MODEL SELECTION AND SIMPLIFICATION USING LATTICES

Jaromír ANTOCH and Jan HANOUSEK[1]

Charles University in Prague

Abstract

This paper shows how to cope with a problem of model selection and simplification using the principle of coherence (Gabriel (1969): *A procedure involving testing a set of models ought not accept a model while rejecting a more general model*). The mathematical lattice theory is used to define a partial ordering over the space of considered models. Several examples of partial ordering in large families of models are given along with a searching algorithm to determine the best model with respect to chosen criteria.

---

[1]Charles University in Prague, CERGE, Politických vězňů 7, POBox 882, CZ−111 21 Praha 1; Tel. (+420 2) 2400 5174; Fax. (+420 2) 2422 7143; jan.hanousek@cerge.cuni.cz (home.cerge.cuni.cz/hanousek)

Motto: "All models are wrong, but some are useful." (G.E.P. Box)

"Models are to be used, but not to be believed." (H. Theil)

1. Introduction

Since the late 1970s econometricians have voiced increasing concern about model specification (see Henry, Leamer and Poirier (1990)). While every textbook on the subject explicitly endorses economic theory as the starting point and the guiding principle of a specification search, the reader is never told that all theories are not only highly abstract, but often quite incompatible. This difficulty is further exacerbated by the very nature of economic research: typically, the design of the experiment is uncontrolled, and hence the data-generating process is exceedingly complicated. These facts should uphold the importance of data analysis in model selection, and should draw attention to the link between specification analysis, sample evidence and/or the evaluation of forecasting models. Indeed, several authors have already voiced their dissatisfaction with attributing excessive significance to economic theory, as opposed to empirical evidence, especially when a time-series approach is utilized (see Diebold (1995) among others). On the other hand, data mining and model hunting are the most heavily criticized methods, perhaps because they represent most precisely what researchers actually do (see Henry and Richard (1982), Kennedy (1998) and Henry, Leamer and Poirier (1990) among others).

Applied econometric research usually proceeds from a widely accepted yet unproven assumption that all data are at hand regardless of the model, and, that the basic criterion for evaluating models is their consistency with these

data[2]. Since the 1980s, however, discussion of the model building problem has been informed by the "encompassing" principle, a requirement that a good model be able to explain results obtained by rival models (see Henry and Richard (1982) and Mizon (1984) for early references)[3].

The goal of this paper is to present an algorithm used in a general searching procedure that will minimize the number of comparisons/tests needed in model specification. As the basic tool for our approach we use the so-called principle of coherence (Gabriel (1969)): "A procedure involving testing a set of models ought not accept a model while rejecting a more general model." The principle of coherence implies the principle of parsimony: because of differences in degrees of freedom, one should use a restricted model that is able to perform as well as the model of which it is a special case. To apply the principle of coherence in a searching procedure, we need to define an ordering over the space of suitable models: that is, we need to impose an algebraic structure (lattice) over the space of the considered models.

Even though proper definition of the coherence principle and implied structures require a lot of algebra, the procedure we propose is easy to use. In fact, we need only define a partial ordering in large families of models, which are usually quite obvious, as we demonstrate in several examples of lattices

---

[2]For the sake of simplicity, in this paper we do not assume that data may be available sequentially, implying that model selection becomes a sequential process.

[3]Even though encompassing was initially introduced in the context of nested models, this standard could prove effective for a comparison of non-nested models as well, see, e.g., West (2000).

corresponding to applied econometric models. Since "encompassing" is asymmetric and transitive, and therefore defines an ordering, our approach lies between a "classical" searching procedure and "encompassing". In addition, using the search results, models can be clustered by similarity into groups that are identical with respect to empirical evidence and a given criteria.

The paper is structured as follows. The principle of coherence and basic rules for model selection are formulated in the Section 2. The third section uses algebraic terminology to formalize the model selection mechanism. An algorithm for searching within introduced general structure is described in Section 4, along with several modifications for the finite structures. Several examples of partial ordering are given in the Section 5. The conclusions of the paper and directions for further research are presented in the final section. Small appendix contains basic notation and results from lattice theory.

2. Principle of coherence and basic rules for model selection

Let us consider model selection from an arbitrary family $\mathcal{M}$ of models. Let us assume, moreover, that some goodness-of-fit test for *testing models* in $\mathcal{M}$ is available and is applied for given data set $\mathcal{D}$ at some particular significance level α (usually constant for all models from $\mathcal{M}$). Thus for each model $m \in \mathcal{M}$, we can apply the test and determine whether it is rejected or accepted[4] based on the data set $\mathcal{D}$. Our main goal is to identify the simplest models in $\mathcal{M}$ that are accepted by our test procedure. The basic tool for this purpose is the so-called *principle of coherence*.

[4]For convenience we say a model is "accepted" instead of the more correct "non−rejected".

Principle of coherence $\big($Gabriel (1969)$\big)$. *A procedure involving testing a set of models ought not accept a model while rejecting a more general model.*

The principle of coherence is equivalent to the idea that if we deem one model compatible with given data, it is absurd to consider that there exists a more general model which is incompatible with the same data. Of course, a given test may or may not obey this principle, so that *coherence* can be regarded both as a property and as a principle. In fact, many commonly used tests exhibit noncoherence, a typical example being the F-test in multiple linear regression.

The model selection procedure is based on the following two rules which are a direct consequence of the principle of coherence.

Rule I. If a model is accepted, then all models that include it are considered accepted without further testing.

Rule II. If a model is rejected, then all of its submodels are considered rejected without further testing.

The use of Rules I and II ensures that the principle of coherence will be respected (provided the coherent test is applied). Their use has one great practical advantage, namely that the number of models to be fitted is drastically reduced.

Clearly, it is preferable to use a coherent test if possible. However, sometimes a coherent test may not be readily available or may have low power or have often undesirable properties. Then it is reasonable to use a noncoherent test. When the violation of the principle of coherence is rare, we call such a test *mildly noncoherent.* For the purposes of our paper we do not need to

quantify this rather vague notion here. Nevertheless, when a "mildly nonco-herent" test is used, there may exist an intermediate region consisting of both accepted and rejected models. As long as the noncoherence is only mild, this region should be small. However, in such a case the procedure should be per-formed several times using different starting points to examine the stability of the solution.

3. MODEL AND MAIN GOALS OF MODEL SELECTION

Let $\left(\mathcal{M}, \preceq \right)$ be the lattice[5] of models, where ordering $\preceq$ is defined in the usual way, i.e., for any two models $m_1, m_2 \in \mathcal{M}$,

$$m_1 \preceq m_2 \quad \text{if} \quad m_1 \text{ is a submodel of } m_2. \tag{1}$$

Analogously we write $m_1 \prec m_2$ iff (if and only if) $m_1 \preceq m_2$ and $m_1 \neq m_2$.

In the sequel we shall work only with the lattices of models $\left(\mathcal{M}, \preceq \right)$ with partial ordering that satisfies condition (1). We shall concentrate on the situ-ations typically considered in applied econometrics, including regression mod-els, loglinear models for contingency tables, and autoregressive time series models.

Let $\left(\mathcal{M}, \preceq \right)$ be the lattice of models. Let us suppose that some *decision rule d* is available for use; for a given data set $\mathcal{D}$ and for every model $m \in \mathcal{M}$, d determines whether to reject or accept m based on the data set $\mathcal{D}$. We shall denote that $d(m, \mathcal{D}) = r$ and $d(m, \mathcal{D}) = a$.

---

[5]The pertinent concepts from the theory of ordered sets and lattices are presented in the Appendix. The main references are Birkhoff (1967) and Grätzer (1978).

