

A Behavioral Model of the German Compound Feed Industry

Functional Form, Flexibility, and Regularity

Fritz Feger

<http://webdoc.sub.gwdg.de/diss/2000/feger/>

1 INTRODUCTION.....	1
1.1 HISTORICAL OVERVIEW	1
1.1.1 Duality	2
1.1.2 Structural Assumptions	3
1.1.3 Flexible Functional Forms.....	3
1.1.4 The Behavior of the Compound Feed Industry.....	5
1.2 PRELIMINARY PROGRAMMATIC REMARKS	6
1.2.1 Part I: Behavioral Model-Building.....	6
1.2.2 Part II: Theory of Flexible Functional Forms.....	8
1.3 SUMMARY OF THE RESULTS.....	9
1.3.1 A Behavioral Model of the German Compound Feed Industry	9
1.3.2 Functional Forms, Flexibility, and Regularity	10

Part I: A Behavioral Model of the German Compound Feed Industry

2 DATA.....	13
2.1 COMPOUND FEED AND FEED COMPONENT DATA.....	13
2.1.1 Quantity Data.....	14
2.1.2 Price Data.....	16
2.1.3 Observation Frequency and Aggregation Compatibility	17
2.2 NON-COMPONENT DATA.....	18
3 A COST FUNCTION MODEL OF THE COMPOUND FEED FIRM.....	20
3.1 GENERAL PREMISES OF THE CONSTRUCTION OF DUAL FEED MODELS	21
3.2 DESIRED SPECIFICATION	23
3.2.1 The Component Cost Function.....	23
3.1.2 The Allocation of Components on Compound Feeds.....	27
3.1.3 Technical Progress.....	28
3.3 STRUCTURAL ASSUMPTIONS.....	29
3.3.1 Weak Separability.....	29
3.3.2 Strong Separability.....	33
3.3.3 Unity Value of the CES Substitution Parameter.....	35
3.1.4 Equal Non-Component Cost – Total Cost Proportions.....	36
3.1.5 Constant Returns to Scale in Component Quantities.....	37
3.1.6 Nonjointness in Component Quantities.....	37
3.1.7 Homothetic Feed Composition.....	38

3.4 RECONSIDERING THE MODEL SPECIFICATION	39
3.4.1 Separability	39
3.4.2 Nonjointness, Homotheticity, and Constant Returns to Scale	47
4 A PROFIT FUNCTION MODEL OF THE COMPOUND FEED FIRM	50
4.1 TURNING FROM A COST TO A PROFIT MODEL	50
4.1.1 Relative Advantages of Dual Feed Cost and Profit Models	50
4.1.2 Differences Suggesting a Successful Profit Model	52
4.2 STRUCTURAL ASSUMPTIONS	53
4.1.1 Weak Separability	54
4.1.2 Strong Separability	56
4.1.3 Unity Value of the CES Substitution Parameter	56
4.1.4 Nonjointness, Homotheticity, and Constant Returns to Scale	57
4.3 MODEL SPECIFICATION	57
4.3.1 Separability	57
4.3.2 Nonjointness, Homotheticity, and Constant Returns to Scale	62
4.4 TOWARDS AN APPLICATION	64
4.4.1 The Epistemological Status of Structural Assumptions	65
4.4.2 Judging the Developed Feed Cost and Profit Model Pragmatically	68
4.4.3 Outlook	70

Part II: Flexible Functional Forms

5 THEORY AND SYSTEMATICS OF FLEXIBLE FUNCTIONAL FORMS	73
5.1 THE CONCEPT OF A FLEXIBLE FUNCTIONAL FORM	75
5.1.1 Definition	75
5.1.2 Flexible Functional Forms in Dual Models	78
5.1.3 The Local Approximation View	80
5.1.4 Semi-flexibility, Third Order Flexibility, and Global Flexibility	86
5.2 THE GENERAL LINEAR FORM AND ITS DERIVATES	90
5.2.1 Special Cases of the General Linear Form	92
5.1.2 Systematics	94
6 PROPERTIES OF FLEXIBLE FUNCTIONAL FORMS	98
6.1 DOMAIN OF APPLICABILITY AND COMPUTATIONAL FACILITY	98

6.2 REGULARITY OF THE GENERAL LINEAR FORM AND ITS DERIVATES	102
6.2.1 Homogeneity	103
6.1.2 Curvature	105
6.1.3 Monotonicity	109
6.1.4 Relative Importance of Regularity Conditions	112
6.3 TRADE-OFF BETWEEN SELECTION CRITERIA AND CONCLUSION.....	114
7 APPLYING GLOBAL FLEXIBILITY.....	120
7.1 SELECTION OF FUNCTIONAL FORM	120
7.1.1 Generalized Leontief.....	121
7.1.2 Trancendental Logarithmic (Translog).....	123
7.1.3 Fourier.....	127
7.1.4 Symmetric Generalized McFadden.....	127
7.1.5 Asymptotically Ideal Production Model (AIM).....	129
7.2 OUTLOOK	131
REFERENCES.....	134

Es gibt Menschen, welche Tatsachen leugnen und das Denken nennen. [...] Dann gibt es Menschen, welche die Schuld unserer Rationalität geben und verlangen, daß wir weniger rational sein sollen. [...] [Erstere] übersehen [...], daß die Tatsachenmenschen, [...] der Ingenieur und der Handelsmann, gar nicht so sehr rational im Sinn einer Präponderanz des Intellektuellen sind, wie sie meinen. [...] An sich ist die Tatsache ja überhaupt nicht rational, sie ist nur ein Regulativ der Rationalität und gewöhnlich nur als Serie für diese wichtig. Und der Ingenieur unterscheidet sich oft gerade dadurch vom Theoretiker, daß er das Denken an irgend einem Punkt abbrechen läßt und seine Konstruktion mit einer Annahme, einem Näherungswert, einem Abkürzungsverfahren, das ein Sprung durch das Unbeweisbare ist, macht, und dies sich durch den Erfolg bestätigen läßt.

Der deutsche Mensch als Symptom (Robert Musil)

Preface

The study in hand has been written at the Institute of Agricultural Economics (IfA) at Göttingen University and financially supported by the IfA and the Deutsche Forschungsgesellschaft (DFG), which I thankfully acknowledge. I would like to express my gratitude to Stefan Tangermann, my „Doktorvater“. He has always been ready for a constructive discussion and has deeply impressed me with his high standards. I am very much indebted to Detlef Müller who has been a competent and supportive companion and, regarding our common publication even more than this study, has helped to bring up a truly interdisciplinary piece of work. Furthermore, Stephan Brosig, Harald Grethe and Thomas Heckelei (both in alphabetical and chronological order) have contributed invaluable by repeatedly reviewing my trials. In particular, the uncomfortably patient questions of Harald Grethe can be tracked on almost every page of this work. This study has also benefitted from comments by Klaus Frohberg and Michael Grings who refereed excerpts, and from comments by the participants of the Doktorandenkolloquium at IfA. I also would like to thank my parents, which unambiguously supported my education and are responsible for my consideration of agriculture. A role which cannot be underestimated has been played by my wife Sabine Döring. In addition to her moral support, she has made me work and, more than anyone else, has coined my view of science.

Essen, spring 2000.

Fritz Feger

1 Introduction

The study in hand deals with the applied modeling of an industry, namely with the construction of a behavioral model of the German compound feed industry and its empirical application. Commonly, such a project includes a presentation of the favored model and its motivation and benefits, a justification of theoretical premises and empirical assumptions presupposed by the model, the choice of a functional specification for the model, an introduction of the estimation method, a description of the utilized statistical data, a presentation of the estimation results, a comment on their statistical significance, and, eventually, the derivation of behavioral parameters and their interpretation, sometimes connected with policy recommendations. Motivation and benefits of an applied model of the compound feed industry are obvious: on the one hand, it allows to forecast the economic behavior of compound feed manufacturers which is useful for policy evaluation and, on the other hand, it constitutes, as any applied model, a possibility to test theories and methods. However, this study is discontinued at an early stage because of the failure of the project to establish a simultaneously consistent and estimable model: it is not possible to construct a dual behavioral model of the compound feed industry which is consistent with economic theory and evident structural assumptions on the one hand and estimable with the available statistical data on the other hand. Nevertheless, after an extensive discussion of this result in the first part of the study, a second part on the theory of flexible functional forms is added. Although considerations on the theory of flexible functional forms are superfluous with regard to the original goal of the study because of the unsuccessful modeling attempt, results on flexible functional forms which were found before it was clear that the modeling would fail will be presented, since these are of significance for any applied economic model: it is shown that the local flexibility concept is obsolete in a globally flexible estimation, and that therefore no estimation result must be rejected because it lacks local flexibility. As a consequence, other important goals like maximal statistical significance and a possibly large region where all theoretical consistency properties apply can be strived for more successfully.

1.1 Historical Overview

At first, a short historical overview will be given which proceeds from the general to the particular, i.e. from developments in economic theory over methodical progress in economics to models of the compound feed sector.

1.1.1 Duality

One of the most significant developments in neoclassical economic theory is the duality concept. From the outset it has been accompanied by the endeavor to create functional forms that are both capable of depicting all economically relevant effects and satisfying the requirements of economic theory.

The duality concept, based on the works of R. Roy and Harold Hotelling, was first formulated as a closed formal theory by Ronald W. Shephard in 1953.¹ Duality theory has been established as a framework both for theoretical considerations and applied model building since then, with contributions from numerous authors.² The basic idea of duality theory is that it is possible – and, for several reasons, advantageous³ – to formulate a dual cost function in terms of input prices and output quantities⁴ or a dual profit function in terms of input and output prices which is equivalent to the primal representation of a technology in terms of input and output quantities.⁵ As the equivalency relationship suggests, the postulated properties of a well-behaved technology correspond to properties of a well-behaved dual cost or profit function, so that it suffices to choose a functional form for a dual cost or profit function which shows these properties to guarantee the existence of a corresponding well-behaved technology.⁶ From a well-behaved cost or profit function there can be derived a complete system of demand and marginal cost or demand and supply functions which inherits theoretical consistency from the master function and thus satisfies the requirements postulated in section 1.1.1 above.

¹ See ROY 1942, HOTELLING 1932, and SHEPHARD 1953/1981.

² See in particular DIEWERT 1982.

³ For arguments for the superiority of the dual approach see GRINGS 1985: 20; BINSWANGER 1977: 377; DIEWERT 1971: 107; and CHAMBERS 1988: 36.

⁴ In consumer demand theory it would be income instead of output quantity.

⁵ More precisely, the equivalence relationship is restricted to the case of a well-behaved technology. If the technology has ill-behaved pieces, these are substituted by a hyperplane. Thus, the duality transformation yields a consistent approximation of the technology.

⁶ The same is valid for systems of demand or supply equations derived from well-defined cost or profit functions.

1.1.2 Structural Assumptions

To empirically apply duality theory, separability assumptions are decisive because they allow the viewing of the behavior of economic agents as a multistage optimization process, where the single stages can be modeled more or less independently on each other. One should think of a separable optimization process of e.g. a cost function as, at a first stage, minimizing the cost of producing one unit intermediate output out of a subset of the inputs. Then, at a second stage, the intermediate outputs are interpreted as aggregate inputs so that the costs of producing one unit aggregate input can be interpreted as marginal costs and consequently as the price of the aggregate input. These input aggregates are processed to the ultimate output in a cost minimizing way, where the price and the composition of the input aggregates are taken as given. The concept of separability was coined independently by Wassilij Leontief and M. Sono.⁷ The transfer of the separability concept to the relation between outputs is due to Robert E. Hall and Ulrich Kohli, who introduced the concept of nonjoint production.⁸ Nonjoint production is present if the quantities of other outputs do not appear in the production function of one output, i.e. there exist neither synergies nor negative external effects between outputs.

1.1.3 Flexible Functional Forms

The insights given by duality theory have considerable implications on empirical analysis, since the application of a microeconomic model to real world data requires the assumption of a concrete functional form for a cost or profit function. Most "classical" functional forms, like the Cobb-Douglas or the CES functional form, and consequently their derived demand and supply systems, exhibit all the properties of a well-behaved cost or profit function but suffer from limited flexibility, i.e. have too few parameters to make possible an independent representation of all economically relevant effects. As an example, the Cobb-Douglas cost function, as well as the Cobb-Douglas production function, restricts the substitution elasticities between all factors and the elasticity of scale and size to unity. In their pioneering works, W. Erwin Diewert on the one hand and Laurits R. Christensen, Dale W. Jorgenson and Lawrence J. Lau on the other hand developed functional forms which have enough parameters to allow each second order effect, e.g. in essence each substitution elasticity, to assume each possible value independent of other parameters: the Transcendental Logarithmic or Translog and the

⁷ See LEONTIEF 1974; SONO 1945; also see BLACKORBY/PRIMONT/RUSSELL 1978.

⁸ See KOHLI 1981; also see HALL 1973.

Generalized Leontief cost function, respectively.⁹ Both forms are still most commonly used despite the fact that neither the Translog nor the Generalized Leontief can be restricted to entertain all properties of a well-behaved cost function without losing its flexibility. However, even the degree of flexibility provided for by the Translog, the Generalized Leontief, and similar forms, is criticized for being not great enough: these functional forms are merely second order flexible, i.e. allow the level, gradient and Hessian values to attain arbitrary values at merely one point, only restricted by economic theory. A more powerful concept of flexibility, namely global flexibility, advocates an approximation of the data generating process at all points simultaneously. This concept was introduced and operationalized by A. Ronald Gallant using a Fourier series expansion and establishing semi-nonparametric estimation techniques, i.e. an estimation where the number of parameters is not fixed a priori but rather determined during the estimation process relying on statistical criteria.¹⁰

An important step on the way towards flexible functional forms behaving regularly in the whole domain is Lau's development of a technique to impose a concavity or convexity constraint in terms of the Hessian matrix of the cost or profit function: he shows that the substitution of the Hessian by its Cholesky decomposition creates the opportunity to guarantee the local concavity or convexity of the original function by restricting the magnitude of the Cholesky values.¹¹ Following up, Diewert, jointly with Terence Wales, proposes a Symmetric Generalized McFadden cost function, which is, as indicated by its name, a generalization of a functional form suggested by Daniel McFadden, who for his part plays an important role in the debate about duality.¹² The Symmetric Generalized McFadden has the property that a restriction for correct curvature using Lau's technique implies global rather than local regularity – with respect to curvature: the monotonicity property can still not be accounted for globally. In addition, this curvature constraint restricts the local flexibility property of the Symmetric Generalized McFadden to one point. The reason that this is possibly the maximum that can be achieved with regard to the project of constructing a second order flexible functional form which is globally regular is stated by Lau's incompatibility theorem: he proves that, at least for linear-in-parameter forms, it is not possible to reach second order flexibility and global

⁹ See DIEWERT 1971; see CHRISTENSEN/JORGENSON/LAU 1973.

¹⁰ See GALLANT 1981 and GALLANT 1982.

¹¹ See LAU 1978.

¹² See DIEWERT/WALES 1987a; see MCFADDEN 1978.

regularity simultaneously.¹³ A way out is shown by Gallant and Gene H. Golub, who propose a procedure to enforce regularity at least for a definite region.¹⁴ Finally, this method is operationalized by William A. Barnett, John Geweke and Michael Wolfe who present a Bayesian technique which allows the performing of a semi-nonparametrical estimation with inequality constraints. Moreover, they apply this estimation technique to a new functional form which is well-suited for economic applications because it is, in contrast to the Fourier series, easy to handle and close to a theoretically consistent behavioral function by construction: the Asymptotically Ideal Production Model (AIM).¹⁵

1.1.4 The Behavior of the Compound Feed Industry

Among the studies that perform system estimation for the compound feed sector, the work of Ludo Peeters and Yves Surry is especially worth mentioning because, firstly, it is the only one to tackle the theoretical foundation of the neglect of the demands for non-component inputs, which all other studies also omit to mention and, secondly, a locally flexible functional form is applied that satisfies more requirements of economic theory than any other.¹⁶ Peeters and Surry use a simultaneously separable and nonjoint dual cost function: they claim that the separability assumption allows the neglecting of the prices of non-component inputs whereas the nonjointness assumption makes it possible to compute how total component demand quantities are distributed on the animal-specific compound feeds. Peeters and Surry estimate the economic model applying their multi-output extension of the Symmetric Generalized McFadden cost function, the concavity restriction implemented by Lau's technique. For an overview of studies dealing with the estimation of feed demand elasticities see their 1997 article "A Review of the Arts of Estimating Price-Responsiveness of Feed Demand in the European Union".¹⁷ All previous studies on the demand for feed components suffer from quite rough input aggregates caused by the shortness of the available time series and the resulting shortness of degrees of freedom. The multiplicity of feed components is reduced to at best five

¹³ See LAU 1986: 1552-1557.

¹⁴ See GALLANT/GOLUB 1984.

¹⁵ See BARNETT/GEWEKE/WOLFE 1991; also see TERRELL 1995.

¹⁶ See PEETERS/SURRY 1993; also see BOYD/BROSEN 1986, MERGOS/YOTOPOULOS 1988, and SURRY/MOSCHINI 1984.

¹⁷ See PEETERS/SURRY 1997.

aggregates like "grains" and "grain substitutes", so that questions of interest for policy evaluation such as for example the magnitude of substitutional relationships between single products, e.g. between soy beans and corn gluten, remain unanswered.

1.2 Preliminary Programmatic Remarks

This book has two parts: the first part deals with the construction of a theoretically consistent dual behavioral model of the German compound feed sector, emphasizing separability assumptions. This is suggested by the asymmetric data availability situation which requires the neglect of non-component price and quantity data to allow an estimation of a system of derived behavioral equations. However, the analysis yields that, in the current state of affairs, a consistent estimation is impossible. In the second part, an overview of the theory of flexible functional forms is given, where a new interpretation of the global flexibility concept is proposed. This part is added in spite of the failure in model building: if no both consistent and estimable model can be established, an estimation cannot be performed, and thus, it is superfluous to consider functional specifications. But the analysis of the theory of flexible forms, which was already undertaken before it became obvious that it would not be finally utilizable in this study, yields a methodical progress which is possibly important for applied economics in general because it implies a change in the standard specification and estimation procedure for applied microeconomic models.

Since the methodical approach pursued throughout this study is somewhat different from both other theoretical treatises and empirical applications in economics, some introductory remarks seem to be appropriate.

1.2.1 Part I: Behavioral Model-Building

The strict conclusion drawn in this study that an estimation which is consistent with economic theory and evident structural assumptions is impossible can generally be objected, as in any other study with the same result: it is always possible to lower theoretical demands until empirical results can be obtained. Methods reach from maintenance of theoretical premises and structural assumptions which are only approximatively true to maintenance of premises and assumptions which are obviously wrong but postulated to be of minor significance for the sake of obtaining results at all – or problematic premises or assumptions are simply concealed. An approach to applied economic models which lowers the theoretic demands for an applied model as far as required to obtain the desired results whatever theoretical cost occur may

be labelled "instrumental": theory is nothing more than an instrument to second the obtainance of results. The central argument for this approach is that any result is better than no result, since one has in any case to deal with the phenomena rather than assuming an indifferent position towards – possibly essential – factual questions. This argument is supported by the observation that absolute intolerance of theoretical inconsistency renders the establishment of empirical hypotheses impossible in virtually any case.

The opposite approach may be labelled "axiomatic" because it claims that any empirical hypothesis which could adequately answer a factual question must be consistently deduced from an universal scientific theory, i.e. all the way up from the axioms, only supplemented by structural assumptions which are scientifically proven elsewhere. The main arguments are the following: first, if strict theoretical consistency is given up, it is not granted that implementing some theoretical properties will yield a better prediction, compared to completely ignoring theory. A successful prediction without theory or with inconsistent theory may not be used as a piece of evidence for an instrumental approach since the reasons for the success is unclear: following the axiomatic approach, theory is insignificant if not accounted for as a whole. Secondly, theory cannot be tested at all with inconsistent empirical hypotheses, so that the question if a theory is superior to another, or empirically significant at all, cannot be answered; scientific progress loses one of its fundamentals.

All science aims at prediction, and thus, even the axiomatic approach is eventually instrumental. The difference merely consists in the idea how to gain empirical knowledge, or in the relative importance which is assigned to prediction quantity on the one hand and prediction reliability on the other hand. As the arguments show, both approaches are equally essential since none of them alone is capable of yielding everything that must be expected from science – followed uncompromisingly, both are equally useless. Thus, in scientific practice, most studies are located somewhere between an instrumental and an axiomatic approach to applied modeling. More fundamental scientific disciplines as theoretic physics tend to the axiomatic approach. In applied economics, however, the instrumental approach prevails. With regard to the important functions of both, it seems fruitful to add a more axiomatic counterpart to the current state of research.

The approach of economic model building followed here, which is claimed to be exemplary for a tendentially axiomatic approach, can be interpreted as an anticipation of the deductivist or falsificationist examination of a scientific theory on an argumentational level. According to the falsificationist theory of a scientific model, a needs-orientated but otherwise unfounded model is invented and then confronted with empirical data in order to reject it – of

course nourished by the hope that all rejection trials will fail.¹⁸ In the present model of the compound feed firm, as in microeconomic models in general, the falsificationist postulate-reject/accept algorithm is pre-iterated: before exposing the model to "hard" empirical data with statistical means, plausibility considerations are undertaken which are not a priori but rather based on "soft" empirical knowledge – general certainties extracted from all kinds of sources like former scientific studies and trial-and-error experiences with the object of the theory. These plausibility considerations lead to the maintenance (or rejection) of assumptions on model structure which are supposed to justify the model specification. This amounts to a justification of neglecting variables or possible relations between these. The epistemological consequences of this paradigmatic method are discussed in section 4.4.1 below.

The following chapter presents the statistical data that will be analyzed by the model. Usually located after the exposition of the behavioral model and the estimation procedure, the presentation of the availability situation for estimation data is put first because the model building chapters explicitly draw upon data availability – just as any researcher actually does when generating her or his model. In any case, a posterior presentation of the data would be inconsistent since it finally becomes evident that a consistently estimable model cannot be established. The order of material in chapter 3 mirrors the falsificationist procedure: a desired cost model specification is proposed in section 3.2, where the desirability corresponds to the data availability, and then all structural assumptions relevant for the postulated specifications are examined in sections 3.3. Finally, in section 3.4, there is a check about how far one can get in justifying the desired specification with the maintained structural assumptions. This procedure is repeated for the profit function in chapter 4. Since the maintainable assumptions do not finally suffice to establish a consistently estimable model, although the model specification tries to incorporate the minimal restrictiveness compatible with the goal of the analysis and the available data, section 4.3 includes considerations about the implications of model failure and the error made if an inconsistent model is estimated.

1.2.2 Part II: Theory of Flexible Functional Forms

In the second part of this study, a conceptual analysis of the theory of flexible functional forms is conducted. A conceptual analysis consists in the exhaustive explication of all possible meanings of a given concept and an examination of its relations to other concepts inside the conceptual system, where, finally, the most powerful consistent notion is defined. Conducting

¹⁸ See POPPER 1961: 40-41.

a conceptual analysis with all concepts of a scientific conceptual (sub-)system yields a closed set of distinct concepts explaining everything the respective theory claims to explain with a minimal number of concepts. This procedure, as in the present case, can well imply the complete rejection of a concept. The method of conceptual analysis is only rudimentarily common in economics, and one might think of a textbook when reading chapters 5 and 6. But, where a textbook merely uses conceptual analysis to enhance understanding, i.e. as a didactic tool, this study is more ambitious. Apart from undeniably trying to present existing knowledge in an easily accessible way, it aims at clarification of concepts in order to examine their consistency and significance. Given methods and widely used concepts are evaluated with the result that one central concept of the theory of flexible functional forms, the concept of local flexibility, is shown to run empty in the modern global flexibility framework. A summary of this result is presented as thesis in section 1.3.2 below.

The rejection of the local flexibility concept has important methodical implications,¹⁹ and thus the putatively surprising result occurs that conceptual analysis yields methodical progress. This should be interpreted as an invitation to promote conceptual analysis in economics.

To avoid confusion of the argumentation for the superfluency of the local flexibility criterion – which consistently occurs in chapters 5 and 6 – with the exposition of the theory of flexible functional forms and the deduction of the favoured method, the former is presented in the form of a condensed thesis in section 1.3.2 below, while the order of material in chapters 5 through 7 follows the logic of the latter.

1.3 Summary of the Results

1.3.1 A Behavioral Model of the German Compound Feed Industry

What is frustrating for the applied researcher may be a thrilling result from a theoretical perspective, namely the failure of the enterprise to establish a consistent model of an industry which is estimable given a certain data availability situation. And this is exactly what happens in this study: three cost modeling approaches and two profit function accounts fail to be both consistent with economic theory and maintained structural assumptions and estimable with regard to the unavailability of sufficient non-component price and quantity data.

¹⁹ See section 1.3.2 below and chapters 5 and 6.

Reflecting the practice of all other available studies on the compound feed sector, a desired cost model is presented which allows for complete neglect of non-component data but is immediately shown to be much more restrictive than the maintained separability assumptions imply. An estimation in spite of this renders depiction of existing effects impossible and is thus inadequate both with regard to a test of the empirical hypothesis expressed by the model and forecast purposes. Then, two cost models are proposed which account for relationships where no separability can be reasonably claimed. The first alternative assumes weak separability of non-component demand shares from output quantities, which is arguable. Nevertheless, this neither leads to a system of estimable demand and marginal cost equations nor to a system of estimable demand and marginal cost share equations because thus maintained structural assumptions in conjunction imply constant returns to scale for the overall cost function, which renders measurement of marginal cost and marginal cost shares by the product price impossible. As second alternative, a cost structure which, more realistically, accounts for an inseparability of the non-component sector from output quantities, is proposed. While this specification is inarguably consistent, it is still not estimable because both the derived demand and marginal cost functions and the respective share equations are not independent on non-component prices, which are not available. With the profit function approach, the problem is more general: both alternatives, which are analogous to the two cost function approaches, are not defined since, in the former case, the whole profit function or, in the latter case, the micro-profit function corresponding to the component combination to compound feed, respectively, exhibit constant returns to scale so that the profit maximization problem is unbounded. Thus, either progress in consistently modeling the profit maximizing behavior of the compound feed industry has to be achieved, or, given that the non-component data availability situation does not improve in future, pragmatic cost function approach solutions relying on generated non-component data or instrumental variables for total non-component cost data must be favored.

1.3.2 Functional Forms, Flexibility, and Regularity

The three chapters on flexible functional forms, chapters 5, 6, and 7, include an extensive presentation of the theory of flexible functional forms and of commonly used specifications, i.e. cover a wide range of known results. Apart from some remarks on these and some epistemological conclusions, there is a twofold central result which is new: by means of a conceptual analysis²⁰ it is shown that, first, requiring a functional form to be locally flexible is obso-

²⁰ See section 1.2.2 above.

lete if it is estimated semi-nonparametrically, i.e. as a globally flexible functional form with a large depth of parameterization determined by the inferential potential of the data, aiming at good approximation of the true data-generating process in its whole course. Secondly, it is shown that this relaxation, coming into effect in a semi-nonparametric estimation function and using a Bayesian estimation technique with an informative prior, allows one to take account of the theoretical consistency requirement much more successfully than within the local flexibility framework.

The neglectability of the local flexibility criterion in a globally flexible framework follows from two arguments in conjunction: First, the local approximation view of flexible functional forms runs empty because there is no way to locate the point of approximation in a statistical context. The analogy to the approximation of a known algebraic function is not available if the approximated structure is merely present through a random sample to which the function is fitted in its whole course. Secondly, if it is possible to find a representation of economic behavior that goes beyond the limited complexity of a local second order approximation, insertion of variable values into this representation is by far superior with regard to forecast precision. Level, first and second order effects only restricted by economic theory by no means provide an exhaustive characterization of a technology, as an axiom of the theory of locally flexible functional forms claims. It must be concluded that local flexibility is irrelevant if the approximation function fits the true function as well as possible at any data point rather than locally.

If it is, according to Lau's incompatibility theorem, impossible to make a locally flexible functional form intrinsically consistent with all postulates of economic theory, i.e. a second order flexible functional form approximating function value and first and second order derivatives of the underlying function at merely one point, this is all the more so with a globally flexible functional form, i.e. with a more complex specification capable of approximating the underlying function as a whole. But a Bayesian estimation technique with an informative prior recently proposed by Barnett, Geweke, and Wolfe provides a solution: drawing samples from the whole sample, estimation results being theoretically inconsistent inside the defined regular region are rejected while all consistent parameter estimates are gathered in a theoretically consistent posterior distribution from which the final parameter estimates are derived. This is, a set of parameters is estimated that incorporates a regular region in the course of the function, independently of the estimated functional form being intrinsically regular.

In contrast to all studies applying this Bayesian technique with an informative prior which lament that enlarging the regular region of a globally flexible functional form severely harms

the local flexibility property and follow that estimation results must therefore be rejected, this study yields the following: a globally flexible parameter estimation with a Bayesian estimation technique with an informative prior yields a parameter set subject to a regular region where any violation of the local flexibility condition is caused by the data alone. The only presupposition for this is that the functional form which is used as estimation function (as opposed to the concrete parameterization subject to the implemented regular region obtained in the estimation) satisfies the local flexibility property, i.e. only those consistency properties are incorporated by construction or by parametric restrictions that do not a priori restrict the ability of the function to depict any theoretically consistent behavior locally. Otherwise, the possibility to depict an arbitrary economic behavior would not only be restricted by economic theory, i.e. by the regular region, but in addition by the intrinsic incapability of the used functional form.

This implies that, in a proper globally flexible estimation, the extension of the regular region is a decision exclusively based on the trade-off between an unlimited forecast region and the best available fit. The increasing local flexibility violations with a growing regular region are irrelevant again, and thus, accounting for maintained theory is freed from the burden of accounting for local flexibility simultaneously – Lau's incompatibility theorem is no longer effective.

Part I: A Behavioral Model of the German Compound Feed Industry

2 Data

Since the reason for the ultimate failure of all modeling efforts undertaken in this study lies in an insufficient data availability, the data will be presented prior to the models of chapters 3 and 4. As an estimation is impossible in the current state of affairs, considerations are discontinued after a description of the asymmetric data availability situation to preserve the consistency of the study: on the one hand, there is quite good data on feed components and compound feed, which will be presented to emphasize how regrettable the inavailability of other essential data is and, on the other hand, the desired but unavailable data is discussed. This is, the questions of how to appropriately edit the data, which aggregation technique to favor, which commodities to aggregate due to which criteria, et c., that is the questions normally completing an exposition of the data material, are not treated.

2.1 Compound Feed and Feed Component Data

The different livestock categories in the German livestock sector are characterized by a different share of self-produced feed. While pig production and in particular poultry production heavily rely on industrially processed feed, the dairy and beef sectors are dominated by feed produced on-farm. However, a consideration of on-farm mixers seems to be unreasonable for three reasons: first, this would complicate theoretical justification of any feed component demand model because self-produced feeds are highly entangled with all other farm activities; two separate behavioral models for industrially supplied compound feed and for on-farm mixers would be required. Secondly, a large portion of self-produced feed, like fodder, corn cob mix, et c., is hardly traded so that a significant market, i.e. significant prices usable as exogenous model variables, does not exist. Following from the large absence of a market for feed produced on-farm, the few available data is of far inferior quality than the data on the compound feed industry. Thirdly, for the remaining easily tradeable feed components produced on-farm, namely grains, feed utilization follows the same ratio as in the compound feed industry, and thus the error made by transferring evidence gained in the compound feed industry on on-farm mixers can be estimated as small. Thus, this study is limited to the com-

pound feed industry's behavior. In the remainder of this section, the data on demand and supply quantities of the German compound feed industry and on component and compound feed prices will be described.

2.1.1 Quantity Data

Data on the produced quantities of compound feeds and the processed quantities of feed components in Germany are published at numerous sites.²¹ All of these publications draw upon a common source, namely the data collected by the governments of the German "Bundesländer" due to the "Getreide-Meldeverordnung" of June 23th 1976. Legal basis of the "Getreide-Meldeverordnung" is the "Gesetz über die Neuorganisation der Marktordnungsstellen" of June 23th 1976, the "Verordnung über Meldepflichten der Getreide-, Stärke- und Futtermittelwirtschaft" of June 26th 1978, and some "Verordnung[en] zur Änderung der Getreide-Meldeverordnung".²² Compound feed producers with an annual production quantity of over 5,000 t compound feed are, due to § 2 section 1 of the "Getreide-Meldeverordnung", obliged to report monthly stocks, receipts and items being which are received and dispatched of the following feeds: wheat, rye, barley, oats, maize, triticale (the latter since July 1st 1991), feed peas, feed beans, other pulses, unmilled rape seed (the latter since July 1st 1991), sun seed (since July 1st 1991), oil cakes, expellers, extraction meal of soy beans, rape seeds, and others, respectively, corn gluten feed, milling by-products, manioc products, fish, meat, and blood meal, citrus and fruit trester (the latter since July 1st 1991), and molasses and sugar beet shred (since July 1st 1991). In addition, produced quantities of compound feed for horses, calves, cattle, pigs, broilers and layers have to be reported. The quantity of purchased additives are explicitly excluded from the reporting obligation, where the term "additive" is a subsumptive concept for all kinds of feed components of non-agricultural origin including minerals, vitamins, preservatives, colorizers, stabilizers, antibiotica, industrially produced amino acids, et c.

The monthly quantity aggregates per "Bundesland" published in the "Statistischer Monatsbericht des BML" constitute a fantastic data set which is, with regard to proportions and quality, surely not often encountered for other economic sectors. Nevertheless, some shortcomings have to be noted: First, as additives gain more and more importance in compound feed pro-

²¹ See e.g. ZMP-Bilanz Getreide und Futtermittel, Mischfutter-Tabellarium, Statistische Monatsberichte des BML.

²² See Bundesgesetzblatt I: 1608, 883, 1408, 1711, and 794.

duction, the lack of the processed quantity of additives severely restricts the significance of any analysis based on this data set. In particular, consider the increasing significance of synthetic amino acids driving out agricultural high protein components. Secondly, the sum of utilized component quantities is smaller than the sum of compound feeds produced, and thus one can only guess about the composition of the difference, presumably apart from the additives including components which correspond to none of the aggregates, shrinkage, and other losses. A solution which accounts for the fact that new components constantly enter the recipes and others are dropped consists in the introduction of residual aggregates, e.g. "other natural components with less than 25% of protein" and "other natural components with more than 25% of protein". Thirdly, not all reported data is in any case published. On the one hand, all available data on the completely insignificant components feed beans, rape seed and sun seed is included, whereas on the other hand the important components oilcakes, expellers, and meals of rape seed, soy, and other oil seeds, which are collected separately in three aggregates according to the respective plant category, are placed together in one huge aggregate. For example, in December 1994, not even 1,000 t of unmilled rape seed were processed, whereas the utilized quantities of oilcakes, expellers, and meals of soy (152,853 t), rape seed (100,363 t) and other oil seeds (98,967 t) exceeded the former by the factor one hundred and more.²³ Furthermore, the aggregate of oilcakes, expellers, and meals of other oil seeds includes the not even unimportant components palm and copra/cocos, which could possibly be reported separately too.

Some data which is not included in the "Getreide-Meldeverordnung" can be found in the monthly international trade statistics. However, this presupposes, first, that the respective component is not produced domestically at all, secondly, that total net imports are utilized as feed component, and in fact by the compound feed industry. These conditions are rarely satisfied, even approximatively. For certain palm and cocos products it can be assumed that the demand is nearly completely due to the compound feed industry, which can also be said of synthetic amino acids and some special antibiotica. With other additives like e.g. vitamins it is clear that a significant share of net imports flows into human nutrition. The only additive which is both included in trade statistics, not produced in Germany, and presumably demanded by the compound feed industry only, is the amino acid lysine.²⁴ Since the market for addi-

²³ See Statistische Monatsberichte des BML. For this study, the latter quantity aggregates are provided by unpublished data base prints of the German ministry of agriculture, fishery and forestry.

²⁴ See Außenhandel: Fachserie 7, Reihe 2.

tives is rather monopolistic and characterized by an extreme division of labor – some amino acids are supplied by not more than one producer in the world –, and since the producers are not at all obliged to report their production quantities, it is not possible to construct a representative data base for the use of additives in the compound feed industry.

2.1.2 Price Data

Information on component prices can be found as cif and fob quotations, stock market quotations, wholesale prices, and producer prices in a number of newspapers and economic and agricultural journals. Furthermore, they can be approximated as unit values from the international trade statistics. Unit values and cif and fob quotations exhibit the shortcoming of being available only for commodities which are internationally traded. In addition, it is questionable whether unit values and cif and fob prices of commodities, which international trade is insignificant compared to domestic trade, correctly reflect domestic prices. Agricultural producer prices exhibit the complementary shortcoming of existing only for commodities produced domestically. Moreover, they differ from the initial cost of raw materials of the compound feed enterprises in different profit margins. Purchasing prices are probably best approximated by wholesale prices and quotations at the stock markets, which also refer to wholesale prices. Since the products of which the wholesale prices are reported are a subset of the products quoted at stock markets, the latter are favored in this study.

The stock markets of Hamburg, Hannover, Brunswick, Dresden, Halle (Saale), Erfurt, Dortmund, Mannheim, and Stuttgart weekly quote prices of numerous agricultural products. The broadest range of products is covered at the Hamburg stock market: wheat, rye, barley, oats, wheat bran, rye bran, tapioka pellets, dry shred, soy meal, rape meal, palm expellers, corn gluten feed, and fish meal. Maize is regularly quoted at the Mannheim stock market only, but beyond that, all feed components quoted at all are quoted in Hamburg. The agribusiness newspaper "Ernährungsdienst" publishes all quotations of all German commodity stock markets, and the brochure "ZMP-Bilanz Getreide, Ölsaaten, Futtermittel" includes the arithmetic means of the weekly quotations of all the German stock markets. The latter can be seen as a sufficiently precise proxy for German domestic prices with regard to the requirements of this study: intertemporal variations of prices affect the estimation parameters, visible in the use of price indices as exogeneous variables, but interregional price level differences vanish in index aggregates in any case, which is appropriate under the assumption that spatial price patterns are due to transportation cost only so that prices vary approximatively proporti-

onally everywhere. Hence, the means published by the "Zentralstelle für Markt- und Preisberichterstattung (ZMP)" can be used.

There is no printed information on prices of any feed additives. A meaningful alternative to the prices of different additives would consist of the price of instant mineral feed for on-farm mixers which contain all additives demanded by the compound feed industry, though on average with an emphasis on minerals and vitamins rather than synthetic amino acids. But these prices are not published either. However, prices for most other compound feeds are published. The "ZMP" calculates average purchasing prices for farmers on a national level from the local price information printed in local agricultural journals. This monthly data is perfectly suited to the goals of this analysis and can thus be employed.

