UNBIASED PREDICTORS

HERMAN O. A. WOLD
UNIVERSITY OF UPPSALA

1. Introduction and summary

1.1. *On the transition from tacit to explicit allowance for residual errors in explanatory models.* This transition, which can be regarded as a subtitle for the present paper, marks a bold raise of aspiration levels in applied statistics. In particular, it is so for model construction in the wide realm of nonexperimental situations. Speaking broadly, there is a lower level where the model is mainly a theoretical construct, and the confrontation with reality consists of *ad hoc* comparisons between theoretical model and empirical evidence. At a higher level the model is built by way of systematic coordination between theoretical approaches on the one hand and statistical data and statistical techniques on the other. Jan Tinbergen's work with multirelation models is an outstanding case in point; his macroeconomic systems for the Netherlands [57], the United States [58], and Great Britain [60] are pioneering in their joint theoretical-empirical approach.

The transition provides clear-cut illustrations of the give and take performance of all deductive theory: *generalization* of the basic assumptions of a model versus *attenuation* of the inference from the model. Disturbances (residuals) in the form of stochastic error terms are introduced to allow for deviations between theoretical relations and actual observations, and this stochastization makes a radical generalization of the model. The attenuation shows up in the fact that the procedures of inference from exact (disturbance-free) relations may or may not extend to inference from stochastic relations. The following procedures are critical cases in point.

(i) The reversal of relationships; or, to paraphrase, the symmetric treatment of explanatory variables and variables to be explained.

(ii) The transformation back and forth from implicit to explicit systems of relationships.

The problems that arise under (i) and (ii) are veritable stumbling blocks on the road of stochastization. They are marked by the long-drawn and partly controversial and confused debate on two key issues: on "choice or regression" in the 1920's and 1930's, and on "interdependent systems" in the 1940's and 1950's. The present paper reviews the situation for models where some or all of the relations are *unbiased predictors*. Being defined in terms of conditional expectations, unbiased predictors are the natural tool for the stochastization of explanatory models, and in particular for the treatment of stimulus-response relations and behavioral relations.

Our review focuses on three types of multirelation models:

Pure causal chain (PCC) *systems*, also known as recursive systems. This is the the approach of Tinbergen in the macroeconomic models referred to above.

Interdependent (ID) *systems*. Initiated by Haavelmo [23] in 1943, this approach was soon thereafter developed under the auspices of the Cowles Foundation [26], [34].

Conditional causal chain (CCC) *systems*. This approach was introduced recently [78] as an alternative to ID systems.

The main incentives for the review are (a) the three approaches, and in particular PCC systems, make the framework for a central line of development in nonexperimental model building, and (b) it is an urgent task to make a comparison and appraisal of the rationale of multirelation approaches, with a view to obtaining better guidance in theoretical and applied work. The allocation of resources is indeed an important problem. To stress one feature on the applied side, the large budgets for big model construction projects are in ample part consumed by the prerequisite compilation of statistical data, and the requirements as to what types of data are needed are partly conflicting, notably for causal chains versus input-output models.

1.2. *Summary*. Sections 2 to 4 deal with three aspects of unbiased predictors: unirelation applications; multirelation applications; estimation by least squares regression. Section 5 adduces an array of comments toward a coordinated view of ends and means in nonexperimental model construction. Section 6 poses the intriguing question whether under favorable circumstances it is possible to assess the direction of causal relations by purely statistical devices.

The upshot of the review is that unbiased predictors, in not being reversible, place the model builder under the burden of designing part or the whole of his model in terms of *directed* (asymmetric) relations, whether the direction is intrinsically implied, as in behavioral relations and other cause-effect relations, or is imputed by specifying the intended operative use of a relation. To "directionize" the explanatory models makes a development program for econometrics and other areas of nonexperimental model construction. If the review can stimulate increased activity along this line it will have fulfilled a main part of its task.

In the emphasis on directed relationships and causal analysis in nonexperimental model building the present review leans heavily on my earlier studies [7], [15], [51], [66] to [82], and links up with [78] to [80] in the systematic use of unbiased predictors. What is novel should partly lie in the arrangement of review material; here we note the emphasis in 4.1 to 4.3 on least squares regression as an unbiased estimation technique that is nonparametric with regard to residual intercorrelations (including autocorrelations) whose assessment in practice would be a problem in itself. Partly it lies in passages where arguments on the basis of typical illustrations like those in (3.8) to (3.10) or those in (3.19) to (3.21) have been carried further by way of general theorems, notably: the explicit formulation of lemma 2.4.4; the conjugate connection between ID and CCC systems shown by theorem 3.2.4; the general interpretation in 3.2.4 of CCC systems in reduced form. As regards the new material in section 6, see the general introduction in that section.