Definition 1. Let $(\mathcal{M}, \preceq)$ be the lattice of models. We call the decision rule *d-coherent* if for any data set $\mathcal{D}$ and any two models $m_1, m_2 \in \mathcal{M}$ cannot occur

$$m_1 \prec m_2 \;\Rightarrow\; d(m_1, \mathcal{D}) = a \quad \& \quad d(m_2, \mathcal{D}) = r. \tag{2}$$

(This means we do not accept a simple model while rejecting a more complex model containing it.)

Let $(\mathcal{M}, \preceq)$ be the lattice of models and $m_1, m_2 \in \mathcal{M}$ any two models. Let d be the coherent rule and $\mathcal{D}$ a given data set. Using definition 1 we can modify Rules I and II for the purpose of model selection/simplification.

Rule I'. If $m_1 \prec m_2$ and $d(m_1, \mathcal{D}) = a$, then we also accept the model m_2 without testing it.

Rule II'. If $m_1 \prec m_2$ and $d(m_2, \mathcal{D}) = r$, then we also reject the model m_1 without testing it.

Evidently, when using Rules I' and II', we have the greatest gain if we accept simple models and reject complex models.

Definition 2. (a) We call a model *weakly accepted* and denote it w-accepted if it includes an accepted model. $\big($That is, we deduce this model as accepted according to Rule I'$\big)$

(b) We call a model *weakly rejected* and denote it w-rejected if it is included in some rejected model. $\big($That is, we deduce this model as rejected according to Rule II'$\big)$

Remark 1. The main (computational) advantage of Rules I' and II' and Definition 2 is that if we know that model $m \in \mathcal{M}$ is either weakly accepted or weakly rejected, we do not compute the rule $d(m, \mathcal{D})$, i.e., we do not use the data to determine/test whether to accept or reject the model.

Main goal. The main goal of our selection procedure is to find two sets of models, denoted $\mathcal{A}$ and $\mathcal{R}$ in the sequel, $\mathcal{A}, \mathcal{R} \subset \mathcal{M}$, such that models in $\mathcal{A}$ are accepted, models in $\mathcal{R}$ are rejected, and any other model in $\mathcal{M}$ is either weakly accepted or weakly rejected. In other words, our aim is to classify all models in $\mathcal{M}$ as either accepted / weakly accepted, or rejected / weakly rejected, respectively. Note that this classification can be done using only classes $\mathcal{A}$ and $\mathcal{R}$.

Figure 1 illustrates how the sets $\mathcal{A}$ and $\mathcal{R}$ divide $\mathcal{M}$ into three disjoint subsets, namely the weakly accepted, weakly rejected, and undetermined models. Four models (filled circles) were fitted. The minimal and maximal models in the undetermined set are shown as hollow circles.

Include Figure 1.

4. COMPUTATIONAL ASPECTS

Now we turn to the computational aspects of the procedure. At first we introduce several notions needed in the sequel. As mentioned above, more details can be found in the Appendix.

Definition 3. Let $\left(\mathcal{M}, \preceq\right)$ be the lattice of models. We call models $m_1, m_2 \in \mathcal{M}$ *incomparable* if neither $m_1 \prec m_2$ nor $m_2 \prec m_1$. Further, we call a subset of

models $\mathcal{S} \subset \mathcal{M}$ *incomparable* if any two models $m_1, m_2 \in \mathcal{S}$ are incomparable (cf also Definition 10).

Definition 4. Let $(\mathcal{M}, \preceq)$ be the lattice of models. For any subset of models $\mathcal{S} \subset \mathcal{M}$, we define $max(\mathcal{S})$ as the set of maximal models in $\mathcal{S}$, and, respectively, $min(\mathcal{S})$ as the set of minimal models in $\mathcal{S}$. This means that

$$max(\mathcal{S}) = \{s \in \mathcal{S} \mid s \prec t \Rightarrow t \notin \mathcal{S}\} \quad \& \quad min(\mathcal{S}) = \{s \in \mathcal{S} \mid t \prec s \Rightarrow t \notin \mathcal{S}\}.$$

Note that $t \notin \mathcal{S}$ denotes that the model t does not belong to the set $\mathcal{S}$.

Evidently, for any subset of models $\mathcal{S} \subset \mathcal{M}$ the sets $max(\mathcal{S})$ and $min(\mathcal{S})$ are incomparable. Moreover, if $\mathcal{S}$ is incomparable, then $\mathcal{S} = max(\mathcal{S}) = min(\mathcal{S})$. The basic constructions used to determine which models to test are the so-called *a-dual* and *r-dual* of a given set of models $\mathcal{S}$.

Definition 5. Let $(\mathcal{M}, \preceq)$ be a lattice of models and $\mathcal{S} \subset \mathcal{M}$. Then the *a-dual* of $\mathcal{S}$, denoted $\mathcal{D}_a(\mathcal{S})$, is defined as the set of simplest models in $\mathcal{M}$ which are not contained in any model from $\mathcal{S}$. If $\mathcal{S}$ is empty, then we define $\mathcal{D}_a(\mathcal{S}) = min(\mathcal{M})$. Formally the a-dual of $\mathcal{S}$ is defined as

$$\mathcal{D}_a(\mathcal{S}) = min\{m \in \mathcal{M} \mid m \npreceq s, \ \forall s \in \mathcal{S}\}. \tag{3}$$

The concept of a-duality is relevant especially when the models in $\mathcal{S}$ have been rejected. More precisely, using the principle of coherence, $\mathcal{D}_a(\mathcal{S})$ consists of the simplest models in $\mathcal{M}$ that could eventually be accepted given that $\mathcal{S}$ contains rejected models. Moreover, note that a set $\mathcal{S}$ and its a-dual $\mathcal{D}_a(\mathcal{S})$

induce a partitioning of $\mathcal{M}$ into two disjoint subsets such that for each model $m \in \mathcal{M}$, either $m \preceq s$ for some $s \in \mathcal{S}$ or $t \preceq m$ for some $t \in \mathcal{D}_a(\mathcal{S})$.

Definition 6. Let $\big(\mathcal{M}, \preceq \big)$ be a lattice of models and $\mathcal{S} \subset \mathcal{M}$. Then the *r-dual* of $\mathcal{S}$, denoted $\mathcal{D}_r(\mathcal{S})$, is defined as the set of the most complex models in $\mathcal{M}$ that do not contain any model from $\mathcal{S}$. If $\mathcal{S}$ is empty, then we define $\mathcal{D}_r(\mathcal{S}) = max\big(\mathcal{M}\big)$. Formally, the r-dual of $\mathcal{S}$ is defined as

$$\mathcal{D}_r(\mathcal{S}) = max\big\{m \in \mathcal{M} \mid s \npreceq m, \; \forall s \in \mathcal{S}\big\}. \tag{4}$$

The concept of r-duality is relevant especially when the models in $\mathcal{S}$ have been accepted. More precisely, using the coherence principle, $\mathcal{D}_r(\mathcal{S})$ consists of the most complex models in $\mathcal{M}$ that could eventually be rejected given that $\mathcal{S}$ contains accepted models. Moreover, note that a set $\mathcal{S}$ and its r-dual $\mathcal{D}_r(\mathcal{S})$ induce a partitioning of $\mathcal{M}$ into two disjoint subsets such that for each model $m \in \mathcal{M}$, either $s \preceq m$ for some $s \in \mathcal{S}$ or $m \preceq t$ for some $t \in \mathcal{D}_r(\mathcal{S})$. If set $\mathcal{S}$ is incomparable, then $\mathcal{D}_r\big(\mathcal{D}_a(\mathcal{S})\big) = \mathcal{D}_a\big(\mathcal{D}_r(\mathcal{S})\big) = \mathcal{S}$. This relationship justifies the term "dual" used for $\mathcal{D}_r(\mathcal{S})$ and $\mathcal{D}_a(\mathcal{S})$.