2.1.3 Observation Frequency and Aggregation Compatibility

It is obvious that monthly data is superior to annual data because of the drastically greater number of degrees of freedom. A possible objection is that some econometric problems increase with the use of monthly data, especially autocorrelation. This can be answered with the argument that the additional degrees of freedom overcompensate the econometric data problems, in particular as effective measures exist against most data problems. Furthermore, the possibility to concentrate on a shorter observation period, following from the rich degrees of freedom, allows the selection of a particularly well-suited starting point, e.g. a change in the "Getreide-Meldeverordnung", in order to minimize data inconsistency problems due to changing aggregate definitions. A focus on a shorter observation period additionally decreases the bias caused by technical change or structural breaks, which can of course be handled econometrically, too. But technical change does, first, not constitute the central interest followed in this study, and, secondly, biases the parameters of interest even if part of it is covered by explicit modeling – even a sophisticated modeling of technical progress and structural breaks will not be able to record it completely. Hence, the data is left in the monthly frequency at which it is observed.

Whereas quantity and price data exhibits the same frequency, another aspect of data compatibility is more problematic: while most reported prices correspond to equally defined quantity aggregates, namely for feed wheat, barley, rye, oats, maize, tapioka, soy meal, rape meal, corn gluten feed and fish meal, this is not always the case. In some cases, it can be presumed that a single-feed price is representative of a whole aggregate. It is, for example, likely that the price of wheat bran adequately describes the price evolution for all milling by-products, and the quantity aggregate of other oil meals, oil cakes, and expellers mainly consists of palm

and cocos products so that an arithmetical mean of the prices for these two feed components will yield a good proxy for the price of the former. The price of triticale is proportional to the feed wheat price, and thus the price of the latter can be assumed as price of triticale without further consideration. However, for feed peas, feed beans, rape seed, sun seed, other oilseeds, and citrus trester, no quotations are available. Whereas feed beans, rape seed, sun seed, other oilseeds, and possibly feed peas can be dropped without consideration for the minor quantitative importance of these, this is not possible with citrus trester. A solution is provided by the international trade statistics: citrus trester is generally imported, and it is exclusively demanded by the compound feed industry so that a price can be approximated by unit values calculated on the basis of traded quantities and values.

For the compound feed quantity aggregates according to the "Getreide-Meldeverordnung", no correspondingly aggregated price is published. Thus, because of the lack of more detailed information on the shares of different compound feeds inside such an aggregate, one compound feed which is believed to be especially significant must be selected to serve as representative for the whole aggregate.

2.2 Non-Component Data

It need not be emphasized that there does not exist an analogy with the "Getreide-Meldeverordnung" for other factors of production demanded by the compound feed industry. Of course, the exact quantities of labor, energy, industrial installations et c. are trade secrets, apart from the thrifty information which has to be published in the balances of joint-stock companies, which, however, most compound feed producers are. The utilization of data published in balances is made more difficult by three aspects: first, the data is heterogeneously composed in different enterprises and highly aggregated with respect to a meaningful use as model variables, i.e. of insufficient quality. Secondly, balances are published annually, but monthly data would be required, i.e. the data exhibits a frequency which is too low. Thirdly, there exist no data aggregates on a national or "Bundesland" level, so that the researcher would have to collect the balances of more than 500 compound feed enterprises or be content with an incomplete data set. These difficulties can well be summarized by stating that non-component quantity data is not available. Price information for aggregate non-component inputs is easier to obtain, e.g. as wage index or price indices for energy, industrial installations, et c. But, since sufficient quantity information is unavailable, these cannot be utilized. The consequence of this regrettable situation will become obvious in the remainder of the first

part of this study: the available data does not suffice to perform a parameter estimation of theoretically consistent models of the behavior of the compound feed industry.

3 A Cost Function Model of the Compound Feed Firm

In this chapter, a dual cost function model of the compound feed firm is constructed. It will allow for quantification of the response of demand for feed components on changes of own- and cross-prices, including information about the allocation of feed ingredients on compound feeds for different livestock categories and about the composition of livestock-specific compound feeds. Unfortunately, it cannot ultimately be established since it lacks the validity of some required structural assumptions.

The following sections are devoted to the trial to develop a cost model of the compound feed firm and thereby of the compound feed sector²⁵ that reconciles three concurring requirements:

1. to be operational with regard to the available data, i.e. sectoral prices and quantities of feed components and compound feeds, but no other input data;
2. to be complex enough to depict all aspects of interest of the compound feed sector's response of component demand and marginal cost behavior, respectively, on input price and output quantity changes; and
3. to be consistent with economic theory and reasonable behavioral assumptions for the compound feed firm.

After a short discussion of general premises of the construction of dual feed models, the operability and complexity demand are treated in section 3.2 on the desired model specification, and the consistency requirement is the subject of section 3.3. Finally, the efforts to justify the desired cost model specification are judged in section 3.4, where, after its unambiguous rejection, alternative specifications are discussed. For an explanation of this somehow unusual proceeding refer to section 1.2.1. An abstract of the results of the first part of this study, including this chapter, is given in section 1.3.1.

3.1 General Premises of the Construction of Dual Feed Models

The behavioral models of the compound feed industry developed in chapters 3 and 4 presuppose the truth of the following:

²⁵ See section 1.4 above.

1. All compound feed producers simultaneously minimize their component cost with respect to actual factor prices and product quantities alone. Besides the crucial importance of the cost minimization postulate for the adequacy of a microeconomic cost function approach in general, this means that optimization based on expected prices and quantities is denied as well as the presence and anticipation of risk and uncertainty and the influence of non-price and non-quantity data.
2. All compound feed manufacturers are price takers on the demand and supply side, i.e. the factor and product markets are characterized by perfect competition. The econometrically estimated parameters of cost function derived demand equal the true parameters of the data generating process only if the factor markets are perfectly competitive.
3. The compound feed manufacturers maximize their profits. This assumption of course implies assumption 1., but it is made separately because it is not essential for the demand side of the cost model considered in chapter 3, but only for the quantification of marginal cost. Together with assumption 2 it is needed to measure marginal cost by observed product prices using the rule "price equals marginal cost", which is only valid for profit-maximizing firms. In contrast, additional revenue maximization which, in conjunction with cost minimization constitutes profit maximization, is essential for the profit function model unfolded in chapter 4.
4. It is appropriate to handle the microeconomic behavioral model of an individual compound feed firm as if it were a model of the whole sector, i.e. the complex problem of aggregation over firms is neglected.
5. Technically, the compound feed producers act efficiently, i.e. do not waste factors of production. Hence, twice differentiable functions denoting the frontier of the production possibility set and its duals adequately describe their behavior, and it is not requisite to adopt a frontier modeling approach.
6. All input and output quantities are perfectly divisible and immediately disposable.
7. Microeconomic theory is empirically meaningful and all model parameters stay constant over time.

These premises can be justified by the following arguments:

- Ad 1. Because the compound feed producers generally utilize computer-aided linear programming methods to optimize the composition of their products, a simultaneous and cost-minimizing calculation can be presupposed without further empirical evidence. To examine whether the compound feed producers actually minimize their total cost shall not be an object of discussion, since this would go beyond the scope of the work in hand.
- Ad 2. Feed components are definitely homogeneous goods,²⁶ traded globally and in large quantities. The component markets are characterized by a very large number of participants on both supply and demand side, the latter not only in compound feed manufacturing, but also in on-farm feeding, in food processing, and in direct human consumption. Hence, any significant market power of a single compound feed firm as a demander is extremely unlikely. The supply side looks a little different, because the compound feed markets are locally segregated to some extent. But even if a supplier covers the whole demand of a region, his possibility to raise prices before less expensive competitors enter his local market is quite restricted: like components, compound feeds are very homogeneous goods. In the German compound feed sector, the biggest supplier has less than a 10% market share, but the plants and the corresponding local markets are spread all over the country, and thus the competition situation does not mirror the overall enterprise size.
- Ad 3. The profit maximization assumption for the compound feed industry is justified as good or bad as for most other competitive industries. Most microeconomic studies cannot avoid to make it, and even in much more suspicious cases than the present one it is usually claimed without comment.
- Ad 4. Although no statement about the extent of the aggregation error is possible here, a sectoral model based on behavioral assumptions for individual firms is entertained in the study in hand. In the absence of disaggregated data the aggregation error is an inevitable factor of uncertainty.
- Ad 5. Inefficiencies must be expected in every aspect of compound feed production. But since the model focuses on the combination of components to compound feed and tries to neglect non-component activities, the relevance of inefficiencies seems to be sufficiently

²⁶ See chapter 2.1 for a listing.

small: apart from shrinkage, which should be significantly smaller than 2%, a waste of components is not very likely.

Ad 6. This assumption is completely unproblematic as long as component inputs and compound feed outputs alone are considered. However, non-component activities are mostly long-run so that models developed which do not abstract from non-component activities would require a revision. In any case, divisibility and disposability problems are neglected here because they open up a new broad field of problems which lie beyond the scope of this study.

Ad 7. These very basic premises are mentioned to complete the picture but will not be subject of discussion.

3.2 Desired Specification

The desired specification of the feed cost model presented in this section is basically identical to the postulated models used in all applied studies at hand which examine the compound feed sector and entertain a dual cost function approach.²⁷ The most elaborate example is provided by Peeters' and Surry's work on the Belgian compound feed industry.²⁸ This section can be seen as a reconstruction of their approach.

3.2.1 The Component Cost Function

Since the only available input data for the compound feed sector are quantities and prices of feed components rather than information on other inputs,²⁹ the first necessary step in fitting the cost model to the available data is to construct a cost function allowing the complete neglect of non-component input data. This is undertaken in this section.

To obtain the desired cost structure, let $C(\mathbf{w}^1, \mathbf{w}^2, \mathbf{y})$ denote a cost function modeling the economic behavior of the compound feed firm where \mathbf{w}^1 is the vector of component prices, e.g. the prices of wheat, barley, soy meal, manioc, et c., \mathbf{w}^2 is the vector of prices of all other inputs like energy, human work force, et c., labelled non-component prices in the following, and \mathbf{y} is

²⁷ See e.g. BOYD/BORSSEN 1986; MERGOS/YOTOPOULOS 1988 and SURRY/MOSCHINI 1984.

²⁸ See PEETERS/SURRY 1993.

²⁹ See chapter 2.

the vector of output quantities, i.e. of poultry feed, cattle feed, and the like. Data availability precludes the evaluation of w^2 , and thus a successful analysis requires component demands x^1 derived according to Shephard's lemma to be independent of w^2 ,³⁰ i.e.

$$\nabla_{w^1} C(w^1, w^2, y) = x^1(w^1, y).$$

Assuring this in turn requires the cost function to be of the form

$$C(w^1, w^2, y) = C^1(w^1, y) + C^2(w^2, y):$$

it is only possible to obtain component demand equations which are independent of w^2 if the cost function assumes the above literally additive form, i.e. the macrofunction with the arguments C^1 and C^2 consists in nothing more than plain addition.

With regard to theoretical justification by means of separability assumptions,³¹ it is advisable to decompose the postulation of a literally additive cost structure into two steps.³² Additivity is in general only justifiable by separability assumptions, and the maximum restrictiveness reachable by appropriate separability assumptions is a CES form, rather than a literally additive form.³³ Thus, first, consider the general CES form

$$C(w^1, w^2, y) = f(y) \left[\tilde{C}^1(w^1, y)^{\rho(y)} + \tilde{C}^2(w^2, y)^{\rho(y)} \right]^{\frac{1}{\rho(y)}} \quad 0 \neq \rho \leq 1.$$

Separability of C is the subject of sections 3.3.1 and 3.3.2 below. From the CES structure, it could be secondly argued why this function should simplify to the literally additive form above by making substantial substitutability postulates. If the CES substitution parameter $\rho(y)$ equals one, the cost function is linear in the two great aggregates \tilde{C}^1 and \tilde{C}^2 ; all exponents equal unity. This assumption is discussed in section 3.3.3. Multiplication then yields the desired structure

³⁰ Insufficient data availability is not made explicit by Peeters and Surry as *the* basic motivation for the following modeling efforts. But, as section 3.4 below shows, estimability hinges crucially on this point.

³¹ For an introduction to separability and other structural assumptions see CHAMBERS 1988. For an extensive formal treatment of separability see BLACKORBY/PRIMONT/RUSSEL 1978.

³² Peeters and Surry do not decompose the postulation of a literally additive form and try to establish it directly by separability assumptions, which is not possible. Their account of justification is discussed in footnotes in section 3.3 below.

³³ See BLACKORBY/PRIMONT/RUSSEL 1978.

$$\begin{aligned} C(\mathbf{w}^1, \mathbf{w}^2, \mathbf{y}) &= f(\mathbf{y})\tilde{C}^1(\mathbf{w}^1, \mathbf{y}) + f(\mathbf{y})\tilde{C}^2(\mathbf{w}^2, \mathbf{y}) \\ &= C^1(\mathbf{w}^1, \mathbf{y}) + C^2(\mathbf{w}^2, \mathbf{y}) \end{aligned}$$

Consider the demand equations for components to see why this cost structure allows the neglect of non-component prices. Application of Shephard's lemma yields conditional compound demand which is independent of non-component prices:³⁴

$$\begin{aligned} \nabla_{\mathbf{w}^1} C(\mathbf{w}^1, \mathbf{w}^2, \mathbf{y}) &= \nabla_{\mathbf{w}^1} (C^1(\mathbf{w}^1, \mathbf{y}) + C^2(\mathbf{w}^2, \mathbf{y})) \\ &= \nabla_{\mathbf{w}^1} (C^1(\mathbf{w}^1, \mathbf{y})) \\ &= \mathbf{x}^1(\mathbf{w}^1, \mathbf{y}) \end{aligned}$$

i.e. the compound demand \mathbf{x} derived from C^1 equals the compound demand derived from C because C^2 is a constant of differentiation with respect to \mathbf{w}^1 .

Partial differentiation of C with respect to output quantities results in marginal cost, the cost of producing one additional unit of aggregate output. Since both C^1 and C^2 contain \mathbf{y} , non-component prices do not vanish in marginal cost. But, because of the literally additive structure of the cost function, the partial differential of component cost alone with respect to output quantity can be interpreted as that portion of marginal cost that results from buying components,³⁵ denoted by z :

$$\nabla_{\mathbf{y}} C^1(\mathbf{w}^1, \mathbf{y}) = z(\mathbf{w}^1, \mathbf{y})$$

Unfortunately, z cannot be observed. But, making an additional assumption, it is possible to obtain a measure for z : in a profit-maximizing industry with perfect competition on both the supply and the demand side,³⁶ marginal cost equals price p , which is observable. To evaluate z , it is requisite to establish a relationship between marginal compound feed cost and feed price, for example

$$z_i(\mathbf{w}^1, \mathbf{y}) = p_i - \zeta_i(\mathbf{w}^1, \mathbf{w}^2, \mathbf{y}) \cdot p_i,$$

³⁴ See PEETERS/SURRY 1993: 111-112.

³⁵ See PEETERS/SURRY 1993: 112.

³⁶ See assumptions 2. and 3. in section 3.1 above.

where z_i denotes the marginal component cost for output i and ζ_i is a real number between zero and unity denoting the marginal cost share that is caused by purchasing non-component inputs. To make z measurable, this share must be assumed to be equal for all compound feeds:³⁷

$$\zeta_i(\mathbf{w}^1, \mathbf{w}^2, \mathbf{y}) = \zeta(\mathbf{w}^1, \mathbf{w}^2, \mathbf{y}) = \frac{\partial C^2(\mathbf{w}^2, \mathbf{y}) / \partial y_i}{\partial C(\mathbf{w}^1, \mathbf{w}^2, \mathbf{y}) / \partial y_i}.$$

Then, marginal component cost can be written as

$$z(\mathbf{w}^1, \mathbf{y}) \stackrel{\text{def!}}{=} \mathbf{p}(1 - \zeta(\mathbf{w}^1, \mathbf{w}^2, \mathbf{y})) = \nabla_y C^1(\mathbf{w}^1, \mathbf{y}).$$

Although ζ obviously depends on all exogeneous variables including \mathbf{w}^2 , z can be obtained without knowledge of \mathbf{w}^2 because the value of ζ – in contrast to the n values of ζ_i – can be calculated directly as quantity-weighted average compound feed price minus marginal component cost, which is equal to average component cost if in addition homogeneity of the component combination process, i.e. the dual of component cost, is assumed:

$$\zeta(\mathbf{w}^1, \mathbf{w}^2, \mathbf{y}) = \frac{\sum_{i=1}^m p_i y_i}{\sum_{j=1}^m y_j} - \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n x_i}.$$

Section 3.3.4 below discusses the assumption of equal marginal non-component cost – marginal component cost proportions, whereas the homogeneity claim is the subject of section 3.3.5 below.

Up to this point, a cost structure is postulated which yields a system of component demand and marginal feed cost equations which is estimable because it is completely independent of the unavailable non-component data. Adequacy of this cost structure is claimed referring to three structural assumptions, namely separability, equal non-component cost-total cost proportions for all compound feeds, and homogeneity of the component combination part of technology.

3.2.2 The Allocation of Components on Compound Feeds

The behavioral model of the compound feed firm developed above provides total component demand quantities, but it lacks information on how total component demands are allo-

³⁷ See PEETERS/SURRY 1993: 112.

cated to the livestock category-specific compound feeds. This information is of interest for policy evaluation since demand and marginal feed cost elasticities by type of livestock are an intuitive tool for assessing the consequences of measures in the livestock sector on component demand quantities and of measures influencing a component price on marginal feed cost, which is an important determinant of feed price and thus influences the livestock sector.

Isolation of the influence of component prices on component demand with respect to only one compound feed y_j requires the derived demand for the i^{th} component allocated to the j^{th} compound feed x_i^j to be independent of any other compound feed y_k :

$$x_i^j(\mathbf{w}^1, \mathbf{y}) = x_i^j(\mathbf{w}^1, y_j).$$

This, in analogy to the independence of component demand from non-components postulated in section 3.2.1 above, is only possible if

$$C^1(\mathbf{w}^1, \mathbf{y}) = \sum_{j=1}^m C_j^1(\mathbf{w}^1, y_j),$$

i.e. component costs are consistent with the assumption of compound feed production being nonjoint in component quantities,³⁸ discussed in section 3.4.5 below. The nonjoint cost function implies

$$x_i(\mathbf{w}^1, \mathbf{y}) = \frac{\partial C^1(\mathbf{w}^1, \mathbf{y})}{\partial w_i} = \sum_{j=1}^m \frac{C_j^1(\mathbf{w}^1, y_j)}{\partial w_i} = \sum_{j=1}^m x_i^j(\mathbf{w}^1, y_j).$$

For $C_j^1(\mathbf{w}^1, y_j)$, i.e. the microfunction depicting the component cost of producing one specific compound feed, it is adequate to assume homotheticity and constant returns to scale of the underlying technology, discussed in section 3.4.6 below, so that it simplifies further:

$$C_j^1(\mathbf{w}^1, y_j) = y_j \cdot c_j^1(\mathbf{w}^1).$$

As this formulation makes apparent, in the presence of constant returns to scale in component combination unit or average cost equal marginal cost a_j :³⁹

³⁸ See PEETERS/SURRY 1993: 112-113.

³⁹ Of course total cost per unit of y_i increases if increasing quantities of other outputs are produced – total cost increase, but the quantity of y_i remains the same –, and thus unit cost does not equal marginal cost with non-zero production of any other output.

$$a_j \stackrel{\text{def.}}{=} \frac{\partial C^1(\mathbf{w}^1, \mathbf{y})}{\partial y_j} = \frac{C^1(\mathbf{w}^1, \mathbf{y})}{y_j} = c_j^1(\mathbf{w}^1) \quad \text{with } y_i = 0 \quad \forall i \neq j,$$

and applying Shephard's lemma now yields

$$x_i = \frac{\partial C^1(\mathbf{w}^1, \mathbf{y})}{\partial w_i} = \sum_{j=1}^m y_j \cdot \frac{c_j^1(\mathbf{w}^1)}{\partial w_i} = \sum_{j=1}^m x_i^j,$$

i.e. the assumptions of nonjointness in component quantities and homotheticity and constant returns to scale in component combination allow to establish livestock-category-specific component demand functions x_i^j , which can be plainly added up to obtain the total component demand function x_i .

3.2.3 Technical Progress

While it is obvious that technical progress primarily affects non-component cost where processing, transportation, management, and other progress-sensitive production factors are gathered, the pure minimization of component cost also leaves some room for progress, e.g. through, first, better programming methods or more sophisticated know-how in animal nutrition; secondly, new or modified components, e.g. synthetic amino acids driving out traditional protein carriers; and thirdly, enhanced processing techniques, e.g. toasting soy beans to increase protein digestability and allow a higher share of soy, or new pelleting techniques. The lack of better indicators – more exactly the diversity of ways in which technical progress acts on compound feed technology, and the incommensurability or insufficient observability of possible indicators – leads to the simplifying hypothesis that it is only related to time, represented by a periodical time index t . Since it is a matter of the choice of a concrete functional form, the discussion of the exact way in which t enters the equation is dealt with in chapter 7, where specifications for the function to estimate are discussed. For now, it will suffice to state that the compound feed manufacturer's cost function assumes the form

$$C(\mathbf{w}^1, \mathbf{w}^2, \mathbf{y}, t) = C^1(\mathbf{w}^1, \mathbf{y}, t) + C^2(\mathbf{w}^2, \mathbf{y}, t).$$

Unfortunately but inevitably, the neglect of C^2 in the estimation has the result that all technical progress in the compound feed industry is explained by C^1 alone, i.e. it is exclusively attributed to the three kinds of technical progress listed above, be it in fact caused by these or by non-component activities. However, according to the results of section 3.2.1 below, this short-

coming is not relevant if component demand shares are regarded because these are separable from non-components. Furthermore, anticipating structural assumption C.12 (constant returns to scale in component quantities) set out in section 3.3.5 below, it has to be noted that technical progress with respect to the component combination cannot increase the “efficiency” of component combination since the constant returns to scale hypothesis is trivially true for this part of technology, whatever progress is made. Technical progress in component combination may only render the production of a unit compound feed cheaper or increase the efficiency of animal production, and hence increase the price of the compound feed that can be realized at the market. This is visible in the cost function in terms of shifting component proportions exclusively; i.e. Hicks-neutral technical change of the component combination technology is necessarily excluded. To save notation, the time variable is omitted in the remainder of this chapter.

3.3 Structural Assumptions

In section 3.2, Peeters' and Surry's cost model of the compound feed industry is reconstructed. This section lists all structural assumptions relevant for their model.⁴⁰

3.3.1 Weak Separability

There are three sectors of economic decision making in the compound feed firm that are immediately distinguishable: the output sector, i.e. the question of which products are supplied at which quantity, the feed component sector, i.e. the question of which components are demanded at which ratios to yield the respective product quantities, and the non-component sector providing the “catalyst” for transforming components into compound feeds, i.e. making possible a “reaction” of components to an identical mass of compound feeds without itself entering the product. The non-component sector includes activities like transportation of factors and products, stockholding, component and compound feed processing, the actual mixing process, marketing, accounting etc. Obviously, separability assumptions between these three sec-

⁴⁰ For a presentation of premises which, as opposed to assumptions on the specific structure of economic behavior, constitute prerequisites of dual behavioral models of the compound feed sector in general, refer to section 3.1 above.

tors are possible both in a cost and profit function context. They are examined in sections 3.3.1 and 3.3.2 for the cost function and in section 4.3.1 and 4.3.2 for the profit function.

But one might wonder whether a more fine-grained separable partition is possible, and more especially inside the non-component sector that includes such different activities. This would generate the prerequisites for an application of the strong separability assumption, which by definition requires at least three distinct sectors,⁴¹ in order to justify an additive cost structure. However, a partition of the non-component sector is contradicted by the following argument: the non-component activities are much more entangled than the sectors of the above binary partition in inputs. It is not even possible to assign all non-component variables to one economic decision exclusively; e.g. wage is relevant for each of the activities listed above, and energy for many. But mutually exclusive variable sectors constitute a necessary condition for a partition. In addition, a further partition is not convincing from an epistemological perspective since Ockham's razor dictates the choice of the most simple a priori reasonable hypothesis.

In general, the assumption of a two stage, separable cost minimization process is used to estimate all micro cost functions first, to interpret micro cost as aggregate prices and then estimate the macro cost function relying on these prices. Apart from tractability considerations, the possibility to individually model each micro cost function constitutes the main reason for such a procedure.⁴² In this study, the reason is different: only one micro function – the component cost function – will be estimated, and the separability assumption will allow inference on the behavior of the compound feed firm without taking into account any input variables outside the compound aggregate. Consequently, the relations between the mutually exclusive and exhaustive sectors \mathbf{w}^1 including all component prices, \mathbf{w}^2 denoting all non-component prices, and output quantity vector \mathbf{y} shall be analyzed with regard to their separability.

Let the index set $I = \{1, \dots, n\}$ of the input price vector $\mathbf{w} \in \mathfrak{R}^n$ be partitioned into mutually exclusive and exhaustive sectors I^r of length n^r with $r = 1, \dots, Q$. With each I^r there corresponds a sub-vector of \mathbf{w} , namely $\mathbf{w}^r = (w_i)_{i \in I^r}$. Let $J = \{1, \dots, m\}$ be the unpartitioned index set of output quantities $\mathbf{y} \in \mathfrak{R}^m$. For the cost function $C(\mathbf{w}^1, \mathbf{w}^2, \mathbf{y}) : \mathfrak{R}^{n+m} \rightarrow \mathfrak{R}$, there are the following six relations relevant for the weak separability assumption, given the binary partition of input prices, i.e. $Q = 2$. A vanishing differential quotient indicates a separability postulate,

⁴¹ See BLACKORBY/PRIMONT/RUSSELL 1978: 142-148.

⁴² See CHAMBERS 1988: 43-44

while a differential quotient being equal to some function $g_{ijk}(\mathbf{w}, \mathbf{y})$ indicates inseparability, i.e. an assumed influence such that the namely differential quotient does not necessarily vanish:⁴³

$$C.1 \quad \frac{\partial}{\partial w_k} \frac{\partial C(\mathbf{w}, \mathbf{y}) / \partial y_i}{\partial C(\mathbf{w}, \mathbf{y}) / \partial y_j} = g_{ijk}(\mathbf{w}, \mathbf{y}) \quad \forall i, j \in J; k \in I^1$$

$$C.2 \quad \frac{\partial}{\partial w_k} \frac{\partial C(\mathbf{w}, \mathbf{y}) / \partial y_i}{\partial C(\mathbf{w}, \mathbf{y}) / \partial y_j} = 0 \quad \forall i, j \in J; k \in I^2$$

$$C.3 \quad \frac{\partial}{\partial w_k} \frac{\partial C(\mathbf{w}, \mathbf{y}) / \partial w_i}{\partial C(\mathbf{w}, \mathbf{y}) / \partial w_j} = \frac{\partial}{\partial w_k} \frac{x_i}{x_j} = 0 \quad \forall i, j \in I^1; k \in I^2$$

$$C.4 \quad \frac{\partial}{\partial y_k} \frac{\partial C(\mathbf{w}, \mathbf{y}) / \partial w_i}{\partial C(\mathbf{w}, \mathbf{y}) / \partial w_j} = \frac{\partial}{\partial y_k} \frac{x_i}{x_j} = g_{ijk}(\mathbf{w}, \mathbf{y}) \quad \forall i, j \in I^1; k \in J$$

$$C.5 \quad \frac{\partial}{\partial w_k} \frac{\partial C(\mathbf{w}, \mathbf{y}) / \partial w_i}{\partial C(\mathbf{w}, \mathbf{y}) / \partial w_j} = \frac{\partial}{\partial w_k} \frac{x_i}{x_j} = 0 \quad \forall i, j \in I^2; k \in I^1$$

$$C.6 \quad \frac{\partial}{\partial y_k} \frac{\partial C(\mathbf{w}, \mathbf{y}) / \partial w_i}{\partial C(\mathbf{w}, \mathbf{y}) / \partial w_j} = \frac{\partial}{\partial y_k} \frac{x_i}{x_j} = g_{ijk}(\mathbf{w}, \mathbf{y}) \quad \forall i, j \in I^2; k \in J$$

These assumptions can be justified as follows:

C.1 The relation between any two marginal costs for specific compound feeds depends on any component price since components are utilized with different intensity in different compound feeds (more than that, this *is* the difference between different compound feed). Thus, a component price change more affects marginal cost of that compound feed in which it is more intensively utilized. A vanishing differential quotient only occurs accidentally if two compound feeds contain equal shares of the respective components.

C.2 To assume weak separability of the output quantities from non-component prices means to maintain that, considering all non-component activities like transportation, stocksholding, feed processing, marketing, accounting etc., any non-component price change is irrelevant for the relative advantageousness of compound feeds. Behind this hypothesis is the claim that it does not matter which compound feed is produced since they are all so-

⁴³ There is no study on the compound feed sector where separability of a cost or profit function is examined for all possible relations. Most authors merely consider one differential equation. See e.g. PEETERS/SURRY 1993 who argue for C.3 only and leave the opposite direction C.5 and all possible relations including the output variables unexamined.

- me component combination processed identically; the compound feeds only differ with respect to their composition. This implies the neglect of differences in processing intensity, e.g. between pellets and meal, and to reduce changes in relative marginal feed costs to component price changes, which is expressed in assumption C.1. The resulting error seems small enough to maintain C.2.
- C.3 Compound feed composition does not depend on non-component prices, neither in a specific compound feed nor at total. The obvious truth of this assumption is supported by the fact that in practice the least-cost component mix is generally calculated using linear programming algorithms where component prices and nutritional restrictions constitute the only restrictions.⁴⁴
- C.4 A change in output of one feed affects component proportions if components are used with different intensity in different compound feeds, which is generally the case. Thus, paralleling C.1, it would be wrong to assume separability here.
- C.5 Paralleling C.2 and C.3, it is of no consequence if a component price alters with regard to the ratio of arbitrary non-component demands since processing, transportation, and marketing requirements of components do not differ. The error made by neglecting e.g. the need to "toast" soy beans to enhance protein accessibility and other component-specific processing demands seems to be sufficiently small.
- C.6 To assume separability of non-components from output level would mean to deny the possibility of changing factor intensities with increasing firm size.⁴⁵ But factor substitution in the non-component sector constitutes an important source of economies of size in the compound feed industry, e.g. the substitution of human working power by technical equipment. Hence, a separability assumption would be phenomenally inadequate. However, it could be argued that economies of size generally reduce to indivisibility problems

⁴⁴ The only exception is an occasional restriction for a minimal share of domestic grains anticipating some farmer's extraeconomic wish to support domestic production although it makes feed more expensive. But this constitutes no non-component influence on compound feed composition either.

⁴⁵ Since according to C.2 processing technology is indifferent towards compound feeds, it makes no difference to analyze the influence of a single output on non-component shares or the influence of the sum of all outputs, i.e. firm size measured in output quantity.

which are excluded by assumption in microeconomic theory.⁴⁶ Then, a separable C.6 could be maintained. But if reality is likely to be imperfect with regard to theory, accounting for the imperfection is heuristically advisable on the one hand, and provides a means to check premises of theory on the other hand: not assuming separability here from the outset allows the testing of the adequacy of the separability assumption.

3.3.2 Strong Separability

In the binary partition, strong separability is conceptually impossible, but also counting the output variable sector leads to three additional equations:⁴⁷

$$C.7 \quad \frac{\partial}{\partial w_k} \frac{\partial C(\mathbf{w}, \mathbf{y}) / \partial y_i}{\partial C(\mathbf{w}, \mathbf{y}) / \partial w_j} = 0 \Leftrightarrow \frac{\partial}{\partial w_k} \frac{\partial C(\mathbf{w}, \mathbf{y}) / \partial w_j}{\partial C(\mathbf{w}, \mathbf{y}) / \partial y_i} = 0 \quad \forall i \in J; j \in I^1; k \in I^2$$

$$C.8 \quad \frac{\partial}{\partial w_k} \frac{\partial C(\mathbf{w}, \mathbf{y}) / \partial y_i}{\partial C(\mathbf{w}, \mathbf{y}) / \partial w_j} = g_{ijk}(\mathbf{w}, \mathbf{y}) \Leftrightarrow \frac{\partial}{\partial w_k} \frac{\partial C(\mathbf{w}, \mathbf{y}) / \partial w_j}{\partial C(\mathbf{w}, \mathbf{y}) / \partial y_i} = g_{ijk}(\mathbf{w}, \mathbf{y}) \quad \forall i \in J; j \in I^2; k \in I^1$$

$$C.9 \quad \frac{\partial}{\partial y_k} \frac{\partial C(\mathbf{w}, \mathbf{y}) / \partial w_i}{\partial C(\mathbf{w}, \mathbf{y}) / \partial w_j} = g_{ijk}(\mathbf{w}, \mathbf{y}) \Leftrightarrow \frac{\partial}{\partial y_k} \frac{\partial C(\mathbf{w}, \mathbf{y}) / \partial w_j}{\partial C(\mathbf{w}, \mathbf{y}) / \partial w_i} = g_{ijk}(\mathbf{w}, \mathbf{y}) \quad \forall i \in I^1; j \in I^2; k \in J$$

These assumptions can be justified as follows:

C.7 Both compound feed quantities and feed component prices are weakly separable from non-component prices, and this is the necessary condition for both sectors jointly being strongly separable from non-component prices. Sufficiency is provided by the catalyst-view of the compound feed technology where all non-component activities of the compound feed firm treat all components and all compound feeds the same, and supply quantities are strictly proportional to compound demand quantities by nature of technology:

⁴⁶ While obvious in cases from e.g. a processing unit up to indivisibility extremes like the design of a firm logo, this thesis can consistently be claimed even for instances which are not that obvious, e.g. where the characteristic iso-transport cost circles around a plant can be reconstructed as a result of the impossibility to divide the plant until one mini-plant is located at each customer's farm so that transport cost for the product would vanish completely. It can be argued analogously on the transport cost of components [considering the role of transport cost, see section 4.4.3 below]. The only remaining benefit of size would then be the realization of better prices, i.e. cheaper purchases and more expensive sales, because of larger batches. But this is exogeneous and correspondingly does not mirror in the structure of the cost function.

⁴⁷ Strong separability implies weak separability, and thus C.1-C.6 apply in any case.

an increase of the price of e.g. energy first causes supply to decrease without affecting relative preferability of compound feed according to C.2. Secondly, this leads to a decrease in aggregated component demand which is strictly proportional to the decrease in aggregated output quantity – this is the part of the price reaction which is not already covered by a weak separability assumption and thus the original object of reasoning with regard to C.7. Thirdly, as it is known by C.3, component shares are not affected by this decrease of total component demand.

C.8 Since output quantities alone are inseparable from component prices, which is known by C.1, they can not be separable from component prices if they are examined jointly with non-component prices, which can be shown by contradiction: cancel down the differential quotient in C.1 by $\partial C(\mathbf{w}, \mathbf{y}) / \partial w_l$ where $l \in I^2$

$$\frac{\partial}{\partial w_k} \frac{\partial C(\mathbf{w}, \mathbf{y}) / \partial y_i}{\partial C(\mathbf{w}, \mathbf{y}) / \partial y_j} = \frac{\partial}{\partial w_k} \frac{\frac{\partial C(\mathbf{w}, \mathbf{y}) / \partial y_i}{\partial C(\mathbf{w}, \mathbf{y}) / \partial w_l}}{\frac{\partial C(\mathbf{w}, \mathbf{y}) / \partial y_j}{\partial C(\mathbf{w}, \mathbf{y}) / \partial w_l}} \quad \forall i, j \in J; k \in I^1, l \in I^2$$

Now assume C.8 did not hold, i.e. the partial derivative of the differential quotient in C.8 for some component price would necessarily vanish. Then, both the numerator and the denominator of the right hand side of the above equation would be independent of w_k by assumption. But this would imply that the complete right hand term would vanish for all output- i, j , component- k and non-component- l , which contradicts C.1. Consequently, C.8 is necessarily true by C.1. Summarized in one line:

$$0 \stackrel{\text{by C.1}}{\neq} \frac{\partial}{\partial w_k} \frac{\partial C(\mathbf{w}, \mathbf{y}) / \partial y_i}{\partial C(\mathbf{w}, \mathbf{y}) / \partial y_j} \stackrel{l \in I^2}{=} \frac{\partial}{\partial w_k} \frac{\frac{\partial C(\mathbf{w}, \mathbf{y}) / \partial y_i}{\partial C(\mathbf{w}, \mathbf{y}) / \partial w_l}}{\frac{\partial C(\mathbf{w}, \mathbf{y}) / \partial y_j}{\partial C(\mathbf{w}, \mathbf{y}) / \partial w_l}} \stackrel{\text{assume C.8 did not hold}}{=} 0$$

C.9 Equation C.9 together with C.1 and C.2, all vanishing, would imply input-output separability or a homothetic technology. But this is obviously not the case, as already C.1 reveals. The proof is the same as the one given for C.8.

3.3.3 Unity Value of the CES Substitution Parameter

The desired cost model postulated in section 3.2 above claims a CES cost structure for the compound feed industry rather than a completely general cost function of output quantities and

input prices, which can in general be established if certain separability conditions apply. In the preceding section, all separability conditions were analyzed which are relevant for the possibility to write the feed cost function as a CES structure. In section 3.4 below it is examined whether this enterprise was successful – with negative result. But for the moment, assume that it was, i.e. that a CES feed cost structure were adequate. Then, if the CES substitution parameter $\rho(\mathbf{y})$ equals one, the cost function shows a linear relationship between component and non-component cost. Due to duality, this implies a Leontief technology characterized by right-angled isoquants between aggregate components and aggregate non-components.⁴⁸ That means that there is no substitution between these aggregates. And, regarding the physical nature of compound feed processing, this is exactly the case: no device and no managerial effort can substitute only the merest portion of any component and vice versa; to obtain one metric ton of compound feed, one metric ton of components has to be fed into the production process. Thus, the restriction

$$C.10 \quad \rho(\mathbf{y}) = 1$$

is perfectly justified.

But, for the aggregate character of the arguments, the interpretation of the Leontief technology function must not be overdone: fixed proportions between the aggregates do not imply fixed proportions between any single disaggregate factors, e.g. a fixed amount of labour or energy per unit compound feed. This inference is not possible because the aggregator functions do not plainly sum up their elements; both $C^1(\mathbf{w}^1, \mathbf{y})$ and $C^2(\mathbf{w}^2, \mathbf{y})$ are arbitrary functions of which only one thing can be said: they are well-behaved cost functions.