The reader is assumed to be oriented in the foundations of regression analysis (see Cramér [10], especially chapters 23 and 37). For a review with emphasis on causal aspects, see [82], chapter 2.

For helpful discussions during the final revision of the manuscript my thanks are due to Ruth Gomulicka (Cambridge, England), E. Robinson (University of Wisconsin), J.-C. Schoenman (Columbia University), and Lyttkens and S. Thore (University of Uppsala).

2. Unbiased predictors: unirelation inference

2.1. *Summing up.* Explanatory relations and the rationale of their operative use are in this section surveyed from the point of view of unbiased predictors. The conclusions are summed up in table I. The results refer primarily to unirelation models, but are valid also for unbiased predictors that make part of multirelation models. Under category C belongs the familiar fact that unbiased predictors are directed relations, that is, the reverse relation never constitutes an unbiased predictor. A symmetric treatment of unbiased predictors is permissible only in the restricted sense indicated under category A, namely, reversal of the direction of inference. Substitutive inference, it will be noted, is the key to the subsequent typification of multirelation models.

TABLE I

BASIC PROCEDURES OF OPERATIVE INFERENCE, CLASSIFIED AS TO POSSIBILITY OF EXTENSION FROM DISTURBANCE-FREE RELATIONS TO UNBIASED PREDICTORS

Category A.	Procedures that are always valid for unbiased predictors.
1.	Direct prediction.
2.	Reversal of the direction of inference from the predictor.
Category B.	Procedures that are valid for unbiased predictors, subject to supplementary assumptions.
1.	Substitutive inference.
Category C.	Procedures that are not valid for unbiased predictors.
1.	Reversal of the direction of the predictor.

2.2. *The notion of unbiased predictor.* A stochastic relation with additive disturbance term (residual), say

$$y = f(x, z, \cdots) + v \tag{2.1}$$

is called an *unbiased predictor* if $f(x, z, \cdots)$ is the conditional expectation of y for known $x, z, \cdots$; in symbols,

$$f(x, z, \cdots) = E(y|x, z, \cdots). \tag{2.2}$$

In situations where (2.1) is a behavior relation or some other explanatory relation, $x, z, \cdots$ are the explanatory variables and y the variable to be explained.

The term *unbiased predictor*, a new name for an old notion, serves to emphasize the operative aspects of conditional expectations. It will suffice for our purpose to consider linear predictors. The definition, given by (2.3) and (2.4), establishes in one stroke (a) the operative use of unbiased predictors (section 2.4), and (b) their estimation by least squares regression (lemma 4.2). By lemma 2.2.1, the residual properties (2.5) need not be assumed; they are implied in the very definition of linear predictors.

2.2.1. *Linear relations.* In the treatment of explanatory relations it is customary to assume that the residuals of linear relations are uncorrelated with the explanatory variables. The approach of unbiased predictors is closely related, as seen from the following elementary lemma and corollary ([78], p. 362).

LEMMA 2.2.1. *Given the linear stochastic relation*

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_h x_h + v, \tag{2.3}$$

suppose that the relation is an unbiased predictor, that is,

$$E(y|x_1, \cdots, x_h) = \beta_0 + \beta_1 x_1 + \cdots + \beta_h x_h \tag{2.4}$$

and that all variables have finite variance. Then

$$E(v) = 0, \qquad E(vx_i) = 0, \tag{2.5}$$

showing that v has zero mean and is uncorrelated with $x_1, \cdots, x_h$.

The proof is immediate, for if we consider (2.3) for any fixed $x_1, \cdots, x_h$, assumption (2.4) implies $E(v|x_1, \cdots, x_h) = 0$, $E(vx_i|x_1, \cdots, x_h) = 0$, which on integration over $x_1, \cdots, x_h$ gives (2.5).

