)f(\mathbf{x}(y')) \geq \alpha y + (1-\alpha)y'$$

Then substituting last inequality into (2.12) provides

$$\alpha C(y) + (1-\alpha)C(y') \geq \mathbf{p}^T\mathbf{x}(\alpha y + (1-\alpha)y') \geq C(\alpha y + (1-\alpha)y'). \quad \square$$

Very useful result is known as **Shephard's lemma**. With help of this lemma factor demand function $\mathbf{x}(\mathbf{p}, y)$ are obtained, which play the main role in my applications. In fact, it is a special property of the cost function (2.11).

Shephard's lemma: Let $x_i(\mathbf{p}, y)$ be the firms conditional factor demand for input i. If the cost function is differentiable at $(\mathbf{p}, y)$ and $p_i > 0$ for $i = 1, \ldots, m$ then

$$x_i(\mathbf{p}, y) = \frac{\partial C(\mathbf{p}, y)}{\partial p_i} \qquad i = 1, \ldots, m \tag{2.13}$$

Proof: Let $\hat{\mathbf{x}}$ be a cost-minimizing bundle that produces y at prices $\hat{\mathbf{p}}$. Then define the function

$$g(\mathbf{p}) = C(\mathbf{p}, y) - \mathbf{p}^T\hat{\mathbf{x}}$$

Since $C(\mathbf{p}, y)$ is the cheapest way to produce y, this function is always nonpositive, at $\mathbf{p} = \hat{\mathbf{p}}; \quad g(\mathbf{p}) = 0$. Since this is a maximum value of $g(\mathbf{p})$ and by deriving it gives

$$\frac{\partial g(\hat{\mathbf{p}})}{\partial p_i} = \frac{\partial C(\hat{\mathbf{p}}, y)}{\partial p_i} - \hat{x}_i = 0 \qquad i = 1, \ldots, m.$$

Hence, the cost-minimizing input vector is just given by the vector of derivatives of the cost function with respect to the prices. $\square$

There are two reasons for investigating the problem in dimension of the cost function. First, the cost function allows us to model the production behaviour of firms without knowing of market price of output. Second, for modeling producer behaviour we use the system of demand equations, derived with help of the Shephard's lemma. Using the classic approach, i.e. profit maximizing s.t. production function generates demands and supplies as implicit function of the relative prices [7]. Using cost minimizing approach we avoid it.

2.3 Duality

In the previous section we have discussed the properties of the cost function. Given any technology we can directly derive its cost function, by solving the cost-minimization problem

$$C(\mathbf{p}, y) \equiv \min_{\mathbf{x}}\{\mathbf{p}^T\mathbf{x} : f(\mathbf{x}) \geq y\}. \tag{2.14}$$

In this section it will be showed that this process can be reversed. Through definition (2.14) production function f determines a cost function C.

The production function can in general be obtained from a cost function satisfying the appropriate regularity conditions as the solution to the following constrained maximization problem:

$$f^*(\hat{\mathbf{x}}) \equiv \max_{y}\{y : C(\mathbf{p}, y) \leq \mathbf{p}^T\hat{\mathbf{x}} \quad \text{for every} \quad p_i \geq 0\} \tag{2.15}$$

where $\hat{\mathbf{x}}^T = (\hat{x}_1, \ldots, \hat{x}_m)$ is given vector of inputs and C is the given cost function. This means that the production function contains essentially the same information that the cost function. This general observation is known as the priciple of **duality**. Given one of these functions under certain regularity conditions, the other can be uniquely determined, this result is summarized in the following theorem.

Samuelson-Shephard duality theorem:

(i) If the production function f satisfies the conditions (2.2) then for $y > 0$ and $p_i > 0$, for $i = 1, \ldots, m$, the cost function defined by (2.14) factors into the following expressions:

$$C(\mathbf{p}, y) = C(\mathbf{p})y,$$

where the unit cost function $C(\mathbf{p})$ also satisfies (2.2).

(ii) If the unit cost function $C(\mathbf{p})$ satisfies (2.2), then for $\hat{x}_i > 0$ for $i = 1, \ldots, m$ the function f^* defined by (2.15) also satisfies (2.2). So f^* can be interpreted as a production function. Equivalent expression of (2.15) is

$$f^*(\hat{\mathbf{x}}) \equiv \frac{1}{\max\limits_{\mathbf{p}}\{C(\mathbf{p}) : \mathbf{p}^T\hat{\mathbf{x}} = 1, p_i > 0 \quad \forall i\}}. \tag{2.16}$$

(iii) Let the unit cost function $C(\mathbf{p})$ satisfy condition (2.2) and define the production function f^* by (2.16). Now define the unit cost function C^* generated by f^* for $p_i > 0$ as

$$C^*(\mathbf{p}) \equiv \min_{\mathbf{x}}\{\mathbf{p}^T\mathbf{x} : f^*\mathbf{x} \geq 1; x_i \geq 0 \quad \text{for} \quad i = 1, \ldots, m\},$$

then for every $p_i^* > 0$ we have $C(\mathbf{p}^*) = C^*(\mathbf{p}^*)$, so the production function f^* which was defined by the original unit cost function C, has a unit cost function C^* which coincides with C.

For proof see [4]. This theoretical result has many modifications, together with the Shephard's lemma (see section 2.2) makes duality theory extremly useful for empirical applications.

2.4 Functional Forms

This section offers an historical overview of the functional forms beginning with the famous Cobb-Douglas form, moving on the constant elasticity of substitution specification, and concluding with flexible functional forms such as the generalized Leontief and logarithmic translog representation.

2.4.1 Survey of the Linear-in-Parameters Forms

Untill now I have spoken just about technology and costs in general. This section offers a survey of common linear-in-parameters functional forms. This forms can be used as a production or cost function depending on which access, direct or dual, is being considered.

Probably the oldest form is known as **Cobb-Douglas**,given by

$$\log y = a_0 + \sum_{i=1}^{m} a_i \log x_i$$

where $\sum_{i=1}^{m} a_i = 1$ is a restriction of homogeneity of degree one in inputs factors, respectively in input prices. This form was developed by Charles Cobb and Paul Douglas in 1928. For our purposes this form is too restrictive because elasticity of substitution (2.7) equals in this case always unity.

Economists who were interested in estimating σ rather then assuming that $\sigma = 1$ have been searching for such form where σ is still constant but not necessary equal to one. This form is called **constant elasticity of substitution** (CES) and is expressed by

$$Y = (a_0 + \sum_{i=1}^{m} a_i x_i^{\rho})^{1/\rho}$$

where $a_0 = 0$ for homogeneity of degree one. Earl Heady and his colleagues at Iowa State University wanted to model some agricultural experiments, but including input combination that resulted in negative marginal product $\mathrm{MP}_i = \partial y/\partial x_i$, which is impossible in the Cobb-Douglas concept. Thus Heady generalized that form [1].

The following two forms are used the most for the empirical implementations because of the flexibility which they offer. First, the **Generalized Leontief** linear functional form can be written as

$$y = a_0 + \sum_{i=1}^{m}\sum_{j=1}^{m} a_{ij}(x_i x_j)^{1/2}$$

where symmetric restriction $a_{ij} = a_{ji} \quad \forall i, j$ are added. The next form is probably most used and will be deeply analysed because it builds the core of my empirical application. In theoretical literature [7], [1] it is known as **Translog functional form**. This concept can be expressed as

$$\ln y = a_0 + \sum_{i=1}^{m} a_i \ln x_i + \sum_{i=1}^{m}\sum_{j=1}^{m} a_{ij}(\ln x_i)(\ln x_j) \tag{2.17}$$

where $\sum_{i=1}^{m} a_i = 1$ and $\sum_{i=1}^{m} a_{ij} = 0 \quad \forall j$ are restriction for homogeneity of degree one. This form can be envisaged as a second-order Taylor's series approximation in logarithms to any arbitrary cost or production function.

We can assume for simplicity that all forms exhibit constant returns to scale. Also for all forms we can generalize $\mathbf{x}^T = (x_1, \ldots, x_m)$ to represent also a vector of inputs or prices and y is output or cost depending on whether direct or dual approach is being considered. As was said in the begining of this section this forms can be used in two ways as cost or production function. So it seems not to matter which use is made of these functional forms. However the dual formulation of production theory has crucial advantage in the developement of econometric modelling: Demands and supplies can be generated as explicit functions of relative prices $\mathbf{x} = \mathbf{x}(\mathbf{p}, y)$ (see section 2.2) without imposing the constraints of production required in case of the direct approach [7].

2.4.2 Deriving the Translog Cost Function

In this part the translog cost function will be derived. Corresponding with (2.17) cost function faces following form:

$$\begin{aligned} \ln C(\mathbf{p}, y) &= \ln \gamma_0 + \sum_{i=1}^{m} \gamma_i \ln p_i + \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \gamma_{ij} \ln p_i \ln p_j \\ &\quad + \gamma_y \ln y + \frac{1}{2}\gamma_{yy}(\ln y)^2 + \sum_{i=1}^{m} \gamma_{iy} \ln p_i \ln y \end{aligned} \tag{2.18}$$

where y is value of output $\mathbf{p}^T = (p_1, \ldots, p_m)$ is $m \times 1$ vector of input factor prices. Note that among exogenous variables of the formula belong the logarithms of input prices and the level of output as well. Rewriting (2.18) into matrix form gives

$$\begin{aligned} \ln C(\mathbf{p}, y) &= \ln \gamma_0 + \boldsymbol{\gamma}'_{\mathbf{p}} \ln \mathbf{p} + \frac{1}{2} \ln \mathbf{p}' \Gamma_{\mathbf{pp}} \ln \mathbf{p} + \\ & + \gamma_y \ln y + \frac{1}{2} \gamma_{yy} (\ln y)^2 + \boldsymbol{\gamma}'_{\mathbf{p}y} \ln \mathbf{p} \ln y, \end{aligned} \tag{2.19}$$

where $\ln \mathbf{p}$ is $m \times 1$ column vector of logarithmus of input prices, $\Gamma_{\mathbf{pp}}$ is $m \times m$ matrix of corresponding parameters (γ_{ij}) $i, j = 1, \ldots, m$ and $\boldsymbol{\gamma}_{\mathbf{p}}$, resp. $\boldsymbol{\gamma}_{\mathbf{p}y}$ are corresponding $m \times 1$ vectors of parameters γ_i, resp. γ_{iy}. Note that here the cost function depends from the prices and the amount of output. We can refer to this form as the translog cost function, indicating the role of the variables, or the constant share elasticity cost function, indicating the role of the parameters [7]. The concept of the share elasticity will be later explained. Corresponding restrictions are

$$\sum_{i=1}^{m} \gamma_i = 1 \qquad \sum_{i=1}^{m} \gamma_{ij} = \sum_{j=1}^{m} \gamma_{ij} = \sum_{i=1}^{m} \gamma_{iy} = 0. \tag{2.20}$$

For the translog cost function to be homothetic there is an adding up restriction that $\gamma_{iy} = 0 \quad \forall i$. If we assume homogeneity of constant degree in output we have additionally to impose the restriction $\gamma_{yy} = 0$ [1].

Now I have extracted from big number of formulas one which can fit to our model. But the econometric approach (which will be explained in the next chapter)is parametric approach, so the result of an empirical application is not only the statement about rejecting or not rejecting of the theretical model, but result contains the value of parameters (in case of 2.18, all γ's) which tell us something about behaviour inside the model. Looking deeper into the derivation of formula (2.18) and (2.20) will explain the meaning of the parameters.

Let us turn back to the concept of cost function. First we start to define some more concepts concerning costs and cost function:

$$\textbf{cost shares:} \qquad v_j = \frac{p_j x_j}{c} \quad j = 1, \ldots, m$$

where $c = \sum_{j=1}^{m} p_j x_j$ are **total cost**. With output fixed, the necessary condition for producer equilibrium when maximizing profit s.t. a production function $f(x)$ are given by equalities:

$$\mathbf{v} = \frac{\partial \ln y / \partial \ln \mathbf{x}}{\imath' (\partial \ln y / \partial \ln \mathbf{x})} \quad \text{where} \quad \imath \quad \text{is a vector of ones}$$

where $\mathbf{v}^T = (v_1, \ldots, v_m)$. Given a definition of total cost and necessary condition for producer equilibrium, we can express total cost c as a function of the input prices and the level of output: (see 2.11). Cost shares of all inputs can be expressed as elasticities of the cost function with respect to the input prices:

$$\mathbf{v} = \frac{\partial \ln C}{\partial \ln \mathbf{p}} (\mathbf{p}, y).$$

Index of returns to scale or **cost flexibility** expressed as elasticity of the cost function with respect to the level of output is like:

$$v_y = \frac{\partial \ln C}{\partial \ln y} (\mathbf{p}, y).$$

The cost flexibility v_y, as derived from the production function, is the reciprocal of the degree of RTS.

$$v_y = \frac{1}{\imath'} \frac{\partial \ln y}{\partial \ln \mathbf{x}}.$$

Next measure is known as **share elasticities**, which are expressed as

$$U_{\mathbf{pp}} = \frac{\partial^2 \ln C}{\partial \ln \mathbf{p}^2}(\mathbf{p}, y) = \frac{\partial \mathbf{v}}{\partial \ln \mathbf{p}}(\mathbf{p}, y)$$

and obtained by differentiating the logarithm of the cost function twice with respect to the logarithms of input prices. This measure gives the response of the cost shares of all inputs to proportional changes in the input prices. Note that $U_{\mathbf{pp}}$ is the $m \times m$ vector. By differentiating the logarithm of the cost function twice with respect to the logarithms of the input prices and the level of output, **biases of scale** measures, are obtained:

$$\mathbf{u}_{\mathbf{p}y} = \frac{\partial^2 \ln C}{\partial \ln \mathbf{p} \partial \ln y}(\mathbf{p}, y) = \frac{\partial \mathbf{v}}{\partial \ln y}(\mathbf{p}, y) = \frac{\partial v_y}{\partial \ln \mathbf{p}}(\mathbf{p}, y).$$

This vector can be employed to derive the implications of economies of scale for the relative distribution of total cost among inputs or to derive the implications of changes in input prices for the cost flexibility. Derivative of the cost flexibility with respect to the logarithm of output provide

$$u_{yy} = \frac{\partial^2 \ln C}{\partial \ln y^2}(\mathbf{p}, y) = \frac{\partial v_y}{\partial y}(\mathbf{p}, y).$$

This measure gives the response of the cost flexibility to proportional changes in the level of output.

Now we want to generate an econometric model of cost and production by assuming that our parameters are the following constants:

$$\Gamma_{\mathbf{pp}} = U_{\mathbf{pp}} \qquad \boldsymbol{\gamma}_{\mathbf{p}y} = \mathbf{u}_{\mathbf{p}y} \qquad \gamma_{yy} = u_{yy}.$$

We can regard this system as a system of second-order partial differential equations. We can integrate this system with respect to $\ln \mathbf{p}$, $\ln y$ to obtain a system of first-order partial differential equations

$$\begin{aligned} \mathbf{v} &= \boldsymbol{\gamma}_{\mathbf{p}} + \Gamma_{\mathbf{pp}} \ln \mathbf{p} + \boldsymbol{\gamma}_{\mathbf{p}y} \ln y \\ v_y &= \gamma_y + \boldsymbol{\gamma}'_{\mathbf{p}y} \ln \mathbf{p} + \gamma_{yy} \ln y \end{aligned}$$

where $\boldsymbol{\gamma}_{\mathbf{p}}$, γ_y are constants of integration and when $\mathbf{p} = \mathbf{1}$, $y = 1$, then $\boldsymbol{\gamma}_{\mathbf{p}} = \mathbf{v}$ and $\gamma_y = v_y$. Now we can integrate this system again with respect to $\ln \mathbf{p}$, $(\ln y)$ to obtain the cost function (2.19).

This derivations require imposing of some restrictions. All restrictions build up the set of restrictions, that need to be considered in the regression model. The complete set of conditions for integrability is as follows [7]:

Homogeneity: The cost shares and the cost flexibility are homogenous of degree zero in the input prices, since the cost function is homogenous of degree one.

$$\Gamma_{\mathbf{pp}} \imath = \mathbf{0} \qquad \boldsymbol{\gamma}'_{\mathbf{p}y} \imath = 0$$

For m inputs there are $m + 1$ restrictions implied by homogeneity.

Cost exhaustion: The sum of the cost shares is equal to unity. Cost exhaustion implies that the value of m inputs is equal to total cost.

$$\boldsymbol{\gamma}'_{\mathbf{p}} = 1 \qquad \Gamma_{\mathbf{pp}}\imath = \mathbf{0} \qquad \boldsymbol{\gamma}'_{\mathbf{p}y}\imath = 0$$

For m inputs there are $m + 2$ restrictions implied by cost exhaustion.

Symmetry: The matrix of share elasticities, biases of scale and the derivative of the cost flexibility with respect to the logarithm of output must be symmetric.

$$\begin{pmatrix} \Gamma_{\mathbf{pp}} & \boldsymbol{\gamma}_{\mathbf{p}y} \\ \boldsymbol{\gamma}'_{\mathbf{p}y} & \gamma_{yy} \end{pmatrix} = \begin{pmatrix} \Gamma_{\mathbf{pp}} & \boldsymbol{\gamma}_{\mathbf{p}y} \\ \boldsymbol{\gamma}'_{\mathbf{p}y} & \gamma_{yy} \end{pmatrix}'$$

Nonnegativity: The cost shares and the cost flexibility must be nonnegative. Since the translog cost function is quadratic in the logarithms of the input prices and the level of output, we cannot impose this restriction, instead, we consider restrictions on the parameters that imply monotonicity of the cost shares wherever they are nonnegative.

Monotonicity: The matrix of share elasticities $\Gamma_{\mathbf{pp}} + \mathbf{vv}'$ is nonpositive definite.