3.3.4 Equal Non-Component Cost – Total Cost Proportions

The pragmatic postulate that the share of marginal non-component cost ζ_i is the same for every compound feed, i.e.

$$C.11 \quad \zeta_i(\mathbf{w}^1, \mathbf{w}^2, \mathbf{y}) = \zeta(\mathbf{w}^1, \mathbf{w}^2, \mathbf{y}) = \frac{\partial C^2(\mathbf{w}^2, \mathbf{y}) / \partial y_i}{\partial C(\mathbf{w}^1, \mathbf{w}^2, \mathbf{y}) / \partial y_i} \quad \forall i.$$

⁴⁸ See e.g. CHAMBERS 1988: 88-92.

is comparably unproblematic.⁴⁹ Although it may be higher in pelleted feeds than in unprocessed meals, higher in concentrate protein feeds for broilers than in pig feed etc., the differences can be supposed to be not too significant. This seems the more justified as more expensively processed feeds commonly contain more expensive components so that the *share* of marginal non-component costs remains almost equal even if non-component cost per unit produced rises.⁵⁰

3.3.5 Constant Returns to Scale in Component Quantities

Constant returns to scale in component quantities or homogeneity of degree one of the primal technology in component composition means that doubling all component quantities doubles the total quantity of compound feed produced, which is trivially true. Consider this technological assumption expressed in terms of the transformation function Y :

$$C.12 \quad Y(\mathbf{x}^1, \mathbf{y}) = 0 \Rightarrow Y(\lambda \mathbf{x}^1, \lambda \mathbf{y}) = 0 \quad \forall \lambda > 0.^{51}$$

⁴⁹ See PEETERS/SURRY 1993: 112

⁵⁰ Peeters and Surry verbally postulate C.11, but in their formal definition they let ζ be the quotient of total non-feed cost and total cost rather than the quotient of their marginal counterparts [see PEETERS/SURRY 1993: 112]. If it is not only a printing error, another severe structural assumption is implied by using the average non-component share ζ_i to measure the share of marginal cost that is caused by buying non-components: these are equal if and only if the whole component feed technology is characterized by constant returns to scale. This is on the one hand, as already stated in section 3.4.1 in the discussion of assumption C.6, phenomenally inadequate, and on the other hand it would mean that the optimization rule "price equals marginal cost" would not apply so that measurability would fail anyway. Furthermore, Peeters' and Surry's definition suggests that ζ is a constant [see PEETERS/SURRY 1993: 112]. Although ζ_i may be identical for all compound feeds at a certain total output level or firm size, respectively, it clearly decreases with rising firm size if economies of scale are not denied. Even more obvious, any component price change influences ζ . This is not problematic since, with the formula proposed in section 3.2.1 above, ζ can be calculated for every observation.

⁵¹ A definition of a linearly homogeneous technology is easy in terms of sets: a technology is homogeneous of degree one if the production possibilities set T can be characterized as follows: $T = \lambda T$ for all positive λ [see FÄRE/PRIMONT 1995: 23]. But it becomes odd in terms of real valued functions like the transformation function above because the general technology then is formulated as implicit function. In the context of dual behavioral functions, linear homogeneity in the multi-output case is straightforward again [see section 3.2.5 below].

It is physically impossible to produce a certain quantity of some compound feed without adding exactly the same quantity of components to the production process, and vice versa. If not, components would be wasted, and inefficiency is excluded by assumption.⁵²

The assumption of constant returns to scale in component quantities may not be confused with the assumption of constant returns to scale in the whole compound feed technology; significant scale effects must be expected in every other aspect of technology. Furthermore, it is important to note that, in the multi-input-multi-output case, a technology homogeneous of degree one does not imply homotheticity.⁵³ Otherwise, it would be impossible to maintain C.12 and C.6 simultaneously.

3.3.6 Nonjointness in Component Quantities

Assuming nonjointness in component quantities means that production of single compound feed y_i may depend on all component quantities, but is claimed to be independent of the production quantities of any other compound feed. Put differently: the primal of component cost C^1 , namely the component-to-feed transformation function $0 = Y^1(\mathbf{x}^1, \mathbf{y})$ solved for y_i does not contain any y_j as an argument and simplifies from

$$y_i = Y_i^1(\mathbf{x}^1, y_1, y_2, \dots, y_{i-1}, y_{i+1}, \dots, y_m) \quad \forall i$$

to

$$C.13 \quad y_i = Y_i^1(\mathbf{x}^1) \quad \forall i.$$

This does not imply that the y_i are produced in physically separate processes, merely the functional relation is that way.⁵⁴ Externalities of one compound feed on another with respect to component utilization do not exist since neither a compound feed is used as component in the production of another, nor does any by-product of one feed enter the production process of another or enhance or inhibit it. This is of course not to deny that, in total, there may exist considerable synergies between the products: it is only component cost that is assumed to be non-joint, not total cost. The claim is that any synergy takes place in the non-component sector,

⁵² See section 1.4 above.

⁵³ See PRIMONT/FÄRE 1995: 139.

⁵⁴ See HALL 1973: 884.

e.g. due to joint transportation and marketing, or due to jointly producing different feeds in one plant. Thus, the assumption of nonjointness in component quantities is completely in accord with the technical nature of the production process.

3.3.7 Homothetic Feed Composition

In a homothetic technology, the expansion paths are straight lines, i.e. a proportional change in output quantities does not alter the cost minimal input ratios. Assuming homothetic feed composition, i.e. an assumption of technology, can equivalently be expressed as an assumption of the cost function, namely component prices and output quantities being mutually weakly separable. Whereas both equation C.1 and equation C.4 do not vanish, i.e. homothetic feed composition for the whole cost function cannot be maintained for reasons named in section 3.4.1 above, the picture changes under the hypothesis of nonjointness in component quantities: the single-product cost functions established by nonjointness⁵⁵ are evidently characterized by homotheticity, i.e.

$$C.14 \quad \frac{\partial}{\partial y_k} \frac{\partial C_k^1(\mathbf{w}, y_k) / \partial w_i}{\partial C_k^1(\mathbf{w}, y_k) / \partial w_j} = \frac{\partial}{\partial y_k} \frac{x_i^k}{x_j^k} = 0 \quad \forall i, j \in I^1; k \in J.$$

If a single compound feed is considered, it is irrelevant for its composition if one ton or 1,000 tons are produced; the recipe is the same for any output quantity. Thus, as this case exemplifies, homothetic nonjoint single-output production functions do not imply a homothetic multi-output technology, even if, according to C.12, this technology is linearly homogeneous.⁵⁶

In contrast to C.1 through C.13, assumption C.14 is not implied by the structure of the desired model specification. But since its adequacy is so evident, it can be introduced without harm to save estimation parameters.

⁵⁵ See section 3.3.2 above and section 3.4.2 below.

⁵⁶ See PRIMONT/FÄRE 1995: 139.

3.4 Reconsidering the Model Specification

3.4.1 Separability

While the assumption of unity value of the CES substitution parameter is adequate so that, given a CES cost function, justification of a literally additive structure of compound feed cost is possible, the maintained separability hypotheses do not suffice by far to establish the required CES specification: even if all equations C.1 through C.9 could realistically be assumed to vanish, i.e. if strong separability of the cost function in the binary and extended partition could be maintained, this would not imply an additive structure.⁵⁷ This result, also utilized in section 4.3.1 below in the profit function context, is commented on in section 4.1.2. Hence, the desired model specification formulated in section 2.1 above is rejected.

Now, the question arises of how far one can get with the maintained separability assumptions. Unfortunately, a cost structure equivalent to assumptions C.1 through C.9, namely a functional representation of C.1 through C.9, could not be established. This constitutes a severe limitation of the model with respect to the enterprise of formulating an applied model without a theoretic gap.⁵⁸ It was not even possible to actually find a general structure that satisfies all nine equations, i.e. to find a non-additive⁵⁹ structure that at least satisfies the necessary condition for equivalence. Two alternatives may be considered: first, a structure might be adopted that is more restrictive than admissible. This is principally unacceptable since not only existing effects cannot be depicted, but moreover the influence of these is burdened onto the remaining

⁵⁷ See FEGER/MÜLLER 1999: 3 and 10-12. See section 4.3.1 below for the opposite case.

⁵⁸ Recall the definition of an equivalence relationship: equivalence means mutual implication. This is, a cost structure is equivalent with a set of separability assumptions if it not only satisfies all maintained equations, but that it is the *only* one to satisfy all equations or to *follow* from the equations. Satisfying all separability equations merely constitutes the necessary condition for equivalence, and if sufficiency cannot be proven one can never be sure if another cost structure does not exist satisfying the same separability assumptions – which is possibly less restrictive and thus more adequate.

⁵⁹ There are additive structures that satisfy the necessary condition for equivalence – for example the desired specification introduced in section 3.2 above. But, as FEGER / MÜLLER 1999 unambiguously show, sufficiency for an additive structure is not even provided for by C.1 through C.9 vanishing. Such a result only holds for the profit function. Thus, an additive structure of a cost function separable in the binary partition, or with no more than three sectors counting input and output variables likewise, respectively, must be completely disregarded.

effects and biases the respective estimation parameters. The only case where one could even think of implying separability in spite of the maintained inseparability is $C.6$ because inseparability of $C.6$ is arguable from a very strict theoretical perspective.⁶⁰ An example is given in the next paragraph. Secondly, it is always possible to postulate a structure that is less restrictive than maintained assumptions would imply. An example of a functional structure that accounts for every maintained inseparability and implies separability where it is assumed with the exception of $C.2$ and $C.7$ is presented further below in this section.

Considering the first strand of argumentation, a cost structure incorporating all maintained separability assumptions and leaving all relations unrestricted where inseparability is supposed, with the exception of $C.6$, i.e. satisfying all maintained equations but one, is

$$C(\mathbf{w}^1, \mathbf{w}^2, \mathbf{y}) = C^{**}[C^1(\mathbf{w}^1, \mathbf{y}), c^2(\mathbf{w}^2)].^{61}$$

Applying Shepard's lemma and the chain rule, the derived component demand equations assume the form

$$x_i = \frac{\partial C(\mathbf{w}^1, \mathbf{w}^2, \mathbf{y})}{\partial w_i} = \frac{\partial C^1(\mathbf{w}^1, \mathbf{y})}{\partial w_i} \cdot \frac{\partial C^{**}[C^1(\mathbf{w}^1, \mathbf{y}), c^2(\mathbf{w}^2)]}{\partial C^1(\mathbf{w}^1, \mathbf{y})} \quad \forall i \in I^1.$$

Hence, component demand is not independent of non-component prices; the second term, i.e. the exterior derivative which is the same for all x_i , still includes \mathbf{w}^2 . The standard chain rule interpretation yields an insight into the nature of the influence: in the first term, a change in a component price w_i results in a change in component costs C^1 where the interaction between component prices takes place. Here, the ratios of component demands are determined. In the second term, the resulting component costs as a whole affect total cost, i.e. the level of component costs, regardless of the component ratios, affects overall feed costs which are only a function of aggregate component cost and aggregate non-component cost. This is, the level of non-component costs influences component demands – not component mix but component demand level. To take it down to the phenomena: if the price of energy or wages rise, this makes feed processing more expensive: c^2 rises. At a second stage, the aggregate "price" of feed processing has an influence on the demand level of each component (which is the same for all

⁶⁰ See section 3.3.1 above.

⁶¹ The small letter c^2 as opposed to the capital letters used for the other cost functions is meant to serve as a reminder that the respective function depends on merely one variable or sector of variables, respectively.

components by $C.3$): fewer components are demanded because it is more expensive to produce one unit of compound feed, resulting in a lower total component demand or a smaller firm size (since output mix is independent from non-component prices by $C.2$).⁶² The amount of firm size reduction induced by higher non-component costs is described by C^{**} . This demand level effect is erroneously neglected if a literally additive form is assumed.

An approach that might be considered is the following: whereas single non-component prices are unavailable, one could argue that estimability of the correct demand equations is already provided for since, rather than non-component prices w^2 , aggregate non-component cost c^2 as a whole is measurable – and thus all explaining variables of the correct demand equations are available. It does not fulfil this hope, although three ways could be proposed.

First, one could claim that c^2 is the difference between total cost and component expenditures. But, on the one hand, total costs are not observable, at least not in the monthly frequency,⁶³ and, on the other hand, although total primal and total dual costs must be equal, c^2 does not equal primal non-component cost. Generally, dual micro functions do not necessarily correspond to the respective primal function. This is obvious in the present case since just the structure of C^{**} shows that total dual costs are *not* the plain sum of component cost and non-component cost, while this is true by definition for the primal cost function. The falsity of this hypothesis is additionally suggested by the independence of c^2 of y : this is hardly likely to be the case for total non-component cost.

Secondly, one could intend to solve Euler's Theorem for c^2 . But Euler's theorem is not applicable since C is only homogeneous in w rather than in w and y , and it is not even necessarily homothetic, so that the generalized Euler theorem could be utilized.⁶⁴

Thirdly, the independence of c^2 from y suggests that c^2 is something like the aggregated price of non-components which can be obtained by subtracting component cost per ton from price that equals marginal cost, particularly as average component cost per ton compound feed equal marginal cost because of constant returns to scale and size in component quantities

⁶² The revenue maximizing output combination is exogenous in a cost function context in any case. However, total component demand can be used as a measure for firm size because it is equal to total compound feed production, i.e. total output, which is a widely used measure for firm size.

⁶³ See section 2.2.

⁶⁴ See section 3.3.1, however. For the generalized Euler theorem, see CHAMBERS 1988: 39.

(C.12).⁶⁵ The result corresponds to output units rather than input units, as would be adequate for an aggregate non-component price. This objection can be overcome by conceding that, c^2 being independent of y , aggregate non-component price is constant with respect to output quantity, and thus it does not matter to which variable it is related. But, and this can not be overcome, marginal costs of the compound feed firm are again not the plain sum of the derivatives of C^1 and c^2 for y : the macro cost function C^* aggregates total cost from C^1 which can be interpreted as component price weighted firm size and from c^2 , but the aggregation rule can be anything, not just plainly adding up C^1 and c^2 . Hence, the project of measuring c^2 directly must be given up.

A partial solution – always keeping in mind that C.6 is erroneously forced to vanish in any case – can be obtained by considering share equations rather than demand equations. Since the exterior derivative is the same for all x_i , it cancels itself out when a component demand equation is divided by the sum of all component (!) demand quantities:

$$\frac{x_i}{\sum_{j=1}^n x_j} = \frac{\frac{\partial C^{**}[C^1(\mathbf{w}^1, \mathbf{y}), c^2(\mathbf{w}^2)]}{\partial C^1(\mathbf{w}^1, \mathbf{y})} \cdot \frac{\partial C^1(\mathbf{w}^1, \mathbf{y})}{\partial w_i}}{\frac{\partial C^{**}[C^1(\mathbf{w}^1, \mathbf{y}), c^2(\mathbf{w}^2)]}{\partial C^1(\mathbf{w}^1, \mathbf{y})} \cdot \sum_{j=1}^n \frac{\partial C^1(\mathbf{w}^1, \mathbf{y})}{\partial w_j}} = \frac{\frac{\partial C^1(\mathbf{w}^1, \mathbf{y})}{\partial w_i}}{\sum_{j=1}^n \frac{\partial C^1(\mathbf{w}^1, \mathbf{y})}{\partial w_j}} \quad \forall i \in I^1.$$

This is not too surprising since it exactly mirrors what assumption C.3 claims: component ratios do not depend on non-component prices. In general, share equations only allow inference on component mix, and thus own and cross price responses of component demand obtained by share equations must not be over-interpreted in using them to make demand level forecasts. They are ceteris paribus-elasticities abstracting from aggregate non-component costs which matter for component demand level. Beyond that, they abstract from the influence of "output quantity weighted aggregate component price" C^1 on total cost. However, in the special case of compound feed production, demand levels can be calculated if demand shares are known because of the equality of total component demand quantity and total output quantity, which is exogenous in a cost function context and thus also known.

The same problems with regard to measurability of c^2 apply to the marginal cost functions

⁶⁵ The independence of c^2 from y as such illuminates again how erroneous a vanishing C.6 is: here, dual non-component cost are a constant with respect to output quantity.

$$\frac{\partial C}{\partial y_i} = \frac{\partial C^1(\mathbf{w}^1, \mathbf{y})}{\partial y_i} \cdot \frac{\partial C^{**}[C^1(\mathbf{w}^1, \mathbf{y}), c^2(\mathbf{w}^2)]}{\partial C^1(\mathbf{w}^1, \mathbf{y})} \quad \forall i.$$

Hence, as with the conditional demand equations above, the only possibility – still keeping in mind that C.6 erroneously vanishes with this cost function specification – consists in dividing through by something that cancels out the exterior derivative. A meaningful divisor is the sum of all marginal cost as the results are marginal cost shares of the respective compound feed:

$$\frac{\frac{\partial C}{y_i}}{\sum_{j=1}^n \frac{\partial C}{y_j}} = \frac{\frac{\partial C^{**}[C^1(\mathbf{w}^1, \mathbf{y}), c^2(\mathbf{w}^2)]}{\partial C^1(\mathbf{w}^1, \mathbf{y})} \cdot \frac{\partial C^1(\mathbf{w}^1, \mathbf{y})}{\partial y_i}}{\frac{\partial C^{**}[C^1(\mathbf{w}^1, \mathbf{y}), c^2(\mathbf{w}^2)]}{\partial C^1(\mathbf{w}^1, \mathbf{y})} \cdot \sum_{j=1}^n \frac{\partial C^1(\mathbf{w}^1, \mathbf{y})}{\partial y_j}} = \frac{\frac{\partial C^1(\mathbf{w}^1, \mathbf{y})}{\partial y_i}}{\sum_{j=1}^n \frac{\partial C^1(\mathbf{w}^1, \mathbf{y})}{\partial y_j}} \quad \forall i, j.$$

The nice side-effect occurs that these equations can be obtained without the assumption of equal non-component – component marginal cost shares for all compound feed, because the "shares" are marginal cost shares rather than marginal component cost shares. However, the decomposition of marginal cost as plain sum of marginal component cost and marginal non-component cost would be wrong in any case if the present non-additive cost structure is considered.⁶⁶

Up to this point, the project of estimating at least a system of share equations, subject to the erroneous C.6 assumption, looks promising. But the constant returns to scale property of the cost function has disastrous consequences for measurability: the optimization rule "price equals marginal cost" does not apply in this case. Hence, marginal cost cannot be measured by feed price,⁶⁷ and the left hand side of a part of the system equations cannot be evaluated so that the system is not estimable.⁶⁸

⁶⁶ See section 3.3.3 above.

⁶⁷ The consideration of homogeneity which lies behind separability in the logical sequence of going through all possible structural assumptions is anticipated here in order to motivate postulation of another separable cost structure, the second alternative below. A treatment of the homogeneity property can be found in section 3.4.2 below.

A similar objection can be raised against the desired specification postulated in section 3.2 above, which has however failed already with respect to separability so that this objection could as well be omitted: constant returns to scale render measurement of marginal cost by price impossible. But in contrast to the alternative considered in this section aiming at an estimation of marginal total cost, the desired specification further above yields marginal feed cost equations. These are characterized by constant returns to scale independent-

A way to escape immeasurability of marginal cost would exist if feed price were not constant, because in this case, marginal cost were measurable as marginal revenue, no matter if returns to scale are constant or not. Two hypotheses are available: first, the premise of perfect competition, introduced as second general premise in section 3.1, could be given up. But, without having obtained a measure of concentration on the feed markets, it seems evident that significant market power on the feed markets does not exist. Secondly, a solution could be found in an elegant redefinition of the compound feed firm: if transportation activities were excluded from the compound feed firm activities, i.e. feeds were sold at the plant rather than delivered to the customer, the compound feed producer faced decreasing prices with rising distance from his plant. Thus, marginal cost were measurable as variable marginal revenue. But, on the one hand, the model were still arguably inconsistent because inseparability of $C.6$ is not accounted for and, on the other hand, one is left with the problem how to model variable prices which depend on the location of the single plant from aggregated sector information. Such a project is beyond the scope of this study and thus not followed.

To summarize the first alternative: even if inseparability of $C.6$ is given up, this neither leads to a system of estimable demand and marginal cost equations nor to a system of estimable demand and marginal cost share equations, because maintained structural assumptions in conjunction with a vanishing $C.6$ imply constant returns to scale for the overall cost function, which renders measurement of marginal cost and marginal cost shares impossible.

Now, consider the second alternative. A cost structure consistent with all maintained inseparable relations, but leaving two equations unrestricted that are supposed to vanish ($C.2$ and $C.7$) is

$$C(\mathbf{w}^1, \mathbf{w}^2, \mathbf{y}) = C^* \left[C^1(\mathbf{w}^1, \mathbf{y}), C^2(\mathbf{w}^2, g(\mathbf{y})) \right].$$

As partial differentiation reveals, this structure accounts for inseparability in $C.1$, $C.4$, and $C.6$, i.e. does not imply constant returns to scale of C , and it superfluously accounts for inseparability

ly of the non-component sector, whereas the marginal cost equations in this section imply constant returns to scale only if the non-component sector corresponds to a linearly homogeneous technology too.

⁶⁸ At this point, one has to resist the temptation of neglecting the marginal cost equations and estimating a system of component demand share equations alone: in the context of a dual cost model, there is no alternative to considering the complete system of derived equations since both the properties of a well-behaved cost function and the maintained structural assumptions imply so many restrictions between any equations that a neglect of some equations would inevitably lead to overall inconsistency.

ty in *C.2* and *C.7*. Thus, it is admissible in that it does not violate maintained inseparability assumptions, i.e. this specification leaves room for any effect which is likely to occur so that weak separability provides no reason against it. But it is again no equivalent representation of either *C.1* through *C.9* or of the same list of assumptions but *C.2* and *C.7* not vanishing.

With regard to the desired specification postulated in section 2.1 above, this structure is much too general: a CES specification merely constitutes a special case of it. That is, if a CES cost function is adopted while the above formula is adequate, existing effects are not depicted. Oddly enough, the CES structure satisfies all nine equations *C.1* through *C.9*, i.e. necessity of a CES structure for equivalence with *C.1* through *C.9* – in contrast to both specifications proposed in this section! But this is no reason to breathe again and rehabilitate the CES structure: the restrictiveness of the CES form, compared with the above structure, is merely more veiled. Exploring the premises of an additive structure in the binary partition reveals that, as set out in the first paragraph of this section, not even separability of all nine equations would supply the sufficient condition to establish a plus sign inside C^* .⁶⁹ This means that, although a CES cost structure does not violate any maintained inseparability assumptions – i.e. satisfies the necessary condition for equivalence –, effects must be expected to exist which are not depicted by the CES structure. If one wants to be sure that all effects which are not excluded by *C.1* through *C.9* are depicted, it is unavoidable to find the functional structure which follows from *C.1* through *C.9* – i.e. the sufficient condition for equivalence.

The ratio behind this second structure is as follows: as in the preceding example, the separable or decentralized decision process consists of a macro cost function, here C^* , that aggregates component cost C^1 which is an arbitrary function of component prices and output quantities, optimized separately from anything else, and non-component cost. But in the present case, non-component cost C^2 is a function of all non-component prices and an aggregate g of output quantities, which could reasonably correspond to the plain sum of all compound feed quantities, i.e. an easy-to-obtain measure for firm size. As suggested by the discussion of *C.6* in section 3.3.1 above, non-component mix now depends on firm size, which can be checked by deriving for a given pair of non-component prices. In addition, this specification gives room for scale effects since C^2 , being an arbitrary function of non-component prices and firm size, can be allowed to increase under-proportionally with rising $g(\mathbf{y})$.

Component demand assumes the form

⁶⁹ See FEGER/MÜLLER 1999: 3 and 10-12. See section 4.3.1 below for the opposite case.

$$x_i = \frac{\partial C(\mathbf{w}^1, \mathbf{w}^2, \mathbf{y})}{\partial w_i} = \frac{\partial C^1(\mathbf{w}^1, \mathbf{y})}{\partial w_i} \cdot \frac{\partial C^* [C^1(\mathbf{w}^1, \mathbf{y}), C^2(\mathbf{w}^2, g(\mathbf{y}))]}{\partial C^1(\mathbf{w}^1, \mathbf{y})} \quad \forall i \in I^1.$$

and marginal cost can be written as

$$\begin{aligned} \frac{\partial C}{\partial y_i} &= \frac{\partial C^1(\mathbf{w}^1, \mathbf{y})}{\partial y_i} \cdot \frac{\partial C^* [C^1(\mathbf{w}^1, \mathbf{y}), C^2(\mathbf{w}^2, g(\mathbf{y}))]}{\partial C^1(\mathbf{w}^1, \mathbf{y})} \\ &+ \frac{\partial C^2(\mathbf{w}^2, g(\mathbf{y}))}{\partial y_i} \cdot \frac{\partial C^* [C^1(\mathbf{w}^1, \mathbf{y}), C^2(\mathbf{w}^2, g(\mathbf{y}))]}{\partial C^2(\mathbf{w}^2, g(\mathbf{y}))} \end{aligned} \quad \forall i.$$

Clearly, one measurability problem with the first alternative regarding non-component cost also applies here since the proposed structure is more general than the above one: like c^2 , non-component cost C^2 cannot be observed. Hence, as both the demand and the marginal cost equations depend on C^2 , these are inestimable in analogy to the first alternative. In contrast to the first alternative however, marginal costs are measurable by product price, since the cost structure allows for non-constant returns to scale so that the optimization rule "price equals marginal cost" applies. But this generality gain leads to another problem: whereas the component demand share equations of the cost structure considered here are equivalent to the ones of the first alternative (since the exterior derivative cancels itself out because of its independence of \mathbf{w}^2), marginal cost, this time more complex, cannot be simplified as above by constructing share equations:

$$\frac{\partial C}{\partial y_i} = \frac{\frac{\partial C^*}{\partial C^1} \cdot \frac{\partial C^1}{\partial y_i} + \frac{\partial C^*}{\partial C^2} \cdot \frac{\partial C^2}{\partial y_i}}{\sum_{j=1}^n \frac{\partial C}{\partial y_j}} = \frac{\frac{\partial C^*}{\partial C^1} \cdot \sum_{j=1}^n \frac{\partial C^1}{\partial y_j} + \frac{\partial C^*}{\partial C^2} \cdot \sum_{j=1}^n \frac{\partial C^2}{\partial y_j}}{\sum_{j=1}^n \frac{\partial C}{\partial y_j}} \quad \forall i, j.$$

Because \mathbf{y} is included both into C^1 and C^2 , the exterior derivatives cannot be singled out of the sum, and thus cancellation fails. Hence, the right hand side of the marginal cost share equations cannot be made independent of the unavailable non-component price vector \mathbf{w}^2 or non-component cost C^2 , respectively. Consequently, the system of demand share and marginal cost share equations cannot be estimated.

To summarize the second alternative: in contrast to the first alternative, the proposed cost structure is theoretically admissible but, just like the former, not estimable – for different reasons: both the derived demand and marginal cost functions and the respective share equations

still depend on non-component prices, which are not available. Thus, the measurability of marginal cost, gained with the second approach, cannot be utilized.

3.4.2 Nonjointness, Homotheticity, and Constant Returns to Scale in Components

Since the maintained separability assumptions do not suffice to establish a cost structure that allows neglect of non-component cost, the remaining structural assumptions of constant returns in component quantities (C.12),⁷⁰ nonjointness in component quantities (C.13), and homothetic feed composition (C.14) are virtually useless because they all relate to component cost only: they simplify component cost, but do not touch the problems left by the preceding section. They will be discussed in spite of this because, in case once sufficient data on non-component prices and quantities becomes available, the second above alternative is consistently estimable so that the implementation of all maintained assumptions is meaningful.

Nonjointness in component quantities implies that the multi-output cost function can be written as plain sum of all single-output cost functions, where each of these contains the respective output y_i alone as an argument rather than the vector of all outputs \mathbf{y} .⁷¹ Thus, by C.13,⁷² component cost becomes

$$C^1(\mathbf{w}^1, \mathbf{y}) = \sum_{i=1}^m C_i^1(\mathbf{w}^1, y_i).$$

Homotheticity, in analogy with the introduction of output aggregator g in non-component cost C^2 in section 3.4.1 above, is equivalent to input-output separability of the cost function. A component cost function representing C.14 is

$$C^1(\mathbf{w}^1, \mathbf{y}) = h(\mathbf{y}) \cdot c^1(\mathbf{w}^1).⁷³$$

Output quantity still influences total component cost, but not the ratio of any derived component demands, since the micro function c^1 is independent of output quantity. Substituting for C_i^1 in the nonjoint component cost function yields

⁷⁰ The assumption of constant returns to scale was already anticipated in the preceding section because it is responsible for the failure of one trial to establish an estimable system of derived equations

⁷¹ See HALL 1973: 884-885; proof *ibid*: 890.

⁷² See section 3.3.3 above.

⁷³ See CHAMBERS 1988: 69-77.

$$C^1(\mathbf{w}^1, \mathbf{y}) = \sum_{i=1}^m h_i(y_i) \cdot c_i^1(\mathbf{w}^1).$$

Under the hypothesis that the underlying component composition technology is not only homothetic but additionally shows constant returns to scale and size, h_i simplifies to

$$h_i(y_i) = y_i,$$

so that component cost assumes the form

$$C^1(\mathbf{w}^1, \mathbf{y}) = \sum_{i=1}^m y_i \cdot c_i^1(\mathbf{w}^1).$$

The component price aggregators $c_i^1(\mathbf{w}^1)$ can now be interpreted as a unit cost of y_i or, referring to the two stage minimization view of a separable cost function, as an internal price of y_i ; component costs increase proportionally with increasing output, which is exactly what the hypothesis of constant returns to scale means.

If the enterprise of neglecting non-component cost due to separability assumptions had succeeded, it would have been possible to test C.12 through C.14 – given that the data provides enough degrees of freedom to do without their implementation – and leave the decision of whether the respective assumptions hold to empiry.

4 A Profit Function Model of the Compound Feed Firm

After the unsuccessful trial, undertaken in the previous chapter, to establish a cost function model of the compound feed firm which is simultaneously consistent with economic theory and reasonable structural assumptions on the one hand and estimable with regard to the available data on the other hand, this chapter considers another approach to depicting its behavior, namely a dual profit function. On the one hand, a profit function approach, including the cost function as a special case, is superior in general, which is explained in section 4.1.1 below, and, on the other hand, some problems responsible for the failure to establish a cost model which is theoretically consistent and simultaneously estimable do not apply in the profit function context. The latter are identified in section 4.1.2 below. Section 4.2 examines the structural assumptions relevant for a profit function model of the compound feed firm or industry, respectively. In section 4.3, possible specifications relying on the assumptions maintained in section 4.2 are discussed with regard to their operationality. Finally, section 4.4 compares the cost function and the profit function account to modeling the compound feed sector, starting at a digression on the epistemological status of structural assumptions in economic theory in general. For an abstract of the results of this chapter and the previous chapter refer to section 1.3.1 above.

4.1 Turning from a Cost to a Profit Model

4.1.1 Relative Advantages of Dual Feed Cost and Profit Models

All available studies on feed component demand are based on a cost function rather than on a profit function approach. This suggests itself because of the common focus on the input side; compound feed supply seems to be found relatively uninteresting.⁷⁴ In addition, a cost function approach is more convenient if output quantity data is better than output price data, or if only output quantities rather than prices are available. Profit maximization yields optimal input and output quantities whereas the cost function assumes the quantities of output produced as given. In the single-output case, the decision on the optimal output quantity reduces to the

⁷⁴ See e.g. PEETERS/SURRY 1993, BOYD/BROSEN 1986, MERGOS/YOTOPOULOS 1988, and SURRY/MOSCHINI 1984.

choice of the point on the production frontier where marginal revenue or price, respectively, equals marginal cost; only the decision on how much single output to produce is not covered by the cost function.

Whereas this does not call for a profit function, the picture changes if multiple outputs are present:⁷⁵ the question of level and shares of output quantities cannot be neglected anymore because the choice of the output combination and level determines the adequate point of evaluation of the demand function derived from a cost function. Since cost-minimizing demand $\mathbf{x}(\mathbf{w}, \mathbf{y})$ derived by Shephard's lemma depends on output quantities, forecasts are based on expectations or scenarios of these quantities, which are only accidentally profit-maximizing. In contrast, the profit-maximizing demand $\mathbf{x}(\mathbf{w}, \mathbf{p})$ derived with Hotelling's lemma uses output prices as exogeneous variables (which are preferable as scenario-tools), and the supply function $\mathbf{y}(\mathbf{w}, \mathbf{p})$, also derived applying Hotelling's lemma, endogeneously provides optimal output quantities. Therefore, in a sector where products are sold in a competitive market, a profit function approach is appropriate.⁷⁶

The difference between cost minimization and profit maximization, which also determines optimal output quantities, can be shown by analogy to the Slutsky-decomposition, familiar from consumer demand theory. It is possible to decompose the profit-maximizing derived demand response to changes in w_i and p_k in the following way:⁷⁷

$$\frac{\partial x_i(\mathbf{p}, \mathbf{w})}{\partial p_j} = \sum_{k=1}^m \frac{\partial x_i(\mathbf{w}, \mathbf{y}^*)}{\partial y_k} \cdot \frac{\partial y_k(\mathbf{p}, \mathbf{w})}{\partial p_j}$$

$$\frac{\partial x_i(\mathbf{p}, \mathbf{w})}{\partial w_j} = \frac{\partial x_i(\mathbf{w}, \mathbf{y}^*)}{\partial w_j} + \sum_{k=1}^m \frac{\partial x_i(\mathbf{w}, \mathbf{y}^*)}{\partial y_k} \frac{\partial y_k(\mathbf{p}, \mathbf{w})}{\partial w_j}$$

with \mathbf{y}^* denoting the profit maximizing output vector. The first fraction on the right hand side of both equations is the cost-minimizing derived demand response to changes in y_k and w_i , respectively. The second right hand side fraction in the first equation corresponds to the change in profit-maximal output caused by a change in that output's price. In the second equation, the effect of a change of an input price on the demand for an input is decomposed into a direct, cost-minimizing factor substitution and an indirect, profit-maximizing output substitution ef-

⁷⁵ This is the reason why there is no profit function analogy in the consumer demand context; there, only one output is produced: utility.

⁷⁶ See FÄRE/PRIMONT 1995: 2.

⁷⁷ See CHAMBERS 1988: 275 and 132-134 for the single-product version.

fect, where demand decreases via the output-decreasing effect of a rising input price. The former can be visualized as a substitution effect along an isoquant due to cost minimization, and the latter corresponds to a movement across isoquants associated with the output change induced by the factor price change.

In the case of the compound feed industry, it seems reasonable to expect considerable influence of changes in both input and output prices on the profit-maximizing demand and supply vector compared to the results of a cost function model. I.e. the magnitudes of the gradients of $x(\mathbf{w}, \mathbf{y})$ in \mathbf{y} and of $y(\mathbf{w}, \mathbf{p})$ in \mathbf{w} and \mathbf{p} must be expected to differ significantly from zero. In favor of a cost function approach it could be objected that the compound feed industry has no leeway in revenue maximization at their disposal, because the livestock industry determines the quantities traded. While this may be true in the short run, since livestock once set up entertains a fixed feed demand for an extremely broad price band, livestock production heavily depends on feed prices in the long run. Thus, the question of whose elasticities are greater and who influences market prices and quantities for compound feed more, respectively, cannot be answered a priori. Empirical evidence would, of course, require the estimation of a profit function model and a statistical test of the respective null-hypotheses, and cannot be gained using a cost function approach.

4.1.2 Differences Suggesting a Successful Profit Model

Three differences between the cost and profit model, relevant with respect to the failure of the cost function approach, can be distinguished. Two of them follow from the fact that the exogeneous variables of the dual profit function are input and output prices rather than input prices and output quantities.

First, the partial derivatives of the profit function for the output variable yield the conditional supply equations rather than marginal cost. This is promising since one – though inconsistent – way of reaching estimability of the cost function model only fails because of problems in measuring marginal cost,⁷⁸ whereas supply quantities are available. Derivation for input prices leads to conditional demand equations in both cases, which are however not equivalent.⁷⁹

⁷⁸ See section 3.4.1 above.

⁷⁹ See section 4.1.1 above.

Secondly, the dual profit function is linearly homogeneous in all its exogenous variables, whereas the dual cost function is generally inhomogeneous in the output variables or homogeneous of degree one in both input prices and output quantities in the case of constant returns to scale, i.e. homogeneous of degree two at total, respectively. This lack of linear homogeneity of the cost function contradicts the possibility to infer from strong separability to an additive structure of the behavioral function, which, as shown elsewhere, exists for the profit function.⁸⁰ Furthermore, from the outset it foils the plan to find a simple decomposition of the profit function by Euler's theorem in order to make aggregate non-component cost measurable on the other hand.⁸¹ Hence, there is enough incentive to work out a profit function model of the compound feed firm. This is the more indicated since, apart from these chances with respect to the concrete problems in modeling compound feed behavior, a profit function approach is generally superior to a cost function approach.⁸²

But, as a third major difference to the cost function approach, a new strand of problems arises since the profit function is not defined for constant returns to scale; the profit maximization problem is unbounded in this case. Thus, since the constant returns to scale hypothesis for the component micro function cannot be avoided, the whole profit function project finally fails. As in the cost function chapter above, "skipping the wrong way" is not favoured for reasons which are explained in the programmatic section 1.3.1 above.

4.2 Structural Assumptions

After the extensive discussion of the previous chapter it is of course no longer possible to assume an innocent perspective and postulate a desired model specification as was undertaken in section 3.2 above before carefully examining the adequacy of relevant structural assumptions and the functional representation consequences of these. Hence, one proceeds immediately to the list of structural assumptions on the profit function, which in any case look quite similar to their cost function analogies. The assumption of equal non-component cost-total cost proportions C.11 has no analogy in the profit function case, and thus II.11 is omitted.

⁸⁰ See FEGER/MÜLLER 1999.

⁸¹ See section 3.4.1 above.

⁸² See section 4.1.1 above.