COROLLARY. *Conversely,* (2.4) *sometimes but not always follows from* (2.5); *it does, for example, if the variables are jointly normally distributed.*

2.3. *Simple illustrations of disturbance-free relations versus unbiased predictors.* We shall consider the relations

$$y = \alpha + \beta x, \tag{2.6}$$

$$y = \alpha + \beta x + v \qquad \text{with} \qquad E(y|x) = \alpha + \beta x, \tag{2.7}$$

where (2.6) is a disturbance-free relation, and (2.7) is a corresponding unbiased predictor.

(i) *The case of several explanatory variables.* Assuming that there are h explanatory variables, as in (2.3), the linear case is covered by (2.7) if we interpret β as a row vector $(\beta_1, \cdots, \beta_h)$ and x as a column vector $\{x_1, \cdots, x_h\}$.

(ii) *The case of time series data.* A simple disturbance-free model in time series analysis is

$$y_t = \alpha + \beta y_{t-1}, \qquad t = 0, \pm 1, \pm 2, \cdots. \tag{2.8}$$

The corresponding model subject to disturbance is

$$y_t = \alpha + \beta y_{t-1} + v_t, \qquad \text{with } E(y_t|y_{t-1}) = \alpha + \beta y_{t-1}. \tag{2.9}$$

Taking the residuals $v_t, v_{t\pm1}, v_{t\pm2}, \cdots$ to be mutually independent and to have zero mean, the left equation in (2.9) defines an autoregressive stochastic process that satisfies the right equation.

2.4. *Basic procedures of operative inference.* Most of the material in section 2.4 belongs to the foundations of regression analysis; see Cramér [10] also for further references. The present exposition is somewhat more general since no restrictions are imposed on the unbiased predictors. For a related treatment see Hurwicz [27], who formulates the restrictive conditions in terms of the conditional distributions that define the residuals.

To establish table I, we shall consider four modes of inference, all of which involve an element of prediction. The procedures will be illustrated by relations (2.6) to (2.9). Speaking generally, the procedures will yield *exact* prediction in the case of disturbance-free relations, *expected* or *average* values in the case of unbiased predictors.

2.4.1. *Direct inference.* This is the fundamental mode of inference from unbiased predictors. Considering the disturbance-free relation in (2.6) and assuming x to be known, the procedure is to predict y directly from the given relation. For the unbiased predictor (2.7), the procedure is the same, except that the residual is ignored and the prediction yields the expected value of y for known x. This gives

$$\text{(2.10)} \qquad \text{Pred}\, y = \alpha + \beta x, \qquad \text{Pred}\, y = E(y|x) = \alpha + \beta x$$

for the two models (2.6) and (2.7), respectively.

2.4.2. *Reverse inference: reversal of relations.*

(i) *Disturbance-free models.* Again considering relation (2.6), now assuming y to be known, the procedure is to predict x from the reverse relation, that is,

$$\text{(2.11)} \qquad \text{Pred}\, x = \frac{1}{\beta}(y - \alpha).$$

The reverse inference (2.11) is or is not unique according to whether x stands for one or more variables.

(ii) *Unbiased predictors.* Ever since the beginning of correlation theory it has been known that unqualified reversal (2.11) is not allowed when dealing with unbiased predictors. In fact, for any joint distribution (x, y) with relation (2.7) and $P\{v = 0\} < 1$, we have

$$\text{(2.12)} \qquad E(x|y) \neq \frac{1}{\beta}(y - \alpha).$$

2.4.3. *Reverse inference: reversal of the direction of inference.* Both for disturbance-free relations and unbiased predictors, as also is well known, reverse inference is allowed in a weaker sense than in (2.11). Given $y = y_0$, if we ask for what value x the predicted (expected) value of y will equal y_0, the answer is

$$\text{(2.13)} \qquad x = \frac{1}{\beta}(\text{Pred}\, y - \alpha),$$

where

$$y_0 = \text{Pred}\, y, \qquad y_0 = \text{Pred}\, y = E(y|x) \tag{2.14}$$

for models (2.6) and (2.7), respectively. As in (2.11) the reverse inference is not unique in the case of several explanatory variables.

2.4.4. *Substitutive inference.* Substitution not always being a permissible procedure for unbiased predictors, lemma 2.4.4 shows that the procedure is valid when one of the explanatory variables is substituted in terms of all the other ones. The situation may seem restrictive, but will suffice for our purposes. In the simplest applications $x_1, x_2, \cdots$ are lagged values of y; the predictor (2.15) is a lagged version of (2.3); and (2.3) is an autoregressive relation with $\beta_{h+1} = \beta_{h+2} = \cdots = 0$.