Taking into account the above definitions, we see that the main computational burden consists of two tasks: fitting the models, and calculating the duals $D_r(\mathcal{A})$ and $D_a(\mathcal{R})$. We do not discuss the first of these tasks here because concept of fit varies from one model to the another and also interfere with chosen criteria. Insted, we concentrate on the second task. It should be note, that for large dimensional problems, the computation of duals can involve considerable calculation. For models specified as graphs, or more generally for any model family with a binary lattice structure, the calculation

of the duals is formally equivalent to the dual representation problem. This procedure was implemented in the analysis of (hierarchic) loglinear models for contingency tables in the program MIM; see Edwards (1991, 1995). Horáková (1989, 1991) contributed considerably to the detailed study of the computational complexity of this problem.

4.1. **General algorithm.**

The basic algorithm follows.

(1) The lattice of models $\left(\mathcal{M}, \preceq\right)$, decision rule d and data set $\mathcal{D}$ are specified.

(2) An initial set of models $\mathcal{S}_o \subset \mathcal{M}$ is chosen and its elements tested. The models are assigned coherently to the classes $\mathcal{A}$ and $\mathcal{R}$, i.e.,

$$\mathcal{A} = \left\{ m \in \mathcal{S}_o \mid d(m, \mathcal{D}) = a \right\} \quad \& \quad \mathcal{R} = \left\{ m \in \mathcal{S}_o \mid d(m, \mathcal{D}) = r \right\}. \quad (5)$$

(3) If $\mathcal{A} = \emptyset$ then go to Step 4. If $\mathcal{R} = \emptyset$ then go to Step 5. Otherwise choose randomly between Step 4 and Step 5. We suggest to choose according to the size of $\mathcal{D}_r(\mathcal{A}) \backslash \mathcal{R}$ and $\mathcal{D}_a(\mathcal{R}) \backslash \mathcal{A}$.

(4) Test the models in $\mathcal{D}_r(\mathcal{A}) \backslash \mathcal{R}$. If these are all rejected, then stop. Otherwise update $\mathcal{A}$ and $\mathcal{R}$. This means that we add accepted models $\mathcal{A}_1$ from $\mathcal{D}_r(\mathcal{A}) \backslash \mathcal{R}$ to $\mathcal{A}$ and rejected models $\mathcal{R}_1$ from $\mathcal{D}_r(\mathcal{A}) \backslash \mathcal{R}$ to $\mathcal{R}$, i.e., we set $\mathcal{A} = \mathcal{A} \cup \mathcal{A}_1$ and $\mathcal{R} = \mathcal{R} \cup \mathcal{R}_1$. Further, it is necessary to eliminate the models redundant with respect to the partial ordering $\preceq$, i.e. to set $\mathcal{A} = min(\mathcal{A})$ and $\mathcal{R} = min(\mathcal{R})$. Then go back to Step 3.

(5) Test the models in $\mathcal{D}_a(\mathcal{R})\backslash\mathcal{A}$. If these are all accepted, then stop. Otherwise update $\mathcal{A}$ and $\mathcal{R}$. This means that we add accepted models $\mathcal{A}_1$ from $\mathcal{D}_a(\mathcal{R})\backslash\mathcal{A}$ to $\mathcal{A}$ and rejected models $\mathcal{R}_1$ from $\mathcal{D}_a(\mathcal{R})\backslash\mathcal{A}$ to $\mathcal{R}$, i.e., we set $\mathcal{A} = \mathcal{A} \cup \mathcal{A}_1$ and $\mathcal{R} = \mathcal{R} \cup \mathcal{R}_1$. Further, it is necessary to eliminate the models redundant with respect to the partial ordering $\preceq$, i.e., to set $\mathcal{A} = min(\mathcal{A})$ and $\mathcal{R} = min(\mathcal{R})$. Then go to 3.

The initial set of models $\mathcal{S}_o$ can be chosen arbitrarily. However, since we require that the assignment of $\mathcal{A}$ and $\mathcal{R}$ be coherent; the simplest way is to use such $\mathcal{S}_o$ which is incomparable. If $\mathcal{S}_o$ is incomparable, then in the first step $\mathcal{A} \cup \mathcal{R} = \mathcal{S}_o$ and $\mathcal{A} \cap \mathcal{R} = \emptyset$. Evidently, if $\mathcal{S}_o$ is close to the solutions, i.e., close to either $\mathcal{A}$ and/or $\mathcal{R}$, the solution will be found more quickly.

When a coherent decision rule d is used, the set of accepted models $\mathcal{A}$ is unique. That is, $\mathcal{A}$ does not depend on the starting subset $\mathcal{S}_o$ and various choices between steps 4 and 5 taken during the procedure. With a noncoherent decision rule d this is not always the case. It is evident that the same is true for the set of rejected models $\mathcal{R}$.

The reason we test either the models from $\mathcal{D}_r(\mathcal{A})\backslash\mathcal{R}$ or $\mathcal{D}_a(\mathcal{R})\backslash\mathcal{A}$ is the following: Let us consider the set of models $\mathcal{T}$ whose status has not been determined yet. That is, the models from $\mathcal{T}$ neither contain a model from $\mathcal{A}$ nor are contained in a model from $\mathcal{R}$. Therefore, we can write

$$\mathcal{T} = \big\{ m \in \mathcal{M} \mid a \not\preceq m, a \in \mathcal{A}, \text{ or } m \not\preceq r, r \in \mathcal{R} \big\}. \tag{6}$$

Then $max(\mathcal{T}) = \mathcal{D}_r(\mathcal{A})\backslash\mathcal{R}$ and $min(\mathcal{T}) = \mathcal{D}_a(\mathcal{R})\backslash\mathcal{A}$. Suppose, moreover, that at some step the models in $\mathcal{D}_r(\mathcal{A})\backslash\mathcal{R}$ are tested and all are rejected.

Then, after updating $\mathcal{R}$, the set $\mathcal{D}_r(\mathcal{A})\backslash\mathcal{R}$ will be empty. This means that the status of all models has been determined and the procedure can be stopped. The same applies if the models in $\mathcal{D}_a(\mathcal{R})\backslash\mathcal{A}$ are tested and all are accepted.

Evidently, it is sufficient to keep (to store) the sets of minimal accepted models $\mathcal{A}$ and maximal rejected models $\mathcal{R}$, since they determine $\mathcal{D}_a(\mathcal{R})$ and $\mathcal{D}_r(\mathcal{A})$. Moreover, $\mathcal{D}_a(\mathcal{R})$ and $\mathcal{D}_r(\mathcal{A})$ can be updated directly after each step of the procedure; it is not necessary to recompute them in each step from scratch.

One modification of the above procedure should be mentioned here. Let us consider as accepted only the models containing a given model $m_o \in \mathcal{M}$ as a submodel, i.e. we restrict our search to the subfamily of the models $\mathcal{M}_o = \big\{m \in \mathcal{M} \mid m_o \preceq m\big\}$. This may be achieved if we initially assign the models from $\mathcal{D}_r(m_o)$ to $\mathcal{R}$ without any testing. Since for any $m \in \mathcal{M}$ it holds that $m_o \preceq m$ iff m is not contained in some model in $\mathcal{D}_r(m_o)$, i.e., $m \not\preceq s, s \in \mathcal{D}_r(m_o)$, we obtain the desired result. In effect we have redefined the minimal model as m_o. We apply the same procedure when we want to perform a restricted search in the subfamilies of the models $\mathcal{M}_1 = \big\{m \in \mathcal{M} \mid m \preceq m_1\big\}$ or $\mathcal{M}_{o,1} = \big\{m \in \mathcal{M} \mid m_o \preceq m \preceq m_1\big\}$.