4.2.1 Weak Separability

Generally, it cannot be inferred from a separable cost structure to a separable profit structure without further considerations. An example is provided by Feger's and Müller's establishment of a CES profit representation of separability in the binary and extended partition, which may not be transferred to the cost function for the different homogeneity properties of the cost and profit function.⁸³ Their proof is utilized in section 3.4.1 above and in section 4.3.1 below. Hence, the relevant equations have to be examined anew. In analogy with the cost function approach, weak separability of the compound feed firm's profit function requires checking the following six equations for the mutually exclusive and exhaustive sectors w^1 (corresponding to index set I^1) including all component prices, w^2 (corresponding to index set I^2) denoting all non-component prices, and output price vector p (corresponding to index set J):⁸⁴

$$\text{II.1} \quad \frac{\partial}{\partial w_k} \frac{\partial \Pi(\mathbf{w}, \mathbf{p}) / \partial p_i}{\partial \Pi(\mathbf{w}, \mathbf{p}) / \partial p_j} = \frac{\partial}{\partial w_k} \frac{y_i}{y_j} = g_{ijk}(\mathbf{w}, \mathbf{p}) \quad \forall i, j \in J; k \in I^1$$

$$\text{II.2} \quad \frac{\partial}{\partial w_k} \frac{\partial \Pi(\mathbf{w}, \mathbf{p}) / \partial p_i}{\partial \Pi(\mathbf{w}, \mathbf{p}) / \partial p_j} = \frac{\partial}{\partial w_k} \frac{y_i}{y_j} = 0 \quad \forall i, j \in J; k \in I^2$$

$$\text{II.3} \quad \frac{\partial}{\partial w_k} \frac{\partial \Pi(\mathbf{w}, \mathbf{p}) / \partial w_i}{\partial \Pi(\mathbf{w}, \mathbf{p}) / \partial w_j} = \frac{\partial}{\partial w_k} \frac{x_i}{x_j} = 0 \quad \forall i, j \in I^1; k \in I^2$$

$$\text{II.4} \quad \frac{\partial}{\partial p_k} \frac{\partial \Pi(\mathbf{w}, \mathbf{p}) / \partial w_i}{\partial \Pi(\mathbf{w}, \mathbf{p}) / \partial w_j} = \frac{\partial}{\partial p_k} \frac{x_i}{x_j} = g_{ijk}(\mathbf{w}, \mathbf{p}) \quad \forall i, j \in I^1; k \in J$$

$$\text{II.5} \quad \frac{\partial}{\partial w_k} \frac{\partial \Pi(\mathbf{w}, \mathbf{p}) / \partial w_i}{\partial \Pi(\mathbf{w}, \mathbf{p}) / \partial w_j} = \frac{\partial}{\partial w_k} \frac{x_i}{x_j} = 0 \quad \forall i, j \in I^2; k \in I^1$$

$$\text{II.6} \quad \frac{\partial}{\partial p_k} \frac{\partial \Pi(\mathbf{w}, \mathbf{p}) / \partial w_i}{\partial \Pi(\mathbf{w}, \mathbf{p}) / \partial w_j} = \frac{\partial}{\partial p_k} \frac{x_i}{x_j} = g_{ijk}(\mathbf{w}, \mathbf{p}) \quad \forall i, j \in I^2; k \in J$$

These assumptions can be justified as follows:

II.1 In analogy to C.1, profit maximizing output mix depends on component prices because the compound feeds are generally characterized by different component intensities; the

⁸³ See FEGER/MÜLLER 1999.

⁸⁴ See section 3.3.1 above.

entrepreneur will *ceteris paribus* always substitute away from the compound feed which becomes more expensive due to a rising component price.

- Π.2 In analogy to C.2 and in opposition to Π.1, output mix is independent of changes in non-component prices because processing is the same for all feeds.
- Π.3 Being identical with the justification of C.3 and paralleling Π.2, it is irrelevant for the feed mixer to know about price changes for factors which are not components.
- Π.4 In analogy with C.4 and paralleling Π.1, knowledge about an output price change will not alter the least-cost recipe for the livestock category-specific compound feed – compound feed cost and profit is nonjoint in inputs –,⁸⁵ but through the product-price induced shift in revenue-maximal product mix the component demand shares shift, too, since in different compound feeds components are utilized with different intensity.⁸⁶
- Π.5 The argument is the same as for C.5: processing technology treats all components the same so that a component price change is meaningless.
- Π.6 Paralleling C.6, an increasing compound feed price yielding a shift in product mix *and*, by Hotelling's lemma, c. p. an increasing total output causes factor substitution in the non-component sector. As in the cost function case, to claim separability here would imply the denial of economies of scale: the non-component sector is the only place where these can be located, because the component sector is, as assumption Π.12 below states, characterized by constant returns to scale.

Thus, it is neither possible to establish a complete weakly separable cost function nor a weakly separable profit function for the compound feed firm. In both approaches, the output and component sector are mutually inseparable so that they could be united in a single sector with all kinds of interactions, i. e. there is no separability in the extended partition. For both approaches, the component and the non-component sector are mutually weakly separable. Hence, both the cost and the production model can be called "separable in the binary partition in inputs".

⁸⁵ See section 4.1.4 below.

⁸⁶ Note that in a single product firm the homothetic compound feed technology would be equivalent with separability of components from output price.

4.2.2 Strong Separability

Now consider the three additional equations constituting strong separability for the profit function:

$$\text{II.7} \quad \frac{\partial}{\partial w_k} \frac{y_i}{x_j} = 0 \Leftrightarrow \frac{\partial}{\partial w_k} \frac{x_j}{y_i} = 0 \quad \forall i \in J; j \in I^1; k \in I^2$$

$$\text{II.8} \quad \frac{\partial}{\partial w_k} \frac{y_i}{x_j} = g_{ijk}(\mathbf{w}, \mathbf{p}) \Leftrightarrow \frac{\partial}{\partial w_k} \frac{x_j}{y_i} = \tilde{g}_{ijk}(\mathbf{w}, \mathbf{p}) \quad \forall i \in J; j \in I^2; k \in I^1$$

$$\text{II.9} \quad \frac{\partial}{\partial p_k} \frac{x_i}{x_j} = g_{ijk}(\mathbf{w}, \mathbf{p}) \Leftrightarrow \frac{\partial}{\partial p_k} \frac{x_j}{x_i} = \tilde{g}_{ijk}(\mathbf{w}, \mathbf{p}) \quad \forall i \in I^1; j \in I^2; k \in J$$

These assumptions can be justified as follows:

II.7 The necessary condition for II.7 expressing strong separability is satisfied by II.2 and II.3. Sufficiency is, as in C.7, again implied by the catalyst character of all non-component activities.

II.8 The argument is the same as for C.8.

II.9 The argument is the same as for C.9.

4.2.3 Unity Value of the CES Substitution Parameter

If, relying on the separability assumptions maintained above, it were possible to establish a CES profit representation, a literally additive profit function would require a unity valued CES substitution parameter, too. In this case, the profit function entertains a linear relationship between component and non-component profit micro-functions. As with the cost function, this implies a Leontief technology characterized by right-angled isoquants between aggregate components and aggregate non-components or no substitution between these aggregates. There is argument in favor of this hypothesis in section 3.3.3 above. Thus,

$$\text{II.10} \quad \rho(\mathbf{y}) = 1$$

can be maintained.

4.2.4 Nonjointness, Homotheticity, and Constant Returns to Scale in Components

Since the assumptions of production nonjoint in component quantities, homothetic feed composition and constant returns to scale in component quantities all correspond to technology rather than dual functions, the argumentation in sections 3.3.5 through 3.3.7 applies without limitation in a profit function context. Hence, constant returns to scale in component combination

$$\text{II.12} \quad Y(\mathbf{x}^1, \mathbf{y}) = 0 \Rightarrow Y(\lambda \mathbf{x}^1, \lambda \mathbf{y}) = 0 \quad \forall \lambda > 0,$$

nonjointness in component quantities

$$\text{II.13} \quad y_i = Y_i^1(\mathbf{x}^1) \quad \forall i,$$

and, formulated in terms of the profit function, homothetic feed composition in the single product case, which is already established by II.13,

$$\text{II.14} \quad \frac{\partial}{\partial y_k} \frac{\partial C_k^1(\mathbf{w}, y_k) / \partial w_i}{\partial C_k^1(\mathbf{w}, y_k) / \partial w_j} = \frac{\partial}{\partial y_k} \frac{x_i^k}{x_j^k} = 0 \quad \forall i, j \in I^1; k \in J.$$

are all valid.⁸⁷

4.3 Model Specification

4.3.1 Separability

While, given the partition established in section 3.3 above, a CES form cost representation would not be established even if all equations C.1 through C.9 could realistically be assumed to vanish, i.e. if strong separability of the cost function in the binary and extended partition could be maintained,⁸⁸ the opposite is the case in a profit function framework: It can be shown that a profit function which is strongly separable in the binary and extended partition in conjunction, i.e. II.1 through II.9 vanish, assumes a CES structure.⁸⁹ Then, by the adequate assumption of

⁸⁷ The ordering number 11 is omitted to ease comparison between cost function and profit function structural assumptions. There does not exist a profit function analogy to C.11, and the remaining ordering numbers refer to the respective analogy.

⁸⁸ See section 3.4.1 above.

⁸⁹ See FEGGER/MÜLLER 1999.

unity value of the CES substitution parameter $\Pi.10$, justification of a literally additive structure of compound feed profit would be possible. But the preceding section reveals that only some separability assumptions can be maintained, whereas elsewhere room has to be left for inseparable behavior. Hence, as with the cost function, the quest for a representation of $\Pi.1$ through $\Pi.9$ or, less demanding, a profit structure that does not violate maintained inseparability, is indicated.

The first observation to be made is that, from the equivalence of vanishing $\Pi.1$ through $\Pi.9$ and a CES structure of the profit function, it follows that less than all nine equations vanishing render the establishment of an additive profit structure impossible. It seems to be not possible to find a genuinely economic argument that makes this result plausible with respect to economic intuition. Algebraic intuition can be approached by the hint that the proof heavily relies on the symmetry of the separability assumptions and the symmetry of the feed output, non-component input, and component input sector with respect to homogeneity of degree one in all variables, so that both an incomplete list of separability assumptions and asymmetric homogeneity, e.g. with the dual cost function, renders the proof impossible.⁹⁰

The analogy of $\Pi.1$ through $\Pi.9$ and $C.1$ through $C.9$ suggests an analogous proceeding: since, just as in the cost function case, a profit structure satisfying all nine equations and thus the necessary condition for equivalence could not be found,⁹¹ the two profit function relatives of the cost structures proposed in section 3.4.1 above are examined.

The first alternative, as in the cost function case, is more restrictive than admissible since it inadequately implies a separable $\Pi.6$:

⁹⁰ On the one hand, an algebraic theory like microeconomic theory is not more than a simplifying model based on some few axioms, trying to depict an infinitely complex world. On the other hand, like algebra itself, such a model leads its own life. For example, it is not possible to make plausible the astonishing properties of the number π or most results of theoretical physics. Considering microeconomic theory, an example for an unintuitive – though at least not contraintuitive! – result is the proof for the additive structure of a profit function which is equivalent with strict separability. If one decides to utilize microeconomic theory, it is eventually not relevant if a proven statement is intuitive or not: a formal proof does not need any intuitive support. In addition, an unintuitive result is not alarming in principle: if our intuition would suffice to understand all phenomena, theory would be superfluous.

⁹¹ As in the cost function case, there is an additive structure that satisfies the necessary condition for equivalence. But, proven in FEGER / MÜLLER 1999, sufficiency for an additive structure requires $\Pi.1$ through $\Pi.9$ to vanish, which is not the case, so that additivity is generally rejected.

$$\begin{aligned}\Pi(\mathbf{w}^1, \mathbf{w}^2, \mathbf{p}) &= \Pi^{**}[\Pi^1(\mathbf{w}^1, \mathbf{p}), \pi^2(\mathbf{w}^2)] \\ &= \Pi^{**}[\Pi^1(\mathbf{w}^1, \mathbf{p}), c^2(\mathbf{w}^2)].\end{aligned}$$

The possibility to interpret π^2 as micro cost function c^2 is provided for by the duality between profit and cost function: if a profit (micro-) function is independent of output price, i.e. an input price sector is weakly separable from \mathbf{p} , the resulting input price aggregate can equivalently be expressed as profit or cost function or internal price of the aggregate.⁹² Applying Hotelling's lemma and the chain rule, the derived compound feed supply and component demand equations assume the form

$$\begin{aligned}y_i &= \frac{\partial \Pi^1(\mathbf{w}^1, \mathbf{p})}{\partial p_i} \cdot \frac{\partial \Pi^{**}[\Pi^1(\mathbf{w}^1, \mathbf{p}), c^2(\mathbf{w}^2)]}{\partial \Pi^1(\mathbf{w}^1, \mathbf{p})} && \forall i; \\ x_i &= -\frac{\partial \Pi^1(\mathbf{w}^1, \mathbf{p})}{\partial w_i} \cdot \frac{\partial \Pi^{**}[\Pi^1(\mathbf{w}^1, \mathbf{p}), c^2(\mathbf{w}^2)]}{\partial \Pi^1(\mathbf{w}^1, \mathbf{p})} && \forall i \in I^1.\end{aligned}$$

Hence, both supply and component demand are, as with the cost function, not independent of non-component prices. The interpretation is completely analogous to the cost function case and is thus not reiterated.⁹³

Now, consider the plan to measure c^2 as a whole: In analogy to the first trial in the cost function context, one could, relying on the primal profit function, claim that c^2 is revenue minus total profit minus component expenditures. But the objections are analogous, too: on the one hand, total profit is not observable in the monthly frequency,⁹⁴ and, on the other hand, the correspondence between primal and dual functions is overtaxed here again, which becomes obvious in the specific structure of Π^{**} that aggregates "component profit" Π^1 and "non-component price" c^2 in an arbitrary, not necessarily additive way.

In contrast, the second approach to measurability, solving Euler's Theorem for c^2 , allows one more step than in the cost function context before it fails: Π is a homogeneous function and thus allows for the application of Euler's theorem. Set up in terms of prices, Hotelling's lemma reveals that the result is exactly the primal profit definition, so that there is nothing gai-

⁹² See CHAMBERS 1988: 154.

⁹³ See section 3.4.1 above.

⁹⁴ See chapter 2 above.

ned with regard to the first trial. But Π is not only a linearly homogeneous function of (\mathbf{p}, \mathbf{w}) , it is also a linearly homogeneous function of Π^1 and c^2 , which are both valid profit functions, and thus itself linearly homogeneous functions of their respective variables. Now, by Euler's theorem in its classical version for homogeneity of degree one,

$$c^2 = \frac{\Pi^{**}[\Pi^1(\mathbf{w}^1, \mathbf{p}), c^2(\mathbf{w}^2)] - \frac{\partial \Pi^{**}[\Pi^1(\mathbf{w}^1, \mathbf{p}), c^2(\mathbf{w}^2)]}{\partial \Pi^1(\mathbf{w}^1, \mathbf{p})} \cdot \Pi^1(\mathbf{w}^1, \mathbf{p})}{\frac{\partial \Pi^{**}[\Pi^1(\mathbf{w}^1, \mathbf{p}), c^2(\mathbf{w}^2)]}{\partial c^2(\mathbf{w}^2)}}.$$

This is a progress with regard to the first way because c^2 is consistently explained by other terms. But these cannot be observed: in the numerator, all terms are known apart from total profit, which could substitute Π^{**} . As in the cost function case, there is only firm-specific data on total profits, cumbersome to collect and published in yearly frequency by the bigger compound feed producers, rather than monthly sector aggregates. The situation is even more desperate with the denominator: there is no data on the profit change induced by changing non-component price, and there is no highly correlated instrument available by which monthly data could be reasonably simulated.

The third trial is not profit function-specific, and thus its rejection in section 3.4.1 above does not have to be repeated.

As with the cost function, a partial solution can be obtained by considering share equations:

$$\frac{y_i}{\sum_{j=1}^n y_j} = \frac{\frac{\partial \Pi^{**}[\Pi^1(\mathbf{w}^1, \mathbf{p}), c^2(\mathbf{w}^2)]}{\partial \Pi^1(\mathbf{w}^1, \mathbf{p})} \cdot \frac{\partial \Pi^1(\mathbf{w}^1, \mathbf{p})}{\partial p_i}}{\frac{\partial \Pi^{**}[\Pi^1(\mathbf{w}^1, \mathbf{p}), c^2(\mathbf{w}^2)]}{\partial \Pi^1(\mathbf{w}^1, \mathbf{p})} \cdot \sum_{j=1}^n \frac{\partial \Pi^1(\mathbf{w}^1, \mathbf{p})}{\partial p_j}} = \frac{\frac{\partial \Pi^1(\mathbf{w}^1, \mathbf{p})}{\partial p_i}}{\sum_{j=1}^n \frac{\partial \Pi^1(\mathbf{w}^1, \mathbf{p})}{\partial p_j}} \quad \forall i, j;$$

$$\frac{x_i}{\sum_{j=1}^n x_j} = - \frac{\frac{\partial \Pi^{**}[\Pi^1(\mathbf{w}^1, \mathbf{p}), c^2(\mathbf{w}^2)]}{\partial \Pi^1(\mathbf{w}^1, \mathbf{p})} \cdot \frac{\partial \Pi^1(\mathbf{w}^1, \mathbf{p})}{\partial w_i}}{\frac{\partial \Pi^{**}[\Pi^1(\mathbf{w}^1, \mathbf{p}), c^2(\mathbf{w}^2)]}{\partial \Pi^1(\mathbf{w}^1, \mathbf{p})} \cdot \sum_{j=1}^n \frac{\partial \Pi^1(\mathbf{w}^1, \mathbf{p})}{\partial w_j}} = - \frac{\frac{\partial \Pi^1(\mathbf{w}^1, \mathbf{p})}{\partial w_i}}{\sum_{j=1}^n \frac{\partial \Pi^1(\mathbf{w}^1, \mathbf{p})}{\partial w_j}} \quad \forall i \in I^1.$$

Again, the respective structures are merely expressions of structural assumptions maintained above: by $\Pi.3$, component ratios do not depend on non-component prices, and, by $\Pi.2$, supply ratios do not depend on non-component prices, either. The inferential potency of share equations derived from the profit function has to be judged more critically than in the cost function

case: only component mix and supply mix can be forecasted, and level forecasts are rendered impossible because, in contrast to the cost function, total production level or firm size is endogenous. This shortcoming is by far compensated for by an advantage with respect to estimability: rather than marginal cost, of which measurement depends on the validity of the price-equals-marginal cost rule, the output share equations show supply quantities on the left hand side, which are available. Unfortunately, this promising strand comes to an end at another point which is demonstrated in section 4.3.2 below.

To summarize the first strand of argumentation: up to the current state of affairs, inadequately giving up inseparability of $\Pi.6$ at least helps to establish a system of estimable supply and demand share equations. Merely the project to formulate a system of estimable supply and demand equations fails because of the impossibility of evaluating either w^2 or c^2 .

Now, turn to the second alternative. A profit structure consistent with all maintained inseparable relations, but leaving $\Pi.2$ and $\Pi.7$ unrestricted in spite of these being supposed to vanish is

$$\Pi(w^1, w^2, p) = \Pi^*[\Pi^1(w^1, p), \Pi^2(w^2, g(p))].$$

The decentralized decision process is depicted by a macro profit function Π^* which optimizes the outcome of two sub-decision processes: component profit function Π^1 , which determines the profit-maximizing component and compound feed combination relying on component and compound feed prices exclusively, and non-component profit function Π^2 , where the optimal non-component combination is calculated relying on non-component prices. In contrast to the first alternative, Π^2 adequately considers an aggregate output price and only ignores the question of output mix. Both micro-functions share an influence on firm size, Π^1 on a disaggregate level and Π^2 on an aggregate one. In analogy with the cost function case, this specification allows for non-constant returns to scale.⁹⁵ Application of Hotelling's lemma yields

$$y_i(w^1, w^2, p) = \frac{\partial \Pi^1(w^1, p)}{\partial p_i} \cdot \frac{\partial \Pi^*[\Pi^1(w^1, p), \Pi^2(w^2, g(p))]}{\partial \Pi^1(w^1, p)} + \frac{\partial \Pi^2(w^2, g(p))}{\partial p_i} \cdot \frac{\partial \Pi^*[\Pi^1(w^1, p), \Pi^2(w^2, g(p))]}{\partial \Pi^2(w^2, g(p))} \quad \forall i$$

⁹⁵ But see section 4.3.2 below.

$$x_i(\mathbf{w}^1, \mathbf{w}^2, \mathbf{p}) = -\frac{\partial \Pi^1(\mathbf{w}^1, \mathbf{p})}{\partial w_i} \cdot \frac{\partial \Pi^*[\Pi^1(\mathbf{w}^1, \mathbf{p}), \Pi^2(\mathbf{w}^2, g(\mathbf{p}))]}{\partial \Pi^1(\mathbf{w}^1, \mathbf{p})} \quad \forall i \in I^1.$$

Clearly, these equations are also inestimable because Π^2 is not measurable, just as c^2 above is not measurable. But, in contrast to the first alternative, constructing share equations does not cure the inestimability: whereas the component demand share equations of the profit structure considered here are equivalent to the ones of the first alternative (since the exterior derivative cancels itself out because of its independence of \mathbf{w}^2), supply shares continue to depend on in-measurable terms:

$$\frac{y_i}{\sum_{j=1}^n y_j} = \frac{\frac{\partial \Pi^*}{\partial \Pi^1} \cdot \frac{\partial \Pi^1}{\partial p_i} + \frac{\partial \Pi^*}{\partial \Pi^2} \cdot \frac{\partial \Pi^2}{\partial p_i}}{\frac{\partial \Pi^*}{\partial \Pi^1} \cdot \sum_{j=1}^n \frac{\partial \Pi^1}{\partial p_j} + \frac{\partial \Pi^*}{\partial \Pi^2} \cdot \sum_{j=1}^n \frac{\partial \Pi^2}{\partial p_j}} \quad \forall i, j.$$

To summarize the second alternative: in contrast to the first alternative, the proposed profit structure is theoretically admissible, but, in contrast to the former at least in the share equations case, not estimable: both the derived supply and component demand functions and the respective share equations still depend on non-component prices, which are not available.

4.3.2 Nonjointness, Homotheticity, and Constant Returns to Scale in Components

As in the cost function case, separability assumptions alone render the estimation of a consistent model, namely the second cost function alternative, impossible because dependence on unavailable variables cannot be overcome, and the same is the case for the arguably inconsistent demand and marginal cost equations and supply and demand equations of the first cost function alternative. However, both the cost and the profit approach, considering separability assumptions alone but neglecting C.6 or $\Pi.6$, respectively, leave the opportunity to evaluate share equations. In both cases the reason why this last hope is also disappointed is the property of constant returns to scale. In the cost function chapter, this was anticipated in section 3.4.1 on separability to provide an additional motivation for the second alternative: it is not exposed to the problem caused by the constant returns to scale and fails for separability reasons alone. Here, in a profit function framework, the situation is different: Both alternatives, be it in the level or share version, are ruled out by constant returns to scale in component quantities, as will be shown immediately.

First, consider nonjointness in component quantities: the multi-output profit function, more precisely the micro function which depends on component quantities, can be written as the plain sum of all single-output profit functions, where each of these contains the respective output price p_i alone as an argument rather than the vector of all output prices \mathbf{p} .⁹⁶ Thus, by $\Pi.13$,⁹⁷ component cost becomes

$$\Pi^1(\mathbf{w}^1, \mathbf{p}) = \sum_{i=1}^m \Pi_i^1(\mathbf{w}^1, p_i).$$

Homotheticity, in analogy with the introduction of output aggregator g in non-component profit Π^2 in section 4.3.1 above, is equivalent to input-output separability of the profit function. A component profit function representing $\Pi.14$ is

$$\begin{aligned} \Pi^1(\mathbf{w}^1, \mathbf{p}) &= h(\mathbf{p}) \cdot \pi^1(\mathbf{w}^1) \\ &= h(\mathbf{p}) \cdot c^1(\mathbf{w}^1) \end{aligned}$$

Substituting for Π_i^1 in the nonjoint component profit function yields

$$\Pi^1(\mathbf{w}^1, \mathbf{p}) = \sum_{i=1}^m h_i(p_i) \cdot c_i^1(\mathbf{w}^1).$$

Now, the assumption of constant returns to scale in component quantities, $\Pi.14$, comes into play.⁹⁸ This inarguable property of the subset of technology depicting the combination of components to compound feed implies that Π^1 is not well-behaved.⁹⁹ In contrast to the cost minimization problem (or the revenue maximization problem), the profit maximization problem is unbounded for a linearly homogeneous technology, and thus the profit function is not defined in this case. To exemplify this result with the decentralized optimization of the compound feed

⁹⁶ See CHAMBERS 1988: 296.

⁹⁷ See section 4.2.3. above.

⁹⁸ See section 4.2.3 above.

⁹⁹ In a separable profit structure, not only the overall optimization function, i.e. the macrofunction, must be well-behaved: every microfunction is required to satisfy all properties of a profit function, and a solution to the optimization problem must exist. Otherwise, a decentralized optimization would be impossible: if only one of the micro- or macrofunctions is ill-behaved or not defined, the whole optimization problem is ill-behaved or not defined.

firm: the decider optimizing Π^1 , i.e. the "component profit maximizer", is not able to determine optimal component demand and optimal compound feed production simultaneously because doubling supply and component demand would always double profit – by definition of constant returns to scale.

A solution can only be obtained if information from outside the constant-returns-to-scale micro function enters the optimization procedure: firstly, it obviously suffices if part of the optimization problem is exogeneous to the respective micro function – in this case, the micro profit function reduces to a cost function where revenue maximizing output quantities are exogeneous, or to a revenue function with exogeneous cost minimizing input quantities, respectively. Secondly, the unboundedness is overcome if firm size is fixed, i.e. if a total amount of compound feed quantity or a total amount of component demand, respectively, taken as firm size measure, is given. But this scenario would always result in a corner solution: the complete firm output would consist in the most profitable output with zero quantities of all other outputs, which is obviously not the case in reality. The question of how the compound feed firm actually manages the profit maximization problem – and no one could seriously deny that it does – leads to a third approach. However, this approach is beyond the scope of this study, since it includes considerations about the structure of the underlying technology and, as is not clear how these should be mirrored in the structure of the profit function, this is therefore merely sketched in the outlook section 4.4.3 below.

4.4 Towards an Application

Possible routes towards an application of a behavioral model of the compound feed industry or towards the development of a both theoretically consistent and simultaneously estimable model, respectively, are approached from the bottom: first, the consequences of the failure of establishing both consistent and estimable models are analyzed theoretically in section 4.4.1. The discussion is meant to provide a general treatment of structural assumptions in the context of an axiomatic approach to applied economics. Then, section 4.4.2 gives a final evaluation of the developed models, whereas section 4.4.3 concludes the modeling part of this study with a discussion of possible ways out.

4.4.1 The Epistemological Status of Structural Assumptions

There are four possible methodical functions that motivate maintaining structural assumptions, all of which play a role in this study. First, structural assumptions save estimation parameters and thus degrees of freedom, so that statistical significance is enhanced and, at the margin, a model is made estimable in spite of scarce data. This is e.g. always true for the separability assumption implied by data aggregation. Secondly, some effects (like the allocation of components on compound feeds) can only be calculated if a specific functional structure is assumed (here outputs being nonjoint in input quantities).¹⁰⁰ Thirdly, some models (like the ones in this study) cannot be estimated at all without some structural assumptions (here separability) due to asymmetric data availability.¹⁰¹ Fourthly, examining the adequacy of a structural assumption may be an end in itself in order to gain general knowledge that can be presupposed in future studies. Another distinction can be made with regard to the justification of a structural assumption. In some cases, a structural assumption is valid a priori. Consider the following example: given that inefficiency can be excluded,¹⁰² the assumptions of constant returns to scale in component composition and nonjointness in components maintained in the present models are true by logical necessity; the constitution of a compound feed as a mixture of components implies their validity, and neither is an empirical test required to establish this result, nor can it be statistically rejected.

But in most cases, structural assumptions are, like economic theory as a whole and thus all deduced hypotheses, of empirical nature. In contrast to conceptually necessary structural assumptions which are true whatever purpose they serve, the methodical function of a structural assumption which cannot be validated without empiry – or, formulated properly within the deductionist approach, a structural assumption that is falsifiable¹⁰³ – has important implications on its epistemological status.¹⁰⁴ In some cases, it is possible to conduct a statistical test on the adequacy of the assumption, in other cases not. If a test is successfully passed, the assumption

¹⁰⁰ See section 3.2.2 above.

¹⁰¹ See section 3.2.1 above.

¹⁰² See section 3.1 above.

¹⁰³ See POPPER 1961: 40-41

¹⁰⁴ For a general treatment of epistemological aspects of economics and social sciences, respectively, see WALLACE 1971.

is justified as well as the whole model. Ironically, an assumption thus justified is superfluous if the first or third methodical function is the motive; testability indicates that there is no need to introduce it. But a statistical test is principally impossible if the model is inestimable without the assumption, which would, in case of a successful modeling, e.g. be the case with the separability assumptions in this study allowing the neglect of non-component price variables.

Such assumptions exhibit a queer in-between status: on the one hand, they do not follow from economic theory like the homogeneity or curvature property of dual behavioral functions, i.e. are not premises of the model that, in conjunction, are the object of the epistemological test implied by any application of a theory, and, on the other hand, they are not part of the empirical hypothesis that is tested with statistical means. They are empirical hypotheses, but stay unexamined. Rather, justification is sought outside the model. In the best case, justification is provided by other scientific studies, ideally significantly more than one, that all unambiguously support the adequacy of the respective assumption, i.e. that satisfy the fourth methodical function of maintaining structural assumptions, and there are no arguments against the adequacy of maintaining the analogy with the model in question. But in many instances, this justification is not available because a test of the respective assumption is generally excluded, if, for example, relevant data is never available.¹⁰⁵ Then, plausibility considerations and pre-theoretical experience with the object of the model are commonly drawn upon. Of course, this is neither theory proper nor empiry proper.

The implied epistemological problem is implied by the influential Duhem-Quine-thesis: if a hypothesis is rejected, it is impossible to infer which premise is responsible for failure.¹⁰⁶ According to this holism of justification, a model can only be rejected as a whole. Thus, if, as in the present study, a, strictly speaking, unjustified structural assumption enters the model, neither any other empirical hypothesis nor the underlying behavioral theory can in fact be tested. It follows that, in order to falsify a hypothesis or a theory, the model must at first be *made* falsifiable by reducing the number of open questions to unity, namely the question of whether the theoretically consistent model is capable of fitting the data.¹⁰⁷ Since this is impossible in the present study, all possible statistical inferences drawn move on thin epistemological ice.

¹⁰⁵ See above.

¹⁰⁶ See QUINE 1981.

¹⁰⁷ See section 6.2 below for another instance where falsifiability of the maintained hypotheses is considered.

The situation that all structural assumptions utilized in a model can be maintained relying on other – proper – scientific studies where in doubtlessly analogous cases the respective assumption was empirically accepted, can be expected to be encountered extremely rarely. The practice, and also the practice of the study at hand, is to be content with less, namely the epistemologically "soft" stage of justification of structural assumptions which correspond to "experience" with the object of the theory, i.e. pre-theoretical knowledge. In the framework of a scientific theory like e.g. neoclassical economic theory, any "knowledge" emanating from a source different from the respective scientific theory has to be considered non-theoretic or pre-theoretic, whatever sophistication of the human mind has brought it about, and in whatever number of cases it has turned out to fit the phenomena.

With respect to a proper scientific practice, the classification of information as pre-theoretic implies that it is irrelevant. But since this kind of knowledge often proves to be quite close to knowledge gained scientifically, such a pragmatic maintainance can be seen as a good proxy. Of course, the proxy use of soft knowledge is only justifiable in an applied context where a test of an empirical hypothesis is not considered, not to talk of an epistemological test of the underlying theory.¹⁰⁸ For applied purposes, structural assumptions which are maintained only relying on pre-theoretic knowledge can be expected to increase the precision of forecasts and are thus adequate since a theory or hypothesis, even if not truly justified, contains information that is, on average, more likely to be true than the null hypothesis of white noise, i.e. the absence of directed causal relationships which could be detected and modeled for use as a basis for forecast. However, even if a theory is closer to the truth than assuming white noise, the introduction of theory or structural assumptions need not necessarily result in an increasing forecast precision.¹⁰⁹

Applied economic practice takes even one step more: pre-theoretic knowledge is commonly used as a posterior credibility benchmark for statistically accepted models, i.e. somehow rated higher than scientific knowledge – because pre-theoretic experience shows that in many cases reliability of pre-theoretic knowledge is at least as high as the reliability of scientifically generated knowledge. One had better refrain from estimating weakly founded models for forecast

¹⁰⁸ This must not be interchanged with the use of pre-theoretic knowledge in inventing theories or models: the genesis of a scientific theory or an empirical hypothesis is not exposed to any criteria; it is only the validity of a theory or hypothesis that is the object of epistemological criticism.

¹⁰⁹ See section 6.2.4 below.

purposes in order to avoid the hypocritical acceptance of an estimation result which matches intuition where a result contradicting intuition would be rejected. If pre-theoretic knowledge is rated this highly, which may be perfectly appropriate with regard to the ends pursued, a scientific study is useless – except maybe to take advantage of scientific devoutness. In contrast, the benchmark use of pre-theoretic knowledge during the construction process of a scientific model, noting the crucial difference between genesis and validity of a theory or hypothesis with respect to justification, needs no defense.

4.4.2 Judging the Developed Feed Cost and Profit Model Pragmatically

Summing up the problems encountered, it is clear that an epistemological test neither of economic theory nor of the deduced empirical hypotheses nor of any structural assumption is possible. The only open question is whether an application of one of the inconsistent models could be of heuristic benefit. There are three candidates for an application. Consider the following table:

Specification	Consistency	Estimability
desired cost model: demand and marginal cost	no	no+
first cost model: demand and marginal cost	no	no
first cost model: share equations	no	no+
second cost model: demand and marginal cost	yes	no
second cost model: share equations	yes	no
first profit model: supply and demand	no	no
first profit model: share equations	no	yes
second profit model: supply and demand	no	no
second profit model: share equations	no	no

The desired cost model specification, proposed in section 3.2 above, is subject to the same measurability problem as the first cost function alternative developed in 3.4.1 above: it implies constant returns to scale and thus renders marginal cost measurement by the optimization rule "price equals marginal cost" impossible. However, and this is denoted by the plus sign, one

could of course ignore this insight and make the totally inconsistent desired cost specification a stage more inconsistent by taking the measured price as marginal cost. The same could be performed with the share equations of the first cost model, which are equivalent to the share equations of the desired cost function in any case. I.e. these two "no" judgements corresponding to measurability could be turned into a "yes" by making the respective "no" judgement corresponding to consistency still more determined. The third candidate for an estimation is the first profit function alternative in its share equations variant. It is the only model which can be estimated without thereby increasing the degree of inconsistency. However, the consistency violation implied by the profit function approach is virtually the same as the violation implied by erroneously interpreting price as marginal cost: the unboundedness of the profit maximization problem in the case of constant returns to scale.

Even this level of inconsistency can be surpassed: one could consider estimating some selected equations rather than the complete system of equations derived from the cost or profit function, respectively. This suggests itself for the second cost model in its share variant since component demand shares are both consistent and estimable, and merely the marginal cost share equations are inestimable because they alone depend on the immeasurable non-component cost. But estimating an incomplete system of derived behavioral functions implies that the inter-equation restrictions expressing the properties of a well-behaved cost or profit function, respectively, cannot be accounted for. Hence, the inconsistency attacks the properties which incorporate economic theory into behavioral functions rather than structural assumptions of which the validity is generally limited to the respective empirical hypothesis. However, the specific situation of the study at hand is different since the above three candidates for inconsistent estimability all suffer from an inconsistency that touches economic theory, namely the unboundedness of the profit maximization problem implied by constant returns to scale: since the constant returns to scale property is a priori valid in the case of component combination,¹¹⁰ the reason for inconsistency of the three candidates is not conditional to empiry, and the result that the profit function is not defined in such a case violates the final justifications of economic theory as basically as the estimation of an incomplete system.

Not even being in the position to estimate a model relying extensively on pre-theoretic premises without violating both maintained structural assumptions and economic theory, i.e. leaving behind everything that could add structure to "blind" real world data, one must doubt that

¹¹⁰ See section 4.4.1 above.

an estimation, even given the availability of excellent data, would yield results which could justify the efforts to do so. Of course, as section 4.4.1 above concluded, the results of inconsistently estimating one of the proposed models are at least likely to lie closer to the truth than white noise, but the reference system for our forecasts and assessments is not white noise but our pre-theoretic knowledge, our intuition. If this is so, one has to expose oneself to the question of whether one would reject a model with contra-intuitive result, and in this case refrain from estimation.

4.4.3 Outlook

There are two possible strands to escape this extremely uncomfortable result: on the one hand, one could try to develop a better model which is capable of yielding consistent estimates, given the limited data availability. Directions of further research are proposed immediately. On the other hand, one could attempt to overcome the limited data availability. Considerations on this topic conclude the model building part of this study.

Model-building progress could start from the observation that the compound feed firm is doubtlessly trying to maximize its profit. Consequently, there must be a clue to get around the unboundedness of the component combination micro profit function without finishing up with corner solutions.¹¹¹ If one looks at the phenomena, the question of an optimal firm size is clearly connected with disadvantages of size in the non-component sector. Moreover, it can be shown that the cost of transportation plays a prominent role in this question. The output mix clearly hinges on the transport situation: depending on the location of the compound feed plant, there is a different number of customers with different respective demands for different compound feeds inside each iso-transportation cost circle around the plant. The optimal supply quantity for each compound feed is – nonjointly – determined as total compound feed demand of the livestock producers that are contingently located inside the frontier where marginal cost of the compound feed producer equals marginal cost of the cheapest concurrent, where marginal cost is rising proportionally to the distance from the plant. Marginal cost may of course vary greatly along this line, depending on where the next concurrent in each direction is located.¹¹²

¹¹¹ See section 4.3.2 above.

¹¹² Another possibility to escape a corner solution of the profit maximization problem would be the assumption that an advantage in marketing is connected with a multi-feed supply. But the implied hypothesis that custo-

But how can one express this in a profit function? The classical structural assumptions, the relevant ones being listed in section 4.2 above, are not affected by this insight: output mix is still weakly separable from non-component prices because a change of the transportation price affects all compound feed quantities the same, as argued in section 4.2.1 above, if, in the absence of a reasonable alternative for all compound feed producers, it is assumed that the producers of different livestock categories are evenly distributed. Thus, the weakly separable profit function specified as second alternative above is correct. But if this structure is correct, the component micro profit function is clearly characterized by constant returns to scale and thus not defined. As it stands, it must be concluded that there is a factor of production, namely the spatial demand quantity distribution around the plant, which plays a crucial role in the determination of profit maximizing supply, that cannot be depicted in a profit function since it lacks a price or a meaningful instrument for a price promising an answer to the question of the density of demand for each compound feed in terms of prices. This is either a challenge for new modeling efforts, or the end of the project of depicting the compound feed firm's behavior in utilizing a profit function.

It is obvious that a successful profit function approach accounting for the behavioral determinants of the compound feed industry identified above would increase the problems which result from the unavailability of non-component cost data, because a disaggregation of non-component cost could no longer be avoided. At least for total non-component cost, however, one could consider generating monthly data using programming techniques or aggregating it pragmatically from diverse non-component information which is available. If it were possible to aggregate e.g. indices for wages, energy cost, capital cost, industrial equipment, et c., relying on cost shares extracted from the balances of selected compound feed enterprises, one could estimate the consistent cost model developed as a second alternative in section 3.4.1 above, using the resulting index as an instrument for C^2 . Depending on the correlation of such an index with the true non-component cost, one could probably expect to establish a reasonable forecasting tool.

Another approach which would utilize generated data corresponds to the proposition made in section 3.4.1 considering the first cost function alternative: if the compound feed firm was

customers want to buy all animal-specific feeds from one hand and are willing to pay for this is unlikely for two reasons: first, many customers, and in particular the significant ones, are specialized on one livestock category. Secondly, farmers seem to be quite marketing-resistant: marketing expenditures in the compound feed industry are said to be extremely low.

defined excluding transportation activities, a variable price depending on the distance from the respective plant could be assumed. Such prices would have to be generated out of aggregate sectoral data additionally utilizing information about number, size distribution and spatial distribution of compound feed plants and transportation cost. However, feed quantities would be exogeneous in such a cost function model, and the arguable assumption of an inseparable $C.6$ would be implied.