LEMMA 2.4.4. *Let* (2.3) *and*

$$x_1 = \alpha_0 + \alpha_2 x_2 + \cdots + \alpha_h x_h + \omega, \tag{2.15}$$

be two unbiased predictors, and let the relation obtained by substituting (2.15) *into* (2.3) *be written*

$$y = L + v + \beta_1 \omega, \tag{2.16}$$

where

$$L = \beta_0 + \alpha_0\beta_1 + (\beta_2 + \alpha_2\beta_1)x_2 + \cdots + (\beta_h + \alpha_h\beta_1)x_h. \tag{2.17}$$

Then (2.16) *is an unbiased predictor with*

$$E(y|x_2, x_3, \cdots, x_h) = L. \tag{2.18}$$

The lemma is a simple corollary to Kolmogorov's general theorem that the expectation of a conditional expectation equals the unconditional expectation [32]. The theorem gives

$$E(v|x_2, \cdots, x_h) = E_{x_1}[E_v(v|x_1, \cdots, x_h)] = E_{x_1}[0] = 0, \tag{2.19}$$

where the subscripts to E indicate that the expectation is formed by integration over the distribution of the variable in the subscript, and where the expectation in the bracket vanishes owing to (2.6). Now since expectations are additive, (2.16) gives

$$\begin{aligned} E(y|x_2, \cdots, x_h) &= E(L + v + \beta_1\omega|x_2, \cdots, x_h) \\ &= L + E(v|x_2, \cdots, x_h) + \beta_1 E(\omega|x_2, \cdots x_h) = L. \end{aligned} \tag{2.20}$$

Here the conditional expectation of v is zero according to (2.19). Further, the expectation of ω vanishes, since (2.15) is an unbiased predictor.

ILLUSTRATION 2.4.4. We apply the procedure of substitution to the disturbance-free relation (2.8). For known y_t, exact prediction of y_{t+k} is provided by

$$\text{Pred}\, y_{t+k} = \alpha(1 + \beta + \cdots + \beta^{k-1}) + \beta^k y_t, \tag{2.21}$$

which is obtained from (2.8) by $k - 1$ iterated substitutions.

Similarly, the substitutive procedure applies to the unbiased predictor (2.9), if interpreted as an autoregressive process. The ensuing formula is again (2.21), where Pred y_{t+k} now gives $E(y_{t+k}|y_t)$, that is, the expected value of y_{t+k} for known y_t.

2.5. *Ceteris paribus, and the concept of operative frame.* This subsection borrows from [78], with some elaboration of the argument.

2.5.1. *Sample, population, frame.* To discuss the rationale of applied work with explanatory models it will be helpful to distinguish three notions:

(i) the sample S of empirical observations used in the model construction for parameter estimation and related statistical purposes,

(ii) the population P, or statistical universe, which is assumed to have generated the observed sample,

(iii) the field of potential applications of the model, or its *operative frame*, or simply its *frame*, say F.

As a rule, the frame F involves an extension in one or more respects relative to the sample S and the population P. In symbols,

$$S \subset P \subset F. \tag{2.22}$$

Usually, the main practical value of the model construction lies in the extension from S and P to F.

A universe U is sometimes defined as a superpopulation that includes P and has the same distribution properties as P. The notion of frame is more general, the point being to emphasize that the potential applications of a model may extend to populations that in some respect or other differ from P.

2.5.2. *Ceteris paribus.* Typically, predictions from an explanatory model involve two assumptions regarding the frame F, namely (a) specified changes in the explanatory variables, and (b) a *ceteris paribus* clause, requiring that there be no changes in other respects. The *ceteris paribus* clause is a crucial point in the models under review. In accordance with customary phrasings of the clause, we have thus far given (b) a vague formulation. In disturbance-free models the clause requires no comment, but for stochastic models it is illuminating to specify two interpretations of the clause,

Ceteris paribus clause A. In the frame F (outside P) the disturbances are absent.

Ceteris paribus clause B. The distribution properties of the disturbances are the same in the frame F as in the population P.

ILLUSTRATION 2.5.2a. In the language of economics, *elasticity* is the logarithmic derivative taken with negative or positive sign. Introduced by Marshall [41] for disturbance-free relations, the definition allows a straightforward extension to unbiased predictors.