However, combining the later restriction to a set of the form $\mathcal{M}_{o,1}$ is a bit trickier. Simply assigning $D_r\big(\{m_1\}\big)$ to $\mathcal{A}$ and $D_a\big(\{m_o\}\big)$ to $\mathcal{R}$ may violate the coherence requirement, which would cause the procedure to fail. To enforce coherence we can assign $D_r\big(\{m_1\}\big)$ to $\mathcal{A}$ and $D_a\big(\{m_o\} \cup D_r(\{m_1\})\big)$ to $\mathcal{R}$. Moreover, using Lemma 3 of Edwards and Havránek (1987) we can simplify

$$D_a\big(\{m_o\} \cup D_r(\{m_1\})\big) = \min\big\{m \wedge m_1 \mid m \in D_a(\{m_o\})\big\},$$

which, with a certain license, can be abbreviated to $D_a\big(\{m_o\}\big) \wedge m_1$. Equivalently we can set $D_a\big(\{m_o\}\big)$ to $\mathcal{R}$ and $D_r\big(\{m_1\}\big) \wedge m_o$ to $\mathcal{A}$.

4.2. **Model selection procedure for (general) finite lattices.**

Evidently, for real applications of the described algorithm the most important question is how to find the duals in an efficient way. As one might expect, the least effective way is to follow the definition directly, while the lattice structure can help us find a more efficient solution. The following three theorems show us how.

Theorem 1. Let $\big(\mathcal{M}, \preceq\big)$ be a lattice of models and $\mathcal{S}$ and $\mathcal{T}$ any two subsets of models from $\mathcal{M}$. Then

$$\mathcal{D}_r(\mathcal{S} \cup \mathcal{T}) = max\big\{s \wedge t \mid s \in \mathcal{D}_r(\mathcal{S}) \text{ and } t \in \mathcal{D}_r(\mathcal{T})\big\} \tag{7}$$

and

$$\mathcal{D}_a(\mathcal{S} \cup \mathcal{T}) = min\big\{s \vee t \mid s \in \mathcal{D}_a(\mathcal{S}) \text{ and } t \in \mathcal{D}_a(\mathcal{T})\big\}. \tag{8}$$

(Here $\mathcal{S} \cup \mathcal{T}$ denotes the union of sets $\mathcal{S}$ and $\mathcal{T}$)

Theorem 2. Let $\big(\mathcal{M}, \preceq\big)$ be a lattice of models. Let $\mathcal{P} \subset \mathcal{M}$ contain all $\wedge$-irreducible models ϵ^i in $\mathcal{M}$ and $Q \subset \mathcal{M}$ contain all $\vee$-irreducible models ϵ_i in $\mathcal{M}$. Then for any model $m \in \mathcal{M}$, it holds that

$$\mathcal{D}_r(m) = max\big\{\epsilon^i \in P \mid m \not\preceq \epsilon^i\big\} \quad \& \quad \mathcal{D}_a(m) = min\big\{\epsilon_i \in Q \mid \epsilon_i \not\preceq m\big\}. \tag{9}$$

Theorems 1 and 2 illustrate how to find duals for any set of models $\mathcal{S} = \{m_1, \ldots, m_k\}$. Let us consider, e.g., the construction of the r-dual $\mathcal{D}_r(\mathcal{S})$. At first we must find $\mathcal{D}_r(m_i)$, $i = 1, \ldots, k$, using (9). The desired solution is then of the form

$$\mathcal{D}_r(\mathcal{S}) = max\big\{s_1 \wedge \cdots \wedge s_k \mid s_i \in \mathcal{D}_r(m_i), \ i = 1, \ldots, k\big\}. \qquad (10)$$

The corresponding a-dual $\mathcal{D}_a(\mathcal{S})$ can be constructed quite analogously, i.e.,

$$\mathcal{D}_a(\mathcal{S}) = min\big\{s_1 \vee \cdots \vee s_k \mid s_i \in \mathcal{D}_a(m_i), \ i = 1, \ldots, k\big\}. \qquad (11)$$

Note that if $card\big(\mathcal{M}\big) = \infty$, then $card(\mathcal{P})$ and $card(Q)$ can also be infinite and we may observe some theoretical and computational problems.

Theorem 3. Let $\big(\mathcal{M}, \preceq\big)$ be a finite lattice of models. For a model $m \in \mathcal{M}$, let there exist other models $x_1, \ldots, x_q \in \mathcal{M}$, and, respectively, $y_1, \ldots, y_h \in \mathcal{M}$, such that

$$m = x_1 \wedge \cdots \wedge x_q, \quad \text{and, respectively,} \quad m = y_1 \vee \cdots \vee y_h. \qquad (12)$$

Then

$$\mathcal{D}_a(m) = min\Big\{\bigcup_i \mathcal{D}_a(x_i)\Big\} \quad \& \quad \mathcal{D}_r(m) = max\Big\{\bigcup_i \mathcal{D}_r(y_i)\Big\}. \qquad (13)$$

$\big($Here $\cup_i$ denotes again the union of subsets $\mathcal{D}_a(x_i)$, and respectively, $\mathcal{D}_r(y_i)\big)$

4.3. Model selection procedure for finite distributive lattices.

The situation can be even more simplified if the set of models forms a finite distributive lattice as shown in the next two theorems.

Theorem 4. Let $\left(\mathcal{X}, \preceq\right)$ be a finite distributive lattice. Then for each $\wedge$-irreducible element ϵ^i there exists a $\vee$-irreducible element ϵ_i such that

$$\epsilon_i = min\big\{x \in \mathcal{X} \mid x \not\preceq \epsilon^i\big\}. \tag{14}$$

Similarly, for each $\vee$-irreducible element ϵ_i there exists a $\wedge$-irreducible element ϵ^i such that

$$\epsilon^i = max\big\{x \in \mathcal{X} \mid \epsilon_i \not\preceq x\big\}. \tag{15}$$

Proof: See Birkhoff (1967), Edwards and Havránek (1987) or Grätzer (1978).

The correspondence between the joint and meet irreducible elements is clearly 1-1 mapping. The construction of the algorithm then follows from the following theorem.

Theorem 5. Let $\left(\mathcal{M}, \preceq\right)$ be a finite distributive lattice of models. Let $\{\epsilon^i, \ i = 1, \ldots, n\}$ denote its irredundant $\wedge$-irreducible elements and $\{\epsilon_i, \ i = 1, \ldots, n\}$ denote its irredundant $\vee$-irreducible elements. Then for any model $m \in \mathcal{M}$ such that

$$m = \bigvee_j \epsilon_j, \ j \in W_1, \ W_1 \subset \{1, \ldots, n\}, \tag{16}$$

we have

$$\mathcal{D}_r(m) = \big\{\epsilon^j, \ j \in W_1\big\}. \tag{17}$$

Similarly, for any model $m \in \mathcal{M}$ such that

$$m = \bigwedge_j \epsilon^j, \; j \in W_2, \; W_2 \subset \{1, \ldots, n\}, \tag{18}$$

we have

$$\mathcal{D}_a(m) = \big\{\epsilon_j, \; j \in W_2\big\}. \tag{19}$$

Let $\big(\mathcal{M}, \preceq \big)$ be a finite distributive lattice of models and $\mathcal{S} = \big\{m_1, \ldots, m_k\big\}$ $\subset \mathcal{M}$. Then the way to find $\mathcal{D}_r(\mathcal{S})$ is to first represent each m_i in its irredundant joint representation. Then $\mathcal{D}_r(\mathcal{S})$ can be calculated combining Theorems 3 and 5 with (10) and (11). Similarly, to find $\mathcal{D}_a(\mathcal{S})$ we first represent each m_i in its irredundant meet representation, and then use the same arguments as above.

5. EXAMPLES

This section demonstrates the lattice structure for some of the most important models frequently used in applied econometrics and data analysis, namely, linear regression models, polynomial regression models, analysis of variance models, hierarchic log-linear models in contingency tables, and ARMA models. The notation used appears in the Appendix. The arrows in the figures show the direction of inequalities; relations following from transitivity are not marked.