Finally, after the experience gathered in this study, and on the basis of working through other studies on similar topics, a proposition shall be formulated which goes beyond the specific modeling efforts undertaken here: it seems highly desirable to make out a checklist-like presentation of known results about behavioral assumptions and the implied functional structure so that the applied researcher is able to properly build models without becoming a methodologist himself, which is possibly neither his passion nor his comparative advantage. The published material is commonly characterized by a formal level which requires the reader to go deeply into mathematical syntax. While this surely enhances understanding, it inhibits a broader distribution of the knowledge. A textbook listing hard and fast rules – which exist! – for standard applications would surely be of great benefit for economic practice.

Part II: Functional Forms, Flexibility, and Regularity

5 Theory and Systematics of Flexible Functional Forms

In spite of the failure of the project to establish a behavioral model of the compound feed industry which is consistent with economic theory and evident structural assumptions and at the same time estimable on the basis of the available data, the discussion of which constitutes the first part of this study, a second part on the theory of flexible functional forms is added. While it is superfluous to consider functional specifications with regard to the unsuccessful modeling efforts, the analysis of the theory of flexible forms, which was already undertaken before it became obvious that it would not be finally utilizable in this study, yields a methodical progress which is possibly of importance for applied economics in general. It implies a change in the standard specification and estimation procedure for applied microeconomic models, and this procedure could be utilized in future studies with the same goal as the one in hand if the modeling problems can be overcome.

There are two ways of quantifying economic behavioral models like the ones developed in the preceding sections: a normative programming approach simulating the production process, or a positive econometric approach measuring the actually observable behavior of market participants with statistical methods. In this work the econometric approach is favored because there the desired model parameters are derived from empirical data rather than a priori postulated.

The first step of the empirical application or validation, respectively, of an economic model consists of the postulation of a specific functional form for the data generation process. Because every dual behavioral function like a cost, revenue, or profit function, corresponds with exactly one primal technology, the assumption of a definite cost or profit function comprises the decision for a definite structure of the implicit technology. This decision is necessarily connected with restrictions: there is a trade-off between universality with complex functional forms and great demands on the data on the one hand and tractability with simple functional forms and many unwanted restrictions on the other hand. The estimation of e.g. a CES production function implies a homogeneous technology with constant elasticities of scale and size and constant substitution elasticities in the entire domain, which are furthermore equal between all factors. All results of statistical tests are valid only under the assumption that the

postulated model is valid; a rejected test for e.g. concavity in a CES-estimation does not necessarily imply that the examined data is inconsistent with a concave technology – the data could simply be inconsistent with the CES specification because the assumption of equal and constant substitution elasticities is not supported by the data.¹¹³ This reveals how crucially the quality of the estimation results depends on the choice of functional form. It is now a question of specifying functional forms that (with descending importance)¹¹⁴

1. can provide information about all economically relevant aspects of the examined process and include no, or as few as possible, unwanted a priori restrictions, i.e. are flexible,
2. are a priori consistent with economic theory to the greatest possible extent,
3. let flexibility and theoretical consistency apply to as large a domain as possible,
4. are easy to estimate with common multivariate estimation procedures, or, highly correlated with this requirement, are estimable in a statistical framework that is theoretically well-founded,
5. are parsimonious, i.e. have no more than the minimum number of parameters needed to maintain flexibility, for superfluous parameters increase multicollinearity problems and reduce degrees of freedom,
6. are robust towards changes in variables not accounted for in the model to allow interpolation and extrapolation even under uncertain conditions,
7. and contain parameters that are easy to interpret, i.e. allow the derivation of elasticities in a simple way.

This and the two following chapters give a systematic overview of flexible functional forms. The focus will be on the dual profit function because it is the most general dual behavioral function, and because it best meets the needs of the present analysis. In spite of that, flexible forms for cost functions are also considered since the cost function approach is the most widespread in literature. In addition, this broader view allows one to showcase the general character of the presented systematics by comparison on the one hand, and to work out specific differences between the approaches and to evaluate them in view of their possible applications on the other hand.

¹¹³ See FUSS/MCFADDEN/MUNDLAK 1978: 223.

¹¹⁴ See FUSS/MCFADDEN/MUNDLAK 1978: 224 – 225; see LAU 1986: 1520.

Rigorously inductive, this chapter starts with fundamental theoretical considerations that only become clear if the reader has a vague idea of what a flexible functional form looks like. If this is not the case, he should leaf through section 5.2 and have a look at the Generalized Quadratic before exposing himself to section 5.1. In addition, it is recommended to take note of the programmatic introduction and the abstract of results of the second part of this study given in sections 1.2.2 and 1.3.2, respectively, because, with respect to the broadness of the material covered in the next three chapters, it is helpful to remain constantly aware of the intention followed throughout the exposition.

5.1 The Concept of a Flexible Functional Form

5.1.1 Definition

A functional form is said to be flexible if its shape is only restricted by the requirements of theoretical consistency. I.e. it must be possible to choose a set of function parameters such that every derived parameter of interest, for example every elasticity, is capable of assuming an arbitrary theoretically consistent value at a certain point in the variable space: limitations of the magnitude of parameters or interdependencies between these may only exist subject to the known properties of the respective function. In a production economics context this absence of unwanted a priori restrictions with respect to all desired parameters is illustrated by the metaphor of "providing an exhaustive characterization of all (economically) relevant aspects of a technology".

The concept of flexibility is commonly associated with a special case of the above general definition that is more precisely labelled "second order flexibility"¹¹⁵ since in most applications the parameters of interest are own and cross price elasticities which are descendant from the second order derivatives of technology functions or behavioral functions like dual cost or profit functions.¹¹⁶ Formally, let $F(\alpha; q)$ be an algebraic form for a real-valued function with variables q and a vector of unknown parameters α . This function F shall approximate function value, gradient and Hessian F, F', F'' of an unknown function $\tilde{F}(q)$ at an arbitrary \bar{q} . Then, flexibility of F implies and is implied by the existence of a solution $\alpha(\bar{q}; \bar{F}, \bar{F}', \bar{F}'')$ – the ver-

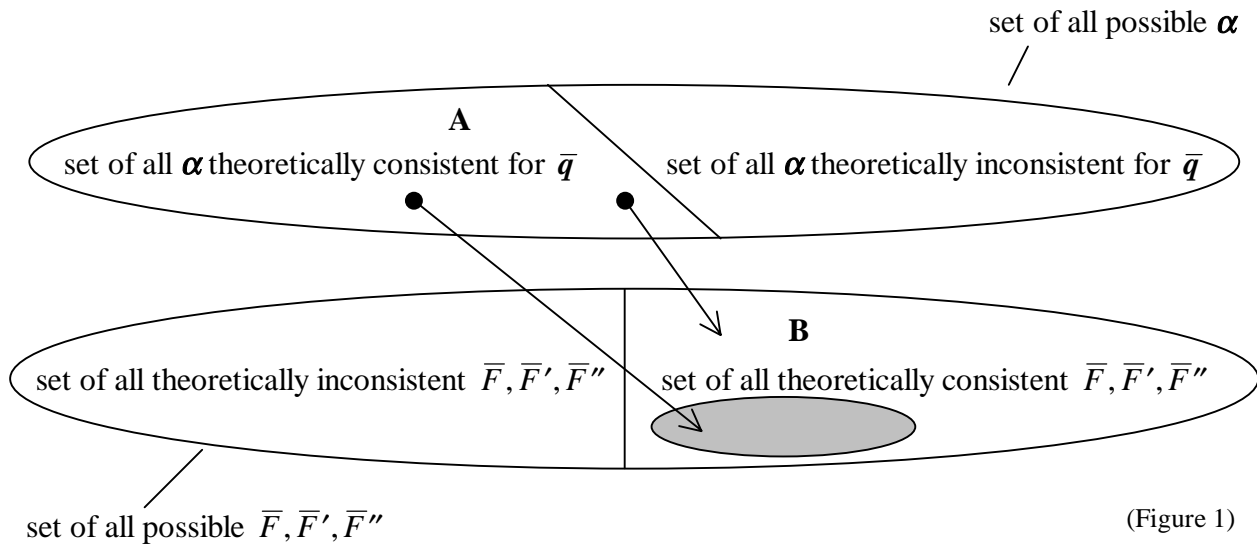
¹¹⁵ Sometimes, second order flexibility is labelled "Diewert-flexibility" after the inventor of the first functional form to exhibit this property, the Generalized Leontief.

¹¹⁶ See LAU 1986: 1540 and DIEWERT 1973.

tical bars denote the numerical value of the respective terms, determined at \bar{q} – to the following set of equations

$$\begin{aligned} F(\alpha; \bar{q}) &= \bar{F} \\ \nabla F(\alpha; \bar{q}) &= \bar{F}' \\ \nabla^2 F(\alpha; \bar{q}) &= \bar{F}'' \end{aligned}$$

under certain consistency conditions on variables q and possible values F, F', F'' that depend on which behavioral function F shall represent. If e.g. F is a cost function, the solution is subject to non-negativity of $\bar{q}; \bar{F}, \bar{F}'$ and semi-definiteness of \bar{F}'' such that $\bar{F} = \bar{q}\bar{F}'$ and $\bar{F}''\bar{q} = 0$.¹¹⁷ In other words: for an arbitrary vector of exogeneous variables \bar{q} , it is possible to find a vector α such that the values of the function, its gradient and its Hessian matrix are equal to some $\bar{F}, \bar{F}', \bar{F}''$, where the set of $\bar{F}, \bar{F}', \bar{F}''$ for which this is true includes all possible theoretically consistent values. Consider the following figure:



The set of all possible $\bar{F}, \bar{F}', \bar{F}''$ – the set where each element of scalar \bar{F} , vector \bar{F}' , and matrix \bar{F}'' can assume arbitrary values between minus and plus infinity – is restricted by economic theory which yields the set of $\bar{F}, \bar{F}', \bar{F}''$ that are consistent with the properties of the respective behavioral function, the set **B**. This restriction implies a corresponding restriction of the set of all possible α because some α produce a theoretically inconsistent $\bar{F}, \bar{F}', \bar{F}''$, where the restriction on α is different for every \bar{q} since F of course depends on the value of the variables. The set of α that satisfies this restriction is labelled **A**. Flexibility means that for

¹¹⁷ See LAU 1986: 1540.

any element of \mathbf{B} there exists a parameter vector α such that, at \bar{q} , $F(\alpha; \bar{q}) = \bar{F}$, $\nabla F(\alpha; \bar{q}) = \bar{F}'$ and $\nabla^2 F(\alpha; \bar{q}) = \bar{F}''$. I.e. a subset of all possible α surjectively maps into the set of theoretically consistent function values.

Consider the counter-example of a Cobb-Douglas cost function: at any \bar{q} , the set of α that yields consistent $\bar{F}, \bar{F}', \bar{F}''$ is the same – any α whose elements sum up to unity cannot produce invalid outcomes, and if α fails to satisfy this condition this results in inhomogeneous values. But only such $\bar{F}, \bar{F}', \bar{F}''$ can be produced that are consistent with unity elasticities of substitution. This subset of all valid $\bar{F}, \bar{F}', \bar{F}''$ is denoted by the shaded area \mathbf{C} in the figure above, and the mapping relation is exemplified by the longer arrow. The shorter arrow is not available with a Cobb-Douglas model; the mapping relation from the set of all admissible α to the set of all valid $\bar{F}, \bar{F}', \bar{F}''$ is not surjective, and thus the Cobb-Douglas model is not flexible.

In this framework, a flexible functional form can provide a local second order approximation of an arbitrary function, either formulated as differential approximation or as Taylor series or numerical approximation, and hence it is called "locally flexible".¹¹⁸ The local approximation property of flexible functional forms is often referred to as the property constituting flexibility, and it is the historical starting point of the theory of flexible functional forms.¹¹⁹ But the local approximation view of a flexible functional form has shortcomings, as will become apparent in sections 5.1.2 and 5.1.3 below, where a superior alternative notion of a flexible functional form is presented.

In a straightforward, degrees-of-freedom-saving application of the given concept of second order flexibility, each relevant aspect, i.e. each level-, gradient- and second derivative-effect is assigned to exactly one parameter which is associated with the respective variables: the level parameter standing alone, the gradient parameters associated with the respective first order variable, and the – as one might say – Hessian-parameters associated with the second order terms, i.e. squared variables or the interaction terms containing two first order variables.¹²⁰ A functional form cannot be second-order-flexible with fewer parameters, and thus the number of free parameters provides a necessary condition for flexibility. For the standard case of a

¹¹⁸ See CHAMBERS 1986: 160-168 and FEGER 1995: 138-139.

¹¹⁹ See DIEWERT 1974: 113. See DIEWERT 1974: 159ff. for tests for flexibility.

¹²⁰ Although more than one parameter is involved in the calculation of substitution elasticities, there is one parameter that plays a more important role than the others.

single-product technology with an n -dimensional input vector, a function exhaustively characterizing all of its relevant aspects should contain information about the quantity produced, all marginal productivities, and all substitution elasticities. Hence, there is one level effect, n gradient effects, and n^2 substitution effects. The latter are symmetric so that, beside the main diagonal with n elements, only half of the off-diagonal effects are needed, i.e. $\frac{1}{2}n(n-1)$. The number of effects an adequate single-output technology function should be capable of depicting independently of each other and without a priori restrictions amounts to a total of $\frac{1}{2}(n+1)(n+2)$, and it follows that a valid flexible functional form must contain at least that number of independent parameters.¹²¹ In the multiple output case, there are m outputs rather than one. Consequently, there are m level effects, $m+n$ gradients and $(m+n)^2$ elements of the Hessian so that, minus half of the off-diagonal elements, a total of at least $\frac{1}{2}(m+n)^2 + \frac{3}{2}(m+n) + m$ independent parameters is required for an "exhaustive characterization" of a technology.

5.1.2 Flexible Functional Forms in Dual Models

The fact that a flexible functional form is capable of locally approximating any technology function and its single valued and invertible transformations up to second order, suggests that not only technology functions but also dual behavioral functions like cost and profit functions can be approximated. It has indeed been shown that the function value and the first and second derivatives of a primal function can be approximated as well by the dual representation of the same technology.¹²² Thus, there is no principal argument against flexible cost and profit functions. With the exception of the Quadratic,¹²³ the duals of technology functions cannot be directly deduced from flexible technology functions conducting a Lagrangian constrained optimization. Frequently, flexible cost and profit functions with the same analytical form as of technology functions are postulated ad hoc instead.¹²⁴ The following discussion will show that it is in any case superfluous to take a detour over primal functions in designing valid and powerful dual flexible functional forms.

¹²¹ See HANOCH 1975: 398-399.

¹²² See BLACKORBY/DIEWERT 1979: 580; proof 588-592.

¹²³ See LAU 1974: 186 and section 6.2.1 below.

¹²⁴ See HASENKAMP 1976: 62.

The equivalence of primal and dual representation again leads one to expect the possibility of depicting all economically relevant aspects of a technology directly in terms of dual flexible functions. But a single-product dual cost function has $n + 1$ variables rather than the n variables of the production function: in addition to the n factor prices, cost depends on output quantity. It seems as if an exhaustive description of all relevant aspects of a cost function would require $\frac{1}{2}(n+3)(n+2)$ independent parameters rather than $\frac{1}{2}(n+2)(n+1)$. The same problem arises in the multi-output context, since a technology function has to be solved for one product to preserve the analogy whereas a multi-output cost function maps from all output quantities (and input prices) to cost. But this is, as again the equivalence leads one to expect, not the case: some of the first and second derivatives of the cost function are not independent of each other due to the homogeneity property of a well-behaved cost function. Euler's theorem yields

$$C(\mathbf{w}, \mathbf{y}) = \sum_{i=1}^n \frac{\partial C(\mathbf{w}, \mathbf{y})}{\partial w_i} \cdot w_i \quad (\text{adding-up condition}^{125}),$$

due to Shephard's lemma identical with the primal definition of cost, so that either the function value or one of the first derivatives of the cost function with respect to a factor price can be expressed as a linear combination of the remaining terms of the equation. By Euler's theorem, too, the first derivatives with respect to \mathbf{w} of the cost function which is homogeneous of degree one are homogeneous of degree zero. Thus,

$$\sum_{i=1}^n \frac{\partial^2 C(\mathbf{w}, \mathbf{y})}{\partial w_j \partial w_i} \cdot w_i = 0 \quad \forall j = 1, \dots, n \quad (\text{Cournot aggregation condition}),$$

and finally, again due to Euler's theorem,

$$\frac{\partial C(\mathbf{w}, \mathbf{y})}{\partial y_j} = \sum_{i=1}^n \frac{\partial^2 C(\mathbf{w}, \mathbf{y})}{\partial y_j \partial w_i} \cdot w_i \quad \forall j = 1, \dots, m \quad (\text{Engel aggregation condition}).$$

Hence, further $n + m$ effects can be obtained as linear combinations of the remaining effects. Consequently, these $n + m + 1$ restrictions altogether reduce the number of required indepen-

¹²⁵ Applying Shephard's lemma and dividing by C yields the classic formulation of the adding-up condition where expenditure shares of all factors or consumer goods must sum up to unity. This shows that the adding-up condition must not be seen in distinction to some homogeneity restrictions inside equations, as is common in demand analysis, but immediately follows from the homogeneity property too [see e.g. PHILIPS 1982: 34-40 where such a distinction is already implied by the section headlines].

dent estimation parameters to exactly $\frac{1}{2}(m+n)^2 + \frac{3}{2}(m+n) + m$, i.e. just the number of parameters required for the exhaustive representation of a technology.¹²⁶

For the dual profit function the argument follows a similar path.¹²⁷

$$\Pi(\mathbf{p}, \mathbf{w}) = \sum_{i=1}^m \frac{\partial \Pi(\mathbf{p}, \mathbf{w})}{\partial p_i} \cdot p_i + \sum_{j=1}^n \frac{\partial \Pi(\mathbf{p}, \mathbf{w})}{\partial w_j} \cdot w_j,$$

according to Hotelling's lemma identical with the primal definition of profit as revenue minus cost,

$$\sum_{j=1}^m \frac{\partial^2 \Pi(\mathbf{p}, \mathbf{w})}{\partial w_i \partial p_j} \cdot p_j + \sum_{j=1}^n \frac{\partial^2 \Pi(\mathbf{p}, \mathbf{w})}{\partial w_i \partial w_j} \cdot w_j = 0 \quad \forall i = 1, \dots, n,$$

and

$$\sum_{j=1}^m \frac{\partial^2 \Pi(\mathbf{p}, \mathbf{w})}{\partial p_i \partial p_j} \cdot p_j + \sum_{j=1}^n \frac{\partial^2 \Pi(\mathbf{p}, \mathbf{w})}{\partial p_i \partial w_j} \cdot w_j = 0 \quad \forall i = 1, \dots, m$$

apply.

5.1.3 The Local Approximation View

To talk of "approximation" suggests that the concept of a flexible functional form is generally understood in distinction to an underlying "true" data generation process whose functional form and parameters are principally unknown. The researcher postulates a micro-theoretic model and a functional form, and then he evaluates the functional form's parameters in a statistical estimation procedure. About the relation between the supposed true function and the corresponding flexible estimation function three concurring hypotheses are possible:

1. The estimation function is a local approximation of the true function.

The approximation properties of flexible functional forms are only locally valid, i.e. restricted to a single point.¹²⁸ Thus, it can in no way be claimed that the assumed true function is approximated in its totality: value, gradients and Hessian of true and estimated function are equal merely at one single point – the point of approximation. Therefore, the estimated para-

¹²⁶ See FUSS/MCFADDEN/MUNDLAK 1978: 232.

¹²⁷ See CHAMBERS 1988: 170-171 for the single-output case.

¹²⁸ See figure 2 below.

meters must be interpreted only locally. This considerably restricts the forecasting capabilities using these parameters if the variable values in the forecast period are relatively distant from the approximation point. In the immediate neighbourhood of the approximation point each flexible functional form provides theoretically consistent parameters if the true structure is theoretically consistent,¹²⁹ because the parameters of the approximating function and those of the true structure are identical at the approximation point, which is exemplified in figure 2 below where a known, globally consistent cost structure is locally approximated at two different points **P** and **Q**. Since e.g. local concavity is a necessary condition for global concavity, i.e. concavity in the entire domain of the estimation function or all possible variable values, respectively, at least the necessary condition of the concavity hypothesis for the postulated true structure can be tested at the approximation point by means of the concavity of the estimation function – which, of course, must not be intrinsically concave.¹³⁰

But which point is the point of approximation, where is it located?¹³¹ Rather than approximating a known algebraic structure at one point as depicted in figure 2, the estimation procedure fits the approximating function to a data sample over a more or less extended range, as figure 3 below elucidates. There is no way to infer from the approximation function to the location of the approximation point, or, more precisely, in a statistics context where the approximated structure is intrinsically unknown the concept of an approximation point runs empty at all. Consequently, since there is no way to infer from an estimated function to the properties of the underlying true structure, if the point where the properties of the approximating function are known is not the approximation point, the properties of the assumed true function remain unknown. From the rejection of the null hypothesis that the approximating function shows all properties of a well-behaved cost function, it can be followed that the examined economic subjects are not cost minimizers – or that the test was not conducted at the approximation point. Hence, testing the null hypothesis is useless.

Commonly, the point of approximation is held to be located at some mean of variables over all observations.¹³² However, this view emanates from erroneously interpreting the point of approximation and the point of expansion of e.g. a Taylor series as synonyms. All second order flexible functional forms can be interpreted as second order differential or Taylor series

¹²⁹ See FUSS/McFADDEN/MUNDLAK 1978: 233-236; CHAMBERS 1988: 177-180.

¹³⁰ See LAU 1978: 418-420.

¹³¹ See MOREY 1986: 227 who raises this question.

¹³² See WHITE 1980: 150. For a recent example, see BARNETT/KIROVA/PASUPATHY 1996: 16.

approximations,¹³³ where the point of expansion is identical with the point of approximation of a known algebraic function. But the point of expansion, which is generally the point where all exogeneous variables assume unity value for a standard formulation of a flexible form, implies nothing about the approximation properties of the resulting specification with respect to a random sample, rather than to a known algebraic structure. Salvanes and Tjøtta prove the obvious: the estimated function is invariant with respect to the point of expansion.¹³⁴ This is, the estimation function, transformed to establish another point of expansion, yields the same graph because the estimated parameters exactly compensate the transformation, which is obvious because after all the same specification fits the same random sample. However, their sorrowful conclusion that second order flexible functional forms do not necessarily perform well at the point of approximation, because this point could be located outside the regular region, is groundless: there is no such point in an econometric estimation of a flexible functional form.

2. The estimation function and the true structure are assumed to be of the same functional form but show the desired properties merely locally.¹³⁵

This hypothesis accounts for the fact that most common flexible cost functions can either not be restricted to a well-behaved cost function without losing their flexibility or cannot be restricted to regularity at all, as will be shown in the next chapter. All flexible functional forms can both be understood as a local approximation of an unknown structure or as a postulated functional form of the true structure. The latter is advantageous insofar as regularity can be tested at any data point, since the whole function is viewed as an approximation. Therefore, points of interest in the true structure can be examined by testing the respective points in the estimation function. But the uncertainty remains whether the estimation function and thereby the true structure is still consistent with the properties of a well-behaved cost function if the data set does not accidentally equal a data set already examined. This problem could only be solved by systematically testing all possible data sets. This procedure, which is hardly elegant, is avoided with the assumption of another hypothesis:

¹³³ See section 5.1.1 above. For a description of flexible functional forms as differential approximation or Taylor series approximation see CHAMBERS 1986: 162-164.

¹³⁴ See SALVANES/TJØTTA 1995.

¹³⁵ See MOREY 1986: 222-227.

3. The estimation function and the true structure are assumed to be of the same functional form and show the desired properties globally.¹³⁶

If one succeeds in finding a flexible functional form that can be restricted to global regularity without losing its flexibility, i.e. a form which has the advantage that local properties are the necessary and sufficient condition for the respective global properties, another possibility arises: only those functional forms allow one to infer from the estimation function to the true structure and hence allow meaningful tests of significance, because otherwise the model lacks theoretical foundation.¹³⁷ Hence, a serious problem arises for the postulates of economic theory if a properly specified flexible cost function which is globally well-behaved is not supported by the data. An unjustified rejection of the null hypothesis could be caused by missing, undersized, badly measured or ill-conditioned data, by a validation method that is inappropriate, or by misspecification. Misspecification includes e.g. omission of relevant variables, a wrong functional form, or wrong modeling assumptions like separability and non-jointness hypotheses.¹³⁸ But, if one insists on the model being properly specified and the data being sufficient, it can be either followed that, claiming unlimited validity of neoclassical economic theory, the economic subjects are not cost minimizers or profit maximizers,¹³⁹ i.e. behave irrationally,¹⁴⁰ or that, allowing for multiple theories of economic behavior, they act rationally according to another rationality concept, i.e. neoclassical economic theory is not applicable in that case (or, taking the easy way out, they behave irrationally although there may exist other theories of economic rationality that are not considered).

The last approach of a flexible functional form does not labour under the illusion that the true structure of the data generating process can be locally approximated in analogy to the

¹³⁶ See MOREY 1986: 219-222.

¹³⁷ See sections 6.2.4 and 6.3 below.

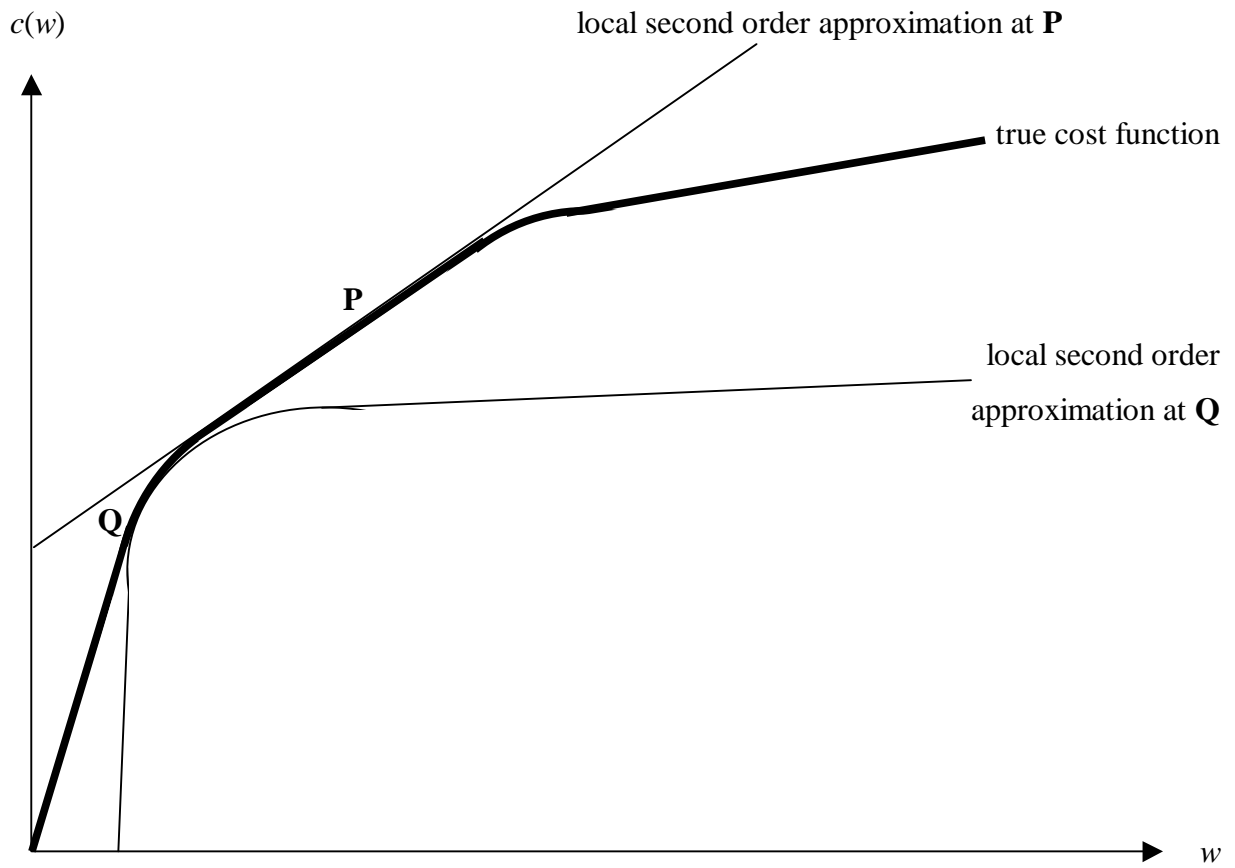
¹³⁸ Another possible misspecification could consist of the assumption that a second order flexible functional form is flexible enough really to provide an exhaustive characterization of technology [see section 5.1.4 below].

¹³⁹ This conclusion not only assumes a well-behaved true profit or cost structure, it moreover presupposes a definite solution of profit or cost optimizing behavior. Thus, an underlying technology with affine pieces and the resulting multiple solutions must be excluded to allow a test of theory.

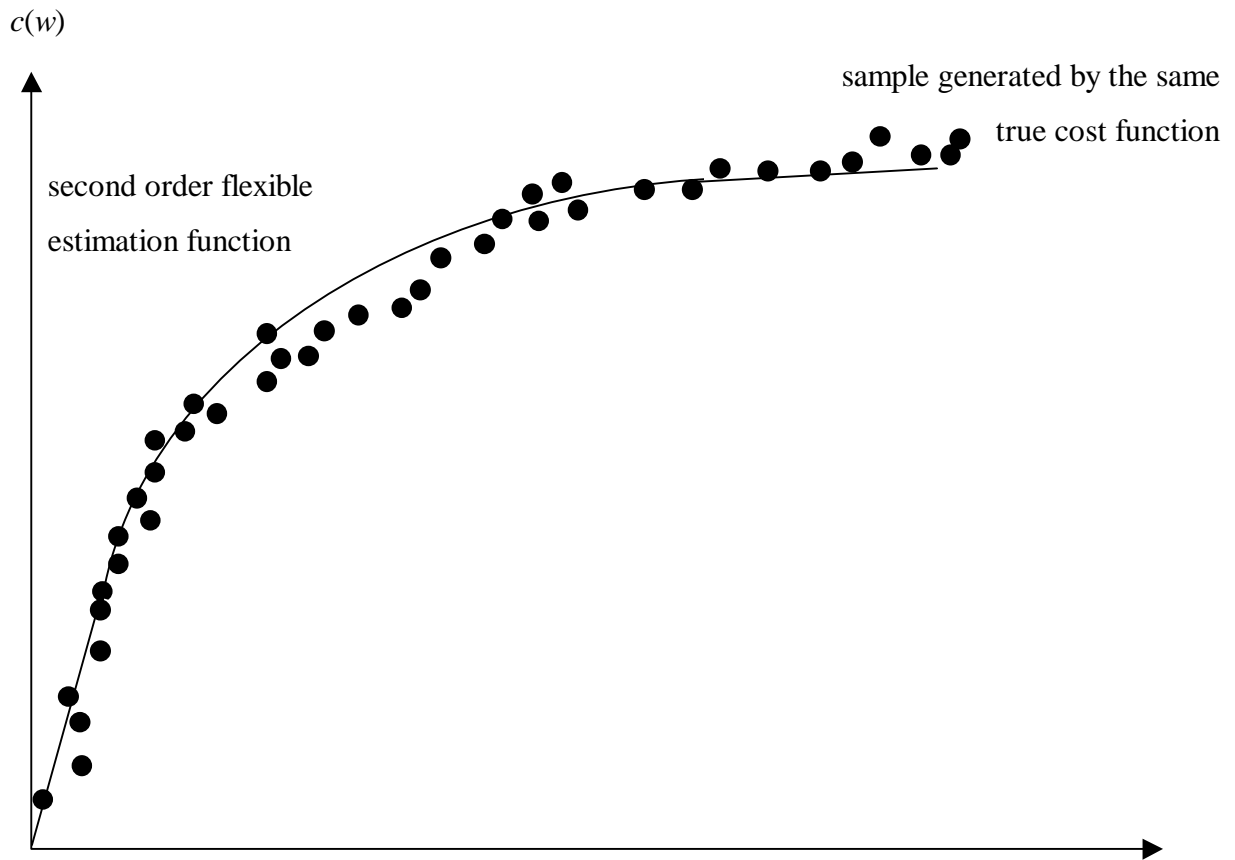
¹⁴⁰ The identification of non-cost-minimizing or non-profit-maximizing behavior with irrationality [see MOREY 1986: 229] is of course only valid if the respective goal function is postulated. However, the argumentation remains the same when other optimization criteria are assumed. These generally imply other specific properties of the respective behavioral function whose violation must be interpreted as irrationality too.

approximation of a *known* complex algebraic structure with the simpler Taylor series approximation as suggested in figure 2 – in econometrically validated models the true structure is merely present through a random sample of observations, which is depicted in figure 3. Rather, it promotes a concept of flexibility where the functional form – whose postulation is, in any case, an unavoidable component of economic modeling – just has to fit the data to the greatest possible extent, subject only to the regularity conditions following from economic theory and otherwise independently depicting all economically relevant aspects. The argument that any flexible functional form can approximate any other flexible functional form and any arbitrary data generation process does not suspend the researcher from the issue of reducing the specification error to the greatest possible extent in selecting the most appropriate functional form for the entire data. In empirical applications and Monte Carlo simulations this has found its manifestation in the different ability of flexible functional forms to fit different technologies.¹⁴¹

¹⁴¹ See TERRELL 1995: 2 and the literature cited there.



(Figure 2)



(Figure 3)

5.1.4 Semi-flexibility, Third Order Flexibility, and Global Flexibility

Pollak and Wales raise another objection to the common notion of flexibility, the second order flexibility defined above: it is misleading to call a functional form flexible in the second order less restrictive than a non-flexible functional form with the same number of independent parameters; to neglect e.g. unimportant cross-price effects in favour of own-price effects with additional third order parameters can be viewed as a net flexibility gain.¹⁴² This objection can be divided into two aspects:

First, one should not overstress the number-of-parameters criterion because not all effects are equally important or some effects can well be quantified a priori, respectively. This leads to the use of so-called semi-flexible functional forms, where estimation parameters are substituted by predetermined values under violation of the number-of-parameters criterion: Shortness of time series frequently precludes parameter estimation on an empirically interesting level of disaggregation due to a lack of degrees of freedom. In semi-flexible versions of second order flexible functional forms, more fine-grained aggregates can be used by determining some parameters a priori and estimating the rest. E.g. it is often possible to claim a priori with sufficient likelihood that some cross effects vanish – if they are not already zero due to separability or nonjointness assumptions, which actually constitute the same means to reduce the degrees-of-freedom problem, although in a more systematic and general manner. In the theoretically stringent semi-flexibility framework, one has to decide which variable(s) are unimportant enough to set the respective parameter column as linearly dependent from the other variables' columns. Thus, the researcher is allowed to choose the degree of flexibility regarding feasibility of estimation while maintaining second order flexibility for all variables, apart from those intentionally restricted, and regularity conditions implemented in the fully flexible relative.¹⁴³

Secondly, there are effects which are possibly not accounted for with sufficient "flexibility" regarding the goal of the analysis. This could be the case for price or quantity effects constituting the basis of the classic flexibility definition as in the aforementioned example of own price effect, which are of greater importance than certain cross price effects on the one hand, or for other exogeneous variables like time or seasonal dummies on the other hand. Se-

¹⁴² See POLLAK/WALES 1992: 64.

¹⁴³ See DIEWERT/WALES 1988.

cond order flexible functional forms yield systems of derived functions with merely one gradient parameter per variable in most instances, which can be seen as rather restrictive. Thus, third order effects are a meaningful extension. A particularly evident case in the consumer demand context is linearity of Engel curves that is yielded by some flexible functional forms, which is theoretically unacceptable.

These considerations eventually lead to a more general approach to the problem: instead of determining and thereby possibly limiting the degree of flexibility prior to estimation, the depth of parameterization can be set depending on the inferential potency of the sample using series expansions or higher order flexible functional forms.¹⁴⁴ This means that, as opposed to parametric estimation where a functional form is specified and an a priori determined number of parameters is estimated, a semi-nonparametrical estimation can be performed, where the basic functional structure is determined a priori but the decision on the optimal number of estimated parameters is left to the data by statistical criteria. Large samples with large explanatory power or, put differently, a model specification fitting well, can allow for the estimation of much more meaningful parameters than provided for by standard flexible functional forms. Furthermore, undersized samples producing overfitting estimates with a second-order-flexible specification can be handled in this framework as well, where the methodology then competes with the semi-flexibility approach.

To find the amount of flexibility and the depth of parameterization that are statistically optimum, comparison of the statistical fit of different parameterization scenarios using an F-test¹⁴⁵ can be applied.¹⁴⁶ The overfitting problem can also be solved by exclusively using globally regular models: assuming that economic theory does not fail, all of the information contained in the data can be extracted by a sequence of increasingly flexible functions without risk, if all possible parameterizations of the model are theoretically consistent; functions thus restricted cannot produce oscillating behavior fitting the noise in the data.¹⁴⁷ I.e. a regular

¹⁴⁴ See GALLANT 1981 and BARNETT/GEWEKE/WOLFE 1991.

¹⁴⁵ JENSEN 1997 uses Eastwood's upward-F-test; see EASTWOOD 1991.

¹⁴⁶ The possible objection that this practice amounts to data mining cannot be raised because, rather than the question of which variables are incorporated into the model and which not, it is only the precision of depiction which is determined by the data [see section 6.2 where this point is treated more extensively in a footnote].

¹⁴⁷ See BARNETT/GEWEKE/WOLFE 1991: 6. An estimation function is said to fit the noise in the data if it has so many parameters compared to points of observation that the course of the function does not only approximately depict the true data generating function but also the error terms. If e.g. a linear true data generating pro-

functional form estimated semi-nonparametrically is capable of extracting the true structure depicted in figure 2 above from the sample presented in figure 3.

With rising depth of parameterization the task of assuring regularity, especially global regularity, becomes more and more intractable with standard methods like parameter restrictions, because gradients and Hessian depend on a rapidly increasing number of parameters,¹⁴⁸ so that recent numerical techniques¹⁴⁹ remain as the only possibility to operationalize this extended notion of flexibility. With this background, one can even think of reviving, even though with a global rather than a local notion, the approximation view of flexibility rejected in section 5.1.3 above: if one succeeds in guaranteeing global theoretical consistency, the estimated function converges to a global, consistent approximation of the data generating process if the depth of parameterization approaches infinity; rather than approximating the underlying function at one point, it is approximated at any point so that the question of *the* point of approximation becomes obsolete. An infinitely large number of parameters with an infinitely high order of effects allows for fitting the sample infinitely flexibly for the whole range of observation, and thus such a concept of flexibility may be called "global flexibility".¹⁵⁰ In such a framework, the achieved amount of flexibility is only limited by the data.