The notion of *demand elasticity* provides material for an instructive case study of stochastization, the approach hovering between symmetric and asymmetric treatment of the variables, between *ceteris paribus* clauses A and B, and between descriptive and causal interpretation of the demand relation. Typical for a symmetric treatment is the unqualified definition of price flexibility as the inverse with respect to demand elasticity with respect to price. Taking (3.8) to be a demand relation with logarithmic variables, suppose that price p_t increases by 1 per cent while the other conditions remain the same in the sense of clause B. Then β_1 is the price elasticity of demand in the sense that the *expected* decrease of demand y is β_1 per cent. It is less in line with the logic of stochastic models to invoke clause A. On clause A the same relation (3.8) is treated as disturbance-free, and the *exact* decrease of demand is β_1 per cent.

ILLUSTRATION 2.5.2b. Linking with formulas (2.8) and (2.9), figure 1 refers to prediction by relation (2.21). The sample is observed up to the time t, the prediction is made at t and refers to the future time point $t + k$. Hence the time up to t belongs to the sample S and the population P, whereas the time after t belongs to the frame F.

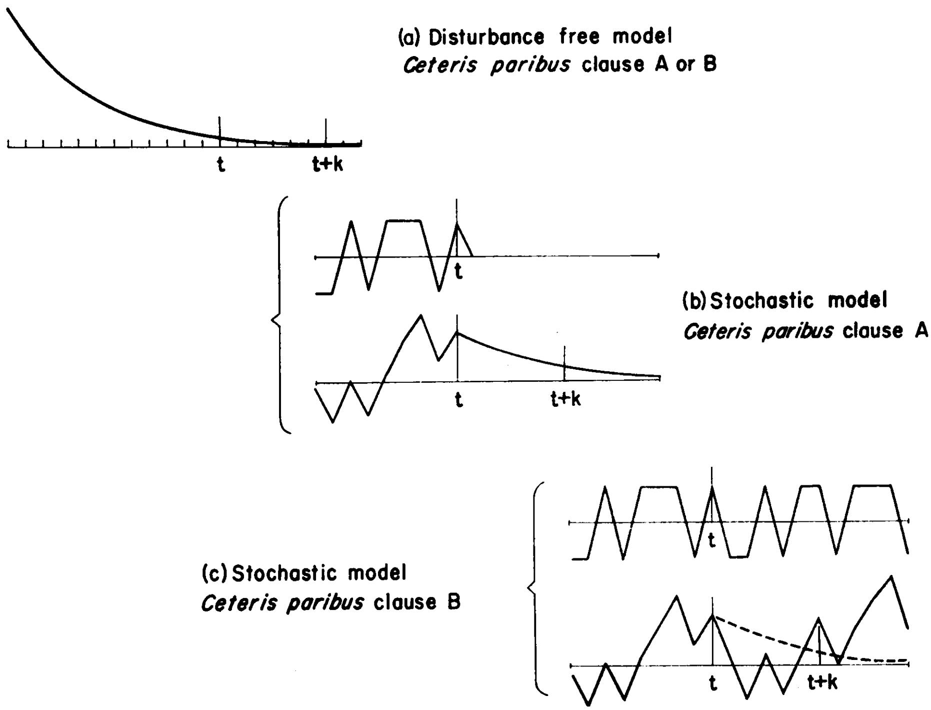

FIGURE 1

Prediction by models (2.8) and (2.9).

Figure 1(a) refers to the disturbance-free relation (2.8) and exact prediction by formula (2.21), valid on *ceteris paribus* clause A (or B). Figures 1(b) and 1(c) refer to the unbiased predictor (2.9). In figure 1(c) the prediction is subject to clause B. The prediction, read from the broken curve, yields the expected value of y_{t+k} for known y_t, and the prediction works in situations where the future disturbances have the same distribution properties as in the past. Figure 1(b) emphasizes that clause A is not satisfactory in stochastic models. The prediction here yields the *exact* value of y for known x, and the prediction works in situations where the disturbances cease to occur after time point t.

2.5.3. *Pitfalls of stochastization.*

(i) Clearly, *ceteris paribus* clause A (or B) is appropriate for prediction on the basis of disturbance-free models, whereas clause B is designed for unbiased predictors.

(ii) An unbiased predictor that works on the "stochastic" clause B is also valid on the "disturbance-free" clause A. The converse is not true in general.