Concavity: The cost function is concave, wherever the cost shares are nonnegative.

To summarize, if one logarithmically differentiates equation (2.18) with respect to input prices and then employs Shephard's Lemma (see section 2.2), one obtains cost share equations of the form.

$$\frac{\partial \ln C}{\partial \ln p_i} = \frac{p_i}{C}\frac{\partial C}{\partial p_i} = \frac{p_i x_i}{c} = \gamma_i + \sum_{j=1}^{m} \gamma_{ij} \ln p_j + \gamma_{iy} \ln y \tag{2.21}$$

for $i = 1, \ldots, m$. Defining the cost shares $v_i \equiv p_i x_i / c$, it follows that $\sum_{i=1}^{n} v_i = 1$. This condition of the share equation system (2.21) has important implication for econometric estimation. I will illustrate this issues in chapter (4).

Chapter 3

Econometric Methods

3.1 Estimation Methods

The objective of econometric modeling is to find numerical values of relevant parameters after an appropriate specification of the economic relations and the device of a suitable estimation technology. An estimator is a formula, method, or recipe for estimating an unknown population parameter, and an estimate is the numerical value obtained when sample data are substituted in the formula.

Economic theory contains plenty of relationships between variables taken in pairs: quantity and price, consumption and income, unemployment and the inflation rate and many more. It suggest the opinion that economists believe that the world can be analyzed only in terms of a collection of bivariate relationships. But nonetheless some bivariate relationships are significant for understanding the basis of statistical and mathematical tools, which are reconstructed on more complicated situations.

The simplest version of the two variable model is:

$$y_t = \alpha + \beta X_t + u_t$$

with u_t being iid $\sim (0, \sigma^2)$ and $t = 1, \ldots, n$ is number of observations. There are thus three parameters to be estimated: α, β and σ^2. The parameters α and β are taken as a pair, since numerical values of both are required to fit a specific line. Once such a line has been fitted, the residuals from that line may be used to form an estimate of σ^2.

For our purpose it is unreasonable to discuss the bivariate relationships any further. Our example gives sense for specifying and analyzing multivariate relations. While the ultimative objective of my econometric research is a system of demand equations, solved as a system of simultaneous equations, first I prefer to restrict the analysis to a single equation including k exogenous variables, where k is in general a number larger than two. The specification of such a relationship is

$$y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \ldots + \beta_k X_{kt} + u_t \qquad (3.1)$$

with the same assumptions like in the bivariate model.

Here it makes sense to mention that Translog cost function is an example of a multivariate relationship and the system of demand equations derived from the cost function gives a typical example of a simultaneous equation system.

3.1.1 Ordinary and Generalized Least Squares

Ordinary least squares (OLS) is one of the basic methods of econometric investigations. Its principle is to find in equation (3.1) vector $\beta' = (\beta_1, \beta_2, \ldots, \beta_k)$ which minimizes the inner product of the disturbance vector u. This approach will be explained as follows.

Let's have $k+1$ variables with n sample observations given: $X_{1t}, X_{2t}, \ldots, X_{kt}, y_t$, $t = 1, \ldots, n$. The aim is to explain y as linear function of $x_1, x_2, \ldots, x_k$, where $x_j' = (X_{j1}, X_{j2}, \ldots, X_{jn})$ and $j = 1, 2, \ldots, k$. Note that $X_{jt} \neq x_{jt}$. It means y is in this case an endogenous variable, explainable by the model and $x_1, x_2, \ldots, x_k$ are independent, or exogenous variables. We are looking for unknown parameters $\beta_1, \beta_2, \ldots, \beta_k$ in the equation system:

$$y_t = \beta_1 + \beta_2 X_{2t} + \ldots + \beta_k X_{kt}, \quad t = 1, \ldots, n \tag{3.2}$$

Rewriting in matrix form will simplify the notation (3.2)

$$y = X\beta \tag{3.3}$$

$$\text{where} \quad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad X = \begin{pmatrix} 1 & X_{21} & \ldots & X_{k1} \\ 1 & X_{22} & \ldots & X_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{2n} & \ldots & X_{kn} \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}$$

y_t and X_{jt}, $(j = 1, 2, \ldots, k)$ represent observations in time t. But this observations do not fulfil (3.3) exactly, there arise some deviation. We indicate them with u_t. Notation in matrix form looks as follows

$$y = X\beta + u \tag{3.4}$$

where $u' = (u_1, u_2, \ldots, u_n)$ is so called vector of random disturbances.

As I have said in the begining of this section, the principle of OLS is to find such vector β that minimizes the square of sum

$$u'u = f(\beta) = (y - X\beta)'(y - X\beta) \tag{3.5}$$

Therefore the title **least squares**. Proceeding in this way: $f(\beta) = y'y - 2y'X\beta + \beta'XX\beta$, derivation with respect to β yields the necessary conditions for existence of the minimum.

$$\frac{df}{d\beta} = -2X'y + 2X'X\beta = 0$$

The solution is the least squares estimate (LSE) of β, denote by $\hat{\beta}$:

$$\hat{\beta} = (X'X)^{-1}X'y. \tag{3.6}$$

Looking at the sufficient condition

$$\frac{d^2 f}{d\beta^2} = 2X'X > 0,$$

$(X'X)$ of full rank and positive definite will fulfil such condition. The following theorem states the properties of $\hat{\beta}$ under additional statistical assumptions.

Gauss-Markoff theorem: Under assumptions

A1: $E(u) = 0$, expected value of u_t is 0
A2: $E(uu') = Cov(u) = \sigma^2 I$, u_t, u_t' for $t \neq t'$ are non-correlated, A1+A2 gives $u \sim \text{iid}(0, \sigma^2 I)$
A3: $E(X'u) = 0$, respectively X-non stochastic
A4: $r(X) = k$, X-full column rank
A5: $k < n$, there are more observations then variables. (We have positive degree of freedom.)

$\hat{\beta} = (X'X)^{-1}X'y$, the LSE of β has following properties:

(i) unbiased: $E(\hat{\beta}) = \beta$
(ii) efficient: $\text{Var}(\hat{\beta}) = \sigma^2(X'X)^{-1}$ has minimum variance in the class of linear unbiased estimators, resp. it is a best linear unbiased estimator (BLUE)
(iii) consistent: $plim(\hat{\beta}) = \beta$

Proof:(ii) A linear estimator of β is $\tilde{\beta} = \sum c_i X_i$ where c_i are to be determined. Unbiasedness requires $E(\tilde{\beta}) = \beta$. Now

$$\tilde{\beta} = \sum c_i(\beta X_i + u_i) = \beta\left(\sum c_i X_i\right) + \sum c_i u_i$$

Thus $\tilde{\beta}^*$ will be a linear unbiased estimator if and only if $\sum c_i = 0$ and $\sum c_i X_i = 1$ When these conditions are satisfied $\tilde{\beta} = \beta + \sum c_i u_i$ and

$$Var(\tilde{\beta}) = E\left[\left(\sum c_i u_i\right)^2\right] = \sigma^2 \sum c_i^2$$

To compare this variance with that of the OLS $\hat{\beta}$, write

$$c_i = w_i + (c_i - w_i)$$

Thus, $\sum c_i^2 = \sum w_i^2 + \sum (c_i - w_i)^2 + 2\sum w_i(c_i - w_i)$ The properties of the w_i and the conditions on the c_i ensure that

$$\sum w_i(c_i - w_i) = 0$$

and so $\text{Var}(\tilde{\beta}^*) = \text{Var}(\hat{\beta}) + \sigma^2 \sum (c_i - u_i)^2$ which proves the theorem. □

There is one more parameter to be estimated: σ^2. Let $e = y - X\hat{\beta}$ is so called residual vector. From it follows $e'e = (y - X\hat{\beta})'(y - X\hat{\beta})$. Expected value of $e'e$ is $E(e'e) = E(u'Mu) = E(\text{tr}Muu') = \text{tr}E(Muu') = \text{tr}M\sigma^2 I = \sigma^2(n-k)$ where
$M = (I - X(X'X)^{-1}X')$ is the idempotent $n \times n$ matrix. Therefore

$$s^2 = \frac{e'e}{n-k} \tag{3.7}$$

is unbiased estimate for variance matrix of estimated parameters. The LS equation is now

$$y_t = \hat{\beta}_1 + \hat{\beta}_2 X_{2t} + \ldots + \hat{\beta}_k X_{kt} + e_t \qquad t = 1, \ldots, n$$

averaging over the sample observations gives

$$\bar{y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2 + \ldots + \hat{\beta}_k \bar{X}_k$$

since $\bar{e} = 0$. Subctracting the second equation from the first gives so called deviation form of the observations.

$$\tilde{y}_t = \hat{\beta}_2 \tilde{X}_{2t} + \ldots + \hat{\beta}_k \tilde{X}_{kt} + e_t$$

where $\tilde{y}_t = y_t - \bar{y}$ and $\tilde{X}_{it} = X_{it} - \bar{X}_i$, $i = 1, \ldots, k$ denote deviations from sample means. Collecting all n observations, the deviation form of the equation may be written compactly using a transformation matrix

$$A = I_n - \left(\frac{1}{n}\right) ii' \tag{3.8}$$

where i is a column vector of n ones. It can be shown that A is symmetric, idempotent matrix, which by multiplication transform a vector of n observations into deviation form. Thus $Ae = e$ and $Ai = 0$

$$y = X\hat{\beta} + e = (i \quad \mathbf{X}_2) \begin{pmatrix} \hat{\beta}_1 \\ b_2 \end{pmatrix} + e$$

where $\mathbf{X}_2$ is the $n \times k - 1$ matrix of observations on the regressors and b_2 is the $k - 1$ element vector containing the coefficients $\hat{\beta}^2, \hat{\beta}^3, \ldots, \hat{\beta}^k$. Premultiplying by A gives

$$\begin{aligned} Ay = (0 \quad A\mathbf{X}_2) \begin{pmatrix} \hat{\beta}_1 \\ b_2 \end{pmatrix} + Ae &= (AX_2)b_2 + e \\ \Rightarrow \qquad y_* &= X_* b_2 + e \end{aligned} \tag{3.9}$$

where $y_* = Ay$ and $X_* = AX_2$. By using (3.9), the decomposition of the sum of squares may be expressed as

$$\begin{aligned} y_*' y_* &= b_2 X_*' X_* b_2 + e'e \\ \mathrm{TSS} &= \mathrm{ESS} + \mathrm{RSS} \end{aligned} \tag{3.10}$$

where TSS,ESS, resp. RSS idicate total, estimated, resp. residual sum of squares. The **coefficient of multiple correlation** R is defined

$$R^2 = \frac{\mathrm{ESS}}{\mathrm{TSS}} = 1 - \frac{\mathrm{RSS}}{\mathrm{TSS}} \tag{3.11}$$

R^2 measures the proportion of the total variation in Y explained by the linear combination of the regressors. R^2 value is used to measure the fit of the estimated model. For this purpose is often used also adjusted R^2, denoted by $\bar{R}^2$.

$$\bar{R}^2 = 1 - \frac{\mathrm{RSS}/(n-k)}{\mathrm{TSS}/(n-1)} \tag{3.12}$$

This statistic takes explicit into account the number of regressors used in the equation.

Many assumptions from A1-A5 are very strong ones and real data does not fulfil them. Therefore theorist search for some more general ways, which gives similar results like OLS. Very typical example is when u_t is not independently distributed. The LSE is still unbiased (i) but not efficient (ii) any more. With respect to this changes OLS is substituted by a so called general least squares (GLS) method.

GLS provide estimation of the linear model

$$y_t = \sum_{i=1}^{k} X_{it}\beta_i + u_t \qquad t = 1, \ldots, n$$

uses the following assumptions:

(i) X is nonstochastic matrix, nonsingular, $k < n$
(ii) u is random variable with zero mean and variance-covariance matrix $E[ee'] = \sigma^2\Omega \neq \sigma^2 I$, Ω is a known symmetric and positive definite matrix.

The linear unbiased, efficient estimate is obtained from corrected square of sum $f_\omega(\beta)$, see (3.5)

$$f_\Omega(\beta) = (y - X\beta)'\Omega^{-1}(y - X\beta) \tag{3.13}$$

differentiating (3.13) with respect to β yield necessary conditions. GLS estimate of β is then

$$\hat{\beta}_\Omega = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y \tag{3.14}$$

(compare with 3.6). The variance-covariance matrix in this case looks as follows:

$$E[(\beta_\Omega - \beta)(\beta_\Omega - \beta)'] = \sigma^2(X'\Omega^{-1}X)^{-1}$$

and an unbiased consistent estimate for σ^2 is:

$$\hat{\sigma}^2 = \frac{e\Omega^{-1}e'}{n-k}, \qquad e = y - X\beta_\Omega$$

In application of GLS arise a crucial problem how to determine Ω. Namely the task is to find estimator for Ω. One possibility arise from parametrisation of Ω with some finite numbers of parameters $\Omega = \Omega(\theta_1, \ldots, \theta_N)$ [6].

3.1.2 Maximum Likelihood Estimation

If the distribution of the disturbance vector is known e.g.

$$u \sim N(0, \sigma^2 I) \tag{3.15}$$

we can estimate the parameter with help of **likelihood function**. The formal definition of likelyhood function is

$$L(\theta, y) = f(y, \theta) \tag{3.16}$$

where θ is some k vector of unknown parameters, $\theta' = (\theta_1, \theta_2, \ldots, \theta_k)$ and $f(y, \theta)$ is the joint density, which indicates the dependence on θ and it may be interpreted as a function of θ, conditional on a set of sample outcomes. Reversing the order of the symbols in (3.16) emphasize the new focus of interest. Maximizing the likelihood function with respect to θ amounts to finding a value $\hat{\theta}$ that maximizes the probability of obtaining the sample values that have actually been observed. $\hat{\theta}$ is called maximum likelihood estimator (MLE) of the unknown parameter vector θ. Often used simplification is to take logarithms of likelihood function, it is denote with $l = lnL$, because of monotonic transformation, there is any problem in maximizing

$$\max_\theta(l) = \max_\theta(lnL) = \frac{\partial l}{\partial \theta} = \frac{1}{L}\frac{\partial L}{\partial \theta} = s(\theta, y) \tag{3.17}$$

The derivative of l with respect to θ is known as the score. The MLE, $\hat{\theta}$ is derived by setting the score to zero.

The widespread use of MLE is due to a couple of desirable properties [6]

(i) Consistency: $plim(\hat{\theta}) = \theta$

(ii) Asymptotic normality: $\hat{\theta} \sim^a N(\theta, I^{-1}(\theta))$. This states, that the asymptotic distribution of $\hat{\theta}$ is normal with mean θ and variance given by inverse of information matrix $I(\theta)$, which is defined by

$$I(\theta) = E\left[\left(\frac{\partial l}{\partial \theta}\right)\left(\frac{\partial l}{\partial \theta}\right)'\right] = -E\left[\frac{\partial^2 l}{\partial \theta \partial \theta'}\right] \tag{3.18}$$

where result is a square, symmetric matrix of second order derivatives or so called Hessian matrix.

(iii) Assyptotic efficiency: If $\hat{\theta}$ is the maximum likelihood estimator of a θ, the previous property means that

$$\sqrt{n}(\hat{\theta} - \theta) \rightarrow^d N(0, \sigma^2)$$

for some finite constant σ^2. If $\tilde{\theta}$ denotes any other consisitent, asymptotically normal estimator of θ, then $\sqrt{n}\tilde{\theta}$ has a normal limiting distribution whose variance is greater than or equal to σ^2. The ML estimate has minimum variance in the class of consistent, asymptotically normal estimators.

(iv) Invariance: If $\hat{\theta}$ is the ML estimate of θ and $g(\theta)$ is a continuous function of θ, then $g(\hat{\theta})$ is the ML estimate of $g(\theta)$.

(v) The score has zero mean and variance $I(\theta)$. To prove zero mean note following property

$$\int \cdots \int f(y_1, y_2, \ldots, y_n, \theta) dy_1 dy_2 \ldots dy_n = \int \cdots \int L dy = 1$$

Differentiating both sides with respect to θ yields

$$\int \cdots \int \frac{\partial L}{\partial \theta} dy = 0$$

$$\text{but} \qquad E(S) = \int \cdots \int \frac{\partial l}{\partial \theta} L dy = \int \cdots \int \frac{\partial L}{\partial \theta} dy$$

$$\Rightarrow Var(S) = E(SS') = E\left[\left(\frac{\partial l}{\partial \theta}\right)\left(\frac{\partial l}{\partial \theta}\right)'\right] = I(\theta)$$

In the case of the linear model the vector of unknown parameters θ is $\theta' = (\beta', \sigma^2)$ and the multivariate density for u is

$$f(u) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{(-\frac{1}{2\sigma^2})(u'u)}$$

the multivariate density for y conditional on X is then

$$f(y \mid X) = f(u)\left|\frac{\partial u}{\partial y}\right|$$

in this case $\left|\frac{\partial u}{\partial y}\right| = 1$. Thus the log-likelihood function is

$$l = lnf(y \mid X) = lnf(u) = -\frac{n}{2}ln2\pi - \frac{n}{2}ln\sigma^2 - \frac{1}{2\sigma^2}u'u =$$
$$-\frac{n}{2}ln2\pi - \frac{n}{2}ln\sigma^2 - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta) \tag{3.19}$$

from (3.17) we have

$$\frac{\partial l}{\partial \beta} = \frac{\partial l}{\partial \sigma^2} = 0$$

it determines ML estimate of β

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1}X'y & (3.20) \\ \text{and} \qquad \hat{\sigma}^2 &= (y - X\hat{\beta})'(y - X\hat{\beta})/n & (3.21) \end{aligned}$$

where X, β and y are as in previous section matrix and vectors.