Two related strands to realize global flexibility and to perform a semi-nonparametric estimation, respectively, are available: series expansions and higher order flexibility. Power series expansions, i.e. sums of power functions or polynoms like the well-known McLaurin series and Taylor series, provide a means not only to approximate value, gradient, and second order derivative of a known complex algebraic structure¹⁵¹ but to approximate a known function at any point if the series is long enough. The extended flexibility is, of course, also of great benefit if the series expansion is utilized to fit samples rather than approximate known

cess generates three points of observation which do, by the random error, not lie on a straight line, a linear estimation function with two parameters yields a better result than a quadratic function with three parameters, because the latter misinterprets the random error as curvature.

¹⁴⁸ See sections 6.2.1 and 6.2.2 below.

¹⁴⁹ See section 6.2.3 below.

¹⁵⁰ See BARNETT/GEWEKE/WOLFE 1991: 6.

¹⁵¹ See section 5.1.3 above.

functions. Examples of series expansions as globally flexible estimation functions are to be found in section 7.1.3 (Fourier series) and in section 7.1.5 (Müntz-Szatz) series.¹⁵²

A slightly different concept is the straightforward generalization of the local flexibility concept. In perfect analogy with the case of local second order flexibility, local third order flexibility of F implies and is implied by the existence of a solution $\alpha(\bar{q}; \bar{F}, \bar{F}', \bar{F}'', \bar{F}''')$ to the set of equations $F(\alpha; \bar{q}) = \bar{F}$, $\nabla F(\alpha; \bar{q}) = \bar{F}'$, $\nabla^2 F(\alpha; \bar{q}) = \bar{F}''$, and $\nabla^3 F(\alpha; \bar{q}) = \bar{F}'''$ under some regularity conditions.¹⁵³ In analogy to a second order function for a single output technology, a third order function contains one level effect, n gradient effects, n^2 second order effects, i.e. the Hessian, and n^3 third order effects, i.e. the n Hessians of the n partial first order derivatives constituting a vector of matrices or a cube, respectively. Thus, a third order flexible functional form allows for any theoretically possible curvature of the derived demand and supply functions (in contrast to curvature of the cost or profit function) at one point. Any Hessian is symmetric according to Young's theorem so that, first, beside the main diagonal with n elements, only half of the off-diagonal elements of the second order Hessian are needed, i.e. $\frac{1}{2}n(n-1)$. Secondly, and in analogy to $f_{ij} = f_{ji}$ for the Hessian, symmetry of the third order derivatives implies interchangeability of the indices, i.e. $f_{ijk} = f_{jik} = f_{ikj} = f_{kij} = f_{kji} = f_{jki}$. The number of independent elements a locally third order flexible functional form must at least contain then amounts to $\frac{1}{6}(n+1)(n+2)(n+3)$. Thus, a third-order-flexible functional form has $\frac{1}{6}n(n+1)(n+2)$ more independent parameters than a second-order-flexible form – a step that may frequently turn out to be too big.

The semi-flexibility method provides a way to reduce this overflow to a measure that fits both the sample size and the model complexity desired.¹⁵⁴ But if the data is capable of supporting the estimation of more parameters, the order of flexibility can be increased until no improvement of statistical significance can be made. With an increasing order of local flexibility the function can at any point better and better approximate the unknown, true function that generated the sample so that it becomes globally flexible. Thus, an increasing order of local flexibility, although not itself meaningful in this context, indicates the increasing ability of the function to approximate globally. However, it has to be noted that the demands on data a-

¹⁵² Beyond that, virtually every locally flexible functional form that is linear in parameters can be seen as second order Taylor series expansion [see FEGER 1995: 44 and the literature cited there].

¹⁵³ See section 5.1.1 above.

¹⁵⁴ But see section 6.3 below.

mount and quality increase drastically with the order of local flexibility: the number of independent parameters included in a functional form locally flexible of order t amounts to

$$\frac{1}{t!} \prod_{i=1}^t (n+i),$$

and even in a very small model of say $n = 4$ – the formula is in any case only valid for single product technologies – the over-proportional rise of the number of independent parameters becomes visible: 5 parameters for $t = 1$, a first order model, 10 for the well-known second order model with $t = 2$, then 20 for $t = 3$, then 35, 56, 84 and so on. The over-proportionality increases substantially with larger models: for $n = 10$, the series is 5, 66, 286, 1001, 3003, 8008 and so on. Facing this, it is not surprising that an empirical application of the higher-order-flexibility strand of global flexibility yet remains to be conducted.

5.2 The General Linear Form and its Derivates

Now, consider a systematic overview over linear flexible functional forms, which is not presented because it would contribute to what shall be shown in this and the next chapter, but rather with view on the application of the new methodical results. As it is argued in the first paragraphs of chapter 7 below, these results open up the opportunity to select among a lot of flexible functional forms much more liberally than previously, so that a systematic of flexible functional forms may be of great benefit in future studies utilizing flexible functional forms and the methodic progress made here.

The most common estimation method is the linear regression. A functional form immediately suited for linear regression has to be linear in its parameters – just like the General Linear Form:

$$F(\boldsymbol{\vartheta}; \boldsymbol{q}) = \sum_{i=1}^n \vartheta_i f_i(\boldsymbol{q})$$

with $f_i(\boldsymbol{q})$ as known, twice continuously differentiable function of vector \boldsymbol{q} and ϑ_i as respective (estimation) parameters.¹⁵⁵ The General Linear Form entertains several advantageous aspects:

1. With its n estimable parameters it can depict n independent effects¹⁵⁶ – and therefore any number of effects since there are infinitely many functions $f_i(\boldsymbol{q})$ with an arbitrarily high n .

¹⁵⁵ See FUSS/MCFADDEN/MUNDLAK 1978: 232.

2. It is linear in parameters ϑ_i but not necessarily linear in variables q .
3. It can approximate any twice continually differentiable function $F^*(q)$ at any point q^0 , i.e. the parameters can be chosen in a way that

$$\begin{aligned} F^*(q^0) &= F(q^0) \\ \nabla_q F^*(q^0) &= \nabla_q F(q^0) \\ \nabla_{qq} F^*(q^0) &= \nabla_{qq} F(q^0) \end{aligned}$$

Hence, there exists a singular vector $\vartheta' = [\vartheta_1, \dots, \vartheta_n]$ for which the system of equations given as second order flexibility definition in section 5.1 is solvable and for which both the value of the function, the gradients, and the Hessian of the General Linear Form equal the respective values of the approximated function at one point; the General Linear Form can provide a local second order differential approximation to an arbitrary function.¹⁵⁷

The third property of the General Linear Form is of particular importance in so far as it shows the suitability of the General Linear Form for serving as a locally flexible functional form: The estimation of level, first order and second order parameters using the General Linear Form supplies parameter values for, as the advocates of second order flexible functional forms put it, all relevant effects of a technology – without knowing about the true technology function, and independently of the General Linear Form itself representing the true technology structure. Since the number of micro functions f_i and their respective parameters ϑ_i is unlimited, the interpretation of the General Linear Form as a globally flexible functional form is possible, too.

Used as an estimation function, the linearity in parameters and corresponding f_i equips the General Linear Form with a far more stringent statistical theory than is available for nonlinear estimators.¹⁵⁸ Even if, as will happen in section 6.2.2, restrictions introduce nonlinearities in parameters, these are only slight and leave the superiority of the General Linear Form with regard to the operationality of estimation and the theoretical foundation of the estimator almost unaffected, compared with intrinsically nonlinear flexible functional forms like the

¹⁵⁶ See CHAMBERS 1988: 161.

¹⁵⁷ See CHAMBERS 1988: 162.

¹⁵⁸ See section 6.2.3 above.

Constant-Difference-of-Elasticities functional form (CDE).¹⁵⁹ Therefore, and to keep the scope of the present study reasonable, the broad field of nonlinear functional forms will not be covered.

5.2.1 Special Cases of the General Linear Form

The General Linear Form is more general than required in virtually every conceivable application, and, therefore, in empirical analysis General Linear Form derivatives are utilized where the micro functions $f_i(\mathbf{q})$ are simplified in a way that they merely depend on a small subvector of \mathbf{q} . If all $f_i(\mathbf{q}) = q_i$, the General Linear Form is linear in the exogeneous variables. Already microfunctions depending on two exogeneous variables allow a large variety of functional structures capable of describing many relevant aspects of a technology, i. e. qualify the derivate to locally approximate an arbitrary function up to second order.¹⁶⁰ The limitations of many such special cases, though severe, do not impair the local flexibility or second order approximation property, respectively. Single valued and invertible transformation of the data, e.g. taking logarithms of all variables, does not touch flexibility either.¹⁶¹

Most linear flexible functional forms belong to the General Linear Form subclass of quadratic functions, because they constitute the least complex functional structure capable of providing a second order approximation and the necessary number of independent parameters, respectively. That is, most linear flexible functional forms are special cases of the Generalized Quadratic¹⁶²

$$\begin{aligned} F(\mathbf{q}) &= \alpha_0 + \sum_{i=1}^n \beta_i f_i(q_i) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \varphi_{ij} f_i(q_i) f_j(q_j) \\ &= \alpha_0 + \beta' f(\mathbf{q}) + \frac{1}{2} f(\mathbf{q})' \Phi f(\mathbf{q}) \end{aligned}$$

where \mathbf{q} is the vector of exogeneous variables and level parameter α_0 , gradient vector $\beta' = [\beta_i]$ and substitution effect matrix $\Phi = [\varphi_{ij}]$ are the estimation parameters with $\varphi_{ij} = \varphi_{ji}$ or Φ being

¹⁵⁹ See JENSEN 1997 for a Monte-Carlo simulation, where a CDE is used to generate the pseudo data and the intrinsically linear Almost Ideal Production Model (AIM) as approximation function.

¹⁶⁰ See MCFADDEN 1978a: 273.

¹⁶¹ See CHAMBERS 1988: 166.

¹⁶² See BLACKORBY/PRIMONT/RUSSELL 1978: 293, 297.

symmetric, respectively. By Young's theorem, the Hessian of any real valued function is symmetric no matter what Φ is, so that Φ can be set symmetric without loss of generality.

As shown in section 5.1.4, dual representations can be depicted by flexible functional forms as well. So let the cost function formulation of the General Linear Form and the Generalized Quadratic, respectively, be denoted by

$$\begin{aligned}
 F(C(\mathbf{w}, \mathbf{y})) &= \sum_{i=1}^n \vartheta_i f_i(\mathbf{w}, \mathbf{y}) \\
 F(C(\mathbf{w}, \mathbf{y})) &= \alpha_0 + \sum_{i=1}^n \beta_i f_i(w_i) + \sum_{i=1}^n \gamma_i f_i(y_i) \\
 &\quad + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \varphi_{ij} f_i(w_i) f_j(w_j) + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \psi_{ij} f_i(y_i) f_j(y_j) + \sum_{i=1}^n \sum_{j=1}^m \theta_{ij} f_i(w_i) f_j(y_j) \\
 &= \alpha_0 + \beta' f(\mathbf{w}) + \gamma f(\mathbf{y}) + \frac{1}{2} f(\mathbf{w})' \Phi f(\mathbf{w}) + \frac{1}{2} f(\mathbf{y})' \Psi f(\mathbf{y}) + f(\mathbf{w})' \Theta f(\mathbf{y})
 \end{aligned}$$

where \mathbf{w} is the vector of exogeneous input prices, \mathbf{y} is the vector of exogeneous output quantities. Costs are formulated as an argument of some function to allow for transformations like taking logarithms. The estimation parameters are level parameter α_0 , input price gradient parameter vector $\beta' = [\beta_i]$, output quantity gradient parameter vector $\gamma' = [\gamma_i]$, square and symmetric input substitution effect parameter matrix $\Phi = [\varphi_{ij}]$, square and symmetric output substitution effect parameter matrix $\Psi = [\psi_{ij}]$, and not necessarily square and generally non-symmetric input-output substitution effect parameter matrix $\Theta = [\theta_{ij}]$. To present the second order parameters in a more intuitive way:

$$\begin{bmatrix} \Phi & \Theta \\ \Theta' & \Psi \end{bmatrix} = \left[\begin{array}{ccc|ccc} \varphi_{11} & \cdots & \varphi_{1n} & \theta_{11} & \cdots & \theta_{1m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \varphi_{n1} & \cdots & \varphi_{nn} & \theta_{n1} & \cdots & \theta_{nm} \\ \hline \theta_{11} & \cdots & \theta_{1n} & \psi_{11} & \cdots & \psi_{1m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \theta_{m1} & \cdots & \theta_{mn} & \psi_{m1} & \cdots & \psi_{mm} \end{array} \right]$$

In perfect analogy, the Generalized Quadratic Profit Function assumes the form

$$F(\Pi(\mathbf{p}, \mathbf{w})) = \alpha_0 + \beta' f(\mathbf{p}) + \gamma f(\mathbf{w}) + \frac{1}{2} f(\mathbf{p})' \Phi f(\mathbf{p}) + \frac{1}{2} f(\mathbf{w})' \Psi f(\mathbf{w}) + f(\mathbf{p})' \Theta f(\mathbf{w})$$

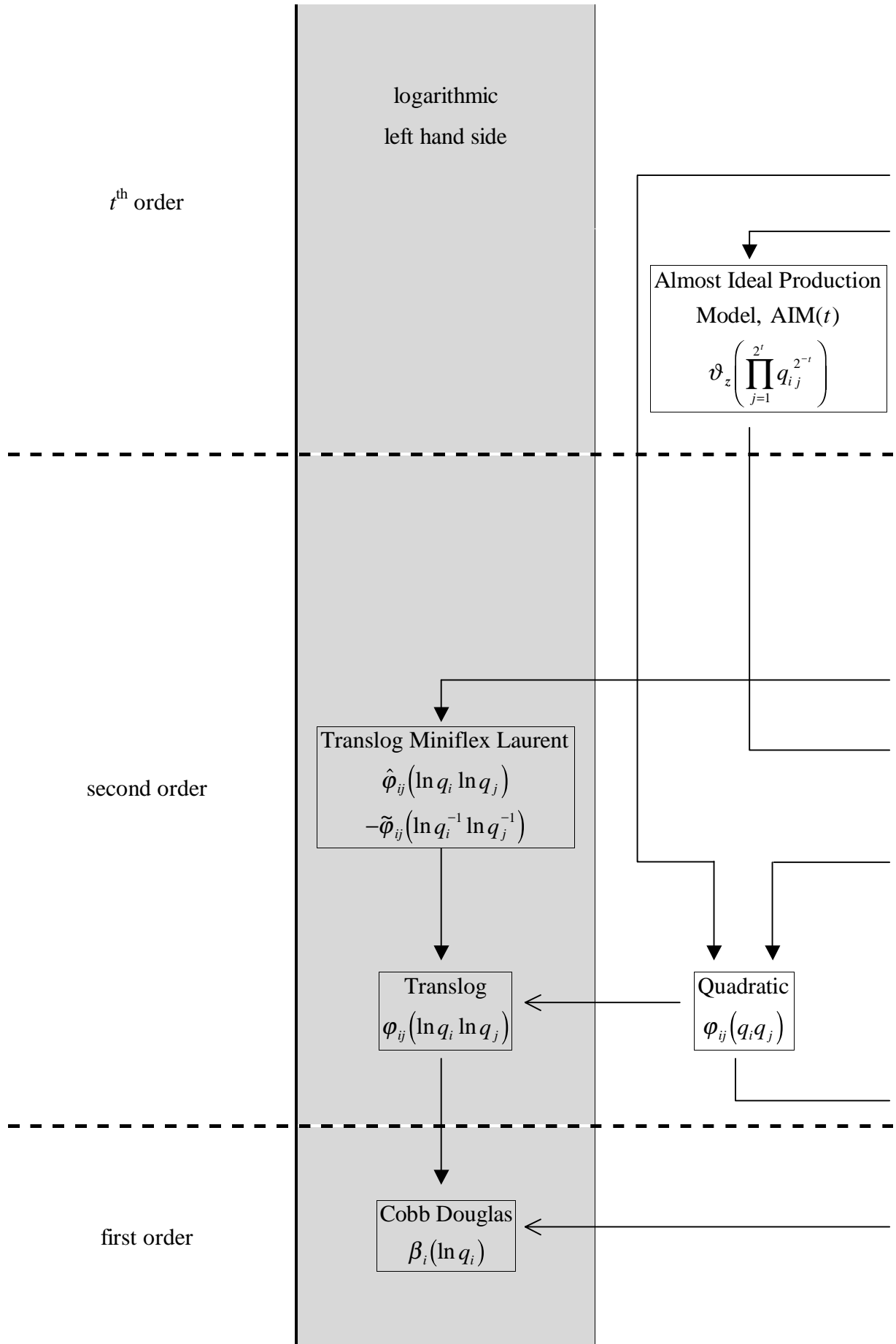
Finally, consider the Generalized Cubic, as the most general formulation of a third order flexible form may be labelled, in the notation-saving non-specific formulation:

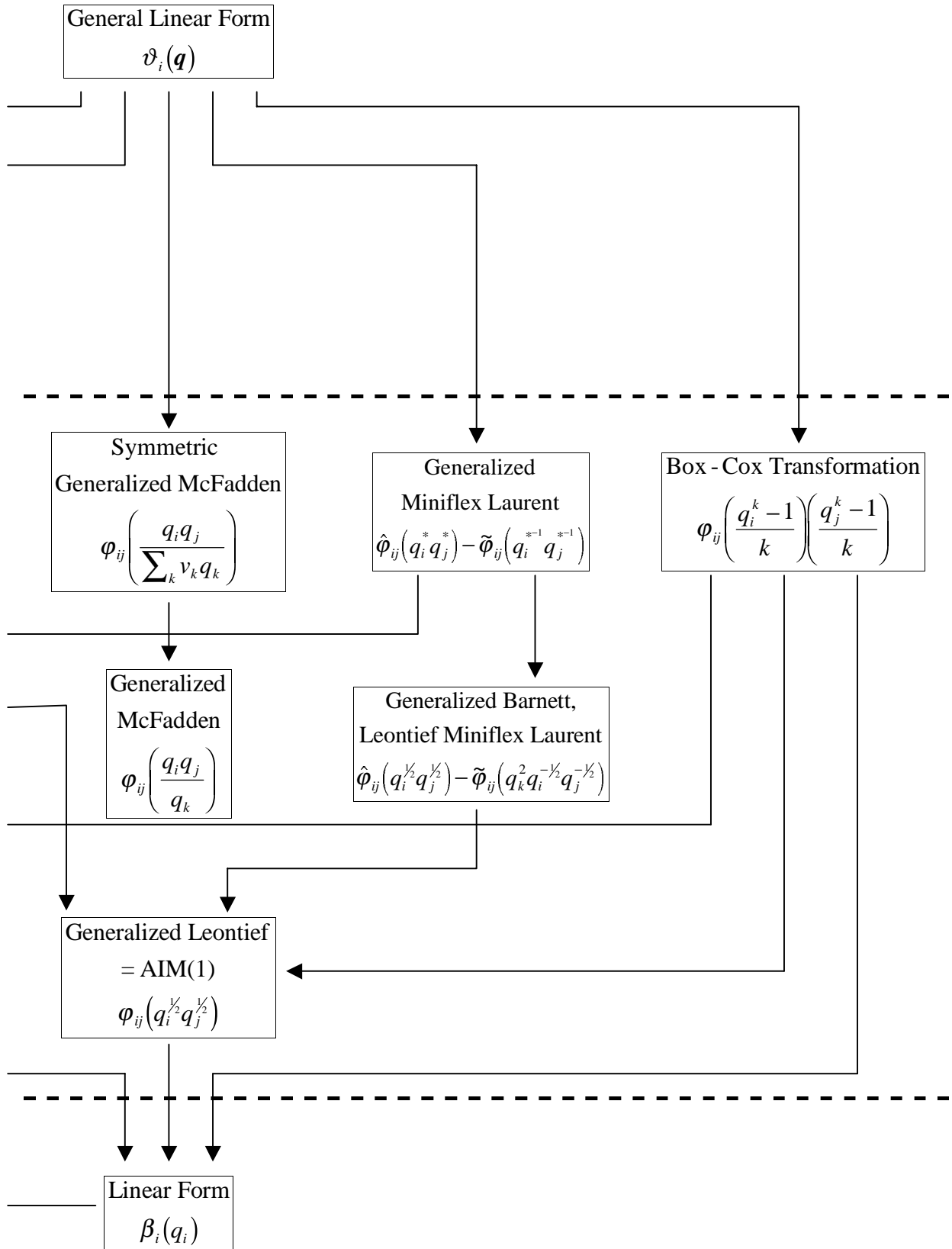
$$\begin{aligned}
F(\mathbf{q}) = & \alpha + \sum_{i=1}^n \beta_i f_i(q_i) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \varphi_{ij} f_i(q_i) f_j(q_j) \\
& + \frac{1}{6} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \xi_{ijk} f_i(q_i) f_j(q_j) f_k(q_k)
\end{aligned}$$

5.2.2 Systematics

The estimation of the parameters of the Generalized Quadratic requires the specification of f_i , which is generally the same for all q_i , at least for each order of effects. This is so not only for aesthetic reasons: first, too many different f_i increase the computational burden, not that much because the variables must be transformed differently, but rather because implementing restrictions quickly becomes tedious. Secondly, heterogeneous microfunctions leave the researcher with insolvable difficulties in parameter interpretation, because point elasticities derived from parameters that are descendent from differently structured microfunctions cannot be compared so easily: in this case, it is no longer possible to distinguish the role of the exogenous variable from the role of the microfunctional form; the *ceteris paribus* principle is violated since the effect of the data on parameter values is not isolated. If parameter interpretation is planned, this argument excludes heterogeneous microfunctions even if empirical evidence, i.e. statistical criteria, would suggest the superiority of such a specification. However, if the goal of the model is to compute forecasts or to test the hypothesis that the maintained model is empirically valid, this objection is obsolete.¹⁶³ Furthermore, note that it is not necessary to find an a priori-argument for differently specified microfunctions: this is an empirical question to the same extent as the question of adequate specification in general.

¹⁶³ See section 6.3 for a discussion of the different goals of an economic behavioral model.





(Arrows with solid tip indicate a special case, arrows with light tip indicate a logarithmic transformation)

The simplest possible case is, of course, all $f_i(q_i) = q_i$, where the resulting flexible functional form is called Quadratic. The historically first flexible functional form, the Transcendental Logarithmic or Translog, which is still widely in use, is based on microfunctions of the form $f_i(q_i) = \ln q_i$. The left hand variable of the Translog is logarithmized too; thus it is sometimes referred to as Double Logarithmic, just like the first order special case of it, the Cobb Douglas functional form. The second early flexible functional form, the Generalized Leontief, is constructed with $f_i(q_i) = q_i^{\frac{1}{2}}$ as characteristic microfunction, i.e. for the second order effects; the first order f_i are linear. For a more intuitive presentation of flexible functional forms with different f_i for first and second order effects it is useful to combine $f_i(q_i)$ and $f_j(q_j)$ to $f_{ij}(q_i, q_j)$. For the Generalized Leontief, $f_{ij}(q_i, q_j)$ then assumes the form $q_i^{\frac{1}{2}} q_j^{\frac{1}{2}}$. Without observing anything about the specific properties of the functional forms, which is the topic of chapter 6, the diagram above depicts the relations between some flexible functional forms. There are two criteria for their selection: they either are or have been among the most commonly used flexible functional forms or, cited despite their empirical insignificance, which is the case with the Generalized Miniflex Laurent and the Generalized Quadratic, they constitute systematic nodes. Consequently, e.g. the pioneering Fourier transformation, which originated a line of research that eventually led to the development of the Almost Ideal Production Model, is not included because it has been applied only a couple of times itself and in addition – apart from simply being a special case of the General Linear Form – stands outside the presented systematics.¹⁶⁴

¹⁶⁴ See section 7.1.3 for a presentation of the Fourier model.

6 Properties of Flexible Functional Forms

As already alluded to in sections 5.1.1 and 5.1.2, economic theory makes some requirements of functional forms that are supposed to depict economically rational behavior, here of dual cost or profit functions. To be a valid or regular economic function, i.e. to be consistent with the behavioral postulates of economic theory, a functional form for a dual cost or profit function must satisfy all properties of a well-behaved cost or profit function which result from the cost minimization or profit maximization hypotheses, respectively.¹⁶⁵ Otherwise it has to be reckoned on estimated parameters inconsistent with economic theory, i.e. the estimated parameters do not match the comparative-static regularities following from economic theory. Before the single regularity properties are addressed, some space will be devoted to the question of where a function can or should entertain those properties, i.e. the question of the domain of applicability, and the question of tractability and theoretical foundation of the estimation procedure.

6.1 Domain of Applicability and Computational Facility

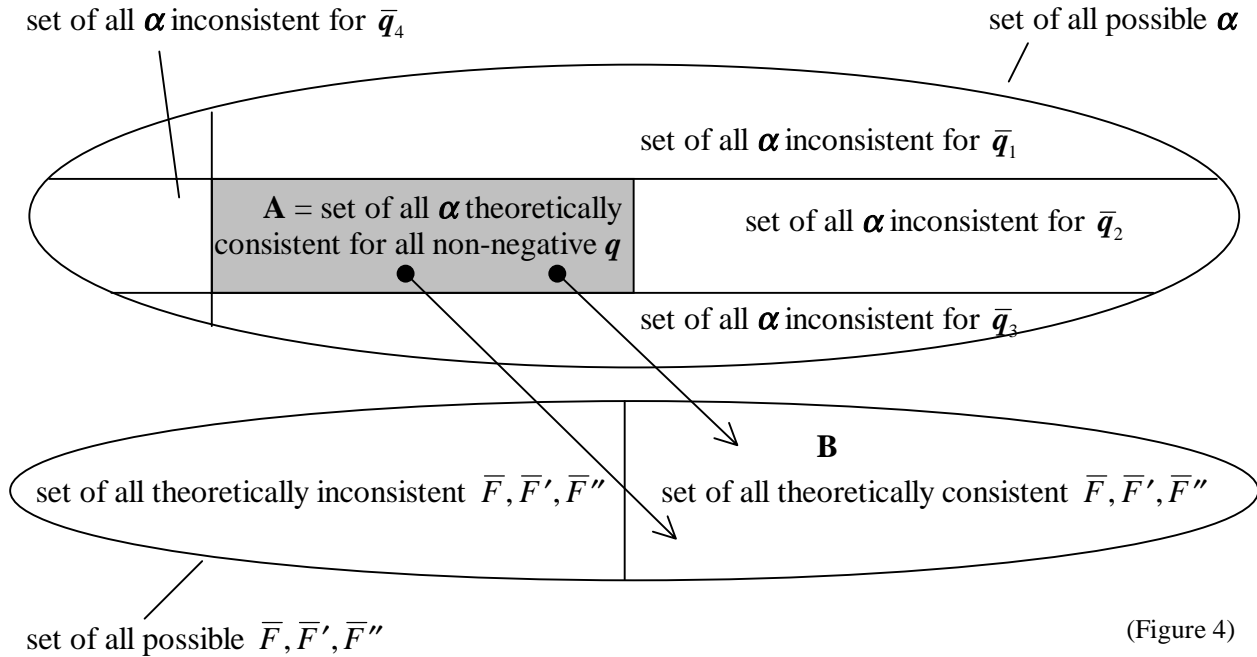
The most common notion of the domain of applicability refers to the set of values of the independent variables q over which the algebraic functional form satisfies all requirements for theoretical consistency, given some parameter vector α .¹⁶⁶ One could think of a flexible function that is globally valid for an arbitrary parameter vector, i.e. that entertains all properties of a well-behaved dual cost or profit function over the complete non-negative orthant of the variable space for an arbitrary α . But this is never the case, not even locally, i.e. near a specific \bar{q} . This can be realized by recalling that the class of well-behaved cost or profit functions is considerably smaller than the class of real-valued functions. The complex restrictions, compared to the class of real-valued functions, of the class of well-behaved cost or profit functions implied by the respective properties of well-behaved cost and profit functions are impossible to reach with an unrestricted parameter vector. Thus, the question is what is the set of parameters for the function to be globally valid?¹⁶⁷ The ideal is clear: to find a set of possible values

¹⁶⁵ See FUSS/McFADDEN/MUNDLAK 1978: 223, or CHAMBERS 1988: 52 and 124, respectively.

¹⁶⁶ See LAU 1986: 1527.

¹⁶⁷ See LAU 1986: 1528.

of α that on the one hand leaves the function globally valid but on the other hand does not restrict its flexibility. In other words, the ideal would be a restriction on α that is not so severe that it renders the solution of the system of equations in the flexibility definition of section 5.1.1 impossible for all theoretically consistent outcomes $\bar{F}, \bar{F}', \bar{F}''$. Consider the following figure 4 depicting the desired state of affairs:¹⁶⁸



The set of all $\bar{F}, \bar{F}', \bar{F}''$ consistent with economic theory is independent of α and thus the same for all points in the variable space. In opposition, the implied restriction on the set of all possible α depends on the choice of \bar{q} , exemplified by \bar{q}_1 through \bar{q}_4 . The shaded area **A** refers to the intersection of the theoretically consistent sets of α for all possible \bar{q} , i.e. the set of parameter vectors that produce only theoretically consistent function values, gradients and Hessians for arbitrary non-negative \bar{q} . Hence, if a fully flexible function is globally valid, there exists a parameter vector α for any element of the set of theoretically consistent $\bar{F}, \bar{F}', \bar{F}''$ such that $F(\alpha; \bar{q}) = \bar{F}$, $\nabla F(\alpha; \bar{q}) = \bar{F}'$ and $\nabla^2 F(\alpha; \bar{q}) = \bar{F}''$ for any non-negative \bar{q} . I.e. a subset of all possible α surjectively maps into the set of theoretically consistent function values, whatever \bar{q} is inserted. Other than for local flexibility, i. e. for only one specific

¹⁶⁸ See figure 1 in section 5.1.1 for comparison.

\bar{q} , this is still quite demanding.¹⁶⁹ Generally, as it follows from Lau's incompatibility theorem,¹⁷⁰ some areas of the set of all consistent $\bar{F}, \bar{F}', \bar{F}''$ cannot be reached from the small remainder of the set of possible α that corresponds with the former one.

But full flexibility in conjunction with theoretical consistency must not generally be given up: there is always a subset of all \bar{q} that is available in the non-negative orthant for which α surjectively maps into the set of consistent $\bar{F}, \bar{F}', \bar{F}''$, i.e. for a restricted domain of applicability of economic theory. The set of all \bar{q} for which it is possible to reach any theoretically consistent $\bar{F}, \bar{F}', \bar{F}''$ via some specific α is called a regular region. Two a priori methods to explore the regular region of a flexible functional form that is kept fully flexible are available, aiming at general knowledge about the behaviour of specific functional forms in specific circumstances, e.g. in the presence of strong complements.

Considering the first method, a grid of variable values is examined for a given parameter vector, and every theoretically consistent point is marked so that the resulting marked area constitutes the regular region. Consistency is examined by explicit solution of the system of derived demand and supply equations with pre-specified elasticities for the parameter vector, the unique parameter vector that solves the system of equations which enables the function to exhibit the pre-specified technology. The respective system of equations becomes insolvable with more than second order effects, since economic theory does not provide enough restrictions.¹⁷¹

The second method does not require an algebraic solution and is based on a Monte-Carlo simulation. To avoid explicit solution for α , the parameter vector for the flexible approximation function is calculated as an arithmetic mean of repeated estimations of data generated by a pre-specified technology – of another functional form than the estimation function, which is either globally valid like the CES or CDE or which is parameterized in a way that all generated values are consistent with theory – with prices randomly drawn from a known distribution. Plotting the regular region then follows the same path as with the first method.¹⁷²

The domain of applicability sometimes also refers to the set of values of the independent variables over which the algebraic functional form is fully flexible (rather than theoretically

¹⁶⁹ See next paragraph and section 6.3 below.

¹⁷⁰ See LAU 1986: 1552-1557 and section 6.3 below.

¹⁷¹ See CAVES/CRISTENSEN 1980.

¹⁷² See e.g. WALES 1977, JENSEN 1997, TERRELL 1995, and the literature cited there.

consistent), given that the function is globally theoretically consistent, but the reverse meaning introduced above is more common.

Another important question connected with the application of flexible functional forms is a practical one: is it possible to determine the function parameters with reasonable expense? The standard econometric procedure, tractable with little methodological and computational requirements and available in many easy-to-use computer applications, is least squares estimation. It only works with functional forms linear in parameters, at least after a known transformation, and linear equality restrictions between model parameters. This turns out to be quite restrictive, since important requirements of economic theory imply inequality restrictions and/or induce the parameter estimation of an originally linear-in-parameters functional form to become a nonlinear optimization problem.¹⁷³

Several problems arise: First, the simple least squares procedure can no longer be applied and one has to switch to nonlinear optimization algorithms, which are more demanding with respect to the technical skills of the user. Second, the risk of non-convergence in the estimation process arises if numerical estimation methods are used. The analytical least squares estimator always has a solution, and linear models almost always converge if estimated using a nonlinear estimator. In contrast, the probability of failure for nonlinear models is at least one order of magnitude higher than for linear models validated numerically. For this reason, even if nonlinearity is inevitable, its degree should be kept at a minimum,¹⁷⁴ which provides argument in favor of staying with intrinsically linear forms even if nonlinear restrictions are present. Third, statistical theory for nonlinear estimators is in most instances far less developed than for the linear least squares estimator, which is not so much a result of historically different states of research but rather a matter of the nature of nonlinearity. The third point is especially serious in so far as it already applies if inequality restrictions are transformed into equality restrictions by the method of squaring introduced in section 6.2.2 below. This method allows the use of standard least squares methods to determine parameter estimates, but statistical theory for least squares estimators is not valid in this case since the underlying

¹⁷³ See section 6.2.2 and 6.2.3 below.

¹⁷⁴ See LAU 1986: 1558-1559.

sampling distribution is truncated.¹⁷⁵ Fortunately, bootstrapping and Bayesian methods provide a solution.¹⁷⁶

6.2 Regularity of the General Linear Form and its Derivates

Regularity is not at all assured with the General Linear Form introduced in section 5.2 above: for the cost function case, it is neither intrinsically homogeneous of degree one nor necessarily everywhere concave and non-decreasing. It is therefore requisite to a priori structure the estimation function according to economic theory as far as possible. To structure the estimation function goes along with its restriction. There are three possible ways of restricting the GLF:

1. The microfunctions f_i can be restricted to depend only on a small subvector of \mathbf{w} and \mathbf{y} or \mathbf{p} and \mathbf{w} , respectively, and f_i can be specified, i.e. given a definite form, with a focus on the resulting properties. This means to determine a GLF derivate or to specify a new flexible functional form, whose properties crucially depend on the properties of the microfunctions.
2. GLF derivates, if not yet theoretically consistent, can be further brought into line with economic theory by imposing parametrical restrictions: the value certain parameters assume is determined a priori or a linear dependance between certain parameters is postulated.
3. The Hessian of GLF derivates can be restricted directly (which eventually amounts to a –complicated – parametrical restriction and can be seen as a special case of 2.). This can be operational since some properties of well-behaved cost functions are immediately mirrored in the Hessian of the respective functions.

With regard to the family tree presented in chapter 3, these three structuring methods can be viewed as the starting point for an alternative categorization of flexible functional forms. Nevertheless, different flexible functional forms only serve as examples for the various ways to fit functions into economic theory or a concrete microeconomic behavioral model here.

¹⁷⁵ See BARNETT/GEWEKE/WOLFE 1991: 7.

¹⁷⁶ See section 6.2.3 below; see e.g. JUDGE et al. 1988 for an introduction to bootstrapping and Bayesian methods in econometrics.

The properties of a well-behaved cost or profit function turn out to be not easily implemented simultaneously in flexible functional forms. The properties of non-negativity, continuity, and no fixed cost or emanation from the origin raise only minor difficulties, and the same is true for the symmetry condition following from differentiability due to Young's theorem. Other than that, monotonicity, curvature, and linear homogeneity in prices concur, as will soon become apparent.

6.2.1 Homogeneity

There are two possibilities to implement linear homogeneity: the choice of microfunctions which guarantee this property globally from the outset, or the imposition of parametric restrictions for functional forms composed of inhomogeneous or not linearly homogeneous microfunctions.

By definition, homogeneity of degree one in prices implies

$$\begin{aligned} C(\lambda \mathbf{w}, y) &= \sum \alpha_i f_i(\lambda \mathbf{w}, y) & \Pi(\lambda \mathbf{p}, \lambda \mathbf{w}) &= \sum \alpha_i f_i(\lambda \mathbf{p}, \lambda \mathbf{w}) \\ &= \lambda \sum \alpha_i f_i(\mathbf{w}, y) & \text{and} & & = \lambda \sum \alpha_i f_i(\mathbf{p}, \mathbf{w}) \\ &= \lambda C(\mathbf{w}, y) & & & = \lambda \Pi(\mathbf{p}, \mathbf{w}) \end{aligned}$$

for all positive λ . From the respective second and third equation, it follows that C and Π are homogeneous of degree one in prices if each f_i is homogeneous of degree one.¹⁷⁷ The pioneering Generalized Leontief Cost Function, specified as a single-output function as in the original article, provides an example:

$$C(\mathbf{w}, y) = \sum_{i=1}^n \beta_i w_i + y \sum_{i=1}^n \sum_{j=1}^n \varphi_{ij} (w_i w_j)^{\frac{1}{2}} + \theta y^2 \sum_{i=1}^n v_i w_i$$

with $\varphi_{ij} = \varphi_{ji}$ and all v_i as predetermined constants to be selected by the researcher, commonly set equal to the respective cost shares to obtain a weighted average price.¹⁷⁸ The first and the third summand are linear in \mathbf{w} and hence linearly homogeneous. But the second summand is also homogeneous of degree one for any values \mathbf{w} may assume, because the microfunctions $(w_i w_j)^{\frac{1}{2}}$ are linearly homogeneous:

¹⁷⁷ See CHAMBERS 1988: 56 and NADIRI 1982: 447.

¹⁷⁸ See DIEWERT 1971: 497.

$$(\lambda w_i \lambda w_j)^{1/2} = (\lambda^2 w_i w_j)^{1/2} = \lambda (w_i w_j)^{1/2}$$

A very operational technique of constructing microfunctions homogeneous of degree one is their normalization. All second order terms, i.e. all terms that will contain the product of two prices, are divided by a definite price called the numéraire. In the most simple case such a microfunction looks like

$$\varphi_{ij} \frac{w_i w_j}{w_k}$$

with w_k as numéraire. This term is clearly homogeneous of degree one:

$$\varphi_{ij} \frac{\lambda w_i \lambda w_j}{\lambda w_k} = \lambda \varphi_{ij} \frac{w_i w_j}{w_k}$$

If other reasons lead to specifications which are not intrinsically homogeneous of degree one, i.e. normalized second order microfunctions or fractional exponents are not wanted, parametric restrictions can force linear homogeneity.