(iii) As applied to models that involve disturbances, clause A may be regarded as a reminiscence of nonstochastic procedures in the treatment of stochastic models. Such reminiscences are dangerous and may give rise to serious bias in the operative procedures, so to speak, a "pitfall of stochastization." Unqualified reversal of unbiased predictors is a case in point, as we know from subsection 2.4.2.

2.6. *Cross section versus time series prediction. Conditional predictors.* Unbiased predictors, as is well known, are used both for cross section and time series inference. In cross section prediction the explanatory variables and the predicted variable refer to the same time period, as in relation (3.8). In time series prediction, the explanatory variables are lagged relative to the predicted variable, as in (3.9) or (3.10).

In practice, cross section prediction often refers to some future time point $t + k$, while the requisite values of the explanatory variables are provided by *ad hoc* assumptions or predictions. For example, suppose that (3.8) is a demand relation in logarithmic variables with $\alpha_1 = 1.28$ and price elasticity $\beta_1 = 0.52$. Then, if we take p_{1970} to be given by means of an assumption or an auxiliary prediction, this relation gives

$$\text{Pred } d_{1970} = 1.28 - 0.52\, p_{1970} \tag{2.23}$$

for the expected value of demand in 1970. We shall speak of *conditional prediction* when an unbiased predictor in this way is applied on the basis of ancillary information on the explanatory variables.

Cross section prediction and time series prediction are in everyday use. Figure 2 serves to emphasize that they are complementary modes of inference.

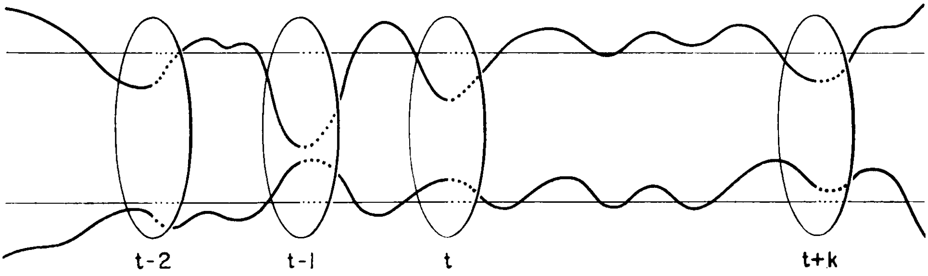

FIGURE 2

Cross section prediction and time series prediction.

The cross sections, drawn as planes perpendicular to the time axis, are symbolic fields of simultaneous observations. The two time series shown in the graph may, for example, represent demand d and price p in model (3.8) to (3.10).

2.7. *Mean square predictors.* Relation (2.1) is called a least squares predictor or, in line with the terminology of Cramér [10], a *mean square* predictor if $f(x, z, \cdots)$ involves unknown parameters and these are chosen so as to make v of minimum variance over the joint distribution of $x, z, \cdots$. If the joint probability distribution of $y, x, z, \cdots$ is normal, then $E(y|x, z, \cdots)$ is linear in $x, z, \cdots$ and the unbiased predictor coincides with the corresponding linear mean square predictor.

With due qualifications, the considerations of our review can be generalized in the direction of mean square predictors. The change in the argument is mainly that conditions of independence will be replaced by conditions of noncorrelation. In lemma 2.4.4, for example, the requisite extension is straightforward (see [82], exercise IV, 9).

3. Unbiased predictors in multirelation models

Our brief review assumes the reader to have some orientation in the large literature on multirelation models. The textbooks of Baumol [4] and Klein [30] emphasize the wide potential scope of the models for economic analysis in general. For PCC systems the basic references are Tinbergen [58], [59]; on the rationale of the approach see also [7], [71], [75], [78]. For ID systems [34] and [26] are authoritative expositions of the theory of the approach, with somewhat revised account of earlier papers; for applied work see [8], [25], [29], [31], [52], [2]. For CCC systems see [78] to [80]. Reference is further made to Leontif's input-output models [35], essentially an approach of comparative statistics working with disturbance-free models. For comparative studies with emphasis on the theoretical rationale of the approaches, see also [6], [49], [43], [51], [50].

3.1. *Introductory.* Speaking generally, a multirelation model contains relations of two types.