5.1. **Linear regression.**

Let us consider the linear regression model of the form

$$Y = \beta_0 + \beta_1 A + \beta_2 B + \cdots + \beta_q Q + \varepsilon, \tag{20}$$

with regressors $A, B, \ldots, Q$. In our search for the set of best regressors we consider the models $\{A\}, \{B\}, \{AB\}, \{ABC\}$ etc., i.e., different subsets of regressors. The situation for three independent variables A, B and C is represented in Figure 2. An analogous situation for four independent variables A, B, C and D is represented in Figure 3. The models in the figures are denoted by A, B, AB, ABC etc.

The ordering $\preceq$ corresponds to the standard inclusion $\subseteq$. Operations *join* and *meet* are defined as union and intersection. This is an example of a (very simple) finite distributive lattice and the results of subsection 4.3 apply here. The $\wedge$-irreducible elements are the subsets containing exactly $q - 1$ variables, i.e., $\epsilon^i = \{A_1, \ldots, A_{i-1}, A_{i+1}, \ldots, A_q\}$, $i = 1, \ldots, q$ (regressor A_i is missing in the model). The $\vee$–irreducible elements are the subsets containing just one variable, i.e., $\epsilon_i = A_i$, $i = 1, \ldots, q$.

Include Figure 2.

Include Figure 3.

5.2. **Polynomial regression.**

Let us consider a polynomial regression model of the form

$$Y = \beta_0 + \beta_1 A + \beta_2 A^2 + \beta_3 B + \beta_4 AB + \cdots + \varepsilon, \tag{21}$$

where the regressors are $A, B, \ldots, Q, A^2, B^2, \ldots, Q^2, AB, A^2B \ldots$. We assume a hierarchic situation, i.e.,

– if A^2 is in the model then A is also included;

– if AB is in the model then both A and B are also included;

– if A^2B is in the model then both A^2 and AB are also included etc.

The models are described by the sets of maximal elements (regressors) contained in the representation (21); the example of three independent variables A, B and C is represented in Figure 4.

The ordering $\preceq$ corresponds to inclusion with respect to the set of all regressors contained in the model, not only to the maximal elements. The operations *join* and *meet* are defined as union and intersection. The $\vee$-irreducible elements consist of the models with just one element that can be (in our shorthand notation) written as $A, A^2, AC, BC, \ldots$. The polynomial regression model is an example of a infinite distributive lattice. It holds that each of its elements can be unambiguously expressed using the $\vee$-irreducible elements.

Include Figure 4.

5.3. Analysis of variance.

Linear regression and variance component models are of central importance in statistical and econometrical theory and practice. Here we shall concentrate on one special case of ANOVA which is used very often. Let us have n parameters $\mu_1, \ldots, \mu_n$ (for example n mean values of n samples). The hypotheses of interest are of the forms $\mu_i = \theta_k$ for $i \in a_k, k = 1, \ldots, q$, where $a_1, \ldots, a_q$ define the partition of the set $a = \{1, \ldots, n\}$. The lattice is composed of the elements x together with the ordering $\preceq$ defined as follows: $x \preceq y$ if y is a refinement of x. The operations *join* and *meet* are introduced as

    – The join of $x = \{a_1, \ldots, a_q\}$ and $y = \{b_1, \ldots, b_s\}$ is given by $x \vee y = \{a_1 \cap b_1, a_1 \cap b_2, \ldots, a_q \cap b_s\}$; empty intersections are omitted.

– The meet-operation requires a two step procedure. In the first step we form $z = \{a_1, \ldots, a_q, b_1, \ldots, b_s\}$, and in the second step we form the nonintersecting sets from z.

The models of the form $\{a_1, a_2\}$ are $\vee$-irreducible and all other models except $\{a\}$ can be composed of them in an unique way. The models of the form $\{a_1, \ldots, a_{n-1}\}$ are $\wedge$-irreducible (one of a_i contains two elements, all other a_i's contain just one element).

The case for $n = 3$ and $n = 4$ is represented in Figures 5 and 6. It should be noted that the case of $n = 5$ is more complicated, as shown in Figure 7. Indeed, it can be easily shown that $(123, 4, 5) = (12, 3, 4, 5) \wedge (1, 23, 4, 5) = (12, 3, 4, 5) \wedge (13, 2, 4, 5)$. This example shows that we do not have unique nonredundant representations of the elements using the $\wedge$-irreducible elements here. Thus, the lattice for $n = 5$ is finite but not distributive.

Include Figure 5.

Include Figure 6.

Include Figure 7.

5.4. **Hierarchic log-linear models in contingency tables.**

Let us consider, for example, a contingency table with four categorical variables A, B, C and D and a model of conditional independence of AB and D given C. Let p_{ijkl} be the probability that A will take the value i, B the value j etc., and assume that $p_{ijkl} > 0$. Let us consider the model of conditional independence of AB and D given C. Then we can use the log-linear representation of probabilities $p_{ijkl} > 0$ to write

$$log\, p_{ijkl} = \theta + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_l^D + \lambda_{ij}^{AB} + \lambda_{ij}^{AC} + \lambda_{jk}^{BC} + \lambda_{kl}^{CD} + \lambda_{ijk}^{ABC}. \quad (22)$$

The models are described by the maximal indexes of the generating classes, e.g., $\{ABC, CD\}$. Then the lattice operations can be formally defined as operations on the generating classes, i.e., $\{ABC, CD\} \vee \{ACD\} = \{ABC, ACD\}$ and $\{ABC, CD\} \wedge \{AB, C, D\} = \{AB, C, D\}$ etc.

The situation with three variables is represented in Figure 8. One can easily check that $\{AB, C\} \vee \{AC, B\} = \{AB, AC\}$ and $\{AB, C\} \wedge \{AC, B\} = \{A, B, C\}$ etc. Thus, the lattice also contains the models corresponding to the collapsed tables, i.e. the models containing only some of the main effects. It is a finite distributive lattice with the minimal element $\{0\}$ and the maximal element $\{ABC\}$.

The distributive sublattice is very important from the practical point of view. This lattice contains only the models which have all the main effects, i.e., $\lambda^A, \lambda^B, \lambda^C$, denoted in Figure 8 by $]*$. In this case we can use for the representation of all models the $\wedge$-irreducible elements $\{AC, BC\}, \{AB, BC\}$ and $\{AB, AC\}$ or the $\vee$-irreducible elements $\{A, BC\}, \{AC, B\}$ and $\{AB, C\}$. In the general case of this special sublattice with general dimension n, the $\wedge$-irreducible elements are of the form $\{a_1, \ldots, a_k\}$, where each a_i contains exactly $n-1$ letters, $k \geq 2$. If $\epsilon^i = (a_1, \ldots, a_k)$ is such a $\wedge$-irreducible element, then the corresponding $\vee$-irreducible element $\epsilon_i = (b, A_1 \ldots, A_s)$, where b is a set of those letters missing in $a_1, \ldots, a_k$, and $A_1, \ldots, A_s$ are the other letters. For example, let $n = 5$ and $\epsilon^i = (ABCD, ABCE, ABDE)$, then $\epsilon_i = (CDE, A, B)$.

Include Figure 8.

5.5. ARMA models.

Let $\{X_t\}_t$ be the autoregressive moving average process $ARMA(p,q)$ of the form

$$X_t + a_1 X_{t-1} + \cdots + a_p X_{t-p} = e_t + b_1 e_{t-1} + \cdots + b_q e_{t-q}, \qquad (23)$$

where $\{e_t\}_t$ is Gaussian white noise. The problem here is to search for the models with the most parsimonious number of parameters. The lattice corresponding to the ARMA models is quite simple provided we use the principle of hierarchy. Namely, the elements of the desired lattice are the pairs $\{i,j\}$, $i = 0, \ldots, p$, $j = 0, \ldots, q$, consisting of the largest indexes included in the model. The case for $p = 3$ and $q = 4$ is represented in Figure 9.