One such reason could be the easy interpretation of the function parameters. In the Trans-log Cost Function,¹⁷⁹ the other important early flexible functional form apart from the Generalized Leontief, the estimated parameters equal the respective elasticities:

$$\ln C(\mathbf{w}, \mathbf{y}) = \alpha + \boldsymbol{\beta}' \ln \mathbf{w} + \boldsymbol{\gamma}' \ln \mathbf{y} + \frac{1}{2} \ln \mathbf{w}' \Phi \ln \mathbf{w} + \frac{1}{2} \ln \mathbf{y}' \Psi \ln \mathbf{y} + \frac{1}{2} \ln \mathbf{w}' \Theta \ln \mathbf{y}$$

where Φ and Ψ are symmetric.¹⁸⁰ This is of course the Generalized Quadratic derivate where $F(C) = \ln C$ and all $f_i(\mathbf{w}) = \ln w_i$ and all $f_i(\mathbf{y}) = \ln y_i$. All f_i are homogeneous, but the degree of homogeneity equals the respective parameter value, i.e. the microfunctions are homogeneous of degree $\beta_i, \gamma_i, \varphi_{ij}, \psi_{ij},$ or θ_{ij} , respectively. According to Euler's theorem, linear homogeneity in \mathbf{w} requires the sum of all elasticities with respect to \mathbf{w} to equal unity:

$$\sum_{i=1}^n \frac{\partial \ln C(\mathbf{w}, \mathbf{y})}{\partial \ln w_i} = \sum_i \beta_i + \sum_{i=1}^n \sum_{j=1}^n \varphi_{ij} \ln w_j + \sum_{i=1}^n \sum_{j=1}^m \theta_{ij} \ln y_j \equiv 1$$

Since both the second and the third summand contain variables that can assume different values, the equation can only be restricted to a definite value if they are both restricted to zero. This is the case if

¹⁷⁹ See CHRISTENSEN/JORGENSON/LAU 1973: 39.

¹⁸⁰ The logarithm of a vector may refer to an element-wise application.

$$\sum_{j=1}^n \varphi_{ij} = 0 \quad \forall i = 1, \dots, n \quad \text{and} \quad \sum_{j=1}^n \theta_{ij} = 0 \quad \forall i = 1, \dots, m.$$

To restrict the entire equation to unity, the remaining summand has to equal unity:

$$\sum \beta_i = 1.$$

Hence, as expected, a total of $m + n + 1$ restrictions establish linear homogeneity.

It has to be emphasized that the homogeneity property of dual behavioral functions results immediately from the economic rationality postulate and does not imply anything about a possible homogeneity property of the underlying primal function, i.e. the technology. Nevertheless, the homogeneity property of a technology, e.g. in the case of constant returns to scale in a linearly homogeneous technology, can of course be expressed in terms of dual functions.

6.2.2 Curvature

Whereas there is no form for microfunctions that would guarantee the required curvature condition for all possible parameter values, a sufficient condition for the General Linear Form to be globally concave or convex is the concavity or convexity of all f_i with all α_i being non-negative.¹⁸¹ From this, Robert G. Chambers infers that, in order to guarantee regularity of a flexible functional form for a dual cost or profit function, it suffices to pick well-behaved cost or profit functions, respectively, for all microfunctions.¹⁸² Unfortunately, it does not fulfill this hope: the condition that all α_i be non-negative has in some cases an unwanted side effect that will become obvious immediately, exemplified with the Generalized Leontief cost function.

The Generalized Leontief cost function is globally concave for non-negative parameters because all its microfunctions are concave in w : linear functions are affine, i.e. concave and convex at the same time, and the second order microfunctions are concave, since their Hessian

$$H\left(w_i^{\frac{1}{2}} w_j^{\frac{1}{2}}\right) = \begin{bmatrix} -\frac{1}{4} w_i^{-\frac{3}{2}} w_j^{\frac{1}{2}} & \frac{1}{4} w_i^{-\frac{1}{2}} w_j^{-\frac{1}{2}} \\ \frac{1}{4} w_i^{-\frac{1}{2}} w_j^{-\frac{1}{2}} & -\frac{1}{4} w_i^{\frac{1}{2}} w_j^{-\frac{3}{2}} \end{bmatrix} \quad \forall i, j$$

¹⁸¹ See CHAMBERS 1988: 311. See also MCFADDEN 1978a: 274-275 for a list of convex microfunctions.

¹⁸² See CHAMBERS 1988: 172.

is negatively semi-definite for positive w_i and w_j , which constitutes the necessary and sufficient condition for concavity.¹⁸³

Unfortunately, a massive shortcoming is connected with a Generalized Leontief restricted to concavity by imposing $\varphi_{ij} \geq 0$: complementarity between factors is ruled out.¹⁸⁴ All cross price effects $\partial x_i / \partial w_j$, i.e. the off-diagonal elements in the Hessian above multiplied with their respective parameter φ_{ij} , are non-negative for non-negative φ_{ij} ; all factors are gross substitutes, which considerably restricts the flexibility of the Generalized Leontief.

Whereas the first order and therefore non-flexible derivative of the Translog, the Cobb-Douglas Cost Function, can easily be restricted to global concavity by imposing $\sum \beta_i \leq 1$ – which is in any case guaranteed by the homogeneity restriction – this is not so with the Translog itself. Robert G. Chambers tries to establish this result by showing that the second partial derivative of a second order microfunction, e.g.

$$\frac{\partial^2 \varphi_{ii} (\ln w_i)^2}{\partial w_i^2} = \frac{2\varphi_{ii} (1 - \ln w_i)}{w_i^2},$$

cannot be restricted to non-positivity, as would be required for global concavity of this microfunction and thus of the Translog as a whole: for positive φ_{ij} , $\ln w_i$ has to be greater or equal unity; for negative φ_{ij} smaller than or equal unity, which is only the case if w_i is greater or smaller than Euler's number $e = 2.7182\dots$ ¹⁸⁵ But this is no valid reason for the Translog to be unqualified for curvature restrictions because the natural logarithm is also taken of C . The curvature properties of $\ln C$ can be destroyed by potentiation with base e , and hence there cannot be inferred from the curvature of $\ln C$ to the curvature of C .¹⁸⁶ The true reason for the unrestrictability of the Translog for curvature becomes obvious when taking anti-logs (here abstracting from output to save notation, as one might say as unit cost function):

¹⁸³ Edward R. Morey mistakenly claims that the Generalized Leontief cannot be restricted to global concavity because its Hessian is still a function of prices, which are subject to change. He infers that concavity at one point does not imply concavity at any other point. But, as an examination of the above Hessian reveals, both Eigenvalues are non-positive for any positive factor prices so that the Hessian is negatively semi-definite in the entire domain. Consequently, as stated, the Generalized Leontief is globally concave for non-negative parameters [see MOREY 1986: 223].

¹⁸⁴ See CAVES/CHRISTENSEN 1980: 425.

¹⁸⁵ See CHAMBERS 1988: 179.

¹⁸⁶ See MOREY 1986: 228.

$$\ln C(\mathbf{w}) = \alpha + \boldsymbol{\beta}' \ln \mathbf{w} + \frac{1}{2} \ln \mathbf{w}' \boldsymbol{\Phi} \ln \mathbf{w}$$

$$C(\mathbf{w}) = e^\alpha \cdot \left(\prod_{i=1}^n w_i^{\beta_i} \right) \cdot \prod_{i=1}^n \prod_{j=1}^n w_i^{(\frac{1}{2} \varphi_{ij} \ln w_j)}$$

First, all terms are connected multiplicatively, and thus the sum-of-concave-functions lemma cannot be applied. Secondly, the second order exponents contain exogeneous variables so that an adding-up restriction between the φ_{ij} in analogy to the Cobb-Douglas would look different for every data point, and a plain restriction of the magnitude of the parameters would not suffice to assure a semi-definite Hessian for all variable values. This means, the Translog can only be restricted locally to concavity or convexity. The only possibility to reach global regularity is to set all φ_{ij} to zero,¹⁸⁷ but this would mean the degeneration of the Translog to the Cobb-Douglas, which is, as already mentioned, not at all flexible.

The inflexibility drawback of the previously shown ways to restrict for curvature can be avoided – or at least relieved – with another method: since the necessary and sufficient condition for a specific curvature consists in the semi-definiteness of its Hessian – if $\nabla^2 C(\mathbf{w}, \mathbf{y})$ is negatively semi-definite, C is concave; if $\nabla^2 \Pi(\mathbf{p}, \mathbf{w})$ is positively semi-definite, Π is convex –, it is possible to restrict the Hessian directly and thereby the underlying function.

Whereas the determinant test is useful for small matrices in algebraic formulation as in the example above, the most operational way of testing numerical matrices for semi-definiteness is the Eigenvalue or spectral decomposition: of all available methods, it requires the fewest calculatory steps and is therefore the method of choice in computer programming.¹⁸⁸ To check for semi-definiteness, it suffices to determine the magnitude of the $m + n$ Eigenvalues. But to restrict a matrix for semi-definiteness, one has to constrain the Eigenvectors to orthonormality in addition, i.e. to normalize each Eigenvector and to take care that they are all orthogonal to each other. Obviously, the Eigenvectors have to be calculated before – and they are a quite complicated function of the Hessian, especially for more than two variables, so that this task can become very tedious for larger models.¹⁸⁹

For the latter purpose, the Cholesky factorization

¹⁸⁷ See LAU 1974: 182-183.

¹⁸⁸ See CROUZEIX/FERLAND 1982: 203.

¹⁸⁹ See LAU 1978: 440.

$$A = LBL'$$

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ l_{21} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ l_{n1} & \cdots & l_{n(n-1)} & 1 \end{bmatrix} \cdot \begin{bmatrix} b_{11} & 0 & \cdots & 0 \\ 0 & b_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & b_{nn} \end{bmatrix} \cdot \begin{bmatrix} 1 & l_{21} & \cdots & l_{n1} \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & l_{n(n-1)} \\ 0 & \cdots & 0 & 1 \end{bmatrix}$$

of a symmetric real valued matrix A with the unit lower triangular matrix L and the diagonal matrix B , containing the Cholesky values, provides an easier solution.¹⁹⁰ A is positively semi-definite if, and only if, all its Cholesky values are non-negative. It is possible to substitute the elements of the Cholesky factorization for the Hessian of the estimation function, more precisely to estimate L and B instead of the elements of the original Hessian A (which is necessarily symmetric according to Young's theorem, and thus the Cholesky factorization always exists). The advantage of the Cholesky factorization is that it contains the same number of elements as the decomposed matrix itself, and that beyond that no restrictions are required. To force positive semi-definiteness, the method of squaring can be applied: auxiliary parameters are substituted against the Cholesky values in the estimation, which are necessarily positive.¹⁹¹

$$b_{ii}^* = (b_{ii})^2 \quad \forall i$$

The only problem connected with the Cholesky factorization lies in the fact that its elements are nonlinear functions of the decomposed matrix, and consequently the resulting estimation function is nonlinear in parameters. Therefore, linear estimation algorithms are ruled out even if the original estimation function is linear in parameters.

Any flexible functional form can be restricted to convexity or concavity with the above method – to local convexity or concavity. The Hessian of most flexible functional forms, for example the Translog or the Generalized Leontief, are not structured in a way that the definiteness property is invariant towards changes in the exogeneous variables.¹⁹² But there are exceptions: e.g. the Hessian of the Quadratic does not contain exogeneous variables at all, and thus a restriction using the Cholesky factorization suffices to impose convexity or concavity at all data points. Unfortunately, the Quadratic is disqualified for its incapability of being re-

¹⁹⁰ See LAU 1978.

¹⁹¹ See LAU 1978: 438.

¹⁹² See JORGENSON/FRAUMENI 1979: 18, and MOREY 1986: 228.

restricted for linear homogeneity. Another exception, the Generalized Symmetric McFadden, will be discussed in section 7.1.4.

6.2.3 Monotonicity

The issue of assuring non-decreasing costs with input prices and output quantities rising was neglected until most recent studies,¹⁹³ and the same is valid for non-increasing profit with input prices rising and non-decreasing profit with output prices rising. But it has appeared that the monotonicity condition is by no means automatically satisfied with most functional forms, and that violations are frequent and empirically meaningful.¹⁹⁴

Since each function with a quadratic specification necessarily has a bliss point, global monotonicity is generally ruled out for all derivatives of the General Linear Form with second order terms if flexibility shall be preserved. Of course, it is always possible to force global monotonicity of e.g. the Generalized Leontief by non-negativity constraints: its gradient

$$\frac{\partial C(\mathbf{w}, y)}{\partial w_i} = \beta_i + y w_i^{-1/2} \sum_{j=1}^n \varphi_{ij} w_j^{1/2} + \theta y^2 v_i \quad \forall i = 1, \dots, n$$

cannot become negative if β_i , all φ_{ij} and θ are bound to non-negativity by the method of squaring presented above.¹⁹⁵ Virtually, the maximum of C with respect to all variables is set to $+\infty$. But this restriction, including a restriction on global concavity,¹⁹⁶ rules out complementarity between factors and thus severely restricts the flexibility of C . Other flexible functional forms restricted this way show a similar loss in flexibility.

To avoid the disturbing choice between inflexible and inconsistent specifications – or at least to make possible a reasonable compromise – an approach imposing theoretical consistency only over the set of variable values where inferences will be drawn was developed recently.¹⁹⁷ The model parameters are restricted in a way that the resulting elasticities meet the requirements of economic theory for the whole range of variable constellations that are a priori likely

¹⁹³ See BARNETT/GEWEKE/WOLFE 1991, BARNETT/KIROVA/PASUPATHY 1996, TERRELL 1996, and KOOP/OSIEWALSKI/STEEL 1994. See also GALLANT/GOLUB 1984.

¹⁹⁴ See BARNETT/KIROVA/PASUPATHY 1996: 2.

¹⁹⁵ See BARNETT/KIROVA/PASUPATHY 1996: 13 and 16-17.

¹⁹⁶ See section 6.2.2 above.

¹⁹⁷ See TERRELL 1996, also see BARNETT/GEWEKE/WOLFE 1991 and KOOP/OSIEWALSKI/STEEL 1994.

to occur, i.e. a regular region is created.¹⁹⁸ Two problems are connected with such a procedure:

First, a decision has to be made about the likeliness of the occurrence of many value combinations of all exogeneous variables, where it may differ largely how precise and how certain such an assumption can be made. A minimum likeliness has to be selected ad hoc to determine the frontiers of the regular region, i.e. the range of variable constellations where theoretical consistency is guaranteed. In other words, there is a lot of contingency involved in a decision which on the one hand affects the remaining flexibility of the estimation function and on the other hand restricts the range of possible forecast scenarios, i.e. the domain of applicability. In addition, for simulation models one has to have an eye on the frontiers of the regular region and to find ways to exclude the non-regular region from the paths of simulation.

Secondly, local restrictions on the parameters are generally inequality restrictions. The method of squaring is not precise enough to provide implementation of a definite regular region and, even more important, standard statistical theory fails to supply a valid foundation of inferences for this kind of restrictions. Few things can be said about the sampling properties of the least squares estimator subject to inequality restrictions: the variance of the estimated parameter vector will be smaller than that of the unrestricted counterpart if the prior information expressed in the restriction is true, but it is always biased.¹⁹⁹ This leads to the use of a completely different econometric framework for inference that integrates the existence of prior information from the outset: the Bayesian approach to econometrics.²⁰⁰ The monotonicity and curvature restrictions can then be imposed using informative priors, and elasticities with a prior that fails to incorporate those restrictions can serve as a reference point. The methodology can be summarized as follows:²⁰¹ all observations are set up in a system of seemingly unrelated regression equations of derived factor demand and marginal cost / product supply equations of the respective functional form with additive errors which are assumed to be jointly normally distributed. Then, the unrestricted prior distributions of the parameters $P_0(\alpha)$ and of the

¹⁹⁸ This approach was preceded by an important study of Gallant and Golub, who imposed regularity on all observed data points [see GALLANT/GOLUB 1984]. This, however, does not necessarily suffice if inference shall be drawn from other data points.

¹⁹⁹ See GREENE 1993: 253.

²⁰⁰ For an introduction to Bayesian econometrics see JUDGE / HILL / GRIFFITHS / LÜTKEPOHL / LEE 1988: 117-156, 275-324, and 806-858.

²⁰¹ For an extensive presentation see TERRELL 1996: 183-185.

error terms $P_0(\Sigma)$ are specified.²⁰² To implement the consistency restrictions, an informative prior distribution is generated using an indicator function $h(\alpha)$ set equal to one if, and only if, all regularity conditions hold²⁰³ for all exogeneous variables which are elements of the subset of positive prices and/or quantities \mathbf{R} over which inference will be drawn, i.e. which are thought to be likely enough to occur, and set to zero otherwise, i.e. for variable constellations outside the regular region:

$$h(\alpha) = \begin{cases} 1 & \text{if } \frac{\partial C}{\partial w_i} \geq 0, \frac{\partial C}{\partial y_j} > 0, \text{ and } \nabla_{ww} C \text{ negatively semi-definite } \forall w_i \in \mathbf{R}^n, y_j \in \mathbf{R}^m \\ 0 & \text{otherwise} \end{cases}$$

here formulated for the cost function. With this indicator function, the informative prior can be defined as

$$P_1(\alpha) = h(\alpha)P_0(\alpha).^{204}$$

Thus, zero value is assigned to parameter vectors implying monotonicity or curvature violations, whereas outside \mathbf{R} the parameter vector is left unrestricted. A very large \mathbf{R} approaches a global restriction on regularity with bad consequences for local flexibility; a smaller \mathbf{R} increases the forecast and simulation problems just mentioned. The posterior density – and thereby the elasticity estimates – are calculated with repeated samples drawn from an unrestricted posterior distribution simulated by a Gibbs sampling algorithm. Then, the restricted estimates are determined using an accept-reject algorithm with respect to $P_1(\alpha)$ for a fine grid over the exogeneous variable space which, under mild regularity conditions, converges to the actual joint density as the number of samples approaches infinity.²⁰⁵ Monotonicity and curvature restrictions are evaluated at each point of the exogeneous variable space, and the respective parameter vector is included in the sample if the conditions in question hold at any points in the grid which is an element of \mathbf{R} , and rejected if violations occur: Violations outside of \mathbf{R} are, of course, ignored. By construction, the informative prior eliminates the portion of the posterior density inconsistent with theory. But the inconsistent part is not simply sliced away; ra-

²⁰² See TERRELL 1996: 184.

²⁰³ The regularity conditions apart from curvature and monotonicity are usually satisfied by construction of the used functional form or, as in the case of the Translog, easily implemented by parametric restrictions without loss of local flexibility.

²⁰⁴ See TERRELL 1996: 184.

²⁰⁵ See CHIB/GREENBERG 1994.

ther the entire distribution shifts towards regularity, which indicates the benefits of this method of imposing theoretical consistency. The decision of how large R is chosen can be supported by comparing posterior odds for different R and applying Jeffreys' criteria for conclusive evidence or by comparing the number of rejections and acceptances during the drawing process for different R .²⁰⁶ Both ways of evaluating the choice of R at the same time give information about the flexibility cost of implementing regularity.

This numerical technique to assure regularity exactly meets the requirements of the global approximation concept introduced in section 5.1.3 above. Higher orders of parameters than second order render the incorporation of theoretical consistency into flexible functional forms impossible, and thus the presented Bayesian technique remains as the only possibility. Of course, functional forms thus estimated must be fully flexible and implementation of consistency must be completely left to the accept-reject algorithm, respectively, because the algorithm can reject inconsistent solutions but not accept consistent and flexible solutions that are excluded from the outset, since they cannot be depicted by the functional form. Hence, only such parametrical restrictions may be introduced to relieve the strain on the estimation procedure that do not at all limit flexibility, e.g. the common homogeneity restrictions.

6.2.4 Relative Importance of Regularity Conditions

It is noticeable how naturally the homogeneity, curvature, and monotonicity property were and are differently accounted for in empirical studies. However, this is completely groundless: none of them is more evident or more meaningful than the others. They all are equally fundamental axioms of economic theory, and any failure has the same consequences from a theoretical perspective. Theoretical foundation of an economic model requires satisfaction of all regularity conditions in conjunction.²⁰⁷

The common neglect of monotonicity seems to be the result of a widespread misinterpretation and trivialization of a violation: e.g. for a linear demand function necessarily intersecting the corresponding input price axis and thus producing negative demand quantities over a certain price, if it is decreasing as dictated by the curvature condition, a kinked shape is assumed where the demand quantities beyond the intersection are set to zero. If input price

²⁰⁶ See JEFFREY 1967: 432.

²⁰⁷ This conclusion, however, is not meant to suggest that cost minimization or profit maximization, respectively, is the only rational decision criterion: other criteria are possible, but any other criterion implies a specific set of regularity conditions of the respective behavioral function, for which the above statement applies.

exceeds a certain level it is prohibitively high, so it is argued.²⁰⁸ But a negative value of a demand function is different from zero; if one holds that a prohibitively high price is empirically evident, the chosen model has to be capable of producing zero or near zero demands over a region. The severe consequences of this misinterpretation which go far beyond the single equation's irregular region may be indicated by the fact that in relations between parameters inside and between equations the invalid negative quantities also come into effect and thereby cause further irregularities. An example is given below.

Three reasons seem to have promoted the implementation of homogeneity as a standard requirement for any application a long time before the importance of the other aforementioned properties was registered. First, homogeneity finds its obvious and significant manifestation in the single equation. In contrast, correct curvature of the cost or profit function reduces to the correct sign of the own substitution effect of the single demand or supply equation or the exclusion of an inverse demand or supply reaction, which is not immediately visible as an incorporation of the concavity condition. In systems of equations, curvature becomes more complex than homogeneity.

Secondly, for this reason homogeneity is easier to implement into many real valued functions even of a higher order. The procedures assuring correct curvature for systems of equations are much more cumbersome, as the above sections reveal.

Thirdly, in part as a result of the other two points and in part as a historical contingency, the techniques to implement homogeneity are more common in economic literature. Even among protagonists of the new efforts to incorporate not only homogeneity and curvature but also monotonicity properties in flexible functional forms, it is common to treat the consistency properties asymmetrically: an arbitrary functional form is transformed into a "cost function", as e.g. Dek Terrell formulates in his recent paper on flexibility and regularity properties of the Asymptotically Ideal Production Model (AIM), if it is made homogeneous of degree one; incorporation of curvature and monotonicity then do not affect the label "cost function".²⁰⁹ A stringent phrasing would be either to call a function a "(valid) cost function" if, and only if, it incorporates all properties of a well-behaved cost function, or – with regard to numerical techniques of assuring regularity posterior to specification of functional form and with regard to the fact that virtually no flexible functional form would then deserve this predi-

²⁰⁸ Beleg?!#

²⁰⁹ See TERRELL 1995: 2.

cate then – to call any function "cost function" if it serves as a cost function in an economic model, no matter which properties it has.

This does not, of course, mean that the implementation of some regularity conditions while leaving others aside is worthless: a homogeneous model failing to satisfy curvature and monotonicity is likely to yield better forecasts in a heuristic sense than a model which is non-homogeneous in addition. But this is something other than claiming proper theoretical foundation, and furthermore heuristics *can* as well suffer from the implementation of some rather than all theoretic requirements, as the example of a monotonicity violation shows: a negative demand quantity which could result from only one badly measured data series for only one exogeneous variable causes only one estimation parameter to be ill-conditioned and to yield wrong forecasts for only one variable in an unrestricted model. Now, the parametrical homogeneity restriction transmits the error on the other parameters of the equation because the other summands have to compensate the error of the one, and concavity, adding-up (emerging from homogeneity too), and symmetry eventually spread the error over the entire system. The disturbing consequence is that there are two rather than three epistemological stages: Apart from models incorporating all regularity conditions in conjunction, which is science proper with falsifiable theories, there is the land of floating pragmatism where the amount of theory used to supplement economic intuition is of purely heuristic status. It remains ultimately uncertain whether additional theory yields better forecasts. The latter is a matter of experience and instinctive feel in combination with empirical long-time evidence – a matter of wisdom rather than a matter of knowledge.

6.3 Trade-Off Between Selection Criteria and Conclusion

Against the background of the enormous difficulties connected with the practical reconciliation of local flexibility, theoretical consistency and domain of applicability demands one might well wonder whether the ideal solution of a fully second-order-flexible functional form that is globally consistent with economic theory – and, in addition, easy to estimate and safeguarded by solid statistical theory – is possible at all. Lawrence Lau proves that at least for functional forms linear in parameters, i.e. for functional forms that ideally meet the requirement of easy computability, this is not the case: his incompatibility theorem states that there can be no linear-in-parameters and parsimonious functional form for a normalized unit cost

function²¹⁰ which can fit arbitrary but theoretically consistent values of a normalized unit cost function and its first and second derivatives at any pre-assigned value of the normalized price and be itself theoretically consistent for all non-negative normalized prices,²¹¹ i.e. which provides a surjective mapping from the parameter set into the set of theoretically consistent function values, gradients, and Hessian.²¹² He concludes that, since one is not likely to give up theoretical consistency and flexibility or even computational facility, the logical area of compromise lies in the domain of applicability. This is exactly in conformance with the Bayesian method outlined in section 6.2.3 above. An additional argument for being content with a restricted domain of applicability is supplied by the fact that the likelihood that the parameter values, estimated on the basis of the available observations, remain the same outside the neighbourhood of the observations is in any case questionable. Furthermore, the confidence band quickly becomes wide so that the extrapolative use of parameter estimates becomes increasingly risky with rising distance from the observed data. Finally, infinity is a purely theoretical concept when it comes to economic data: economic processes always happen inside a finite range of variable constellations over extension of which one can reason.²¹³

Whereas Lau assigns equal weight to local flexibility and theoretical consistency,²¹⁴ one could argue that local flexibility is tendentially more important because the data is primary and economic theory is secondary. We had better allow the data to be inconsistent with economic theory than previously restrict the possible information we allow the data to give us: Our theory of economically rational behavior could be wrong or phenomenologically inadequate, respectively, but the data cannot – measurement and specification error excluded –; it is our only source of information about the nature we want to describe with our theories. Other than reconstructing data information that is not depictable by the restricted estimation function, we can test consistency with a functional form not restricted for consistency.

²¹⁰ Lau proposes several dimensions of generalization for his theorem which remain to be proved: the number of independent variables could be increased, the number of parameters could be increased, and the functional form could be linear-in-parameters including monotonic transformations [see LAU 1986: 1557]. Further, he states without proof that his theorem equally applies to the dual profit function and dual revenue function [see LAU 1986: 1558].

²¹¹ See LAU 1986: 1552-1557.

²¹² See section 6.1 above.

²¹³ See LAU 1986: 1557.

²¹⁴ See LAU 1986: 1557.

Implementing consistency can be seen from two completely different aspects. On the one hand, if the analysis aims at the extraction of parameters for forecast and simulation use, i.e. as applied theory, validity of the theory is generally presupposed, and its implementation is led by the hope that this will result in better heuristic quality. In order to increase forecast quality, it seems advisable to include all available information – and the assumption that economic theory is valid constitutes extremely valuable information about existing relations between effects that held in the past and will hold in future. However, a counter example is given in section 6.2.4 above. On the other hand, estimating a theoretically consistent model – or validating it,²¹⁵ as parameter estimation is also called characteristically enough – always includes an epistemological test: is a statistically significant estimation based on the respective theory possible?

If not, it is questionable whether – in case, of course, we do not prefer to assume that bad data or a misspecification caused the failure – economic theory is applicable to the problem in hand. If general validity of theory is claimed, and this is the case for neoclassical economic theory and scientific theories in general, the excuse of inapplicability is not available, and thus the neoclassical researcher is left with the burden of proof that something else has to be blamed for the failure of his validation trial; otherwise the theory would be falsified.²¹⁶ However, following from the Duhem-Quine-thesis,²¹⁷ the only possibility to actually prove that it is not the invalidity of economic theory but something else that causes model rejection is to successfully estimate an alternative model where the blamed problem is solved. Otherwise, theory is neither rejected nor remains unrejected – the model is simply epistemologically meaningless. It is questionable if one has to go as far as Barnett, Geweke, and Wolfe, who claim that models violating maintained theory are virtually useless in policy applications²¹⁸ since a heuristic forecast quality may still exist.²¹⁹ But one has to realize that one has completely left the (puta-

²¹⁵ The term "validation" implies that after a successful empirical test the examined theory may be seen as "true". It would be more precise to label a successfully tested theory "unfalsified" since no test can positively prove the validity of a theory but rather can constitute a (further) example where the theory escapes rejection [see POPPER 1961 and section 1.2.1 above]. But these epistemological considerations do not touch the above argumentation.

²¹⁶ See section 4.4.1 for another area where unambiguous falsification of theory is made impossible.

²¹⁷ See QUINE 1981.

²¹⁸ See BARNETT/GEWEKE/WOLFE 1991: 15.

²¹⁹ See section 6.1.4 above.

tively) safe grounds of neoclassical economic theory, with all bad consequences for argumentation.

Considering simulation models, the demand is a little higher for a practical reason: global consistency for the applied parameter vector (but not for any possible parameter vector) is highly desirable for an application of estimated functional forms in simulation models since convergence of its optimization algorithm is guaranteed with a globally well-behaved goal function. This is especially important if the starting values are relatively distant from the optimum. A globally well-behaved optimization problem significantly reduces the difficulties connected with the choice of starting values, i.e. it eliminates the sensitivity of the results to starting values because it precludes multiple optima, and it reduces the likelihood of non-convergence because the optimization algorithm cannot be placed into local irregularities.²²⁰

The picture changes if global flexibility is considered. It provides a conceptual framework where both the applied and the epistemological aspect of estimation largely benefit. Given that the statistical data is rich enough to allow for the evaluation of a large number of parameters in a globally flexible estimation, a locally flexible function is, of course, quite restrictive: it entertains the same degree of complexity as point elasticities, i.e. it depicts level, first, and second order effects of the cost or profit function, and that means level and first order effects of the derived demand or supply functions. With rising deviation of the demand or supply function from linearity or a transformation of linearity, respectively, and with rising distance between the the forecast scenario and the point of approximation or the point of elasticity evaluation, respectively, the prediction error increases, and a globally flexible estimation makes increasingly more sense.²²¹

²²⁰ These considerations do not apply for simulation models that can be solved analytically: these are in any case intrinsically consistent, otherwise an analytical solution would not exist. However, the restrictiveness of globally regular models has to be taken into account when using analytically solvable models.

²²¹ Sometimes, point elasticities are, adherent to the local flexibility approach described above, erroneously misunderstood as the ultimate scenario tool which has to be favored over insertion of variables into the estimated function. An impressive example is provided by Perroni and Rutherford [see PERRONI/RUTHERFORD 1995]: with regard to the current practice of using point elasticity estimates taken from the literature to calibrate computable equilibrium models, they introduce a procedure to test the property of second order flexible functional forms to preserve local calibration information away from the point for which the elasticities are known. Obviously, the elasticity information is only locally valid, and, given no additional information is available, the hypothesis that the true function entertains the same elasticity (or the same share or difference of elasticities) everywhere is as probable to be true as the hypothesis that a certain second order flexible functional form is the correct specification with regard to the true data generating process. Additional information

The severe restrictions on local flexibility reported for all common flexible functional forms when expanding the regular region with numerical techniques²²² are easy to cope with from an applied perspective: the lack of local flexibility, i.e. the impossibility of reaching some theoretically consistent level, gradient, and Hessian values *at one point* is of purely academic meaning in a globally flexible context. Consequently, the relative importance of local flexibility and theoretical consistency changes fundamentally for functional forms approaching global flexibility, i.e. for series expansions or higher-than-second-order flexible functional forms that fit any technology data arbitrarily well with a number of parameters approaching infinity. Consistency, even global consistency, then appears to take absolute priority over local flexibility, whatever local flexibility costs occur. Lau's incompatibility theorem is no longer effective. In addition, the compulsion to set complete columns of parameters linearly dependent to retain local flexibility for the remaining variables demanded by the semi-flexibility approach ceases to apply: any parameters appearing to contribute to the significance particularly few can be set to zero without further considerations. The decision of which parameters to estimate and which to neglect, as well as the choice between the series expansion and higher-order-flexibility approach in general and between functional forms in particular is merely due to statistical criteria.²²³

The objection that the empirical meaningfulness of such an approach is limited since one is not likely to encounter data rich enough to allow for the estimation of that many parameters very frequently, and that consequently standard second order flexible forms remain superior, can be handled easily: second order flexible functions are generally included as a special case.

could e.g. consist in knowledge about the functional specification from which the elasticities are derived and which is accepted in the source study. If this is e.g. a CES, it would of course be of benefit to use a functional form for simulation with minimal deviation from the CES hypothesis, but if it is e.g. a Generalized Leontief or any other non-CES form, this selection criterion would be wrong.

²²² See TERRELL 1996.

²²³ The accusation of data mining misses the point because setting one parameter to zero and introducing another due to statistical criteria does not amount to the questionable practice of letting the data decide whether a variable has an influence on the outcome of the equation or not. As in standard second order flexible forms, the maintained hypotheses clearly state *which* variables are relevant. In contrast to second order flexible forms, which assign equal weight to each variable, a globally flexible estimation leaves the question of *how precisely* the influence of each variable can be depicted to the data. In any case there is no approach to formulate an a priori hypothesis on this, and thus there is nothing reprehensible about asking the data – just like it is a proper practice to let the estimation parameter values be determined by the data rather than formulating a hypothesis on their magnitude, apart, of course, from parameter restrictions implied by economic theory.

If the data is not capable of providing information for more than level, first order and second order parameters, the global flexibility approach will favour exactly this depth of parameterization. But if there is more or less inferential potential, the globally flexible estimation as a semi-nonparametric procedure will react while the statical second order flexible functional form will withhold information or fail statistically, respectively.

The epistemological benefit can be seen in the increased likelihood of a consistently modeled hypothesis – and thus the underlying theory – of escaping empirical falsification, since a variable depth of parameterization, subject only to consistency restrictions but free from the burden of assuring local flexibility, leaves more room to fit the data. Fitting the data arbitrarily well while maintaining theoretical consistency inside the observed region only, or even globally, constitutes the ideal of testing theory with statistical means, and each additional parameter which can be estimated significantly increases the credibility of the specified model and the applied theory – functional form as source of specification error is now excluded.²²⁴ From this, it follows that, both from an applied and a theoretical perspective, both with respect to an instrumental and an axiomatic approach to applied science,²²⁵ research should focus on globally flexible functional forms and semi-nonparametric estimation methods with the critical remainder of consistency restrictions that damage flexibility when incorporated parametrically implemented not before the estimation procedure.

²²⁴ A "wrong" functional specification merely amounts to a waste of degrees of freedom. Again, this does not imply data mining [see previous footnote]. How could the hypothesis that the data fits e.g. a translog specification rather than a Generalized Leontief specification be justified a priori? This question remains necessarily an empirical one.

²²⁵ See section 1.2.1 for a discussion of the instrumental and the axiomatic approach to applied science.

7 Applying Global Flexibility

Rather than a reiteration of known results about the properties of common locally flexible functional specifications, the goal of the overview given in the first section of this chapter is to evaluate the suitability of these to serve as basis for a globally flexible or semi-non-parametric estimation with a constrained regular region in the sense outlined in the two preceding chapters. Since neither a simulation experiment nor an empirical application is conducted in the course of this study, the considerations remain speculative and are all exposed to the objection of arm-chair science. Correspondingly, this part of the study is rather cursory. Afterwards, an outlook will be given which identifies remaining weak spots of the presented approach and shows directions of further research, marking the way from a theoretical foundation of the global flexibility approach – which is hopefully provided by the study at hand – towards a broadly applicable methodical tool.

7.1 Selection of Functional Form

In a semi-nonparametric estimation with a constrained regular region, there is only one reason to incorporate theoretical requirements in a functional form: to save degrees of freedom and to relieve the accept-reject algorithm from the burden of accounting for them, respectively,²²⁶ i.e. the reason to ease estimation.²²⁷ If the statistically accepted regular region is large enough both from a hypothesis test and a forecast perspective, it is irrelevant if a theoretical requirement is accounted for by construction, by parametrical restriction or during the estimation process. Thus, the ability of a functional form to allow for the incorporation of possibly many properties of a well-behaved cost or profit function while remaining locally flexible, having dominated the whole functional form literature since the seventies, loses weight as a selection criterion. In turn, another criterion, having stayed in the background since then, gains importance: is the data generating process following a quadratic or a log-

²²⁶ See section 6.2.3 above.

²²⁷ Note that an implementation of e.g. homogeneity by parametrical restriction always applies globally. Thus, different regular regions imposed during the estimation process do not include properties thus enforced. This may have negative implications on estimability if the respective property is not supported by the data outside the regular region. However, there is no way to find this out but an estimation where the respective property is implemented not before the estimation process so that it can be included in variations of the regular region.

linear course? How far does the approximating function behave like the true structure in a purely statistical sense? How well does the function actually approximate the data, theoretical properties left aside? This revival of the old question of an adequate specification leaves more room for choice between functional forms, since the ability to incorporate theory has led to very few functional forms – mostly not more than one – which were considered state of the art and thus acceptable as estimation functions. Hence, the following discussion of a small selection of functional forms has likewise to account for recent examples and specifications that are outdated in the context of second order flexible functional forms for their inferior ability to incorporate theoretical properties. This is not to imply that the extensive literature on regularity of second order functional forms is worthless for the purpose of this study – the opposite is the case: for the global flexibility account advocated here, it is of essential significance to draw upon knowledge of functional forms that allow a maximum of incorporated regularity while remaining fully flexible.

All functional forms discussed below in a historical order can be found in the systematics diagram in section 5.2.2 above. The choice of the functional forms presented here is somewhat arbitrary, but tries to emphasize both forms which are or were broadly used in applied studies, and/or can be seen as pathbreaking methodical progress. For the notation, please refer to section 5.2.1 above.

7.1.1 Generalized Leontief

The pioneering Generalized Leontief Cost Function, leading off the extensive literature on second order flexible functional forms motivated by the endeavour to make the progresses of duality theory empirically utilizable,²²⁸ was introduced as

$$C(w, y) = \sum_{i=1}^n \beta_i w_i + y \sum_{i=1}^n \sum_{j=1}^n \varphi_{ij} (w_i w_j)^{\frac{1}{2}} + \xi y^2 \sum_{i=1}^n v_i w_i$$

with $\varphi_{ij} = \varphi_{ji}$ and all v_i as predetermined constants to be selected by the researcher.²²⁹ Since it does not treat input and output variables symmetrically, several multi-output generalizations are possible. Consider for example

²²⁸ See section 1.1.2 and 1.1.3 above.