(1) *Relations that constitute the fundamental hypotheses of the model.* This is the model in original form, also known as the structural form, or, as we shall say, the *primary form.* In PCC, ID, and CCC systems the primary form contains, first of all, the behavior relations of the decision-taking units whose actions it is the system's fundamental purpose to explain. The variables that enter the model as variables to be explained, explicitly or implicitly, are called *endogenous.* The system may further involve *exogenous* variables; these represent outside influences. In a behavior relation the variable to be explained thus is a current endogenous variable, and the explanatory variables may involve (a) other current endogenous variables, (b) lagged values of the endogenous variables, and (c) current or lagged values of the exogeneous variables.

(2) *Relations that by formal procedures are derived from the primary form.* For PCC, ID, and CCC systems these relations include the *reduced form.* Obtained

by regarding the primary form as an implicit system for the current endogenous variables, the reduced form gives the current endogenous variables explicitly in terms of (a) lagged values of the endogenous variables and (b) current or lagged values of the exogenous variables.

In disturbance-free systems the primary and reduced forms are equivalent for purposes of prediction. Thus if we know the past of the endogenous variables, and the past and the future of the exogenous variables, the two forms give the same results when predicting the future development of the endogenous variables. For example, this is so in the simple cobweb model with primary form

$$(3.1) \qquad q_t = \alpha_1 - \beta_1 p_t, \qquad q_t = \alpha_2 + \beta_2 p_{t-1},$$

and reduced form

$$(3.2) \qquad p_t = \frac{\alpha_1 - \alpha_2}{\beta_1} - \frac{\beta_2}{\beta_1} p_{t-1}, \qquad q_t = \alpha_2 + \beta_2 p_{t-1}.$$

(See [14], [56] for the cobweb theory of the early 1930's with its disturbance-free models for the demand-supply balance.) The free use of both the primary form and the reduced form for predictive inference may or may not extend to systems constructed in terms of unbiased predictors. The ensuing parting of the ways is shown in table II, where categories A to C refer to table I.

TABLE II

PCC, ID, AND CCC SYSTEMS CLASSIFIED WITH REGARD TO DIRECT PREDICTION FROM THE SYSTEM IN PRIMARY AND/OR REDUCED FORM

1. *Direct prediction from the model in primary form.* PCC and CCC systems belong to category A. ID systems belong to category C.
2. *Direct prediction from the model in reduced form.* PCC and ID systems belong to category B. CCC systems belong to category C.
3. *Direct prediction from the model both in primary and reduced form.* PCC systems belong to category B. ID and CCC systems belong to category C.

3.2. *The formal structure of* PCC, ID, *and* CCC *systems.* In the subsequent specifications of PCC, ID, and CCC systems we do not press for generality. Our intention is, rather, to set forth what is typical for the three approaches and to give a set of simple illustrations. For all three types of models we shall use the notation

n is the number of relations and endogenous variables;

$y = \{y_1, \cdots, y_n\}$ denotes the current endogenous variables;

z denotes lagged endogenous and current or lagged exogenous variables.

We adduce two definitions and an auxiliary lemma to prepare a key argument in the mathematics behind table II.

DEFINITION 3.2.1a. *A square matrix $A = [a_{ik}]$ is called recursive if there are mere zeros in and above the main diagonal, giving*

$$A = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ a_{21} & 0 & 0 & \cdots & 0 \\ a_{31} & a_{32} & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ a_{n1} & a_{n2} & a_{n3} & \cdots & 0 \end{bmatrix},$$

(3.3)

$$I - A = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -a_{21} & 1 & 0 & \cdots & 0 \\ -a_{31} & -a_{32} & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ -a_{n1} & -a_{n2} & -a_{n3} & \cdots & 1 \end{bmatrix}.$$

DEFINITION 3.2.2b. *A system of linear relations* (3.4) *is called recursive if and only if* (a) *the system contains one and only one explanatory relation for each of the* (*current*) *endogenous variables, and* (b) *the relations can be so ordered that the matrix A is recursive.*

LEMMA 3.2. *The system* (3.4) *is recursive if and only if its reduced form* (3.6) *can be obtained from* (3.4) *by a sequence of substitutions.*

The terminology for PCC systems is fluent [59], [7], [51], [78]. In this review we shall say that a system (3.4) is recursive if the matrix A is recursive in the sense of definition 3.2.1, and that it is a pure causal chain if it satisfies the additional assumptions (3.5) and (3.7), which ensure that the relations of the model are unbiased predictors both before and after transformation to the reduced form.