The ordering is naturally defined as $\{i,j\} \preceq \{k,l\}$ iff $i \leq k$ & $j \leq l$. Operations *join* and *meet* are defined as follows, namely $\{i,j\} \vee \{k,l\} = \big\{ \max\{i,k\}, \max\{j,l\} \big\}$ and $\{i,j\} \wedge \{k,l\} = \big\{ \min\{i,k\}, \min\{j,l\} \big\}$. The $\vee$-irreducible elements ϵ_i are of the form $\{i,0\}$ and $\{0,j\}$. The $\wedge$-irreducible elements ϵ^i are of the form $\{i,q\}$ and $\{p,j\}$. Moreover, each model can be unambiguously expressed using the $\vee$- and $\wedge$-irreducible elements and because $\{i,j\} = \{i,0\} \vee \{0,j\} = \{i,q\} \wedge \{p,j\}$.

The correspondence between $\vee$- and $\wedge$-irreducible elements is straightforward. It holds, e.g., that $\epsilon_i = \min \big\{ x \mid x \not\preceq \epsilon^i \big\}$ $\big($cf $(14)\big)$, where $\epsilon^i = \{i,q\}$ and $\{p,j\}$. Thus $\min \big\{ x \mid x \not\preceq \{2,4\} \big\} = \{3,0\}$ etc. More generally, $\{i+1,0\}$ corresponds to the model $\{i,q\}$ while, quite analogously, $\{0,j+1\}$ corresponds to the model $\{p,j\}$. Similar relations hold also for ϵ^i.

Include Figure 9.

6. Conclusions

The goal of this paper is to present an algorithm used in a general searching procedure to minimize the number of comparisons/tests needed in model specification. The procedure introduced typically selects multiple models consistent with the data. It can be argued that this approach faithfully reflects inherent ambiguity and, therefore, is advantageous.[6] Similarly, Tukey (1985) argues that "Science is the holding of multiple working hypotheses..." when urging that "...methods that give multiple answers would be adopted." Data may be ambiguous and any selection procedure should reflect this in its conclusions. This ambiguity may – or may not – be subsequently resolved through the use of sensitive econometric/statistical analysis appropriate to the subject matter. The main advantage of our approach is that the models can be clustered by similarity into groups that are observationally equivalent with respect to selected criteria (i.e., a given test). In addition, one can use overall goodness of fit as a decision rule in searching procedures rather than nested test procedures to search within non-nested structures. (see, for example, Antoch and Hanousek (1999) in which predictive accuracy is used in discrete choice models.)

---

[6]Although in some contexts a formal model is not of primary concern, econometricians feel obliged to choose one in order to be able to apply the standard arsenal of statistical methods.

7. Appendix: Basic notation, definitions and theorems from the lattice theory

This appendix summarizes the basic elements of lattice theory used in our paper. For more details the reader can refer to any textbook about algebra, Birkhoff's or Grätzer's monographs are also excellent sources.

Definition 7. Set $\mathcal{X}$ with partial ordering $\preceq$, i.e. the pair $(\mathcal{X}, \preceq)$, is called a *partially ordered set (poset)* if the partial ordering is reflexive, antisymmetric and transitive, i.e., if $\forall x, y, z \in \mathcal{X}$ hold $x \preceq x; x \preceq y \ \& \ y \preceq x \Rightarrow x = y$ and $x \preceq y \ \& \ y \preceq z \Rightarrow x \preceq z$.

Definition 8. The *Lattice* is the partially ordered set $(\mathcal{X}, \preceq)$ such that:

(1) For any two elements $x, y \in \mathcal{X}$, there exists a unique greatest lower bound denoted by $x \wedge y$.

(2) For any two elements $x, y \in \mathcal{X}$, there exists a unique lowest upper bound denoted by $x \vee y$.

Remark 2. Let $(\mathcal{X}, \preceq)$ be a lattice. We can consider $\wedge$ and $\vee$ as binary operations and, as is usual, call them *meet* and *join*. Operations *join* and *meet* have the following properties, i.e., for any $x, y, z \in \mathcal{X}$:

$$x \wedge x = x \qquad\qquad x \vee x = x; \tag{24}$$

$$x \wedge y = y \wedge x \qquad\qquad x \vee y = y \vee x \qquad \text{comutativity;} \tag{25}$$

$$(x \wedge y) \wedge z = x \wedge (y \wedge z) \quad (x \vee y) \vee z = x \vee (y \vee z) \quad \text{associativity;} \tag{26}$$

$$x \wedge (x \vee y) = x \qquad\qquad x \vee (x \wedge y) = x \qquad \text{adsorbility.} \tag{27}$$

Definition 9. Lattice $(\mathcal{X}, \preceq)$ is called a *finite lattice* if $card\ \mathcal{X} \prec \infty$.

Definition 10. Let $(\mathcal{X}, \preceq)$ be a lattice. If for two elements $x, y \in \mathcal{X}$ neither $x \not\preceq y$ nor $y \not\preceq x$, we call elements x and y *incomparable*. We call subset $Y \subset \mathcal{X}$ *incompatible* if any two elements $x, y \in Y$ are incompatible. {Here $x \not\preceq y$ means that element x does not lie before element y and they are not identical.}

Definition 11. Lattice $(\mathcal{X}, \preceq)$ is called *distributive lattice* if for any $x, y, z \in \mathcal{X}$ it holds that :

$$x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z) \quad \text{and} \quad x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z). \quad (28)$$

These two identities are equivalent; see Birkhoff (1967), page 11.

Definition 12. Distributive lattice $(\mathcal{X}, \preceq)$ is called a *boolean distributive lattice* if it contains *minimal element* $\emptyset$, *maximal element* I and, moreover, if for every element $x \in \mathcal{X}$ there exists just one element, $-x$, called a *complement*, such that

$$x \wedge -x = \emptyset \quad \& \quad x \vee -x = I. \quad (29)$$

Definition 13. Let $(\mathcal{X}, \preceq)$ be a lattice. Element $x \in \mathcal{X}$ is called *meet-irreducible* ($\wedge$-irreducible) if

$$x = y \wedge z \quad \Rightarrow \quad x = y \quad \text{or} \quad x = z. \quad (30)$$

Definition 14. Let $(\mathcal{X}, \preceq)$ be a lattice. Element $x \in \mathcal{X}$ is called *join-irreducible* ($\vee$-irreducible) if

$$x = y \vee z \quad \Rightarrow \quad x = y \quad \text{or} \quad x = z. \tag{31}$$

We shall denote by $\epsilon^1, \ldots, \epsilon^n$ the $\wedge$-irreducible elements and by $\epsilon_1, \ldots, \epsilon_n$ the $\vee$-irreducible elements of the lattice $(\mathcal{X}, \preceq)$. It is clear that in a finite lattice $(\mathcal{X}, \preceq)$ there exists a finite number of $\wedge$- and $\vee$-irreducible elements, but their number generally need not be equal. In an infinite lattice their number can be infinite. Moreover, let $(\mathcal{X}, \preceq)$ be a finite lattice. Then each element $x \in \mathcal{X}$ can be represented either as a join of $\vee$-irreducible elements or as a meet of $\wedge$-irreducible elements.

The next theorem shows that even more holds for finite distributive lattices.

Theorem 6. Let $(\mathcal{X}, \preceq)$ be a finite distributive lattice. Then every element $x \in \mathcal{X}$ has a unique representation as an irredundant meet of $\wedge$-irreducible elements. This means that there exist $\wedge$-irreducible elements $\epsilon^1, \ldots, \epsilon^p \in \mathcal{X}$ such that

$$x = \epsilon^1 \wedge \cdots \wedge \epsilon^p \quad \text{and} \quad \forall i \quad x \neq [\epsilon^1 \wedge \cdots \wedge \epsilon^{i-1} \wedge \epsilon^{i+1} \wedge \cdots \wedge \epsilon^p]. \tag{32}$$

Similarly, every element $x \in \mathcal{X}$ has a unique representation as an irredundant join of $\vee$-irreducible elements, i.e., there exist $\vee$-irreducible elements $\epsilon_1, \ldots, \epsilon_q \in \mathcal{X}$ such that

$$x = \epsilon_1 \vee \cdots \vee \epsilon_q \quad \text{and} \quad \forall i \quad x \neq \epsilon_1 \vee \cdots \vee \epsilon_{i-1} \vee \epsilon_{i+1} \vee \cdots \vee \epsilon_q]. \tag{33}$$

Proof: Birkhoff (1967), Chapter 9.