ML estimator of β is seemed to be the OLS estimator $\hat{\beta}$ (see 3.4) and $\hat{\sigma}^2$ is $e'e/n$ where $e = y - X\hat{\beta}$, where e is the vector of OLS residuals. From OLS theory we know that $E[e'e/(n-k)] = \hat{\sigma}^2$. Thus $E(\hat{\sigma}^2) = \sigma^2(n-k)/n$, so that $\hat{\sigma}^2$ is biased for σ^2, while $\hat{\beta}$ is unbiased for β. By computing the second-order derivatives with respect to parameters we obtain information matrix (see 3.18)

$$I(\theta) = I\begin{pmatrix} \beta \\ \sigma^2 \end{pmatrix} = \begin{bmatrix} \frac{1}{\sigma^2}(X'X) & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix} \tag{3.22}$$

and its inverse is

$$I^{-1}\begin{pmatrix} \beta \\ \sigma^2 \end{pmatrix} = \begin{bmatrix} \sigma^2(X'X)^{-1} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{bmatrix}$$

The zero off-diagonal terms indicate that $\hat{\beta}$ and $\hat{\sigma}^2$ are distributed independently of one another. Substituting the ML estimator values (3.15) and (3.21) in the log-likelihood function and exponentiating gives the maximum value of likelihood function

$$L(\hat{\beta}, \hat{\sigma}^2) = \left(\frac{2\pi e}{n}\right)^{-\frac{n}{2}} (e'e)^{-\frac{n}{2}} \tag{3.23}$$

3.1.3 Seemingly Unrelated Regression

Many econometric applications involve the question of the solution the system of equations, which are somehow related to each other. My investigation presented below utilizes a system of the demand equations and thus I need additional techniques for single equation estimators. Very popular approach is called seemingly unrelated regression (SUR), respectively Zellner estimator after its inventor. Simply said it is an extension of GLS estimator on a multi-equation system.

Suppose that ith equation in a set of m equations is

$$y_i = X_i\beta_i + u_i \qquad i = 1, \ldots, m \tag{3.24}$$

the set of equations can be written also in matrix form

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} X_1 & 0 & \ldots & 0 \\ 0 & X_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & X_m \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{pmatrix} \tag{3.25}$$

where y_i is an $n \times 1$ vector of observations on the ith endogenous variable, X_i an $n \times k_i$ matrix of observations of exogenous variables, β_i a $k_i \times 1$ vector of coefficients and u_i

an $n \times 1$ vector of disturbances. Note, that in the previous sections β_i was the unique parameter, here it is held as vector of parameters.

In our case y variables are cost shares of individual energy inputs and X's are prices for energy goods and value of output. The main question is whether the equations should be treated separately or as a set. The explanation of this problem lies in the assumptions of the Gauss-Markoff theorem (GMT) (see section 3.1.1). In fact they are not absolut and sufficient for system of equation, because correlation among disturbances of various equations are missing. Second assumption of the GMT looks $E(u_i u_i') = \sigma_{ii}^2 I$, $(i = 1, \ldots, m)$ but by considering the system is assumed that there exists some between-equation relations, that the equations are only **seemingly unrelated**. Summarizing this considerations we use the assumptions of the GMT

$$\begin{aligned} E(u_i) &= 0 \\ E(u_i u_j') &= \sigma_{ij}^2 I \\ E(X_i u_i) &= 0 \end{aligned} \tag{3.26}$$

First assumption is clear, second is consequence of the previous paragraph and the last says that the disturbance and explanatory variables in each equation are assumed to be uncorrelated.

By definition, the variance-covariance matrix for u is

$$\Sigma = E(uu') = E \begin{pmatrix} u_1 u_1' & u_1 u_2' & \ldots & u_1 u_m' \\ u_2 u_1' & u_2 u_2' & \ldots & u_2 u_m' \\ \vdots & \vdots & \ddots & \vdots \\ u_m u_1' & u_m u_2' & \ldots & u_m u_m' \end{pmatrix} \tag{3.27}$$

Each term in the principal diagonal of Σ is an $n \times n$ variance-covariance matrix. Thus $E(u_i u_i')$ is the variance-covariance matrix for the disturbances in the ith equation. Each off-diagonal term in Σ represent an $n \times n$ matrix whose elements are covariances between disturbances from a pair of equations. Substituting (3.26) into (3.27) gives:

$$\Sigma = \begin{pmatrix} \sigma_{11} I & \sigma_{12} I & \ldots & \sigma_{1m} I \\ \sigma_{21} I & \sigma_{22} I & \ldots & \sigma_{2m} I \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1} I & \sigma_{m2} I & \ldots & \sigma_{mm} I \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \ldots & \sigma_{1m} \\ \sigma_{21} & \sigma_{22} & \ldots & \sigma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1} & \sigma_{m2} & \ldots & \sigma_{mm} \end{pmatrix} \otimes I = \Sigma_c \otimes I \tag{3.28}$$

where I is the identity $n \times n$ matrix and $\otimes$ denotes Kronecker multiplication that is, each element in Σ_c is multiplied by I. The GLS estimator of β is then unbiased and looks: (compare with 3.14).

$$\hat{\beta}_{GLS} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y$$

where

$$\Sigma^{-1} = \Sigma_c^{-1} \otimes I = \begin{pmatrix} \sigma^{11} I & \sigma^{12} I & \ldots & \sigma^{1m} I \\ \sigma^{21} I & \sigma^{22} I & \ldots & \sigma^{2m} I \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^{m1} I & \sigma^{m2} I & \ldots & \sigma^{mm} I \end{pmatrix}$$

The variance-covariance matrix for the GLS estimator is

$$\mathrm{Var}(\hat{\beta}_{GLS}) = (X'\Sigma^{-1}X)^{-1}$$

In general Σ is unknown and the question is how to construct an estimator for $\hat{\Sigma}$ while estimate for β stay consistent. One possibility is to use the estimate: $\hat{\Sigma} = \hat{\Sigma}_c \otimes I$, respectively provide an inverse. The Σ depends only on $m(m+1)/2$ different elements of Σ_c, however Σ is $mn \times mn$ matrix, it is possible by estimating each of the m equation by OLS and using residuals to estimate σ_{ij}. The residuals computed from OLS-single equation are:

$$\tilde{u}_i = y_i - X_i\tilde{\beta}_i = (I - X_i(X_i'X_i)^{-1}X_i')u_i \qquad i = 1, \dots, m$$

Substituting into $\frac{\tilde{u}_i'\tilde{u}_j}{n}$ we have

$$\frac{u_i'u_j}{n} - \frac{u_i'X_j}{n}\frac{(X_j'X_j)^{-1}}{n^{-1}}\frac{X_j'u_j}{n} - \frac{u_i'X_i}{n}\frac{(X_i'X_i)^{-1}}{n^{-1}}\frac{X_i'u_j}{n} + \frac{u_i'X_i}{n}\frac{(X_i'X_i)^{-1}}{n^{-1}}\frac{X_i'X_j}{n}\frac{(X_j'X_j)^{-1}}{n^{-1}}\frac{X_j'u_j}{n}$$

Denote $s_{ij} = \tilde{u}_i'\tilde{u}_j/n$ then (s_{ij}) indicate matrix S and $\hat{\Sigma} = S \otimes I$ is a consistent estimator for Σ. It can be shown that SUR estimate of β

$$\hat{\beta} = (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}y$$

as OLS estimate of β, $\tilde{\beta} = (X'X)^{-1}X'y$ is efficient too.

It is noteworthy, that if $\sigma_{ij} = 0, \quad i \neq j$, the SUR estimator (SURE) reduces to the application of OLS to each equation separately. If the disturbances are also normally distributed the OLS estimate is also ML estimate.

3.1.4 Estimation under Restrictions

Economic theory offers many attractive results, however for their consideration we often have to resort to restrictive conditions. It brings to econometric models a new dimension. In my model estimation methods, without the possibility of including restriction are completely useless, because a well behaved cost function has to be homogenous of degree one in input prices and the system of demand equations has to fulfil some kind of symmetry restrictions. Reformulating the conditions in mathematical notation can provide an extension of the methods mentioned so far, which can be covered by **estimation under restriction**.

From econometric point of view restrictions mean that there exists some prior information about parameters. Therefore by minimizing (3.5), restrictions have to be taken into the consideration. Besides further assumptions, GLS estimator is valid. Let us assume m additional restrictions for β:

$$\sum_{j=1}^{k} R_{ij}\beta_j = r_i \qquad i = 1, \dots, m \tag{3.29}$$

or $R\beta = r$ in matrix form, where R is a nonstochastic $m \times k$ matrix and r is $m \times 1$ vector. The idea of this aproach is to derive a GLS restricted estimate from the unrestricted one [6]. A GLS estimator under restrictions (3.29) is derived with help of the Lagrange function, where restriction are implemented as

$$f_R(\beta) = f(\beta) - (\beta'R' - r')\lambda = (y - X\beta)'\Omega^{-1}(y - X\beta) - (\beta'R' - r')\lambda \tag{3.30}$$

where R is the restriction matrix and λ is vector of the Lagrange multiplicator associated with restrictions. Extract β from (3.30), is obtain β_R. Proceeding in this way gives

$$\beta_R = \hat{\beta} + (X'\Omega^{-1}X)^{-1}R'[R(X'\Omega^{-1}X)^{-1}R']^{-1}(r - R\hat{\beta}) \tag{3.31}$$

where $\hat{\beta}$ is unrestricted GLS estimate of β.

β_R always fulfils the restrictions (3.29) also when the unrestricted parameter vector does not. When the latter fits to the restriction equation (3.29) then $E(R\hat{\beta}) = r$. It means that GLS estimate $\hat{\beta}$ is compatibile with the restriction.

3.2 Statistical Inference

Statistical inference on the validity of parameter restriction can be undertaken in a number of alternative ways. The first part of this section introduce you with the most common ones. The second one analyses the problem of autocorrelated disturbances and simultaneously explains the test statistic specific to this issue.

3.2.1 Test of Linear Restrictions

In section (3.1.1) we have established the properties of the LS estimator of β. It remains to show how to use this estimator to test various hypotheses about β. Suppose that in the model (3.1) are assumed some more informations about β and we test the reliability of this assumptions. Most common hypotheses are

(i) $H_0 : \beta_i = 0$. This sets up the hypothesis that the regressor X_i has no influence on y. This type of test is known as the significance test.
(ii) $H_0 : \beta_i = \beta_{i0}$. Here β_{i0} is some specified value. If, for instance, β_i denotes a price elasticity one might wish to test that the elasticity is -1.
(iii) $H_0 : \beta' = (\beta_2, \beta_3, \dots, \beta_k) = (0, \dots, 0)$. This sets up the hypotheses that the complete set of regressors has no effect on y. It tests the significance of regression as a whole. The constant parameter does not enter into the hypothesis, since interest centers of the variation of y around its mean and the level of the series is usually of no specific relevance.

All examples fit into the general linear framework

$$R\beta = r$$

Where R is a $q \times k$ matrix of known constants, with $q < k$, and r is a q-vector of known constants. Each null hypothesis determine the relevant elements in R and r. For the foregoing examples we have

(i) $R = (0 \quad \dots \quad 0 \quad 1 \quad 0 \quad \dots \quad 0) \qquad r = 0 \qquad q = 1$
with 1 in the ith position
(ii) $R = (0 \quad \dots \quad 0 \quad 1 \quad 0 \quad \dots \quad 0) \qquad r = \beta_{i0} \qquad q = 1$
with the 1 in ith position
(iii) $R = (\mathbf{0} \quad \mathbf{I}_{k-1}) \qquad r = \mathbf{0} \qquad q = k - 1$
where $\mathbf{0}$ is a vector of $k - 1$ zeroes

The efficient way to proceed is to derive a testing procedure for the general linear hypothesis

$$H_0 : R\beta - r = 0 \tag{3.32}$$

This general test is applicable to any hypothetic specification. Given the LS estimator $\hat{\beta}$ (see 3.6) an obvious step is to compute the vector $R\hat{\beta} - r$. This vector measures the discrepancy between expectation and observation. If this vector is relative large it tends to forget the existence of the null hypothesis and conversely, if it is relative small it tends not to contradict the H_0. To distinguish between small and large relevant sampling distribution help. In this case it is the distribution of $R\hat{\beta}$ when $R\beta = r$. From $E(\hat{\beta}) = \beta$ (see Gauss-Markoff Theorem, section 3.1.1) it follows directly

$$E(R\hat{\beta}) = R\beta.$$

Therefore from $\mathrm{Var}(\hat{\beta}) = \sigma^2(X'X)^{-1}$ we have

$$Var(R\hat{\beta}) = RVar(\hat{\beta})R' = \sigma^2 R(X'X)^{-1}R'$$

We thus know the mean and variance of the $R\hat{\beta}$ vector. Since $\hat{\beta}$ is a function of the vector u see (3.5) the distribution of $R\hat{\beta}$ will be determined by the distribution of u. First two assumption of the GMT (see section 3.1.1) plus assumption that the u_is are normally distributed can be combined in the (3.15) statement. Since linear combinations of normal variables are also normally distributed, it follows directly that

$$\hat{\beta} \sim N[\beta, \sigma^2(X'X)^{-1}]$$

$$\text{then} \qquad R\beta \sim N[R\beta, \sigma^2 R(X'X)^{-1}R']$$

$$\text{and so} \qquad R(\hat{\beta} - \beta) \sim N[0, \sigma^2 R(X'X)^{-1}R'] \tag{3.33}$$

If the null hypothesis $R\beta = r$ is true then

$$R\hat{\beta} - r \sim N[0, \sigma^2 R(X'X)_{-1}R']$$

This relation gives us the distribution of $R\hat{\beta}$.

Suppose now that $X \sim N(0, \sigma^2 I)$, where $X = (X_1, X_2, \ldots, X_k)$ and each X_i is still independent and has zero means. Thus

$$\frac{X_1^2}{\sigma^2} + \frac{X_2^2}{\sigma^2} + \ldots + \frac{X_k^2}{\sigma^2} \sim \chi^2(k) \tag{3.34}$$

has χ^2-distribution with k degrees of freedom. (3.34) can be written in matrix form

$$\frac{1}{\sigma^2}X'X \sim \chi^2(k) \tag{3.35}$$

rewriting in quadratic form gives $X'(\sigma^2 I)^{-1}X$. This allows us to write

$$(R\hat{\beta} - r)'[\sigma^2 R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r) \sim \chi^2(q) \tag{3.36}$$

it is easy to show from (3.35) that

$$\frac{e'e}{\sigma^2} \sim \chi^2(n-k) \tag{3.37}$$

(3.36) and (3.37) may be combined to form a computable statistic, which has an F distribution under the null hypothesis

$$\frac{(R\hat{\beta} - r)'[R(X'X)^{-1}R'](R\hat{\beta} - r)/q}{e'e/(n-k)} \sim F(q, n-k) \tag{3.38}$$

The test procedure is then to reject the hypothesis $R\beta = r$ if the computed F value exceed a preselected critical value. Now it will be shown what this test procedure amounts to the three specific applications indicated previously.

First, we rewrite (3.36) as

$$(R\hat{\beta} - r)'[s^2 R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r)/q \sim F(q, n-k) \tag{3.39}$$

because $s^2 = e'e/(n-k)$, see (3.7). Thus, $s^2(X'X)^{-1}$ is the estimated variance-covariance matrix $\hat{\beta}$, indicating i, jth element of the $(X'X)^{-1}$ by c_{ij} then

$$s^2 c_{ii} = \mathrm{Var}(\hat{\beta}_i) \qquad \text{and} \qquad s^2 c_{ij} = cov(\hat{\beta}_i, \hat{\beta}_j) \qquad i, j = 1, 2, \ldots, k$$

In each applications are specific forms of R, which are then substituted in (3.38) and (3.39)

(i) $H_0 : \beta_i = 0$ Equation (3.39) looks

$$F = \frac{\hat{\beta}_i^2}{s^2 c_{ii}} = \frac{\hat{\beta}_i^2}{\mathrm{Var}(\hat{\beta}_i)} \sim F(1, n-k)$$

because $R\hat{\beta}$ picks out $\hat{\beta}_i$ and $R(X'X)^{-1}R'$ picks out c_{ii}. Taking the square root of previous one

$$t = \frac{\hat{\beta}_i}{s\sqrt{c_{ii}}} = \frac{\hat{\beta}_i}{s.e(\hat{\beta}_i)} \sim t(n-k) \tag{3.40}$$

where $s.e$ is standard error, $s.e = \sqrt{\mathrm{Var}(\hat{\beta}_i)}$. Thus the null hypothesis that X_i has no assotiation with y is tested by dividing the ith estimated coefficient by its estimated standard error and referring the ratio to the t distribution.

(ii) $H_0 : \beta_i = \beta_{i0}$, t-distribution in this case looks

$$t = \frac{\hat{\beta}_i - \beta_{i0}}{s.e(\hat{\beta}_i)} \sim t(n-k)$$

Instead of testing specific hypothesis about β_i one may compute, say, a 95% confidence interval for β_i. It is given by

$$s_i \pm t_{0,025} \quad \mathrm{s.e}(\hat{\beta}_i)$$

(iii) $H_0 : \beta_2 = \beta_3 = \ldots = \beta_k = 0$ The first two examples have each involved just a single hypothesis, therefore F and t procedures were equivalent in this cases. Now $R(X'X)^{-1}R'$ picks out the square submatrix of order $k-n$ in the bottom right-hand corner of $(X'X)^{-1}$. To evaluate this submatrix, we divide X as $(\imath \quad X_2)$ where X_2 is the matrix of observations on all $k-1$ regressors. Then

$$X'X = \begin{pmatrix} \imath' \\ X_2' \end{pmatrix} (\imath \quad X_2) = \begin{pmatrix} n & \imath' X_2 \\ X_2'\imath & X_2' X_2 \end{pmatrix}$$

inverse of such matrix

$$(X_2'X_2 - X_2'\imath n^{-1}\imath' X_2)^{-1} = (X_2' A X_2)^{-1} = (X_*' X_*)^{-1}$$

where A is transformation matrix (3.8), A transforms observations into deviation form (see 3.9). With help of the (3.10) the F statistic for testing the complete set of regressors is

$$F = \frac{\mathrm{ESS}/(k-1)}{\mathrm{RSS}/(n-k)} \sim F(k-1, n-k)$$

By using (3.11), this statistic may be expressed as

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \sim F(k-1, n-k) \tag{3.41}$$

3.2.2 Likelihood Ratio, Wald and Lagrange Multiplier Test Statistics

In this part three basic tests for validity of the linear restriction will be assumed. Every test has common context of linear hypotheses about β. Null hypothesis of linear relation of β take the form (3.32).