²²⁹ See DIEWERT 1971: 497.

$$\begin{aligned} \Pi(\mathbf{p}, \mathbf{w}) &= \alpha_0 + \sum_{i=1}^n \beta_i w_i + \sum_{i=1}^n \gamma_i p_i \\ &+ \frac{1}{2} \left(\sum_{i=1}^n v_i^p p_i \right) \sum_{i=1}^n \sum_{j=1}^n \varphi_{ij} (w_i w_j)^{1/2} + \frac{1}{2} \left(\sum_{i=1}^n v_i^w w_i \right) \sum_{i=1}^m \sum_{j=1}^m \psi_{ij} (p_i p_j)^{1/2} \\ &+ \sum_{i=1}^n \sum_{j=1}^m \theta_{ij} \ln w_i \ln p_j \end{aligned}$$

which emphasizes the weighting of the second order terms with the respective other variables, here incorporated by an average of these, but simultaneously reveals the ugly asymmetry by means of the impossibility of formulating a reasonable weight for the mixed second order summand and the neglect of the last term of the original single-product cost function formulation – which is, again not exactly straightforward, a third order term, and it could be moreover argued that the original formulation does not include any proper second order terms because of the γ weight. Another possibility, aesthetically more convincing, is

$$\begin{aligned} \Pi(\mathbf{p}, \mathbf{w}) &= \alpha_0 + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \varphi_{ij} (w_i w_j)^{1/2} + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \psi_{ij} (p_i p_j)^{1/2} + \sum_{i=1}^n \sum_{j=1}^m \theta_{ij} (w_i p_j)^{1/2} \\ &= \alpha_0 + \frac{1}{2} \mathbf{w}' \Phi \mathbf{w}^{1/2} + \frac{1}{2} \mathbf{p}' \Psi \mathbf{p}^{1/2} + \mathbf{w}' \Theta \mathbf{p}^{1/2} \end{aligned}$$

which assumes an indifference with respect to the question of whether a variable is an input or output variable. The β_i and γ_i terms can be dropped because they are equal to the φ_{ii} and ψ_{ii} terms, but this implies that the resulting functional form is no longer second order flexible by the number-of-parameters criterion. The Generalized Leontief is linearly homogeneous in prices by construction, but curvature and monotonicity can either be implemented locally only,²³⁰ or, if restricted for globally, the second order flexibility property is lost.²³¹

Similar to the multi-output generalization, the generalization for higher order flexibility is not straightforward. One solution is the Asymptotically Ideal Production Model (AIM), of which the Generalized Leontief is the special case where the order of expansion equals one.²³² Another approach would be to continue the construction principle guaranteeing homogeneity with microfunctions including more parameters for higher order flexibility. In this case, the microfunctions assume the form

²³⁰ See sections 6.1.1, 6.1.2, and 6.1.3 above.

²³¹ See LAU 1986: 1531-1539.

²³² See section 7.1.5 below.

$$\frac{1}{6} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \xi_{ijk} (w_i w_j w_k)^{1/3}$$

for third order terms.²³³ Symmetry of all Hessians by Young's theorem implies that $\xi_{ijk} = \xi_{jik} = \xi_{ikj} = \xi_{kij} = \xi_{kji} = \xi_{jki}$ does not restrict the generality of the function, and is hence assumed.²³⁴

For terms of order t , the microfunctions can be written as

$$\frac{1}{t!} \sum_{i \in I^t} \vartheta_i \prod_{j=1}^t w_{i_j}^{1/t}$$

where i denotes the multi-index over which summation is performed, and I^t is the set of t -tuples $I^t = \{i_1, i_2, \dots, i_t\}$ with $i_j \in \{1, \dots, n\}$ and $i_1 \leq i_2 \leq \dots \leq i_t$. For a comparison of the expansion and the higher order flexibility approach to increasing the depth of parameterization of locally flexible functional forms please refer to section 5.1.4.

Methods of incorporating curvature and monotonicity are uninteresting for a semi-nonparametric estimation based on the Generalized Leontief since they either work merely locally – as opposed to the requirement of implementing them for a regular region –, or, if global, destroy the flexibility property of the Generalized Leontief. This means that, in order to establish a functional form that satisfies the requirements of a globally flexible estimation, a Generalized Leontief which is linearly homogeneous by construction and otherwise unrestricted is perfectly suitable. This is so because the Generalized Leontief entertains only theoretical properties, namely homogeneity, which leave its capability unaffected to depict any consistent behavior without a priori restrictions not implied by economic theory – i.e. because it is locally flexible.

7.1.2 Transcendental Logarithmic (Translog)

The locally flexible functional form following the Generalized Leontief is the Transcendental Logarithmic or Translog.²³⁵ The microfunctions of the Generalized Quadratic are of the form $f_i(q_i) = \ln q_i$, and the left hand variable, i.e. cost or profit, is logarithmized, too:

²³³ In contrast to the second order case, a simplifying matrix notation is not available here.

²³⁴ See section 5.1.4 and 5.2.1 above.

²³⁵ See CHRISTENSEN/JORGENSON/LAU 1973: 256.

$$\begin{aligned} \ln \Pi(\mathbf{p}, \mathbf{w}) &= \alpha_0 + \sum_{i=1}^n \beta_i \ln w_i + \sum_{i=1}^m \gamma_i \ln p_i \\ &+ \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \varphi_{ij} \ln w_i \ln w_j + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \psi_{ij} \ln p_i \ln p_j + \sum_{i=1}^n \sum_{j=1}^m \theta_{ij} \ln w_i \ln p_j \\ &= \alpha_0 + \beta' \ln \mathbf{w} + \gamma' \ln \mathbf{p} + \frac{1}{2} \ln \mathbf{w}' \Phi \ln \mathbf{w} + \frac{1}{2} \ln \mathbf{p}' \Psi \ln \mathbf{p} + \ln \mathbf{w}' \Theta \ln \mathbf{p} \end{aligned}$$

where the \ln operator denotes taking logarithms element-wise, and with Φ and Ψ being symmetric. Setting Φ , Ψ , and Θ to null matrices reveals that the Translog is a generalization of the Cobb-Douglas functional form, which was introduced in 1928,²³⁶ and, apart from the Constant Elasticities of Substitution (CES) form,²³⁷ which entertains the Cobb-Douglas as limiting case, dominated applied economics until the development of the Translog, and is still widely in use. The Translog is probably the best investigated second order flexible functional form, and surely the one with the most applications. In contrast to the Generalized Leontief, the third and higher order extension is straightforward and exhibits the form

$$\frac{1}{6} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \xi_{ijk} \ln w_i \ln w_j \ln w_k$$

for third order terms.²³⁸ Again, symmetry of all Hessians by Young's theorem implies that $\xi_{ijk} = \xi_{jik} = \xi_{ikj} = \xi_{kij} = \xi_{kji} = \xi_{jki}$ does not restrict the generality of the function and is hence assumed.²³⁹ Higher order terms assume the form

$$\frac{1}{t!} \sum_{i \in I'} \vartheta_i \prod_{j=1}^t \ln w_{i_j},$$

where the notation is the same as with the Generalized Leontief in section 7.1.1 above.

The theoretical properties of the second order Translog are well-known:²⁴⁰ it is easily restrictable for global homogeneity, correct curvature can be implemented only locally if local flexibility shall be preserved, and monotonicity, as with all other derivatives of the Generalized Quadratic, is impossible to maintain globally without losing second order flexibility. For hig-

²³⁶ See COBB/DOUGLAS 1928.

²³⁷ See ARROW/CHENERY/MINHAS/SOLOW 1961.

²³⁸ In contrast to the second order case, matrix notation neither simplifies presentation nor makes the way in which the parameters and variables interact more obvious, and is thus omitted.

²³⁹ See section 5.1.4 and 5.2.1 above.

²⁴⁰ See LAU 1986: 1530-1533, and sections 6.2.1 through 6.2.3 above.

her orders, homogeneity can be incorporated in analogy with the second order case without flexibility loss, as will be demonstrated with a third order unit cost function to save notation:

$$\ln c(\mathbf{w}) = \alpha_0 + \sum_{i=1}^n \beta_i \ln w_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \varphi_{ij} \ln w_i \ln w_j + \frac{1}{6} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \xi_{ijk} \ln w_i \ln w_j \ln w_k$$

is globally linearly homogeneous by Euler's theorem if

$$\sum_{i=1}^n \frac{\partial \ln c(\mathbf{w})}{\partial \ln w_i} = \sum_i \beta_i + \sum_{i=1}^n \sum_{j=1}^n \varphi_{ij} \ln w_j + \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \xi_{ijk} \ln w_j \ln w_k \equiv 1$$

Since both the second and the third summand contain variables that can assume arbitrary values, the equation can only be restricted to a definite value if they are both restricted to zero. This is the case if

$$\begin{aligned} \sum_{j=1}^n \varphi_{ij} &= 0 \quad \forall i = 1, \dots, n \\ \sum_{k=1}^n \xi_{ijk} &= 0 \quad \forall i, j = 1, \dots, n \end{aligned}$$

To restrict the entire equation to unity, the remaining summand has to equal unity:

$$\sum \beta_i = 1.$$

The Translog is perfectly suitable as a functional form for a semi-nonparametric estimation aiming at global flexibility because, if merely restricted for global homogeneity, it has only theoretical properties which leave its capability to depict any behavior without a priori restrictions not implied by economic theory unaffected, i.e. is locally flexible. As the examination of the Generalized Leontief above yielded, and as it will reveal below, the same can be said about other second order flexible functional forms and their extensions to higher degrees of flexibility.

But the Translog entertains two advantages over all other specifications: first, it is extremely convenient to estimate, and secondly, it is likely to be a particularly good specification for economic processes. In addition, the unambiguous multiple-output generalization may be seen as an advantage. With regard to the first point, all parameters of the Translog are evaluated when estimated as a system of $n - 1$ cost share equations and the logarithmic cost function itself, where the form of the share equations, here formulated for the third order unit cost function above, is agreeably simple:

$$s_i = \frac{w_i x_i}{c(\mathbf{w})} = \frac{\partial \ln c(\mathbf{w})}{\partial \ln w_i} = \beta_i + \sum_{j=1}^n \varphi_{ij} \ln w_j + \sum_{j=1}^n \sum_{k=1}^n \xi_{ijk} \ln w_j \ln w_k$$

Considering the appropriacy for depicting economic processes, the Translog can be shown to possess a comparably large region of theoretical consistency when restricted for consistency for only one point.²⁴¹ This can be interpreted as the desirable property of being relatively close to theoretically consistent behavior by construction, i.e. by the way the variables interact in this specification prior to any trial to rule out theoretically inconsistent behavior by parametrical restrictions. Empirical evidence supports this mathematical finding: throughout applied economic literature, a relative superiority of the Translog with respect to statistical fit can be reported, which already occurred with its predecessor, the Cobb-Douglas form. A recent example is Terrell's study where he applies a Translog, Generalized Leontief, and Symmetric Generalized McFadden cost function to the classical Berndt and Wood data, utilizing exactly the technique which is appropriate for the global flexibility concept advocated in this study. This result is especially interesting since the Generalized Leontief constitutes the second order flexible case of the AIM which is today's state of the art, suggesting that Translog extensions to higher order could frequently outdo the AIM too. One is tempted to conclude that the natural logarithm is closer to the true data generating processes of economic decision making than other transformations. In any case, the Translog is a particularly promising candidate for an application in a globally flexible estimation.

The possible objection that the Translog is more restrictive than other second order flexible functional forms, since it exhibits constant share elasticities is based on a misconception: all alternative candidates have not more than one parameter per effect either, so that their identical restrictiveness is merely veiled by their "ability" – also being a constraint in fact – to produce varying elasticities. More flexibility requires more parameters, as with any other functional form. In contrast, the iso-elasticity property must be seen as an advantage over other forms since it allows an immediate parameter interpretation. However, this advantage is lost with higher order specifications because the elasticities then depend on more than one parameter.

²⁴¹ See LAU 1986: 1538. Also see CAVES/CHRISTENSEN 1980.

7.1.3 Fourier

The first series expansion to be introduced as an economic functional form is the Fourier model.²⁴² Due to the extensive notation which is requisite to formulate a Fourier cost function, and anticipating that it is in any case an inappropriate choice to depict cost minimizing behavior, as will become obvious immediately, it shall be refrained from listing it here. As a series expansion, the Fourier model is the first specification which allows semi-nonparametric estimation methods and was shown to be capable of a globally flexible estimation.²⁴³ In addition, the basic principles of the recent techniques of imposing consistency on functional forms during the estimation process, set out in section 6.2.3 above, have been developed with it.²⁴⁴ Nevertheless, the Fourier model is not an ideal candidate for approximating economic behavior since it consists of sines and cosines, i.e. is a periodic function which is not exactly close to a well-behaved economic function. Thus, it is not surprising that it is difficult to implement regularity without thereby restricting the flexibility of the Fourier form in an unacceptable manner. In addition, the technical expenditures of estimating its parameters seem to be that high that Barnett, Geweke, and Wolfe recommend leaving application of the Fourier model to professional econometricians. Both problems are to some extent overcome with the AIM presented in section 7.1.5 below.

7.1.4 Symmetric Generalized McFadden

Since its introduction in 1987, any econometric study of demand or supply behavior relying on another specification is exposed to the question of why the Symmetric Generalized McFadden was not utilized: this functional form is commonly considered state of the art, questioned not until the presentation of the AIM and the global flexibility approach, respectively.²⁴⁵ The single-product cost formulation of the Symmetric Generalized McFadden, a close relative of the Generalized Leontief treated in section 7.1.1 above, is

²⁴² See GALLANT 1981.

²⁴³ See GALLANT 1982, see section 7.2 below.

²⁴⁴ See GALLANT/GOLUB 1984.

²⁴⁵ See section 7.1.5 below.

$$\begin{aligned}
 C(\mathbf{w}, y) &= \sum_{i=1}^n \beta_i w_i + \frac{1}{2} y \left(\sum_{i=1}^n v_i w_i \right)^{-1} \sum_{i=1}^n \sum_{j=1}^n \phi_{ij} w_i w_j + y \sum_{i=1}^n \theta_i w_i + y^2 \xi \sum_{i=1}^n v_i w_i \\
 &= \boldsymbol{\beta}'\mathbf{w} + \frac{1}{2} y \frac{\mathbf{w}'\boldsymbol{\Phi}\mathbf{w}}{\mathbf{v}'\mathbf{w}} + y \cdot \boldsymbol{\theta}'\mathbf{w} + y^2 \xi \mathbf{v}'\mathbf{w}
 \end{aligned}$$

with $\phi_{ij} = \phi_{ji}$ and all v_i as predetermined constants to be selected by the researcher.²⁴⁶ Possible multi-output generalizations and higher order extensions are analogous to the Generalized Leontief presented in section 7.1.1 above. Like the Generalized Leontief, the Symmetric Generalized McFadden is linearly homogeneous in prices by construction, and monotonicity can either be implemented locally only or, if restricted for globally, the second order flexibility property is lost.²⁴⁷ But there is one important difference which provides the reason for the common distinction of the Symmetric Generalized McFadden as state of the art: if it is restricted for correct curvature by Lau's technique using the Cholesky decomposition,²⁴⁸ the constrained curvature property applies globally. Unfortunately, the second order flexibility property is in this case restricted to only one point.²⁴⁹ This drawback, recently reported as empirically meaningful against the original expectations,²⁵⁰ and the incapability of incorporating the global monotonicity property could well constitute the best possible result that can be obtained with regard to the project of a globally regular and locally flexible functional form, noting that Lau's incompatibility theorem states that this ideal is in any case impossible to reach with linear functional forms.

The results of this study qualify the progress implied by the invention of the Symmetric Generalized McFadden. First, noting that theoretical consistency requires all properties to apply in conjunction, the ability to incorporate one more regularity property, but still not all, is irrelevant from an epistemological perspective. Secondly, in an applied context, it is likely that also accounting for curvature increases forecast credibility, but this must not necessarily be the case, in particular as the local flexibility property is harmed by this restriction.²⁵¹ Thirdly, the Bayesian method of incorporating regularity eliminates the relative attractiveness of the Generalized Symmetric McFadden since, on the one hand, with this method the expendi-

²⁴⁶ See DIEWERT/WALES 1987a.

²⁴⁷ See section 6.1.3 above.

²⁴⁸ See section 6.1.2 above.

²⁴⁹ See DIEWERT/WALES 1987a: 54.

²⁵⁰ See RYAN/MAH 1994.

²⁵¹ See section 6.2.4 above.

ture of restricting second order flexibility to one data point is way too high compared to the advantage of being globally restrictable on curvature and, on the other hand, Lau's technique does not yield a globally correct curvature for higher order generalizations of the Symmetric Generalized McFadden, because the structure of the Hessian becomes more complex when third and higher order effects are considered so that it is not applicable in any case. Nevertheless, it is possible that the Symmetric Generalized McFadden, not restricted for correct curvature to preserve full flexibility, turns out to be a good specification to utilize in a globally flexible estimation simply because of the way the variables interact with it, i.e. that it overdoes other forms like the Translog or the AIM in a statistical sense, because it is a specification that is closer to the true data generating process.

7.1.5 Asymptotically Ideal Production Model (AIM)

The decisive step towards a broad applicability of semi-nonparametric methods and the progress of Gallant's and Golub's technique to account for theoretical consistency, namely the imposition of a regular region using numerical techniques, consists of the introduction of the Asymptotically Ideal Production Model, the AIM. It is a behavioral function based on the multivariate version of the Müntz-Szatz series expansion.²⁵² The AIM(t), i.e. with a series expansion of order t ,²⁵³ formulated as a single-product, constant returns to scale cost function formulation as in the original article,²⁵⁴ assumes the form

$$C(\mathbf{w}, y) = y \sum_{i \in I^t} \vartheta_i \prod_{j=1}^{2^t} w_{i_j}^{2^{-t}},$$

with the index notation as introduced in section 7.1.1 above. To gain the feeling of its construction, note that the AIM(1) reduces to the Generalized Leontief, presented in section 7.1.1 above, and that the AIM(2) constant returns to scale cost function for a single output and, with respect to the last formula, for three inputs, can be written as

²⁵² See BARNETT/JONAS 1983.

²⁵³ Note the difference between the order of expansion and the order of terms or local flexibility considered above.

²⁵⁴ See BARNETT/GEWEKE/WOLFE 1991.

$$\begin{aligned}
 C(\mathbf{w}, y) &= y \sum_{i_1=1}^n \sum_{i_2=1}^n \sum_{i_3=1}^n \sum_{i_4=1}^n \vartheta_{\{i_1, i_2, i_3, i_4\}} \prod_{j=1}^4 w_{\{i_1, i_2, i_3, i_4\}_j}^{1/4} \\
 &= y \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n \vartheta_{ijkl} w_i^{1/4} w_j^{1/4} w_k^{1/4} w_l^{1/4} \\
 &= y \sum_{i=1}^n \beta_i w_i + y \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \varphi_{ij} w_i^{1/2} w_j^{1/2} + y \underbrace{\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n \xi_{ijkl} w_i^{1/4} w_j^{1/4} w_k^{1/4} w_l^{1/4}}_{\forall i, j, k, l \text{ where at least one index differs from all others}} \\
 &= y(\beta_1 w_1 + \beta_2 w_2 + \beta_3 w_3 \\
 &\quad + \varphi_{12} w_1^{1/2} w_2^{1/2} + \varphi_{13} w_1^{1/2} w_3^{1/2} + \varphi_{23} w_2^{1/2} w_3^{1/2} \\
 &\quad + \xi_{1112} w_1^{3/4} w_2^{1/4} + \xi_{1113} w_1^{3/4} w_3^{1/4} + \xi_{1222} w_1^{1/4} w_2^{3/4} + \xi_{1333} w_1^{1/4} w_3^{3/4} + \xi_{2223} w_2^{3/4} w_3^{1/4} + \xi_{2333} w_2^{1/4} w_3^{3/4} \\
 &\quad + \xi_{1123} w_1^{3/4} w_2^{1/4} w_3^{1/4} + \xi_{1223} w_1^{1/4} w_2^{3/4} w_3^{1/4} + \xi_{1233} w_1^{1/4} w_2^{1/4} w_3^{3/4})
 \end{aligned}$$

where the notation approaches step by step the conventions used in this study. Considering the indices of the last formulation reveals that the "own-effects" reduce to the respective lower order effects

$$\begin{aligned}
 \varphi_{11} w_1^{1/2} w_1^{1/2} &= \beta_1 w_1 \\
 \xi_{1122} w_1^{1/4} w_1^{1/4} w_2^{1/4} w_2^{1/4} &= \varphi_{12} w_1^{1/2} w_2^{1/2}
 \end{aligned}$$

and can thus be omitted. Global homogeneity of degree one is guaranteed by the exponents summing up to unity by construction, and global concavity, as with the Generalized Leontief, can be enforced by restricting all parameters to non-negativity, thereby sacrificing local flexibility and disqualifying it for a globally flexible estimation.²⁵⁵

Terrell performs a simulation study to examine the performance of the AIM where he utilizes Bayesian techniques to create a regular region following the Gallant and Golub approach of accounting for consistency. It is not surprising that he encounters a superiority of the AIM(3) against the AIM(1) or Generalized Leontief and a second order Translog. A more important result is that the AIM(3) significantly overcomes the problem of the AIM(1) to depict strong complements. Furthermore, he finds strong evidence that non-negativity constraints in order to enforce global concavity render the AIM incapable of approximating most of the simulated data, whatever order of expansion is chosen.²⁵⁶ It yet remains to be checked whether the superiority of the AIM still exists if compared to a Translog specification with the same number of parameters or, in a semi-nonparametrical estimation, with that depth of pa-

²⁵⁵ See section 6.3 above.

²⁵⁶ See TERRELL 1995.

of parameterization the data is able to support. Barnett, Geweke, and Wolfe suspect that the AIM has a particularly large natural regular region,²⁵⁷ but this can well be a consequence of the higher degree of flexibility in general rather than a property of the AIM specification in particular as opposed to e.g. the Translog specification: possibly, many regularity violations encountered with second order flexible functional forms are a result of their limited flexibility rather than actually caused by the data. These considerations are somehow contradicted by the findings of Jensen, who plots regular regions for the AIM using data generated with a CDE functional form and, in contrast to Terrell's study, additive errors. He reports the smaller natural regular regions the higher the order of expansion is, and he observes that statistical fit is sometimes better with lower order AIMs.²⁵⁸ However, Jensen, following his project of determining the natural regular regions of the AIM, does not restrict for consistency, so that his recommendation to be careful with higher order AIMs in the presence of noisy data is based on a wrong premise: his estimation, if it is misinterpreted as a model for an empirical application rather than a test of how large the natural regular regions are, is exposed to the objection that it does not account for regularity violations caused by overfitting by imposing regularity over a region, which is adequate no matter whether one estimates generated data which is known to be consistent but noisy, or real world data of which nothing is known a priori.

7.2 Outlook

There are several open questions with the presented approach, most of which can only be answered empirically or by performing simulation experiments, respectively. At least one question, however, is a purely mathematical one: for the Fourier expansion and the AIM, it is already shown that they possess the global flexibility property in the sense that they asymptotically can reach any continuous function. This is, however, not a trivial matter where it suffices to refer to the possibility of adding infinitely many parameters. Instead, this property must be verified in four steps: first, a matrix norm must be found to measure the distance between the approximated function and the approximation function, i.e. to measure the approximation error. In the case of the Fourier model and the AIM, the Sobolev norm was used. Then, it must be checked whether the examined functional form is, secondly, continuous, and thirdly, dense with respect to the norm. Finally, it has to be shown that the norm is continuous with

²⁵⁷ See BARNETT/GEWEKE/WOLFE 1991: 41.

²⁵⁸ See JENSEN 1997.

respect to the estimation method. For the details of this verification procedure see Gallant.²⁵⁹ This procedure has to be performed for any functional form prior to its use in a globally flexible estimation, that is, with regard to the recommendation of this study to try the Translog, in particular for the higher-degree flexible Translog specification.

With regard to an application of the global flexibility concept, there are four approaches which could be considered: first, although there exist two studies on the performance of the AIM, neither of these provides the information which is needed to evaluate its suitability for the global flexibility approach proposed in this study. Rather, an analysis has to be performed which, first, uses consistently generated data with an additive error, like Jensen's study but unlike Terrell's, because otherwise, important properties like the behavior with regard to overfitting cannot be observed. Secondly, regularity must be enforced for a sufficiently large region, like Terrell's study but unlike Jensen's, who discontinues his efforts at an early stage because of local inflexibility. Thirdly, a comparison of different depths of parameterization with regard only to statistical criteria has not been performed, i.e. ignoring the share of rejections caused by consistency violations in implementing the regular region or the amount of binding restrictions, respectively, because it is impossible to find out whether these are caused by an inconsistent data generation process or by overfitting the noise in the data and thus irrelevant: Superiority in the context of the global flexibility account of this study requires a superior statistical fit only.

Secondly, another fruitful direction of further research would be to conduct a simulation experiment similar to the one proposed above with a semi-nonparametric Translog, given that it turns out that the Translog is theoretically suitable for a globally flexible estimation according to the criteria sketched in the first paragraph of this section. Reasons for a possible superiority of the Translog over the AIM are presented in section 7.1.2 above.

Thirdly, there is another aspect of the studies on semi-nonparametric estimation of dual behavioral functions, i.e. the three studies on the AIM mentioned in section 7.1.5 above, which calls for a revision: in all these studies there is discrimination between AIM(1), AIM(2), and AIM(3), whereas mixed parameterizations are neglected. As the considerations in section 5.1.4 above imply, this is by no means requisite. Instead, a statistically optimal approximation of the data would allow for an inhomogeneous depth of parameterization which is the greater the more an estimation of the respective parameters increases overall statistical significance.

²⁵⁹ See GALLANT 1982.

Fourthly, a positive answer to one of the questions above would, of course, suggest an empirical application of either the AIM or the Translog model or both with the methods outlined here. A particularly attractive choice would be the well-known Bernd-Wood data, which was already used in a number of comparative analyses of flexible functional forms. Another desirable application would be the compound feed cost model proposed in the outlook of the first part of this study, so that the two parts of this study could finally be united into one model.

References

- ARROW, K.J. / CHENERY, H. B. / MINHAS, B. S. / SOLOW, R.M. 1961: *Capital-Labour Substitution and Economic Efficiency*. Review of Economics and Statistics, pp. 225-250.
- Außenhandel: Fachserie 7, Reihe 2, div. issues, Statistisches Bundesamt Metzler-Poeschel, Stuttgart.
- BARNETT, William A. / GEWEKE, John / WOLFE, Michael 1991: *Seminonparametric Bayesian Estimation of the Asymptotically Ideal Production Model*. Journal of Econometrics, pp. 5-50.
- BARNETT, William A. / JONAS, Andrew 1983: *The Münzt-Szatz Demand System: An Application of a Globally Well Behaved Series Expansion*. Economic Letters, pp. 337-342.
- BARNETT, William A. / KIROVA, Milka / PASUPATHY, Meenakshi 1996: *Technology Modeling: Curvature is not Sufficient for Regularity*. Economics Working Paper Archive ewp-em/9602002;
- BINSWANGER, Hans P. 1974: *A Cost Function Approach to the Measurement of Elasticities of Factor Demand and Elasticities of Substitution*. American Journal of Economics, pp. 377-386.
- BLACKORBY, C. / PRIMONT, D. / RUSSEL, R.R. 1978: *Duality, Separability, and Functional Structure: Theory and Economic Applications*. North Holland; New York, Oxford, Shannon.
- BLACKORBY, Charles / DIEWERT, Walter E. 1979: *Expenditure Functions, Local Duality and Second Order Approximations*. Econometrica, pp. 579-603.
- BOYD, M.S. / BRORSEN, B.W. 1986: *Dynamic Price Relationships for U.S. and EC Corn Gluten Feed and Related Markets*. European Review of Agricultural Economics, pp. 199-215.
- Bundesgesetzblatt I, div. issues, Bundesanzeiger-Verlags-GmbH, Bonn.
- CAVES, D.W. / CHRISTENSEN, Laurits R. 1980: *Global Properties of Flexible Functional Forms*. American Economic Review, pp. 422-432.
- CHAMBERS, Robert G. 1988: *Applied production analysis: A dual approach*. Cambridge University Press; Cambridge, New York, New Rochelle, Melbourne, Sydney.
- CHIB, S. / GREENBERG, E. 1994: *Hierarchical analysis of SUR models with extensions to correlated serial errors and time varying parameter models*. Journal of econometrics, forthcoming.
- CHRISTENSEN, L.R. / JORGENSON, D.W. / LAU, L.J. 1973: *Transcendental Logarithmic Production Frontiers*. American Economic Review, pp. 367-383.
- COBB, C.W. / DOUGLAS, P.C. 1928: *A Theory of Production*. American Economic Review, pp. 139-165.
- CROUZEIX, J. / FERLAND, J.A. 1982: *Criteria for quasi-convexity and pseudo-convexity: relationships and comparisons*. Mathematical Programming, pp. 193-205.
- DIEWERT, Walter E. / NAKAMURA, Alice O. 1993: *Essays in Index Number Theory*, vol. I, North Holland; Amsterdam, London, New York, Tokyo.

- DI EWERT, Walter E. / WALES, Terence J. 1987a: *Flexible Functional Forms and Global Curvature Conditions*. *Econometrica*, pp. 43-68.
- DI EWERT, Walter E. / WALES, Terence J. 1987b: *A Normalized Quadratic Semiflexible Functional Form*. *Journal of Econometrics*, pp. 327-342.
- DI EWERT, Walter E. / WALES, Terence J. 1988: *A Normalized Quadratic Semiflexible Functional Form*. *Journal of Econometrics*, pp. 327-342.
- DI EWERT, Walter E. 1971: *An Application of the Shepard Duality Theorem: A Generalized Leontief Production Function*. *Journal of Political Economy*, pp. 481-507.
- DI EWERT, Walter E. 1974: *Applications of Duality Theory*. In: INTRILIGATOR, Michael D. / KENDRICK, D.A. (Hrsg.): *Frontiers of Quantitative Economics*, vol. II. North Holland; Amsterdam, Oxford, New York, pp. 106-171.
- DI EWERT, Walter E. 1982: *Duality Approaches to Microeconomic Theory*. In: ARROW, K. J. / INTRILIGATOR, M.D. (Ed.): *Handbook of Mathematical Economics*, vol. II. North Holland; Amsterdam, New York, Oxford, pp. 535-599.
- Ernährungsdienst, div. issues, Verlag Alfred Strothe, Frankfurt.
- FÄRE, Rolf / PRIMONT, Daniel 1995: *Multi-Output Production and Duality: Theory and Applications*. Kluwer Academic Publishers; Boston, London, Dordrecht.
- FEGER, Fritz / MÜLLER, Detlef 1999: *Strong Separability in the Binary and Extended Partition and the Functional Structure of the Dual Profit Function*. Diskussionsbeitrag 9904 des Instituts für Agrarökonomie der Universität Göttingen.
- FEGER, Fritz 1995: *Schätzung einer vollständigen Matrix von Nachfrageelastizitäten für Futtermittel*. Diplomarbeit, Institut für Agrarökonomie der Universität Göttingen.
- FUSS, Michael / MCFADDEN, Daniel / MUNDLAK, Yair 1978: *A Survey of Functional Forms in the Economic Analyses of Production*. In: FUSS, Michael / MCFADDEN, Daniel (ed.): *Production Economics: A Dual Approach to Theory and Application*. North Holland; Amsterdam, pp. 219-268.
- GALLANT, A. Ronald / GOLUB, Gene H. 1984: *Imposing Curvature Restrictions on Flexible Functional Forms*. *Journal of Econometrics*, pp. 295-321.
- GALLANT, A. Ronald 1981: *On the Bias in Flexible Functional Forms and an Essentially Unbiased Form: The Fourier Flexible Form*. *Journal of Econometrics*, pp. 211-245.
- GALLANT, A. Ronald 1982: *Unbiased Determination of Production Technologies*. *Journal of Econometrics*, pp. 285-321.
- GREENE, William H. 1993²: *Econometric Analysis*. Maxwell Macmillan; New York, Oxford, Singapore, Sydney.
- GRINGS, Michael 1985: *Ein Abgebotsmodell für den Agrarsektor der Bundesrepublik Deutschland: Ökonometrische Schätzung auf der Grundlage der Dualitätstheorie*. Schriften zur angewandten Ökonometrie, Haag + Herchen Verlag; Frankfurt/Main.
- HALL, Robert E. 1973: *The Specification of Technology with Several Kinds of Output*. *Journal of Political Economy*.
- HASENKAMP, Georg 1976: *Specification and Estimation of Multiple-Output Production Functions*. Springer-Verlag; Berlin, Heidelberg, New York.
- HOTTELING, Harold 1932: *Edgeworth's Taxation Paradox and the Nature of Demand and Supply Functions*. *Journal of Political Economy*, pp. 577-616.

<http://econwpa.wustl.edu/eprints/em/papers/9602/9602002.abs>

- JEFFREYS, H. 1967³: *A Theory of Probability*. Oxford University Press; Oxford.
- JENSEN, Mark J. 1997: *Revisiting the Flexibility and Regularity Properties of the Asymptotically Ideal Production Model*. *Econometric Reviews*, pp. 179-203.
- JUDGE, George G. / HILL, R.Carter / GRIFFITHS, William E. / LÜTKEPOHL, Helmut / LEE, Tsoung-Chao 1988²: *Introduction to the Theory and Practice of Econometrics*. John Wiley & Sons, New York.
- KOHLI, Ulrich 1981: *Nonjointness and Factor Intensity in U.S. Production*. *International Economic Review*, pp. 3-18.
- KOOP, Gary / OSIEWALSKI, Jacek / STEEL, Mark 1994: *Bayesian Efficiency Analysis with a Flexible Form: The AIM Cost Function*. *Journal of Business and Economic Statistics*, pp. 339-346.
- Landwirtschaftliches Wochenblatt Westfalen-Lippe, div. issues, Landwirtschaftsverlag GmbH, Münster.
- LAU, Lawrence J. 1972: *Profit Functions of Technologies with Multiple Inputs and Outputs*. *Review of Economics and Statistics*, pp. 281-289.
- LAU, Lawrence J. 1974: *Comments on Applications of Duality Theory*. In: INTRILIGATOR, Michael D. / KENDRICK, D.A. (ed.): *Frontiers of Quantitative Economics*, vol. II. North Holland; Amsterdam, Oxford, New York, pp. 176-199.
- LAU, Lawrence J. 1978: *Testing and imposing monotonicity, convexity and quasi-convexity constraints*. In: FUSS, M. / MCFADDEN, D. (ed.): *Production Economics: A Dual Approach to Theory and Applications*, vol. I. North Holland; Amsterdam, New York, Oxford, pp. 409-453.
- LAU, Lawrence J. 1986: *Functional Forms in Econometric Model Building*. In: GRILICHES, ZVI / INTRILIGATOR, M.D. (Ed.): *Handbook of Econometrics*, vol. 3. Elsevier; Amsterdam, New York, Oxford, pp. 1515-1566.
- LEONTIEF, Wassilij. W. 1947: *A note on the interrelation of subsets of independent variables of a continuous function with continuous first derivatives*. *Bulletin of the American Mathematical Society* no. 53, pp. 343-350.
- MCFADDEN, Daniel 1978: *The General Linear Profit Function*. In: FUSS, Michael / MCFADDEN, Daniel (ed.): *Production Economics: A Dual Approach to Theory and Applications*, vol. 1. North Holland; Amsterdam, New York, Oxford, pp. 269-286.
- MERGOS, G.J. / YOTOPOULOS, P. A. 1988: *Demand for Feed Inputs in the Greek Livestock Sector*. *European Review of Agricultural Economics*, pp. 1-17.
- Mischfutter-Tabellarium, Fachverband der Futtermittelindustrie (ed.), div. issues, Bonn.
- MOREY, Edward R. 1986: *An Introduction to Checking, Testing, and Imposing Curvature Properties: The True Function and the Estimated Function*. *Canadian Journal of Economics*, pp. 207-235.
- NADIRI, M. Ishaq 1982: *Producers Theory*. In: ARROW, Kenneth J. / INTRILIGATOR, Michael D. (ed.): *Handbook of Mathematical Economics*, vol. II. North Holland; Amsterdam, Oxford, New York, pp. 431-490.
- PEETERS, Ludo / SURRY, Yves 1993: *Estimating Feed Utilisation Matrices Using a Cost Function Approach*. *Agricultural Economics*, pp. 109-126.

- PEETERS, Ludo / SURRY, Yves 1997: *A Review of the Arts of Estimating Price-Responsiveness of Feed Demand in the European Union*. Journal of Agricultural Economics, pp. 379-392.
- PERRONI, Carlo / RUTHERFORD, Thomas F. 1995: *A Comparison of the Performance of Flexible Functional Forms for Use in Applied General Equilibrium Analysis*. University of Warwick Discussion Paper 441.
- POPPER, Karl 1961: *The Logic of Scientific Discovery*. Science Editions; New York.
- QUINE, William O. 1981: *Five Milestones of Empiricism*. In: *Theories and Things*. Belknap; Cambridge (Mass.), pp. 67-72.
- ROY, R. 1942: *De l'utilité – Contribution à la théorie des choix*. Hermann; Paris.
- RYAN, D. L. / MAH, T. W. 1994: *Resolving curvature violations using flexible functional forms: an empirical examination of alternative approaches*. Unpublished Working Paper.
- SALVANES, Kjell G. / TJØTTA, Sigve 1995: *Local Properties of Estimated Flexible Functional Forms*. Working Paper 1795, Department of Economics, University of Bergen.
- SHEPHARD, Ronald W. 1953/1981: *Cost and Production Functions*. Princeton University Press / Springer Verlag; Berlin, Heidelberg, New York.
- SONO, M. 1945: *The Effect of Price Changes on Supply and Demand of Separable Goods* (auf japanisch). Kokumin Keisai Zasshi, pp. 1-51 (English 1961 with the same title in International Economic Review, pp. 239-271).
- Statistische Monatsberichte des BML, Bundesministeriums für Ernährung, Landwirtschaft und Forsten (ed.), div. issues, Bonn.
- SURRY, Y. / MOSCHINI, G. 1984: *Input Substitutability in the EC Compound Feed Industry*. European Review of Agricultural Economics, pp. 455-464.
- TERRELL, Dek 1995: *Flexibility and Regularity Properties of the Asymptotically Ideal Production Model*. Econometric Reviews, pp. 1-17.
- TERRELL, Dek 1996: *Incorporating Monotonicity and Concavity Conditions in Flexible Functional Forms*. Journal of Applied Econometrics, pp. 179-194.
- WALLACE, Walter 1971: *The Logic of Science in Sociology*. Aldine, Atherton; Chicago, New York.
- WHITE, H. 1980: *Using Least Squares to Approximate Unknown Regression Functions*. International Economic Review, pp. 149-170.
- ZMP-Bilanz Getreide, Ölsaaten, Futtermittel, Zentrale Markt- und Preisberichtsstelle für Erzeugnisse der Land-, Forst- und Ernährungswirtschaft GmbH (ed.), div. issues, Bonn/Bad Godesberg.