It can be shown that the meet of the form (32) is irredundant iff the set $\{\epsilon^i, i = 1, \ldots, p\}$ is incomparable. Analogously, the join of the form (33) is irredundant iff the set $\{\epsilon_i, \; i = 1, \ldots, q\}$ is incomparable.

References

[1] Anderson, S. A., 1990, The lattice structure of orthogonal linear models and orthogonal variance component models, Scandinavian Journal of Statistics, Theory and Applications 17, 287-319.

[2] Antoch, J., 1986, Algorithmic development in variable selection procedures, in: A. Rizzi, ed., Proceedings of the COMPSTAT'86 (Physica Verlag, Vienna) 83-90.

[3] Antoch, J., 1987. Variable selection based on trimmed least squares estimator, in: Y. Dodge ed., Statistical data analysis based on ℓ_1-norm and related methods (North Holland, Amsterdam), 231-246.

[4] Antoch, J. and J. Hanousek, 1999, A specification test for discrete choice models, submitted.

[5] Bearse, P. M., B. Hamparsum and A. M.Schlottmann, 1997, Empirical econometric modelling of food consumption using a new informational complexity approach, Journal of Applied Econometrics 12, 563-586.

[6] Birkhoff, G., 1967, Lattice theory (American Mathematical Society Colloquium Publications 25)..

[7] Cox, D. R. and E. Spjotvoll, 1982, On partitioning means into groups, Scandinavian Journal of Statistics 9, 147-152.

[8] Damiani, M. and L. Panattoni, 1992, Optimal simulation with econometric models, Journal of Economic Dynamics and Control 16, 93-108.

[9] David, F., H. E. E. Leamer and D. J. Poirier, 1990, The ET dialogue: A conversation on econometric methodology, Econometric Theory 6, 171-261.

[10] Dorfman, J. H. and A. M. Havenner, 1992, A bayesian approach to state space multivariate time series modeling, Journal of Econometrics 52, 315-346.

[11] Edwards, D., 1990, Hierarchic mixed interaction models (with discussion), Journal of the Royal Statistical Society B 52, 3-20.

[12] Edwards, D., 1991, MIM – Program for PC designed to fit hierarchic interaction models.

[13] Edwards, D., 1993, Some computational aspects of graphical model selection, in: J. Antoch, ed., Computational Aspects of Model Choice (Physica Verlag, Heidelberg) 187−210.

[14] Edwards, D., 1995, Introduction to graphical modelling (Springer Verlag, Berlin).

[15] Edwards, D. and T. Havránek T, 1985, A fast procedure for model search in multidimensional contingency tables, Biometrika 72, 339-351.

[16] Edwards, D. and T. Havránek, 1985. A fast model selection procedure, Journal of the American Statistical Association 82, 205-213.

[17] Gabriel, K. V., 1969 Simultaneous test procedures − some theory of multiple comparisons, Annals of Mathematical Statistics 40, 224-250.

[18] Grätzer, G., 1978, General lattice theory (Birkhäuser, Basel).

[19] Havránek T., 1984, A procedure for model search in multidimensional contingency tables, Biometrics 40, 95-100.

[20] Henry, D. F., E. E. Leamer and Poirier D. J., 1990, The ET Dialogue: A conversation on econometric methodology, Econometric Theory 6, 171-261.

[21] Henry, D. F. and J. F.Richard, 1982, On the formulation of empirical models in dynamic econometrics, Journalof Econometrics 20, 3-33.

[22] Holmes, J. M. and P. A. Hutton, 1989, "Optimal" model selection when the true relationship is weak and occurs with a delay, Economics Letters 30, 333-339.

[23] Horáková, M., 1989, Structures of data dependencies, PhD Thesis, Czech Academy of Sciences.

[24] Horáková, M., 1991, Implementation of fast model selection for graphical mixed interaction models, Computational Statistics Quarterly 6, 360-378.

[25] Kennedy, P., 1998, A guide to econometrics, 4th ed. (MIT Press, Cambridge).

[26] Mizon, G. E., 1984, The encompassing approach in econometrics, in: D. F. Henry and K. F. Wallis, eds, Econometrics and Quantitative Economics (Basil Blackwell) 135-172.

[27] Pagan, A., 1987, Three econometric methodologies: A critical appraisal, Journals of Economic Surveys 1, 3-24.

[28] Taylor, L. W., 1997, An R^2 criterion based on optimal predictors, Econometric Reviews 16, 109-118.

[29] Theil, H., 1971, Principles of econometrics (J. Wiley, New York).

[30] Tukey, J. W., 1985, Comment on "More ... directions" by G. Hahn, The American Statistician 39, 12-14.

[31] Weiss, A. A., 1996, Estimating time series models using the relevant cost function, Journal of Applied Econometrics 11, 539-560.

[32] Wermuth, N., 1976, Model search in multiplicative models, Biometrics 32, 253-263.

[33] Whittaker, J., 1985, Additive elements of ARMA models, Journal of Time Series Analysis 2, 135-140.

[34] Whittaker, J., 1990, Graphical models in applied multivariate statistics, (J. Wiley, Chichester).
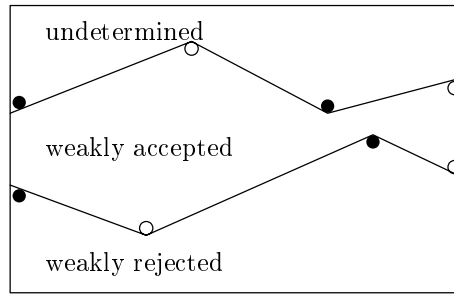
Figure 1. A step in the procedure.
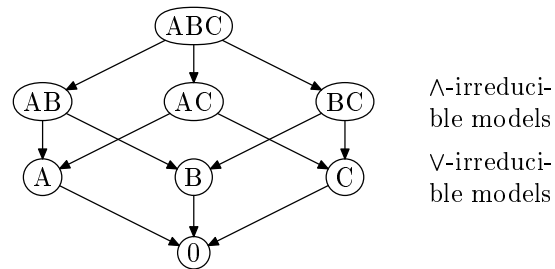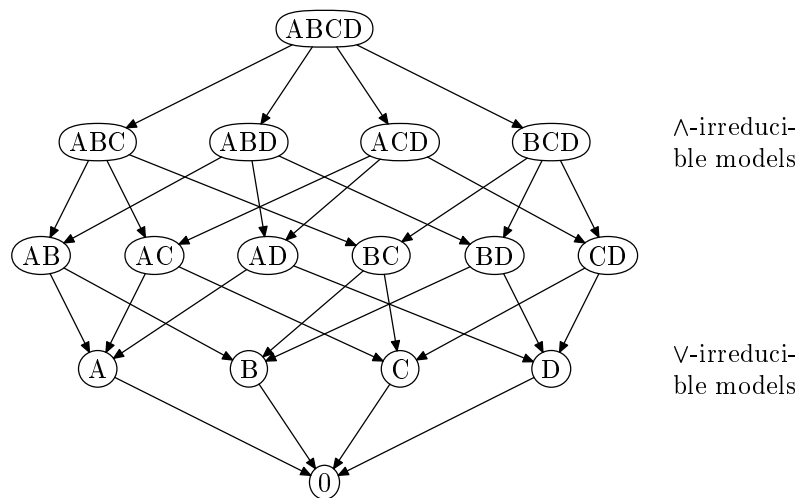
Figure 2. Linear regression model (20), $q = 3$.