Likelihood ratio test statistics is derived from ML estimator. The resultant value of likelihood function $L(\hat{\beta}, \hat{\sigma}^2)$, see section 3.1.2, is the unrestricted maximum likelihood and is expressible as a function of the unrestricted residual sum of squares $e'e$ (see 3.23). The model may also be estimated in restricted form by maximizing the likelihood subject to the restriction (3.32). Let the resultant estimators be denoted by $\tilde{\beta}$ and $\tilde{\sigma}^2$. Then maximum likelihood estimator is $L(\tilde{\beta}, \tilde{\sigma}^2)$. The restricted maximum cannot exceed the unrestricted maximum, but if the restrictions are valid, we would expect the restricted maximum to be relative close to the unrestricted maximum. According to this consideration likelihood ratio is defined as

$$\lambda = \frac{L(\tilde{\beta}, \tilde{\sigma}^2)}{L(\hat{\beta}, \hat{\sigma}^2)} \tag{3.42}$$

and intuitively we expect to reject the null hypothesis of binding restriction if λ is relative small. A large-sample test of general applicability is available for (3.42) in this form

$$LR = -2\ln\lambda = 2[\ln L(\hat{\beta}, \hat{\sigma}^2) - \ln L(\tilde{\beta}, \tilde{\sigma}^2)] \sim^a \chi^2(q)$$

The restricted ML is derived by maximizing

$$l^R = l - \mu'(R\beta - r)$$

where μ is an $q \times 1$ vector of Lagrange multipliers and $l = \ln L$. It can be shown that $\tilde{\beta}$ is simply the restricted β already derived in (3.31). If we denote the corresponding residuals by

$$e_R = y - X\beta_R$$

the restricted ML estimator of σ^2 is $\tilde{\sigma}^2 = e_R' e_R / n$ and so

$$L(\tilde{\beta}, \tilde{\sigma}^2) = const(e_R' e_R)^{-n/2} \tag{3.43}$$

Substituting (3.43) into (3.42) gives LR test statistic as

$$LR = n(\ln e_R' e_R - \ln e'e)$$

The calculation of the LR statistics thus requires both models, restricted and unrestricted.

The next test statistic requires fitting only one of the restricted model. By **Wald** procedure the vector $(R\hat{\beta} - r)$ indicates the extent to which the unrestricted ML estimate fits the null hypothesis. From asymptotic normality follows $\hat{\beta} \sim^a N(\beta, I^{-1}(\beta))$ (see section 3.1.2). Therefore for hypothesis (3.32), $R(\hat{\beta} - \beta)$ is asymptotically distributed as multivariate normal with zero mean vector and variance-covariance matrix $RI^{-1}(\beta)R'$, where $I^{-1}(\beta) = \sigma^2(X'X)^{-1}$, compare with (3.33). As shown in (3.22) the information matrix $I(\theta)$ for the linear regression model is block diagonal, so we can concentrate on the submatrix relating to β. It gives

$$(R\beta - r)'[RI^{-1}(\beta)R']^{-1}(R\hat{\beta} - r) \sim^a \chi^2(q)$$

The assymptotic distribution still holds when the unknown σ^2 in $I^{-1}(\beta)$ is replaced by consistent estimator $\hat{\sigma}^2 = e'e/n$. It gives

$$W = \frac{(R\beta - r)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r)}{\hat{\sigma}^2} \sim^a \chi^2(q)$$

where W indicate the Wald statistics.

The last test from this trinity is based on the score vector (3.17). The unrestricted estimator, $\hat{\theta}$ is found by solving $s(\hat{\theta}) = 0$. When the score vector is evaluated at $\tilde{\theta}$-the restricted estimator, it will in general not be zero. However, if the restrictions are valid $l(\tilde{\theta})$ will be close to the $l(\hat{\theta})$. As shown earlier, the score vector has zero mean and variance-covariance matrix given by $I(\theta)$ (see section 3.1.2). The quadratic form $s'(\hat{\theta})I^{-1}(\hat{\theta})s(\hat{\theta})$, will then have a χ^2-distribution. Evaluating this form for $\theta = \tilde{\theta}$ provides a test of the null hypothesis. Therefore under the null hypothesis

$$LM = s'(\tilde{\theta})I^{-1}(\tilde{\theta})s(\tilde{\theta}) \sim^a \chi^2(q)$$

where LM is the **Lagrange multiplier** test statistic. In comparison with the Wald test, it is necessary to calculate only the restricted estimator rather then the unrestricted one, which is much easier in many cases.

There also exists some comparison of this three tests. It looks as

$$W \geq LR \geq LM.$$

It is possible to derive this inequality with help of some analytical tools, see [6]. The tests are assymptotically equivalent, but in finite samples give different numerical results.

3.2.3 Durbin-Watson test

Suppose that in the model $y = X\beta + u$ (see 3.4) it is assumed that disturbance terms fulfil following pattern

$$u_t = \varphi u_{t-1} + \varepsilon_t \tag{3.44}$$

where ε_t is white noise process: $\varepsilon_t \sim N(0, \sigma_t^2 I)$, this process is known as first-order autoregressive proces, (AR(1)). Simply speaking it means that disturbance in time point t depends on previous disturbance term. The task is to provide some test for the hypothesis about relation (3.44). The null hypothesis of zero autocorrelation is then

$$H_0 : \varphi = 0$$

against the alternative hypothesis

$$H_1 : \varphi \neq 0$$

One of the many who have investigated this issue are Durbin and Watson [6].

The **Durbin-Watson** (DW) test statistic is computed from the vector of OLS residuals $e = y - X\hat{\beta}$. It is denoted like d or DW and is defined as:

$$d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} \tag{3.45}$$

The mean residual is zero, so the residuals will be scattered around the zero line which represents $E(e) = 0$. If the e's are positively autocorrelated, succesive values will tend to be close to each other, runs above and below the expected values and the first differences will tend to be numerically smaller than the residuals themselves. Alternatively, if the e's have a first-order negative autocorrelation, there is a tendency for the next observation to be on opposite side of $E(e) = 0$ axis, therefore first differences tend to be numerically larger than the residuals. Thus d will tend to be relative small for positive autocorrelated e's and for negative is then relative large. If the e's are really random, or non correlated, there is no tendency for runs above and below or for alternate jumps across horizontal axis and d will have an intermediate value. Expanding (3.45) we have

$$d = \frac{\sum_{t=2}^{n} e_t^2 + \sum_{t=2}^{n} e_{t-1}^2 - 2\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2}$$

For large n the different ranges of summation in numerator and denominator have a negligible effect and

$$d \simeq 2(1 - \hat{\varphi}) \tag{3.46}$$

where

$$\hat{\varphi} = \frac{\sum e_t e_{t-1}}{\sum e_t^2}$$

is the coefficient in the OLS regression of e_t on e_{t-1}. Thus (3.46) gives various states of d:

$0 < d < 2$ for positive autocorrelation of the e's
$4 > d > 2$ for negative autocorrelation of the e's
$d \simeq 2$ for zero autocorrelation of the e's

However the hypothesis under test is about the properties of the unobservable u's. For a random u series the expected value of d is

$$E(d) = 2 + \frac{2(k-1)}{n-k} \tag{3.47}$$

where k is the number of coefficients in the regression. From (3.47) is clear, that any computed d is associated with the matrix X and therefore particular d's are not tabulated. Durbin-Watson established upper (d_U) and lower (d_L) frontiers for the critical values. These frontiers depend only on the sample size n and the number of regressors k.

d_U and d_L are used to test the hypothesis of zero autocorrelation against the alternative of positive first-order autocorrelation. The testing recipe is

(i) If $d < d_L$, reject the hypothesis of nonautocorrelated u in favor of the hypothesis of positive first-order autocorrelation.
(ii) If $d > d_U$, do not reject the null hypothesis.
(iii) If $d_L < d < d_U$, the test is inconclusive.

If one wishes to test the null hypothesis against the alternative of negative first-order autocorrelation, this test provides a $4 - d$ value. There are two important qualifications to the use of the DW test. First, it is necessary to compute with a constant term. And second, it is strictly valid for a nonstochastic X matrix. Thus, DW test is not useful for the model where lagged dependent variables are employed as regressors [6].

3.3 Problem of Vector Autocorrelation

Now we will discuss an extension of the model (3.25). This part is of my special interest, because estimation with autocorrelated disturbances brings many practical complications and I don't know if there exist in each case a satisfactory solution.

Consider the multivariate regression model (3.25). A comprehensive expression is given in simple matrix form

$$y_t = Bx_t + u_t \qquad t = 2, \dots, n \tag{3.48}$$

where y_t is an $m \times 1$ vector of dependent variables, B is an $m \times k$ matrix of unknown parameters, x_t is a $k \times 1$ vector of exogenous variables u_t is an $m \times 1$ vector of random disturbances. We assume that $(u_2, \dots, u_n)$ is a sample from a stationary vector stochastic process which satisfies the stochastic difference equation

$$u_t = Qu_{t-1} + \varepsilon_t \qquad t = 2, \dots, n \tag{3.49}$$

where $\varepsilon_t \sim \text{iid}\, N[0, \Omega]$ and $Q = (Q_{ij})$ is an $m \times m$ matrix of unknown parameters. It is noteworthy that the first observation is lost owing to the presence of lagged variables u_{t-1}. We assume here the adding up condition (because this still be the relevant case in the application further below).

$$\imath' y_t = 1 \qquad t = 1, \dots, n \tag{3.50}$$

where $\imath$ is an $n \times 1$ vector of ones. From (3.48),(3.50)

$$\begin{aligned} \Rightarrow \quad & \imath' B = (1 \quad 0 \quad \dots \quad 0) \\ \textit{and} \quad & \imath' u_t = 0 \qquad t = 1, \dots, n \end{aligned} \tag{3.51}$$

Since u_{t-1} and ε_t are statistically independent then from (3.49) and (3.51) follows that

$$\begin{aligned} \imath' Q &= a' \\ \textit{and} \quad \imath' \varepsilon_t &= 0 \qquad t = 1, \dots, n, \end{aligned} \tag{3.52}$$

where a is an arbitrary constant.

Hence in the context of an autoregressive model the adding up condition (3.50) implies that each column of Q must sum to the same unknown constant a and that $\Omega = 0$, which means that Ω is singular. This restriction is a strong one, because in the case of diagonal matrix it imposes all diagonal elements to be equal.

Since $\varepsilon_t \sim \text{iid}\, N(0, \Omega)$ we consider the ML estimation of the model defined by specifications (3.48),(3.49),(3.50). Since the covariance matrix Ω is singular ε_t cannot have a density. We assume that Ω has only one zero root so that when one component of ε_t is deleted the resulting vector has a nonsingular distribution. Let us denote by ε_t^m the vector ε_t with the last element deleted. The density of this vector can be written as

$$f(\varepsilon_t^m) = 2\pi^{-\frac{1}{2}(m-1)} |\Omega_m|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \varepsilon_t^{m\prime} \Omega_m^{-1} \varepsilon_t^m \right\}$$

where Ω_m is the covariance matrix with the last row and column deleted.
$\varepsilon_2^m, \dots, \varepsilon_n^m \sim N[0, \Omega_m]$ Therefore the likelihood function is:

$$L = (2\pi)^{-\frac{1}{2}(n-1)(m-1)} |\Omega_m|^{-(n-1)/2} \exp\left\{ -\frac{1}{2} \sum_{t=2}^{n} \varepsilon_t^{m\prime} \Omega_m^{-1} \varepsilon_t^m \right\}$$

Now consider ML estimation of a system of $m-1$ equations. Deleting the last equation from (3.48) and (3.49) gives:

$$\begin{aligned} y_t^m &= B_m x_t + u_t^m \qquad t = 2, \ldots, n \qquad (3.53) \\ u_t^m &= Q_m u_{t-1}^m + \varepsilon_t^m \quad t = 2, \ldots, n \qquad (3.54) \end{aligned}$$

Where y_t^m and u_t^m are the vectors y_t and u_t with the last element deleted and B_m and Q_m are the parameter matrices B, Q with the last row deleted. Since Q_m is not a square matrix (it has order $m-1\times$ m), the ML estimation procedure is not applicable to (3.53) and (3.54). However, this difficulty can easily be remedied. Since $\imath' u_t = 0$, we can rewrite the stochastic difference equation (3.49) as:

$$\begin{pmatrix} u_{1t} \\ \vdots \\ u_{mt} \end{pmatrix} = \begin{pmatrix} Q_{11} - Q_{1m} & \cdots & Q_{1,m-1} - Q_{1m} \\ \vdots & \ddots & \vdots \\ Q_{m1} - Q_{mm} & \cdots & Q_{m,m-1} - Q_{mm} \end{pmatrix} \begin{pmatrix} u_{1,t-1} \\ \vdots \\ u_{m-1,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{mt} \end{pmatrix}$$

or more compactly:

$$u_t = \bar{Q} u_{t-1}^m + \varepsilon_t \qquad t = 2, \ldots, n \qquad (3.55)$$

where $\bar{Q}_{ij} = Q_{ij} - Q_{im} \quad i = 1, \ldots, n, \quad j = 1, \ldots, n-1$. From (3.52) and (3.55) $\Rightarrow \bar{Q}_{1j} + \bar{Q}_{2j} + \ldots + \bar{Q}_{mj} = 0$ and the computable system is then

$$\begin{aligned} y_t^m &= B_m x_t + u_t^m \qquad t = 2, \ldots, n \\ u_t^m &= \bar{Q}_m u_{t-1}^m + \varepsilon_t^m \quad t = 2, \ldots, n. \end{aligned}$$

where $\bar{Q}_m$ is the $\bar{Q}$ with last row deleted can be estimated by ML procedure. Hence the parameter matrices B_n,$\bar{Q}_m$ and Ω_m have a unique ML estimate and using these estimates we can obtain ML estimates of the full parameter matrices B, $\bar{Q}$ and Ω. Two issues arise to discuss:

(i) Invariance: Is the ML estimate of the parameter matrices, B, $\bar{Q}$ and Ω the same regardless of which equation is deleted?
(ii) Identification: Can an ML estimate of Q be derived from that of $\bar{Q}$?

Invariance: For the case Q and B unrestricted Barten?? has shown that the ML estimate of B is invariant to the equation deleted. Barten's result also holds for the case where $Q = 0$ and B is restricted and more when B and Q are suitably restricted then the estimation result is also invariant.

Identification: As was said before the matrix $\bar{Q}_m$ always has a unique ML estimate. Hence if there exist a unique nonsingular linear transformation of $\bar{Q}_m$ into Q, then Q has a unique ML estimate, too. To derive Q given knowledge of $\bar{Q}_m$ we require prior information. For example it is very often that certain elements of Q are assumed to be zero. For a matrix 3×3 it can be as

$$\begin{pmatrix} Q_{11} & 0 & Q_{13} \\ 0 & Q_{22} & Q_{23} \\ Q_{31} & 0 & 0 \end{pmatrix}$$

We order the prior informations into linear restrictions associated with the elements of Q.

$$c = MQ^v \qquad (3.56)$$

where c is $J \times 1$ known vector, M is $J \times m^2$ matrix and Q^v is the $m^2 \times 1$ vector obtained by stacking the columns of Q. Vector c represent the knowledge of prior information. In addition to the prior restriction (3.56) the model recquires the restriction (3.52). This is expressed by $m-1$ linearly independent equations

$$\mathbf{0} = KQ^v \tag{3.57}$$

where $(\mathbf{0})$ is a $m-1 \times 1$ column of zeroes, K is known $(m-1) \times m^2$ matrix of rank $m-1$. Finaly, from (3.55) the elements of $\bar{Q}_m$ generate $m-1$ linearly independent equations.

$$Q^v_m = LQ^v \tag{3.58}$$

where $\bar{Q}^v_m$ is the vector obtained by stacking the columns of $\bar{Q}_m$ and L is known $(m-1)^2 \times m^2$ matrix of rank $(n-1)^2$. Putting (3.56), (3.57) and (3.58) together we see that the matrix Q is identified if and only if the system

$$d \begin{pmatrix} 0 \\ \bar{Q}^v_m \\ M \end{pmatrix} = \begin{pmatrix} K \\ L \\ M \end{pmatrix} Q^v = DQ^v$$

can be uniquely solved for Q^v. Thus the rank of D is m^2. Since the matrix

$$\begin{pmatrix} K \\ L \end{pmatrix}$$

has full row rank, the rank of M must be at least $m^2 - [(m-1) + (m-1)^2] = m$. Simply said the number of linearly independent prior restriction J has to be greater or equal to the number of equations in the full model -m. The matrix Q is said to be underdefined when $J < m$, just identified when $J = m$ and overidentified when $J > m$ [2].

Chapter 4

Applications

This chapter provides an application of the theory surveyed. I have divided it into three parts. In the first part I will solve the basic model, which means that all details will be explained in this chapter. Then, on the basis of various tests I will outline possible misspecification of the basic model. In the last part I will solve a corrected model, where the assumption of autocorrelated disturbances will be added. The aim of each model of producer behavior is the computation of elasticities and their discussion. Therefore, this will be done for both models and their intercomparison as well.

4.1 Translog Cost Function for the Paper Industry

In this section we will employ a nonhomothetic translog cost function (2.18) for modeling the data of the Austrian paper industry. The classic cost function expresses the relationship between prices of all inputs and total input costs. Nonhomothetic cost functions allow the influence of the level of output. Our model of the paper industry states the hypothesis that energy costs can be determined envisaged just by the prices of energy inputs and the aggregated level of output [5].

Let me start with a brief introduction about the data. The observations describe the development of the Austrian paper industry over the years 1972-1996. For our purposes we need the price developement of energy inputs mostly used in paper industry. They are: coal, oil, gas and electricity. On the other side we need energy inputs levels entering the production process. They reflect the answer of producers on changes in prices. Since we are involving the nonhomothetic cost function we also need the quantity of paper produced. To ilustrate the observations for one given year are

where Y indicates quantity of paper produced, measured in 1000t; X_c, X_o, X_g, X_e are

$Year$	Y	X_c	X_o	X_g	X_e	p_c	p_o	p_g	p_e
1990	2932	$4079,7$	$5203,26$	$16765,3$	$5039,52$	$154,0$	$144,0$	$159,0$	$240,0$

observations of the energy inputs, given in TJ (initials c, o, g, e indicate energy inputs as follow: coal, oil, gas, electricity) and p_c, p_o, p_g, p_e are prices for corresponding energy inputs, given in Austrian Shillings/Mwh. Note that observations of quantities are in aggregated form, i.e. the quantity of paper produced is amount of paper produced by all Austrian producers of paper. Related to the energy input quantities, aggregated means the sum of the particular energy input over all Austrian producers. Due to this issue we state the next hypothesis, that the cost function describes not only the behavior of the particular producer, but could be generalized to the whole industry branch. Simply said, the response of all producers on price development is more or less the same. The whole data set is available in the appendix, where energy inputs prices are recalculated to the Austrian Shillings/TJ as they are used in this thesis.

The nonhomothetic translog cost function of the paper industry corresponding with (2.18) is written as

$$\begin{aligned} \ln C(\mathbf{p}, y) &= \gamma_0 + \gamma_y \ln y + \sum_{i=1}^{4} \gamma_i \ln p_i + \frac{1}{2} \sum_{i=1}^{4} \sum_{j=1}^{4} \gamma_{ij} \ln p_i \ln p_j + \\ & \quad + \sum_{i=1}^{4} \gamma_{iy} \ln p_i \ln y + \frac{1}{2} \gamma_{yy} \ln y \ln y \end{aligned} \tag{4.1}$$

where symmetry restrictions are imposed ($\gamma_{ij} = \gamma_{ji}$), for all $i, j = 1, \ldots, 4$. For simplicity we state that numbers $1, 2, 3, 4$ indicate the four energy inputs c, o, g, e, in the same order. One could of course estimate the translog cost function (4.1) directly, but gains in efficiency can be realized by estimating the optimal, cost-minimizing input demand equations [1]. Employing Shephard's lemma (2.13), i.e. deriving the cost function (4.1) with respect to $\ln \mathbf{p} = (\ln p_c, \ln p_o, \ln p_g, \ln p_e)'$

$$\frac{\partial \ln C}{\partial \ln p_i} = \frac{p_i}{C} \frac{\partial C}{\partial p_i} = \frac{p_i X_i}{C} = S_i \quad i = 1, 2, 3, 4 = c, o, g, e$$

we derive the corresponding system of four input cost shares equations

$$\begin{aligned} S_c &= \gamma_c + \gamma_{cc} \ln p_c + \gamma_{co} \ln p_o + \gamma_{cg} \ln p_g + \gamma_{ce} \ln p_e + \gamma_{cy} \ln y \\ S_o &= \gamma_o + \gamma_{oc} \ln p_c + \gamma_{oo} \ln p_o + \gamma_{og} \ln p_g + \gamma_{oe} \ln p_e + \gamma_{oy} \ln y \\ S_g &= \gamma_g + \gamma_{gc} \ln p_c + \gamma_{go} \ln p_o + \gamma_{gg} \ln p_g + \gamma_{ge} \ln p_e + \gamma_{gy} \ln y \\ S_e &= \gamma_e + \gamma_{ec} \ln p_c + \gamma_{eo} \ln p_o + \gamma_{eg} \ln p_g + \gamma_{ee} \ln p_e + \gamma_{ey} \ln y. \end{aligned} \tag{4.2}$$

In language of econometric methods S's are endogenous variables, i.e explainable by the model and the logarithm of input prices and of output are exogenous variables. γ's are parameters to be estimated. Since the cost function is assumed to be well behaved, we have to impose the restrictions (2.20). Rewriting for our example, homogeneity of degree zero in prices and cost exhaustion restrictions (compare with section 2.4.2) are as

$$\begin{aligned} \gamma_c + \gamma_o + \gamma_g + \gamma_e &= 1 \\ \gamma_{cc} + \gamma_{co} + \gamma_{cg} + \gamma_{ce} &= 0 \\ \gamma_{oc} + \gamma_{oo} + \gamma_{og} + \gamma_{oe} &= 0 \\ \gamma_{gc} + \gamma_{go} + \gamma_{gg} + \gamma_{ge} &= 0 \\ \gamma_{ec} + \gamma_{eo} + \gamma_{eg} + \gamma_{ee} &= 0 \\ \gamma_{cy} + \gamma_{oy} + \gamma_{gy} + \gamma_{ey} &= 0. \end{aligned} \tag{4.3}$$

This model will be the main object of the following investigation. To provide an accomplished model means searching for the levels of parameters, optimal values of the statistical inference measures and measures of fit. There does not exist any prescription for obtaing such perfect model and you can never say that some model is the best one. Nonetheless, the rest of my thesis is dealing with this, say basic model and afterwards with a respecified model, or say, corrected model.

4.1.1 Methods

In this section the methods for estimating the basic model will be described. To implement the share equation system (4.2) empirically, it is necessary to specify a stochastic framework. We add to each equation a random disturbance term, to express theoretical error of regression. We indicate it by u_i where $(i = c, o, g, e)$ depending on equation. Here it is assumed that the random disturbance vector $\mathbf{u}' = (u_c, u_o, u_g, u_e)$ is multivariate normally distributed with mean vector zero and constant covariance matrix Σ (3.28). Rationale for stochastic specification could be that there exist some informations, which are known to producers, but are unobservable for econometricians [1].

Since the sum of the cost shares is equal to one, $\sum_{i=1}^{4} S_i = 1$, the system (4.2) is singular. Thus one equation could be expressed as linear combination of remaining ones. With help of the homogeneity restriction (4.3) and symmetry we can delete one arbitrary equation from the system (4.2) without loss of any parameter of the deleted equation. For instance, we consider to delete the electricity equation, thus parameters associated with electricity equation could be calculated like

$$\begin{aligned}
\gamma_e &= 1 - \gamma_c - \gamma_o - \gamma_g \\
\gamma_{ce} &= -\gamma_{cc} - \gamma_{co} - \gamma_{cg} \\
\gamma_{oe} &= -\gamma_{oc} - \gamma_{oo} - \gamma_{og} \\
\gamma_{ge} &= -\gamma_{gc} - \gamma_{go} - \gamma_{gg} \\
\gamma_{ey} &= -\gamma_{cy} - \gamma_{oy} - \gamma_{gy} \\
\gamma_{ee} &= \gamma_{cc} + \gamma_{oo} + \gamma_{gg} + 2(\gamma_{co} + \gamma_{cg} + \gamma_{og}),
\end{aligned}$$

and the model with one equation deleted is then

$$\begin{aligned}
S_c &= \gamma_c + \gamma_{cc}\ln(p_c/p_e) + \gamma_{co}\ln(p_o/p_e) + \gamma_{cg}\ln(p_g/p_e) + \gamma_{cy}\ln y \\
S_o &= \gamma_o + \gamma_{oc}\ln(p_c/p_e) + \gamma_{oo}\ln(p_o/p_e) + \gamma_{og}\ln(p_g/p_e) + \gamma_{oy}\ln y \\
S_g &= \gamma_g + \gamma_{gc}\ln(p_c/p_e) + \gamma_{go}\ln(p_o/p_e) + \gamma_{gg}\ln(p_g/p_e) + \gamma_{gy}\ln y
\end{aligned} \tag{4.4}$$

The **stacked system** (4.4) consists of the 15 parameters to be estimated but imposing the 3 symmetry restrictions $(\gamma_{ij} = \gamma_{ji})$ we reduce the number of unknown parameters to 12.

The discussion about singularity of the system (4.2) we have to extend into stochastic specification as well. For each observation the sum of the disturbances across equations must always equal to zero. This implies that the disturbance covariance matrix Σ is nondiagonal and singular [1]. To avoid this complication we will search for a methodology where singular Σ could be replaced with Σ_i. Subscript i means that original covariance matrix has ith row and ith column deleted. The indicator i is corresponding with deleted equation in stacked system (in this case - e).

Finally, there remains the question of possible estimation methods, which cover the above to the considerations. We will refer to the section 3.1.3, where seemingly unrelated regression is explained. First, we have to decide about the technique and define the system for software packages. One could apply the restrictions (symmetry and homogeneity ones) directly to the model. Then there exist two approaches usually applied to such restricted singular systems (4.2) and (4.3). They are in principle the same and under some specific conditions they yield the same results, and both allow for the possibility of cross correlation among the disturbances in different equations of the stacked system. The first is known as feasible GLS, which uses the seemingly unrelated regression technique. The identified structural equations are first estimated by two-stage least squares (2SLS), what is a single equation estimator, i.e each equation in system is estimated as single equation. The resultant residuals are then used to estimate the disturbance covariance matrix, which is then used to estimate all identified structural parameters jointly [6]. The second is the maximum likelihood (ML) technique, see section 3.1.2, which is suitable for direct constrained, singular system, such as the translog demand system. The ML estimator, when a constrained system is being considered, is known as full information maximum likelihood estimator (FIML). If the estimation process of 3SLS is iterated rather then stopped at the third stage, the estimates converge on the FIML estimates of the structural model. The first of these techniques is used by Limdep7.0 and Eviews as well. The second, FIML technique is available only in Limdep. However, the results are more or less the same, thus we will indicate it like ML/SURE.

The second possibility is, first to define our system nonrestricted, with restrictions added afterwards. For system (4.4) it gives following specification

$$\begin{aligned} S_c &= \gamma_c + \gamma_{cc}\ln p_c + \gamma_{co}\ln p_o + \gamma_{cg}\ln p_g + \gamma_{ce}\ln p_e + \gamma_{cy}\ln y \\ S_o &= \gamma_o + \gamma_{oc}\ln p_c + \gamma_{oo}\ln p_o + \gamma_{og}\ln p_g + \gamma_{oe}\ln p_e + \gamma_{oy}\ln y \\ S_g &= \gamma_g + \gamma_{gc}\ln p_c + \gamma_{go}\ln p_o + \gamma_{gg}\ln p_g + \gamma_{ge}\ln p_e + \gamma_{gy}\ln y \end{aligned} \tag{4.5}$$

where symmetry restriction imposed are $\gamma_{og} = \gamma_{go}$, $\gamma_{oe} = \gamma_{eo}$, $\gamma_{ge} = \gamma_{eg}$ and homogeneity of degree zero in input prices are

$$\begin{aligned} \gamma_{ce} &= -\gamma_{cc} - \gamma_{co} - \gamma_{cg} \\ \gamma_{oe} &= -\gamma_{oc} - \gamma_{oo} - \gamma_{og} \\ \gamma_{ge} &= -\gamma_{gc} - \gamma_{go} - \gamma_{gg} \end{aligned}$$

The methodology used is restricted GLS, which utilizes the seemingly unrelated regression too. We will indicate it RGLS/SURE. This procedure is computed using the restricted least squares formula (see 3.31), after the unrestricted estimates are obtained. Therefore, the RGLS estimator is a function of the unrestricted estimator, not an iterative estimator in its own right. Thus we do not regard it as ML estimator even it is allowed to iterate to convergence [8].

One interesting issue arises, whether the parameter estimates are invariant to the choice of which equation is deleted. If such invariance were lacking, it would be a trouble feature, since it is imposible to estimate without deleting. Fortunately, if ML/SURE or RGLS/SURE is implemented to the arbitrarily stacked system, all parameter estimates, log-likehood values and standard errors will be invariant to the choice of which 3 equations of the system (4.2) is being considered [1]. The proof of this statement is based first on the linearity of imposed restrictions and second on the linear dependency of the whole system (4.2). With respect to this I will present in the next section estimation results. Note, that the stacked system could be obtained by deleting an arbitrary equation, just rearranging depends on the equation deleted.

4.1.2 Results

This section offers an overview of the estimation results and discussion about them. Table 4.1 gives the estimated values of the parameters of model (4.4) estimated by iterative

SURE in Eviews-software, but as was noted before, gives the same results as ML/SURE. What I present is provided by Limdep 7.0-software.

Note, that the number of degrees of freedom is observations minus regressors (parameters

	Coefficient	Sdt.Error	t-Statistic	Prob.
$\hat{\gamma}_c$	0.129645	0.090399	1.434140	0.1565
$\hat{\gamma}_o$	1.732435	0.165644	10.45876	0.0000
$\hat{\gamma}_g$	0.187264	0.253560	0.138539	0.4629
$\hat{\gamma}_e$	−1.049222	0.274023	−3.828957	0.0003
$\hat{\gamma}_{cc}$	0.115533	0.029479	3.919155	0.0002
$\hat{\gamma}_{co}$	0.010204	0.017169	0.594315	0.5544
$\hat{\gamma}_{cg}$	−0.073470	0.014696	−4.999410	0.0000
$\hat{\gamma}_{ce}$	−0.052232	0.019708	−2.650242	0.0102
$\hat{\gamma}_{oo}$	0.182044	0.036426	4.997685	0.0000
$\hat{\gamma}_{og}$	−0.105669	0.028741	−3.676557	0.0005
$\hat{\gamma}_{oe}$	−0.086588	0.017584	−4.924207	0.0000
$\hat{\gamma}_{gg}$	0.202601	0.036703	5.0519981	0.0000
$\hat{\gamma}_{ge}$	−0.023497	0.033188	−0.707996	0.4816
$\hat{\gamma}_{ee}$	0.162270	0.039231	4.136263	0.0001
$\hat{\gamma}_{cy}$	0.004329	0.010239	0.422799	0.6739
$\hat{\gamma}_{oy}$	−0.181200	0.022003	−8.235187	0.0000
$\hat{\gamma}_{gy}$	0.029515	0.033701	0.875790	0.3845
$\hat{\gamma}_{ey}$	0.147351	0.036599	4.026040	0.0002

Table 4.1: Estimated parameters with ML/SURE, 13 degrees of freedom

to be estimated). Let us discuss the results. The intercepts are positive, in the case of electricity negative. This terms stay in the cost function attached the logarithms of prices. Coefficients γ_{ii} are all positive, it reflects that increasing in prices means increasing in the respective cost share. The influence of an increase of production on energy cost shares are shown by the γ_{iy}'s. Corresponding to this, the only cost share of oil inputs is decreasing when output grows, at constant prices. The values of γ_{ij} reflect the response of energy input cost share i to a proportional changes in price of the jth input.

Visual inspection of the next table, table 4.2 we can compare ML/SURE with the RGLS/SURE method. This estimation was provided by Limdep 7.0. The small difference is given by different internal process, as was noted before. The degrees of freedom are now observations minus number of parameters in particular equation, i.e. 6. Note, that the changes are only in decimal places, not in the number sign.

The important statistics of the estimated parameter is the value of t-statistics (see 3.40) and the corresponding probability. The large values of the test statistics will reject the hypothesis that parameter is equal to zero. We can state the hypothesis for all parameters to be simultan equal to zero. To this hypothesis underline the F-statistics (3.41) and was rejected for both models, on the 5% significance level.

Table 4.3 reports about the quality of the fitted model. Expressions for R^2, respectively $\bar{R}^2$, see expression (3.11), resp.(3.12). The measure $\tilde{R}^2$ is a generalized R^2 measure and indicates the proportion of the generalized variance in *left-hand-side* variables explained by variation in the *right-hand-side* variables in the system of equations, and is computed

	Coefficient	Sdt.Error	t-Statistic	Prob.
$\hat{\gamma}_c$	0.1596597132	0.082824752	1.928	0.0539
$\hat{\gamma}_o$	1.730680812	0.16355457	10.582	0.0000
$\hat{\gamma}_g$	0.1783862552	0.19501522	0.915	0.3603
$\hat{\gamma}_e$	−1.068726781	0.21833495	−4.895	0.0000
$\hat{\gamma}_{cc}$	0.1371074717	0.025720977	5.331	0.0000
$\hat{\gamma}_{co}$	0.009691637783	0.015915150	0.609	0.5426
$\hat{\gamma}_{cg}$	−0.08051970762	0.013382888	−6.017	0.0000
$\hat{\gamma}_{ce}$	−0.06627940187	0.017298199	−3.832	0.0001
$\hat{\gamma}_{oo}$	0.1823785448	0.035475528	5.141	0.0000
$\hat{\gamma}_{og}$	−0.1058429481	0.027286504	−3.879	0.0001
$\hat{\gamma}_{oe}$	−0.08622723448	0.017054971	−5.056	0.0000
$\hat{\gamma}_{gg}$	0.2043803673	0.02933551	6.967	0.0000
$\hat{\gamma}_{ge}$	−0.01801771162	0.025122724	−0.717	0.4733
$\hat{\gamma}_{ee}$	0.1705243480	0.03198216	5.332	0.0000
$\hat{\gamma}_{cy}$	0.003446080603	0.0094966465	0.363	0.7167
$\hat{\gamma}_{oy}$	−0.1810461684	0.021715289	−8.337	0.0000
$\hat{\gamma}_{gy}$	0.02951683889	0.025858540	1.141	0.2537
$\hat{\gamma}_{ey}$	0.1480832489	0.029295766	5.055	0.0000

Table 4.2: Estimated parameters with restricted GLS/SURE, 19 degrees of freedom, for every particular equation

as follows [1]:

$$\tilde{R}^2 = 1 - \frac{|E'E|}{|(S-\bar{S})'(S-\bar{S})|} \tag{4.6}$$

where E is the $n \times 3$ matrix of residuals of the stacked system, S is the $n \times 3$ matrix of observations of corresponding energy cost shares and $\bar{S}$ is the $n \times 3$ matrix, where columns are corresponding sample means.