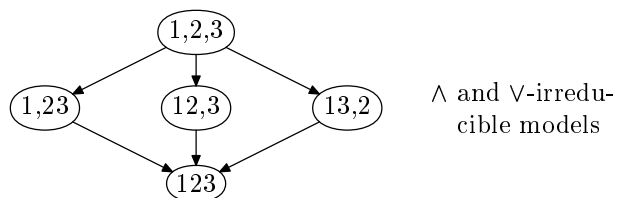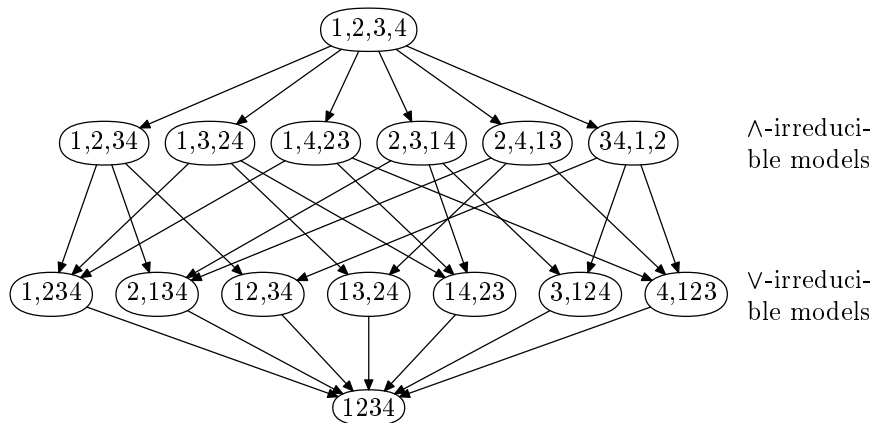Figure 3. Linear regression model (20), $q = 4$.
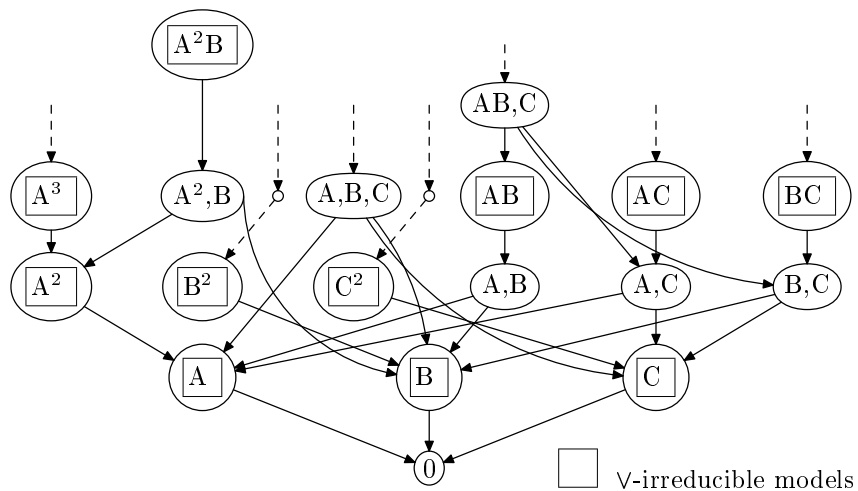
Figure 4. Polynomial regression.

Figure 5. Irreducible lattices in analysis of variance, $n = 3$.

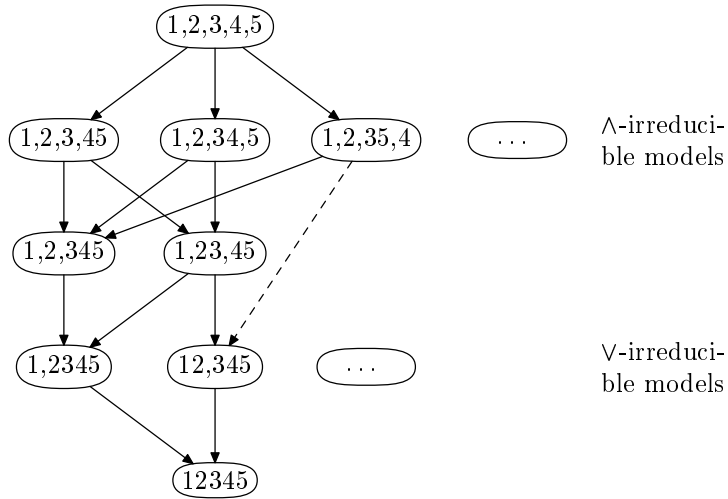Figure 6. Irreducible lattice in analysis of variance, $n = 4$.

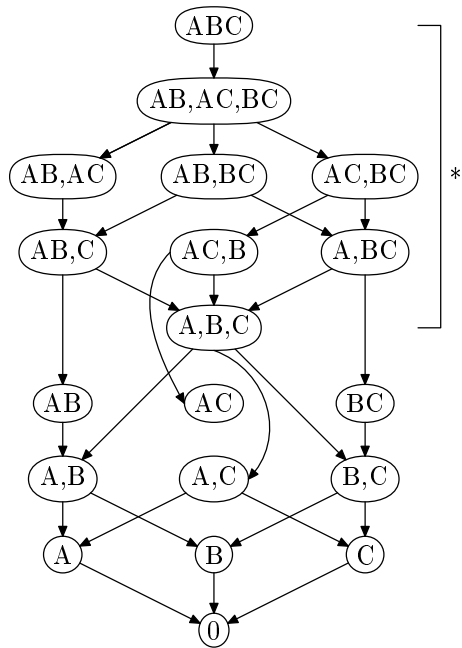Figure 7. Lattice in analysis of variance, $n = 5$.

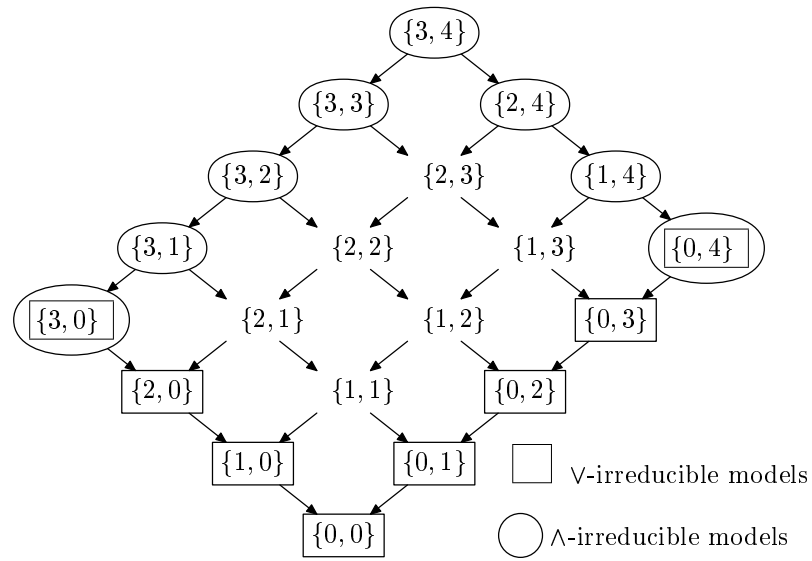Figure 8. Hierarchic log-linear models.

Figure 9. Lattice for ARMA(3,4) model.

List of legends for figures

Figure 1. A step in the procedure.

Figure 2. Linear regression model (20), $q = 3$.

Figure 3. Linear regression model (20), $q = 4$.

Figure 4. Polynomial regression.

Figure 5. Irreducible lattices in analysis of variance, $n = 3$.

Figure 6. Irreducible lattice in analysis of variance, $n = 4$.

Figure 7. Lattice in analysis of variance, $n = 5$.

Figure 8. Hierarchic log-linear models.

Figure 9. Lattice for ARMA(3,4) model.