	Restricted GLS/SURE				Iterative SURE or ML/SURE			
	S_c	S_o	S_g	S_e	S_c	S_o	S_g	S_e
R^2	0.521305	0.943594	0.449028	0.621635	0.547901	0.943591	0.443095	0.628363
$\bar{R}^2$	0.39533	0.92875	0.30403	0.52206	0.457481	0.932309	0.331714	0.554036
$\tilde{R}^2$	0.981954643				0.9822			

Table 4.3: Measures of goodness of fit

The reason, why also $\tilde{R}^2$ is reported, is because R^2 of particular equations are not appropriate for estimating a system. In single equation we try to minimize the $e'e$ and thus maximize the value R^2, resp. $\bar{R}^2$. But system estimation does not necessarily minimize the RSS. For example ML estimator minimizes the determinant of the residual cross-product matrix $|E'E|$. Notice that since ML estimator minimizes $|E'E|$, then also maximizes $\tilde{R}^2$, (4.6) [1].

As I have said before, the main purpose of the econometric modelling is to provide some model of producer behaviour and report about parameters and measures like substitution elasticities, technical changes and economics of scales in translog cost function.

The Hicks-Allen partial elasticities of substitution between inputs i and j for a general dual cost function C having n inputs are computed as

$$\sigma_{ij} = \frac{C.C_{ij}}{C_i.C_j} \tag{4.7}$$

where the i, j subscripts refer to first and second partial derivatives of the cost function C with respect to inputs prices p_i, p_j [1]. It is possible to show that expression (4.7) is in fact substitution elasticity (2.7) applied to the cost function. For our translog cost function (4.1) these elasticities turn out to be

$$\begin{aligned} \sigma_{ij} &= \frac{\gamma_{ij} + S_i S_j}{S_i S_j}, \qquad i,j = 1,\ldots,n, \qquad \text{but} \qquad i \neq j \\ \sigma_{ii} &= \frac{\gamma_{ii} + S_i^2 - S_i}{S_i^2}, \qquad i = 1,\ldots,n \end{aligned} \tag{4.8}$$

If one wants to express price relationship $\partial \ln X_i / \partial \ln p_i$, when output quantity and all other input prices are fixed, i.e price elasticity could be done by calculating $\epsilon_{ij} = S_j \sigma_{ij}$. As in the substitution elasticity case, this concept for the translog cost function appears as

$$\begin{aligned} \epsilon_{ij} &= \frac{\gamma_{ij} + S_i S_j}{S_i}, \qquad i,j = 1,\ldots,n, \qquad \text{but} \qquad i \neq j \\ \epsilon_{ii} &= \frac{\gamma_{ii} + S_i^2 - S_i}{S_i}, \qquad i = 1,\ldots,n. \end{aligned} \tag{4.9}$$

Several important comments should be made concerning these elasticities. Parameter estimates and fitted shares can replace the theoretical γ's and real observations S's when computing estimates of the σ_{ij} and ϵ_{ij}. This implies that the estimated elasticities will vary across obsevations and that estimated elasticities have thus stochastic distribution. In praxis we can calculate some unique elasticity value, which is representative over time. Therefore we substitute for the fitted S_i the average value of this fitted column. The tables 4.4 and 4.5 offers the overview of the computed substitution, resp. price elasticities based on the ML/SURE translog parameter estimates. On the basis of these outputs we can

σ_{cc}	σ_{co}	σ_{cg}	σ_{ce}	σ_{oo}
10.031409	1.654772	−1.662019	−1.334088	0.204367
σ_{og}	σ_{oe}	σ_{gg}	σ_{ge}	σ_{ee}
−0.223291	−0.236287	−0.232355	0.810567	−0.539801

Table 4.4: Survey of the estimated substitution elasticities

say that pairs coal and oil, and gas and electricity are substitutable energy inputs (where the values of σ are positive) and other pairs are complements (i.e. since the share of oil increase, the share of electricity decrease). The value of γ_{cc} is inappropriatelly high. I could be caused by the very low level of coal cost share in general. Concerning price elasticities, let us discuss the value of ϵ_{ge}. It seems that if the price of electricity increases then the value of gas cost shares increase. Price elasticities give the response of the cost shares to the changes in prices of energy comodities. Note, that even they are not symmetric the signs stay the same. Thus we can say that the response of the ith, resp. jth inputs to the price change of jth, resp. ith energy input is qualitatively the same.

ϵ_{ij}	coal	oil	gas	electricity
coal	0.707852	0.365458	−0.650061	−0.423081
oil	0.116767	0.045135	−0.087335	−0.074934
gas	−0.117278	−0.049314	−0.090880	0.257056
electricity	−0.094138	−0.052184	0.317035	−0.171188

Table 4.5: Survey of the estimated price elasticities

4.1.3 Estimating the Cost Function

The original cost function (4.1) is in general nonhomothetic cost form. This formula implies that returns to scale, represented by γ_y and $\gamma_y y$ are not constrained a priory. See section 2.1, where concept of RTS is explained. As has been shown by Giora Hanoch [1], returns to scale μ are computed as the inverse of the elasticity of costs with respect to output.

$$\mu = \frac{1}{\epsilon_{Cy}} \qquad \epsilon_{Cy} \equiv \frac{\partial \ln C}{\partial \ln y}$$

Involving this into our translog cost function context we write

$$\epsilon_{Cy} = \gamma_y + \gamma_{cy} \ln p_c + \gamma_{oy} \ln p_o + \gamma_{gy} \ln p_g + \gamma_{ey} \ln p_e + \gamma_{yy} \ln y$$

By estimating the stacked system (4.4), rather than directly the translog cost function (4.1) we loose γ_y and γ_{yy}, because they disappear by the derivation process. But previous formula requires them. One possible way could be estimating the stacked system together with the general cost specification, of course appropriately restricted. As in the case of the stacked system (4.4), the ML estimator even as iterative SURE procedure are invariant to the equation deleted, since the translog cost function is linear in paramenters as well as the restrictions. Note that involving the symmetry, homogeneity of degree one and cost exhaustion restrictions (4.3) the original cost function (4.1) has to be rearranged as

$$\begin{aligned}
\ln(C/p_e) &= \gamma_0 + \gamma_c \ln(p_c/p_e) + \gamma_o \ln(p_o/p_e) + \gamma_g \ln(p_g/p_e) + \gamma_y \ln y \\
&\quad + \gamma_{co} \ln(p_c/p_e) \ln(p_o/p_e) + \gamma_{cg} \ln(p_c/p_e) \ln(p_g/p_e) + \gamma_{og} \ln(p_o/p_e) \ln(p_g/p_e) \\
&\quad + \frac{1}{2}\gamma_{cc}[\ln(p_c/p_e)]^2 + \frac{1}{2}\gamma_{oo}[\ln(p_o/p_e)]^2 + \frac{1}{2}\gamma_{gg}[\ln(p_g/p_e)]^2 \\
&\quad + \gamma_{cy} \ln(p_c/p_e) \ln y + \gamma_{oy} \ln(p_o/p_e) \ln y + \gamma_{gy} \ln(p_g/p_e) \ln y + \frac{1}{2}\gamma_{yy}(\ln y)^2.
\end{aligned}$$

in practise we deal with the stacked system, where the electricity equation is deleted. The results of this system are very close to those which have been provided before with the singular system. A reason could be that the system is relatively well estimated already by the share system alone. Thus for saving the paper and energy, it is enough to write down the three remaining paramters $\gamma_0 = 35.80502103$, $\gamma_y = -6.820591605$ and $\gamma_{yy} = 0.909030075$. The γ_{yy} gives the response of the cost flexibility to the proportional changes in the level of output. Since the number positive is the cost function is convex in the level of output. The γ_0 is the constant of integration. The computed values of the return to scale and ϵ_{Cy} are as follows

$$\mu = 2.914539 \quad \text{and} \quad \epsilon_{Cy} = 0.343107.$$

There was not enough time to discuss this result with my supervisor, but in my opinion and corresponding with the RTS concept by production function, the μ reflects the degree to which a proportional incerase in output increase costs. Since the value is positive, our cost function refer to the increasing RTS.

4.2 Testing

In this section the possible misspecifications of the well behaved basic model, estimated with the ML/SURE, will be investigated and tested. In modelling of producer behaviour we meet two different concepts. The imaginations of economists about producer behaviour are on one side, and the underlying theoretical statements of econometricians about fitting the model on the other side. Thus the researchers, who are modelling microeconomic data, are looking for possible explanation of non rational behaviour or to develop such analytical tools which better describe the real behaviour.

4.2.1 Testing the Constraints

Microeconomic theory imposes that the well behaved cost function has to be homogenous fo degree one in input prices and has to fulfil symmetry restrictions, (compare with the section 2.4.2). This assumptions are converted in our energy-cost function case into well known restrictions (4.3). The question which has to be answered in this section is whether the restrictions describe the real situation among paper producers.

The rejection of the hypothesis of existing restrictions, means for econometricians that we can proceed by rearranging the model and fitting it again, till we obtain the best results as possible. But from microeconomic point of view it means that the producers do not behave rational and rearranging is not a possible way because it is not any more explainable by tools of microeconomic theory.

We will test the relevance of restrictions in the obvious way, with help of the section (3.2.2), where basic tests for restrictions are described. The basic methodology is to provide the estimation of unrestricted system and then test the restrictions in unrestricted system. Testing the homogeneity restrictions and symmetry restrictions at the same time, we will obtain Wald statistic (W) value and likelihood ratio (LR), both statistics are naturally invariant with equation deleted. The values are:

$$W = 30.6624 \qquad LR = 22.763 \qquad > 12.582 = \chi^2_{crit(6)}$$

Both values have $\chi^2(q)$ distribution, where q is the number of restrictions, in our case 6. Three symmetry restrictions and three for homogeneity restrictions. The values give an answer to our question. The restrictions in the model are rejected by the data. Contrary to earlier studies, rationality does not seem to dominate in the paper industry.

4.2.2 Recursive Estimates

Recursive estimation is a comfort procedure utilizing proceeding time, since time gives an unique ordering of the data [6]. The aim of this acces is to give a statements about parameter constancy or its reverse across the time. Our model in time schedule may be written

$$\begin{aligned}
S_{ct} &= \gamma_c + \gamma_{cc}\ln(p_{ct}/p_{et}) + \gamma_{co}\ln(p_{ot}/p_{et}) + \gamma_{cg}\ln(p_{gt}/p_{et}) + \gamma_{cy}\ln y_t + u_{ct} \\
S_{ot} &= \gamma_o + \gamma_{oc}\ln(p_{ct}/p_{et}) + \gamma_{oo}\ln(p_{ot}/p_{et}) + \gamma_{og}\ln(p_{gt}/p_{et}) + \gamma_{oy}\ln y_t + u_{ot} \\
S_{gt} &= \gamma_g + \gamma_{gc}\ln(p_{ct}/p_{et}) + \gamma_{go}\ln(p_{ot}/p_{et}) + \gamma_{gg}\ln(p_{gt}/p_{et}) + \gamma_{gy}\ln y_t + u_{gt},
\end{aligned}$$

where $t = 1, \ldots, n$. Subscript t indicate the tth observation of the variables. The idea behind recursive estimates is very simple: Fit the model to the first r observations. Next use the first $r+1$ data points and compute all coefficients again and then again till the endpoint of the sample. This process generate in our case row vectors of γ's. First possible estimation I have provided with observations till year 1986, thus in this case $r = 14$. The standard errors of the various coefficients may be computed at each stage of the recursion, exept the case when $r = k$, where k is number of regressors since the fit is in this case perfect ($RSS = 0$). However most computer programs doing this estimation in some time point $t > k$. Graphs showing the evolution of each coefficient plus and minus two standards errors can be prepared. Visual inspection of this graphs could give the incentive of possible inconstancy of parameters. As data are added, graph may sometimes display some vertical movements. Such movements reflect the instability of parameters.

Since in our system appear 18 various parameters and more or less they reflect some instabilities, but gain in such investigation can give also recursive shares values, where recursive estimated parameters are gradually substituted. Graphs of recursive cost shares are available in appendix. Visual inspection of them show that cost shares of coal and oil are relative stable in proceeding time, and a narrow, nonchanging band of statistical error assure that the parameters estimated there are stable too. Another consideration gives the view of the last two energy inputs shares. In both cases we can see that around year 1993 the standard error band became wider. This can have several explanation. What is most plausible is that the last four observations, were added to the previous data set *ex post* and were taken from another source, than the previous 21 observations. Therefore the new data probably do not fit to the microeconomic framework and create a distortion, which is also reflected in relative small values of R^2 of this equations. Nonetheless the cutting out of the last four observations means less observations and less degrees of freedom in a model as a whole.

4.2.3 One-Step Ahead Prediction Errors

The next approach works on the basis of recursive estimation. By using all data up to and including period $t-1$, the one step ahead prediction of y_t, i.e endogenous variable, is $\mathbf{x}_t'\hat{\boldsymbol{\beta}}_{t-1}$. The **one step ahead prediction error** is thus [6]

$$v_t = y_t - \mathbf{x}_t'\hat{\boldsymbol{\beta}}_{t-1}$$

The only difference in comparing with residuals is that the difference between observed and estimated value is computed with parameters obtained in previous step of regression. Since the vector is obtained in recursive way, it is also referred to as **recursive residuals**.

Again we have to implement it into our system frame. Since we have four equations, we will obtain four vectors of one-step ahead prediction errors. In practical examples it is obvious to plot this vector together with plus or minus twice with recursively estimated standard errors. This estimated standard errors are obtained from first $t-1$ observations. We will start this procedure in year 1987, thus starting vector $\boldsymbol{\beta}$ will be estimated with 1972-1986 sample. Residuals lying outside the standard error bands are suggestive of parameter inconstancy [6]. Looking on our four plots (available in the appendix) is clear that again the last two equations, i.e. gas and electricity, do not fulfil our expectations and stress that the observations after 1993 are not in conformity with the rest of the estimation, at least.

One question of interest arises, we deal here with the system, not with the particular equation, thus we cannot neglect whatever happens in one equation without disturbing another ones. Therefore we have to consider about a measure which will generalize this procedure to a system specification.

4.2.4 Chow Test for Constancy of Parameters

The following test belongs to the class of tests which search for statement about constancy of parameters outside of the sample set, i.e if it is possible to generalize the validity of the model. In this chapter we have had many possibilities to affirm the statement that our basic model (4.2) and corresponding restrictions (4.3) are not appropriate in each case.

The **Chow test** for constancy of parameters is also referred as **test for predictive accuracy**. This test is based on calculating of the vector of prediction errors $\mathbf{d}$ and can be decomposed into several steps.

(i) Divide the sample set into two parts, where first n_1 observations will provide the estimation and last n_2 observations will be after estimation utilized for testing. In our case the ratio n_1/n_2 is 21/4. The reason is quite simple, as was noted before, the last four observations were added later and we will monitor the constancy of parameters in these *ex post* observations. I suppose you probably guess the result. Subscripts 1,resp. 2 we will use corresponding with n_1, or n_2 sample set.

(ii) Estimate the OLS vector from the first n_1 observations, obtaining

$$\hat{\boldsymbol{\beta}}_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y_1}$$

is the primary estimation.

(iii) Use $\hat{\boldsymbol{\beta}}_1$ to obtain a prediction of the $\mathbf{y}_2$ vector, namely,

$$\hat{\mathbf{y}}_2 = \mathbf{X}_2\hat{\boldsymbol{\beta}}_1$$

(iv) Obtain the vector of prediction errors and analyze its sampling distribution under the null hypothesis of parameter constancy. The vector of prediction errors is

$$\mathbf{d} = \mathbf{y}_2 - \hat{\mathbf{y}}_2 = \mathbf{y}_2 - \mathbf{X}_2\hat{\boldsymbol{\beta}}^1$$

In search for the sampling distribution, we assume that $E(\mathbf{uu}') = \sigma^2\mathbf{I}$, holds for both data sets. Thus $E(\mathbf{d}) = \mathbf{0}$ and it may be shown, see [6] that the variance-covariance matrix for $\mathbf{d}$ is

$$\mathrm{var}(\mathbf{d}) = \sigma^2[\mathbf{I}_{\mathbf{n}_2} + \mathbf{X}_2(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_2']$$

Therefore, under the hypothesis of parameter constancy we use following test statistic

$$F = \frac{\mathbf{d}'[\mathbf{I}_{\mathbf{n}_2} + \mathbf{X}_2(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_2']^{-1}\mathbf{d}/n_2}{\mathbf{e}_1'\mathbf{e}_1/(n_1-k)} \sim F(n_2, n_1-k)$$

Large values of this F statistic would reject the hypothesis that the same β vector applies within and outside the the estimation data (compare with [6]).

Note, that this test is defined for single equation specification, thus we will compute the upper F test for each equation. The estimated $\hat{\boldsymbol{\beta}}_1$ will be obtained from ML/SURE estimation of our system. The table 4.6 gives the values of Chow F-test in each share equation. Since the values of test statictic are rather big in case of the gas and electricity equation, we reject the hypothesis, that paramters in this equations are constant outside the sample. As obvious we encounter the problem of system estimation. In system of

	S_c	S_o	S_g	S_e
F-test	0.465531	0.516173	8.419764	36.99304

Table 4.6: Values of the Chow F test

demand equations all elements have the unique irrecoverable place which is given by the underlying microeconomic theory. Thus, we can not remove some elements without disturbing other ones. Concerned with this test we can say that parameters estimated have general validity for coal and oil equations, but for the remaining two not. It does not mean, that we can throw out gas and electricity equation. Doing it we would loose our starting hypothesis, that energy costs are explainable by all energy inputs and the level of output. Simply said, the failure of constancy of the models, does not mean that the model is notappropriate in general.

This methodology, and also some of the previous ones challenge the exploration of tests, which are applicable to system specification.

4.3 Estimating with Vector Autocorrelation

The purpose of econometric modelling of producer behaviour is not to support the underlying theory by the real situation in production sphere, but to show with the analytical or econometric tools the comparison with actual behaviour of real producers. In this section we will solve the basic model (4.2) by adding the assumption of autocorrrelated disturbances. First, we will search for a possible method, which allows to compute the system of factor demands with vector of autocorrelated disturbances. Second, we will introduce the results and the last task will be providing a comparison with previous results. Since the basic model has been changed, one important question arises, if respecifying better reflects the real situation of paper producers. Hopefully, this part gives a reasonable answer.

4.3.1 Methods

We turn now our attention to empirical implementation of the autocorrelated disturbances assumption (3.49) into basic model (4.2). First, note that we have reason to assume its relevance. For testing the hypothesis of zero autocorrelation against the alternative of positive first-order autocorrelation (see paragraph 3.2.3) we use lover d_L and upper d_U bounds of the Durbin-Watson test statistic tabulated. The table 4.7 gives the values of d_L and d_U for 25 numbers of observation and 4 regressors in the case of ML/SURE regression and for 25 observations and 5 regressors what is the case of restricted GLS/SURE, where the parameters are *ex-post* restricted. Note that the number of regressors is excluding intercept. The number of regressors is after excluding the intercepts. Comparing the

observations	d_L	d_U
$25, 4$ regressors	1.038	1.767
$25, 5$ regressors	0.953	1.886

Table 4.7: Critical bounds of the Durbin-Watson test statistic

critical bounds of the DW test statistic for a particular share equation, 4.8,

equation	DW in ML/SURE	DW in RGLS/SURE
S_c	1.932	1.9031
S_o	1.575	1.5754
S_g	0.422	0.4252
S_e	0.298	0.2978

Table 4.8: Survey of the DW values for particular equations

we can say that at least in two cases (S_g and S_e) the value of DW statistic is much smaller than d_L not rejecting the existence of positive autocorrelation in disturbance terms. This statement might us affirm the residual graphs of S_o, S_g and S_e, which are available in appendix. Comparing the graphs of electricity and gas equation with oil equation we can observe the difference. The rational for such behavior can be explained by small flexibility of producers when there happen unexpected changes in prices or technologies. They are not prepared to respond immediately, they need time to decide. DW of coal share equation rejects the hypothesis about AR(1) and by oil equation the test is inconclusive, because the value is between d_L and d_U.

Supported by this investigation we can say that the basic model does not cover the whole problem because we neglect that disturbances follow an AR(1) process. We will proceed in the same way like in paragraph (3.3). Suppose that our model

$$\begin{aligned}
S_{ct} &= \gamma_c + \gamma_{cc}\ln p_{ct} + \gamma_{co}\ln p_{ot} + \gamma_{cg}\ln p_{gt} + \gamma_{ce}\ln p_{et} + \gamma_{cy}\ln p_{yt} + u_{ct} \\
S_{ot} &= \gamma_o + \gamma_{oc}\ln p_{ct} + \gamma_{oo}\ln p_{ot} + \gamma_{og}\ln p_{gt} + \gamma_{oe}\ln p_{et} + \gamma_{oy}\ln p_{yt} + u_{ot} \\
S_{gt} &= \gamma_c + \gamma_{gc}\ln p_{ct} + \gamma_{go}\ln p_{ot} + \gamma_{gg}\ln p_{gt} + \gamma_{ge}\ln p_{et} + \gamma_{gy}\ln p_{yt} + u_{gt} \\
S_{et} &= \gamma_c + \gamma_{ec}\ln p_{ct} + \gamma_{eo}\ln p_{ot} + \gamma_{eg}\ln p_{gt} + \gamma_{ee}\ln p_{et} + \gamma_{ey}\ln p_{yt} + u_{et},
\end{aligned} \tag{4.10}$$

has autocorrelated vector of disturbances u_t

$$\begin{aligned}
u_{ct} &= Q_{cc}u_{ct-1} + \varepsilon_{ct} \\
u_{ot} &= Q_{oo}u_{ot-1} + \varepsilon_{ct} \\
u_{gt} &= Q_{gg}u_{gt-1} + \varepsilon_{ct} \\
u_{et} &= Q_{ee}u_{et-1} + \varepsilon_{ct}
\end{aligned} \tag{4.11}$$

where $t = 2, \ldots, 25$. Note that Q is assumed to be diagonal

$$Q = \begin{pmatrix} Q_{cc} & 0 & 0 & 0 \\ 0 & Q_{oo} & 0 & 0 \\ 0 & 0 & Q_{gg} & 0 \\ 0 & 0 & 0 & Q_{ee} \end{pmatrix}$$

and cross correlation relationships do not occur as they do not have any practical explanation. As was discussed before Q_{cc} is zero too. The sum of the cost shares has to be one, that is in fact adding up condition (3.50). This condition says that the system (4.10, 4.11) is singular and with constraints for parameters

$$\gamma_c + \gamma_o + \gamma_g + \gamma_e = 1$$

$$\gamma_{cc} + \gamma_{oc} + \gamma_{gc} + \gamma_{ec} = 0$$
$$\gamma_{co} + \gamma_{oo} + \gamma_{go} + \gamma_{eo} = 0$$
$$\gamma_{cg} + \gamma_{og} + \gamma_{gg} + \gamma_{eg} = 0$$
$$\gamma_{ce} + \gamma_{oe} + \gamma_{ge} + \gamma_{ee} = 0$$
$$\gamma_{cy} + \gamma_{oy} + \gamma_{gy} + \gamma_{ey} = 0$$

and for u_t

$$\imath'\mathbf{u}_t = 0 \qquad \imath'\varepsilon_t = 0 \qquad \text{and} \qquad \imath'Q = k',$$

where $\mathbf{u}_t$ and ε_t are corresponding vectors of the system (4.11). The last condition is very strong and for our assumption (4.11) it gives

$$Q_{cc} = Q_{oo} = Q_{gg} = Q_{ee} = k,$$

where k is an unknown constant. This constraint is too restrictive because of various values of DW in particular equations. Therefore it could create greater distortion than neglecting the autocorrelated disturbances at all. But estimation of the model (4.10, 4.11) where diagonal elements of Q are not equal, is not invariant to the equation deleted. So we will have four different results, one for each deleted equation. We wonder whether such result can improve our basic model.

We will attempt to estimate the autocorrelated model. In fact there are two possibilities. First is to let the ML procedure run without deleting any equation. Since the diagonal elements are assumed not to be equal, system is not singular any more. But problematic issue is that computer packages do not recognize such small "nonsingularity" and there does not exist any tool to force,to compute with near singular matrix.

One interesting idea, introduced by Cochrane-Orcutt as **iterative estimation procedure** could be employed to our estimation. First, let us reformulate our system to clarify Cochrane-Orcutt procedure. We will take the stacked system where coal equation is deleted, since in this equation the existence of autocorrelated disturbances could be neglected. The system may be written in unrestricted form

$$\begin{aligned} S_o &= \gamma_o + \gamma_{oc}\ln p_c + \gamma_{oo}\ln p_o + \gamma_{og}\ln p_g + \gamma_{oe}\ln p_e + \gamma_{oy}\ln y \\ S_g &= \gamma_g + \gamma_{gc}\ln p_c + \gamma_{go}\ln p_o + \gamma_{gg}\ln p_g + \gamma_{ge}\ln p_e + \gamma_{gy}\ln y \\ S_e &= \gamma_e + \gamma_{ec}\ln p_c + \gamma_{eo}\ln p_o + \gamma_{eg}\ln p_g + \gamma_{ee}\ln p_e + \gamma_{cy}\ln y \end{aligned} \tag{4.12}$$

where symmetry restriction imposed are $\gamma_{og} = \gamma_{go}$, $\gamma_{oe} = \gamma_{eo}$, $\gamma_{ge} = \gamma_{eg}$ and homogeneity of degree zero in input prices are

$$\begin{aligned} \gamma_{oc} &= -\gamma_{oo} - \gamma_{og} - \gamma_{oe} \\ \gamma_{gc} &= -\gamma_{og} - \gamma_{gg} - \gamma_{ge} \\ \gamma_{ec} &= -\gamma_{oe} - \gamma_{ge} - \gamma_{ee} \end{aligned}$$

This notation will help us to understand, how the iterative estimation procedure works. Now, taking the arbitrary equation from the stacked system (4.12) with assumed autocorrelated disturbances we have

$$S_{it} = \gamma_i + \gamma_{ic}\ln p_{ct} + \gamma_{io}\ln p_{ot} + \gamma_{ig}\ln p_{gt} + \gamma_{ie}\ln p_{et} + \gamma_{iy}\ln y_t + u_{it}$$
$$\text{and} \qquad u_{it} = Q_{ii}u_{it-1} + \epsilon_{it}$$

The nature of this restriction may be seen by rewriting in following form

$$\begin{aligned} S_{it} &= \gamma_i(1-Q_{ii}) + \gamma_{ic}\ln p_{ct} + \ldots + \gamma_{ie}\ln p_{et} + \gamma_{iy}\ln y_t \\ &\quad -Q_{ii}(\gamma_{ic}\ln p_{ct-1} + \ldots + \gamma_{ie}\ln p_{et-1} + \gamma_{iy}\ln y_{t-1}) + Q_{ii}S_{it-1} + \epsilon_{it} \end{aligned}$$

where $t = 2, \ldots, 25$, in both expressions. However, this expression is nonlinear in parameters, therefore nonlinear least squares is required. But in order to stay in the same, say dimension or compatibility with previous estimations, we will proceed not in this direction. Rearranging the previous formulation as

$$\begin{aligned} (S_{it} - Q_{ii}S_{it-1}) &= \gamma_i(1-Q_{ii}) + \gamma_{ic}(\ln p_{ct} - Q_{ii}\ln p_{ct-1}) + \ldots + \gamma_{ie}(\ln p_{et} - Q_{ii}\ln p_{et-1}) \\ &\quad +\gamma_{iy}(\ln y_t - Q_{ii}\ln y_{t-1}) + \epsilon_{it} \end{aligned} \quad (4.13)$$

or equivalently

$$\begin{gathered} (S_{it} - \gamma_i - \gamma_{ic}\ln p_{ct} - \ldots - \gamma_{ie}\ln p_{et} + \gamma_{iy}\ln y_t) = \\ Q_{ii}(S_{it-1} - \gamma_{ic}\ln p_{ct-1} - \ldots - \gamma_{ie}\ln p_{et-1} - \gamma_{iy}\ln y_{t-1}) + \epsilon_{it} \end{gathered} \quad (4.14)$$

where again $t = 2, \ldots, 25$ and i could be arbitrary subscript of a (o, g, e) set. Cochran-Orcutt procedure is based on a simple consideration. If Q_{ii} were known in equation (4.13), the γ's could be estimated by an OLS regression and conversely, if the γ's were known in equation (4.14), the Q_{ii} could be estimated by an OLS procedure. Therefore we need just a start guess of $\hat{Q}^1_{ii}$ and iterative procedure can start. The iterations continue until a satisfactory degree of convergence is reached. Note that we will loose the first observation due to a lagged expression. We will use this procedure, to estimate parameter system (4.12) where the homogeneity and symmetry restrictions are imposed. Results, obtained with help of Limdep 7.0 software are presented in the next section.

4.3.2 Results

Before I outline some smart tables with estimation results, I would like to discuss my ideas concerning the autocorrelation in equation system of paper producers. The autocorrelation coefficients have sense just in three equations (oil, gas, electricity). Thus, I decided to estimate just this system, with the adding up assumption of autocorrelated disturbances. The econometric tool was, as just explained, iterative (Cochran-Orcutt) procedure applied to a system specification. After it, γ's for deleted coal equation have been computed from homogeneity restriction patterns (4.3). This could be done under the strict assumption, that homogeneity is valid over the whole sample. The values of Wald and Likelihood ratio tests (see section 4.2.1) are not encouraging though. However, the main reason was, that in coal equation we can freely neglect the autocorrelation assumption, because the equation does not require it. And the second reason, as was noted before, various values of autocorrelated coefficients make the system not invariant to the equation deleted. The restricted model (4.12) with the appropriate equations of autocorrelated disturbances (4.11) we will refer to as corrected model.

After this introduction we can turn our attention to the discussion of the results. Note that the iterative procedure could be stopped after first iteration of Cochrane-Orcutt procedure. Naturally the results with converged procedure are even better, thus I will state just them. Limdep-software use as the first guess of autocorrelated coefficient the value

$$\hat{Q}^1_{ii} = 1 - \frac{\mathrm{DW}_i}{2}$$

where DW_i is the Durbin-Watson statistic computed using the single equation OLS residuals [8]. Table 4.9 offers the starting values of the autocorrelation coefficients.

	S_o	S_g	S_e
$\hat{Q}^1_{ii}$	0.2105	0.4317	0.6195

Table 4.9: Starting values of autocorrelation coefficients

Finally, we can view the estimated parameters. In the table 4.10 some statistical informations are missing. They are associated with the parameters of the deleted equation and thus they have no underlying statistical results.

I think the first question is eliminating the problem of autocorrelation. The table 4.11

	Coefficient	Sdt.Error	t-Statistic	Prob.
$\hat{\gamma}_c$	0.027298361			
$\hat{\gamma}_o$	1.835653728	0.17696322	10.373	0.0000
$\hat{\gamma}_g$	0.195866252	0.25349267	0.773	0.4397
$\hat{\gamma}_e$	−1.058818341	0.31208749	−3.393	0.0007
$\hat{\gamma}_{cc}$	0.11430377			
$\hat{\gamma}_{co}$	0.006303849	0.021640664	0.291	0.7708
$\hat{\gamma}_{cg}$	−0.072555607	0.018123459	−4.003	0.0001
$\hat{\gamma}_{ce}$	−0.048052012	0.031278163	−1.536	0.1245
$\hat{\gamma}_{oo}$	0.16917057	0.034989005	4.835	0.0000
$\hat{\gamma}_{og}$	−0.094996688	0.027251599	−3.486	0.0005
$\hat{\gamma}_{oe}$	−0.080477731	0.016844121	−4.778	0.0000
$\hat{\gamma}_{gg}$	0.217922624	0.035588989	6.123	0.0000
$\hat{\gamma}_{ge}$	−0.050370329	0.029450453	−1.710	0.0872
$\hat{\gamma}_{ee}$	0.178900071	0.047090075	3.799	0.0001
$\hat{\gamma}_{cy}$	0.018365257			
$\hat{\gamma}_{oy}$	−0.195577821	0.022915802	−8.535	0.0000
$\hat{\gamma}_{gy}$	0.034781759	0.033549551	1.037	0.2999
$\hat{\gamma}_{ey}$	0.142430805	0.041415883	3.439	0.0006

Table 4.10: Estimation results of corrected model

provide a survey of DW test statistic, R^2 and $\bar{R}^2$ in the corrected model. The values of

	S_o	S_g	S_e
DW	1.9134	0.7969	0.6457
R^2	0.944272	0.671346	0.851033
$\bar{R}^2$	0.92961	0.58486	0.81183

Table 4.11: Survey of the DW, R^2, $\bar{R}^2$ measures in corrected model

DW reflect that AR(1) process of the residual has improved the oil equation, but the remaining two equations in spite of the small improvement (compare with the table 4.8) are still suspicious with respect to autocorrelated disturbances. The goodness of fit measures may be compared with the basic model values given in the table 4.3 in columns restricted

GLS/SURE. From the first view we see that the corrected model has a better fit. As conclusion we can say that for oil, gas and electricity equations the corrected model reflect more the behaviour of producers than the basic model. Additionally the gas equation is appropriately explained with autocorrelated disturbances.

One could object against such "nonscience" approach, which has been introduced. I can not say that it is really the best and most correct solution of the described problem. However, let us look who, in fact, is interested in substitution and price elasticities and measures of scale? In my opinion, mostly economists. They would like to know the producers response on changing prices or changing technologies. For them the best possible result is more worthy than any perfect theoretical result.

4.3.3 Comparing with the Basic Model

This section has the aim to compare the basic and the corrected model from the point of view of elasticities. The elasticity measure gives us two kind of informations, first to make some statement about relationships between components in the cost function, second, to tell us the power of influence. The first is represented by the plus, or minus sign and the second by the value of the measure. Corresponding to this we will make a comparison with basic model. Table 4.12 shows that for the substitution elasticities values, there does not appear any change in the signs, unless in decimal places. This holds for the price elasticities as well. The only change in sign is happen. σ_{oo} and afterwards ϵ_{oo}

	Basic model	Corrected model
σ_{cc}	10.031409	5.914313
σ_{co}	1.654772	1.345061
σ_{cg}	-1.662019	-1.216653
σ_{ce}	-1.334088	-1.007855
σ_{oo}	0.204367	-0.091829
σ_{og}	-0.223291	-0.057539
σ_{oe}	-0.236287	-0.225343
σ_{gg}	-0.232355	-0.138630
σ_{ge}	0.810567	0.571951
σ_{ee}	-0.539801	-0.329901

Table 4.12: Comparing table of the estimated substitution elasticities

	Basic model	Corrected model		Basic model	Corrected model
ϵ_{cc}	0.707852	0.482547	ϵ_{gc}	-0.117278	-0.099266
ϵ_{co}	0.365458	0.301173	ϵ_{go}	-0.049314	-0.012884
ϵ_{cg}	-0.650061	-0.488095	ϵ_{gg}	-0.090880	-0.055615
ϵ_{ce}	-0.423081	-0.295625	ϵ_{ge}	0.257056	0.167765
ϵ_{oc}	0.116767	0.109743	ϵ_{ec}	-0.094138	-0.082231
ϵ_{oo}	0.045135	-0.020562	ϵ_{eo}	-0.052184	-0.050457
ϵ_{og}	-0.087335	-0.023084	ϵ_{eg}	0.317035	0.229454
ϵ_{oe}	-0.074934	-0.066098	ϵ_{ee}	-0.171188	-0.096767

Table 4.13: Comparing table of the estimated price elasticities

became negative in the corrected model. The reason is that probably with assumption of autocorrelated disturbances, the oil cost share response conversely when the oil share and oil price proportional increase. Since the oil equation was the only, where autocorrelated disturbances have had success, this change might be concerned with this issue.

Here I stop with my investigation of paper producer behaviour. The corrected model does not substantially improve the problematic results in gas and electricity equation. For correcting the model we would need either more knowledge about the market with energy input commodities, or to investigate alternative models. Probably the missing information on better data could help us to explain the non-smart behaviour of recursive residual curve, mainly in the last years. I believe there could exist other possible and more involved model variants but that I want to leave for others to do.

Chapter 5

Summary

This chapter concludes my thesis. Since econometric modelling is a methodolody where one can never stop improving, some remarks about the possible ways to proceed will be in order. Finally, I will present my personal view on the topic I have investigated.

5.1 Challenges for Future Research

Among the useful econometric tools to improve the model specification one might consider lagged variables in the model. Lagged means in terms of time point t one of the previous observations, i.e. $t-1, t-2, \ldots$, of some variable. The rationale for such behaviour could be, for example, that the previous prices influence also current decisions of producers or consumer.

Let us implement this concept into the basic model (4.2). We will consider only a simple case, where besides original regressors also the $t-1$ values of the endogenous input shares are included. The renewed model is then

$$\begin{aligned}
S_c &= \gamma_c + \gamma_{cc}\ln p_c + \gamma_{co}\ln p_o + \gamma_{cg}\ln p_g + \gamma_{ce}\ln p_e + \gamma_{cy}\ln y + \rho_{cc}S_{c-1} \\
S_o &= \gamma_o + \gamma_{oc}\ln p_c + \gamma_{oo}\ln p_o + \gamma_{og}\ln p_g + \gamma_{oe}\ln p_e + \gamma_{oy}\ln y + \rho_{oo}S_{o-1} \\
S_g &= \gamma_g + \gamma_{gc}\ln p_c + \gamma_{go}\ln p_o + \gamma_{gg}\ln p_g + \gamma_{ge}\ln p_e + \gamma_{gy}\ln y + \rho_{gg}S_{g-1} \\
S_e &= \gamma_e + \gamma_{ec}\ln p_c + \gamma_{eo}\ln p_o + \gamma_{eg}\ln p_g + \gamma_{ee}\ln p_e + \gamma_{ey}\ln y + \rho_{ee}S_{e-1}.
\end{aligned} \tag{5.1}$$

The sign -1 means $t-1$th observation of S_i. We loose one observation with the imposition of the $t-1$ term and thus have $t = 2, 3, \ldots, 25$. Since the sum of input shares is equal to one, the upper expression states that ρ's have to be zero too. To overcome this, we can implement also cross-equation lagged variables. Thus rewriting (5.1) we have

$$\begin{aligned}
S_c &= \ldots + \rho_{cc}S_{c-1} + \rho_{co}S_{o-1} + \rho_{cg}S_{g-1} + \rho_{ce}S_{e-1} \\
S_o &= \ldots + \rho_{oc}S_{c-1} + \rho_{oo}S_{o-1} + \rho_{og}S_{g-1} + \rho_{oe}S_{e-1} \\
S_g &= \ldots + \rho_{gc}S_{c-1} + \rho_{go}S_{o-1} + \rho_{gg}S_{g-1} + \rho_{ge}S_{e-1} \\
S_e &= \ldots + \rho_{ec}S_{c-1} + \rho_{eo}S_{o-1} + \rho_{eg}S_{g-1} + \rho_{ee}S_{e-1},
\end{aligned} \tag{5.2}$$

where dots indicate the original regressors from basic model (4.2). Taking into the consideration that the sum of input cost shares is one, following restrictions to the ρ's have to hold

$$\begin{aligned}
\rho_{cc} + \rho_{oc} + \rho_{gc} + \rho_{ec} &= 0 \\
\rho_{co} + \rho_{oo} + \rho_{go} + \rho_{eo} &= 0 \\
\rho_{cg} + \rho_{og} + \rho_{gg} + \rho_{eg} &= 0 \\
\rho_{ce} + \rho_{oe} + \rho_{ge} + \rho_{ee} &= 0.
\end{aligned}$$

Proceeding with deleting an arbitrary (for instance electricity) equation from the singular system (5.3) we obtain

$$\begin{aligned}
S_c &= \ldots + (\rho_{cc} - \rho_{ce})S_{c-1} + (\rho_{co} - \rho_{ce})S_{o-1} + (\rho_{cg} - \rho_{ce})S_{g-1} + \rho_{ce} \\
S_o &= \ldots + (\rho_{oc} - \rho_{oe})S_{c-1} + (\rho_{oo} - \rho_{oe})S_{o-1} + (\rho_{cg} - \rho_{ce})S_{g-1} + \rho_{oe} \\
S_g &= \ldots + (\rho_{gc} - \rho_{ge})S_{c-1} + (\rho_{go} - \rho_{ge})S_{o-1} + (\rho_{gg} - \rho_{ge})S_{g-1} + \rho_{ge}
\end{aligned} \tag{5.3}$$

where again dots indicate the part of the model (4.4).The next problem of this system arises from the unsuitable large number of unknown parameters. Summing the original number of 12 from (4.4) plus 12 new parameters of the lagged variables we have 24 unknown parameters to estimate. Comparing it with the 25 observations, we have encountered the frontier of econometric investigation.

Inspite of this, there are still things which are worthy to consider. We have added to the basic system (4.2) lagged variables as regressors. Note, that the basic system is in fact a product of the derivation from translog cost function (4.1). The next task is therefore to investigate the reverse process to derive such a cost function, which after employing Shephard's lemma would yield system (5.2).

Here we are encountering to the limitations of the microeconomic analysis. This problems has several possible directions. Either the cost function has to be respecified into the dynamic framework and then by integrating the system (5.2) it has to be appropriate involved. Or, to assume that S_{i-1} is simply independent from the $\ln p_i$ and to hold it as constant by integration process.

5.2 Concluding Remarks

Arriving at the end of my thesis I briefly want to summarize. My aim at the start was to provide, test and improve an empirical model of producer behaviour. In three core chapters (Microeconomic analysis, Econometric methods and Applications) I have introduced you to the issue of econometric modelling in the field of microeconomic theory and have shown how it can be applied to data of the Austrian paper industry. The following story might help you to find out how succesfull I was from my point of view.

When I was a small child as well as during my teen-ager time I wanted to help people. On some occasions, my parents have reminded me that I was a social subject. After this time, getting older I have lost some of my enthusiasm, however I wanted to be useful at least. Later when my university studies began I have recognized that to be useful is much harder than to be wise, so I wanted to know as much as possible. And in this state the question of finding a topic of my master thesis has come up. I did not know, what was a good topic or a bad one, but more or less by accident I have met my supervisor and afterwards settled on my topic. And here again I encountered a similar story. In the beginnig (i.e. aproximately one year ago) I wanted to invent something quite new and possibly monumental. It was the time when I had to introduce myself into my topic. Since then I have recognized that to find out some new idea is not so simple. I have turned my attention to modelling the data, with the hope that I will manage to provide an accomplished model at least. It has taken some time, but again I have experienced that to find out a perfect model is not really a simple task and, in fact, you cannot say that in any point of your investigation that you have provided a really good model. Thus my last strategy was to provide at least some reasonable model which can be established and examined according to available approaches and which helps nevertheless to learn as much as possible about the relation between mathematical economics and the reality. I think that this last aim was fulfilled and that makes my thesis successful.

Now I will outline some comments associated with the problems which I had to confront during the process of working on my thesis. The problems could be divided into two areas.

The first category consists of problems with data. For econometric modelling we need the biggest sample which is available. To satisfy this condition in the microeconomic field is not so easy. In most cases you have data available for the last decade, especially in the case of post-comunistic countries. Ten observations could not explain such a broad system as the theory of the cost function offers. This problem could be avoided by using more producers or countries, say units, over the same time period. This builds up a **panel data set**. Such set consists of many small sets where the same behaviour (in sense of parameters) is assumed. I belive that panel data models are the future for modeling producer behaviour. The next problem is concerned with the information about the concrete data set. If we do not see irrational movements behind actual observation, we equally cannot explain some irrational response as may be reflected in estimated parameters, derived elasticities, the resulting disturbances and anything else which is connected with the estimation output.

The second area of problems could be specified as the conflict of interest between microeconomic theory and econometric methods. The microeconomic visions are cast into functions, systems and restrictions, which cannot be neglected. Estimation results, from the econometric point of view, are strictly categorized. Either the parameters, test or restriction are significant - and then they are recommended to enter the model - or not, and then they have to be excluded from the model. The recipe used to solve this conflict is the *give-and-take* methodology. Since the microeconomic theory is a normative theory, neglecting some of their restrictions or assumptions generates a new theoretical aspect or component, we usually have to explain the failure of the econometric result to comply with theory in all aspects by apologising for the inadequacy of data on the one hand, or by the inadequacy of the theory to explain the real world on the other.

Finally, I would like to state my personal attitude to the presented topic. I think that econometric modelling is a challenging and quick developing methodology. The reason is that it finds possible applications in various branches of the modern world. Everywhere around we can observe the relationships between issues concerning our life. The econometric models can give an answer about the approximative extent of the real dependency, can reject or not-reject our hypothesis, can measure the strength of relations and helps to gain probably much more insight into the issues.

I am very pleased that I could be introduced to this topic. I have to say, that my gain was much bigger. As a by product of preparing my thesis I have come to know a new counry and the people there. More or less I can speak their language. In this way, I would like to thank to all people, who enabled me to spend one semester in Austria. I have learned to work with two appropriate econometric software packages and there is surely much more to mention, but here is not the adequate place to write it down.

Returning to the considerations at the beginning, now I think that first of all, you have to know as much as possible, then you have a higher probability to be useful and further, if you are useful and appropriately wise you have not such a small chance that you might really be able to help people. But this speculations require the strict assumption to be convinced that the wish to help people is your starting incentive to become wise and useful.

Chapter 6

Resume

Ekonometrické metódy zohravajú dôležitú úlohu pri modelovaní skutočnosti. Ich úlohou je vyjadriť pomocou známych premenných kvalitatívne a kvantitatívne vzťahy medzi nimi. Svoje uplatnenie ekonometria našla aj v mikroekonómii. Mikroekonómia skúma a popisuje racionálnu podstatu správania sa ekonomických subjektov ako sú napríklad výrobcovia alebo spotrebitelia. Proces výroby je popísaný produkčnou funkciou, ktorá odzrkadľuje vzájomný vzťah medzi vstupmi výrobného procesu a jeho výstupmi. Racionalita výrobcu je založená na maximalizácii zisku vzhľadom k produkčnej funkcii. Dá sa však popísať aj ako snaha o minimálne náklady na výrobu. Tento proces popisuje nákladová funkcia. Keďže tieto vzťahy sú ekvivalentné, čo sa týka racionality výrobcu, môžme použit ľubovolný postup a dosiahneme tie isté výsledky. Tento záver však dokazuje istý vzťah duality medzi nákladovou a produkčnou funkciou. Pre ekonómov to znamená, že stačí vedieť jednu z týchto dvoch funkcií a tá už nesie informáciu aj o druhej.

V ekonometrii sa tento poznatok veľmi využíva. Je oveľa výhodnejšie využit nákladovú funkciu, pri ktorej dopytové a ponukové funkcie sa dajú vyjadriť explicitne derivovaním nákladovej funkcie podľa vektoru cien. V prípade maximalizácie zisku vzhľadom k produkčnej funkcii, by takýto systém bol iba implicitne závislý od cien, čo je pre ekonometrické modelovanie nevýhodné.

Ekonometrický model správania sa výrobcu, ktorý je popísaný a aplikovaný v tejto práci, má formu systému dopytových rovníc, resp. systém rovníc nákladových podielov (cost shares), odvodených z nákladovej funkcie. Opísaný model je trochu špecifický. Zakladá sa na predpoklade, že energetické náklady na výrobu sú závislé iba od cien energetických komodít a veľkosti výroby. Dáta obsahujú informácie o množstvách štyroch energetických vstupov, ktoré vstupujú do výrobného procesu, vývoj ich cien a veľkost výroby papierenského priemyslu ako takého, počas rokov 1972 - 1996. Ďalším špecifikom modelu je, že budeme modelovať systém dopytových funkcií pre celé priemyselné odvetvie a nie pre konkrétneho výrobcu. Predpokladáme teda, že správanie výrobcu papiera sa dá zovšeobecniť. Štyri energetické vstupy, ktoré budeme uvažovat sú: uhlie, ropa, plyn a elektrina. Pod veľkosťou výroby rozumieme celkové množstvo vyrobeného papiera bez ohľadu na jeho typ.

Ekonometrickým nástrojom na modelovanie súvisiacich systémov je *seemingly unrelated regression*, resp. Zellnerov odhad. Je to vlastne zovšeobecnená metóda najmenších štvorcov, kde predpokládame, že vektory reziduí jednotlivých rovníc môžu byť závislé od reziduí ďalších rovníc v systéme. Takéto uvažovanie je opodstatnené, pretože dopytové rovnice navzájom súvisia. Sú odvodené z rovnakej nákladovej funkcie a majú rovnaké endogénne premenné. Najdôležiteším argumentom je skutočnosť, že nákladová funkcia musí byť homogénna prvého stupňa a jej prvá derivácia podľa vektora cien musí byť homogénna stupňa nula. Dopytové funkcie sú ešte navyše uvalené touto reštrikciou.

Ako som už spomínala nákladová funkcia, teda aj produkčná spolu súvisia a môžu nadobúdať rozličnú funkcionálnu formu. V predkladanej práci sú dáta modelované pomocou Translogovanej nákladovej funkcie, ktorá je považovaná za najflexibilnejšiu funkcionálnu formu. Avšak pre niektoré jej paramtre platí , že musia byť symetrické. Pre odvodený systém dopytových funkcií pravidlá symetrie tiež platia. Sú však umiestnené medzi rovnicami systému (cross restrictions), čo tiež podporuje myšlienku odhadu systému ako celku.

Predkladaná práca je rozčlenená na úvodnú kapitolu, 3 jadrové kapitoly a záverečnú kapitolu. Druhá a tretia kapitola obsahujú niektoré teoretické poznatky z mikroekonómie a ekonometrických metód. V kapitole o mikroekonomickej analýze som sa zamerala hlavne na poznatky o produkčnej a nákladovej funkcii, dualitnom vzťahu medzi nimi a na záver je odvodená translogovaná nákladová funkcia, lebo je hlavným predmetom tejto práce. Kapitola o ekonometrických metódach poskytuje základné informácie o odhadoch, testoch a iných javoch bohato využívaných vo štvrtej kapitole, kde je aplikácia tohoto prístupu ukázaná na konkrétnych dátach. Dáta sú modelované dvoma rozličnými spôsobmi. Prvý spôsob predpokladá ideálne správanie sa výrobcu, t.j. že model presne odzrkadľuje výrobcov papiera. V druhom navyše predpokládame, že vektor chýb opisuje AR(1) proces. Na záver je porovnanie týchto dvoch modelov. Odhady týchto modelov boli urobené pomocou dvoch ekonometrických softvérov: Eviews a Limdep 7.0. Záverečná kapitola uzatvára celú prácu a naznačuje ďalšie možné smery v modelovaní správania výrobcu.

Bibliography

[1] Berndt E.R., *The practise of econometrics, Classic and contemporary*, Ch.9, Addison-Wesley Publishing Comp., Inc., 1991

[2] Berndt E.R., Savin N.E., *Autoregressive disturbances*, Econometrica, Vol.43, 5-6, p.937-957, Sept./Nov. 1975

[3] Böhm B., *Ökonometrische modelle-Behelf zur Vorlesung*, TU-Wien, 1998

[4] Diewert W.E., *Applications of duality theory*, Handbook of Mathematical Economics, Vol.2, p.106-171, North Holland, Amsterdam, 1974

[5] Haider F., *Diplomarbeit "Energiekosteneffekte in der österreichischen Papierindustrie"*, unter der Anleitung von Univ.Doz.Dr.Bernhard Böhm, TU-Wien, 1996

[6] Johnston J., DiNardo J.: *Ecometric Methods*, Fourth edition, McGraw Hill, 1997

[7] Jorgenson D.W., *Econometric methods for modelling producer behavior*, Handbook of Econometrics, Vol.3, p.1842-1915, Elsevier Science Publishers BV, 1986

[8] Limdep 7.0 - help manual

[9] Partl, H.; Schlegl E.; Hyna I., LaTeX-*Kurzbeschreibung*, EDV-Zentrum der TU, 1990

[10] Varian H.R., *Microeconomic analysis*, third edition, W.W.Norton & Comp., 1992

Appendix A

1. Complete Data Set
2. Residual Graphs
3. Recursive Estimates of the Share Equations
4. Recursive Residuals of the Share Equations