3.2.1. *Pure causal chain* (PCC) *systems.* Linear PCC systems in primary form may be written

(3.4) $$y = Ay + Bz + v,$$

where

(i) the system is recursive;

(ii) all relations of the primary form are unbiased predictors; hence

(3.5) $$E(y_i|z, y_1, y_2, \cdots, y_{i-1}) = \{Ay + Bz\}_i, \qquad i = 1, \cdots, n,$$

where the ith element of a column vector $\{\cdot\}$ is denoted $\{\cdot\}_i$;

(iii) in the reduced form of the system, that is,

(3.6) $$y = (I - A)^{-1} Bz + (I - A)^{-1} v,$$

all relations are unbiased predictors; hence

(3.7) $$E(y|z) = (I - A)^{-1} Bz.$$

REMARK 3.2.1. By lemma 2.2.1, assumption (3.5) implies that the residual v_i of the ith relation is uncorrelated with the explanatory variables $z, y_1, \cdots, y_{i-1}$ of the same relation. In the customary design of PCC systems a typical assumption is, furthermore, that all current residuals $v_1, \cdots, v_n$ are mutually uncorrelated.

On the assumption of jointly normal distributions the two sets of residual noncorrelation assumptions will, under general conditions of stationarity, be equivalent with assumptions (3.5) and (3.7). We shall come back to these matters in subsection 5.3.2.

ILLUSTRATION 3.2.1. We quote the following PCC system, a simple demand-supply model for a market under free competition,

$$d_t = \alpha_1 - \beta_1 p_t + v_t \qquad \text{demand relation} \tag{3.8}$$

$$s_t = \alpha_2 + \beta_2 p_{t-1} + v_t' \qquad \text{supply relation} \tag{3.9}$$

$$p_t = p_{t-1} + \gamma(d_{t-1} - s_t) + v_t'' \qquad \text{price mechanism} \tag{3.10}$$

(See [71], [75], [78] for more detailed comment on this and related models.)

The system has no exogenous variables. Taking the endogenous variables in order s_t, p_t, d_t the nonzero elements of matrix A are

$$a_{21} = -\gamma, \qquad a_{32} = -\beta_1, \tag{3.11}$$

which shows that A is recursive.

The system in (3.8) to (3.10) is the primary form of the model. The reduced form is given by

$$s_t = \alpha_2 - \beta_2 p_{t-1} + v_t', \tag{3.12}$$

$$p_t = \gamma(\alpha_1 - \alpha_2) + (1 - \gamma\beta_1 - \gamma\beta_2)p_{t-1} + v_t^*, \tag{3.13}$$

$$d_t = \alpha_1 - \beta_1\gamma(\alpha_1 - \alpha_2) - \beta_1(1 - \gamma\beta_1 - \gamma\beta_2)p_{t-1} + v_t^{**}, \tag{3.14}$$

where the residuals v_t^*, v_t^{**} are linear expressions of v_t, v_t', v_t''.

According to 3.2.1(ii) and (iii), all relations in (3.8) to (3.10) and in (3.12) to (3.14) are unbiased predictors, and further unbiased predictors can be obtained by iterated substitutions in the primary or the reduced form. We note, for example, that (3.8) and (3.14) imply

$$\begin{aligned} E(d_t|p_t) &= \alpha_1 - \beta_1 p_t, \\ E(d_t|p_{t-1}) &= \alpha_1 - \beta_1\gamma(\alpha_1 - \alpha_2) - \beta_1(1 - \gamma\beta_1 - \gamma\beta_2)p_{t-1}. \end{aligned} \tag{3.15}$$

The directional structure of the unbiased predictors in (3.8) to (3.10) is illustrated in figure 3 by an *arrow scheme*, a type of graph used by Tinbergen [59] for illustration of PCC systems, in particular for the causal interpretation of the behavior relations.

3.2.2. *Interdependent* (ID) *systems.* Linear ID systems in primary (structural) form may be written

$$Cy = Bz + \nu, \tag{3.16}$$

where

(i) the coefficient matrix $C = [c_{ik}]$ of the current endogenous variables is nonsingular, which implies that the reduced form of the system is well defined, being given by

$$y = C^{-1} Bz + C^{-1} \nu; \tag{3.17}$$

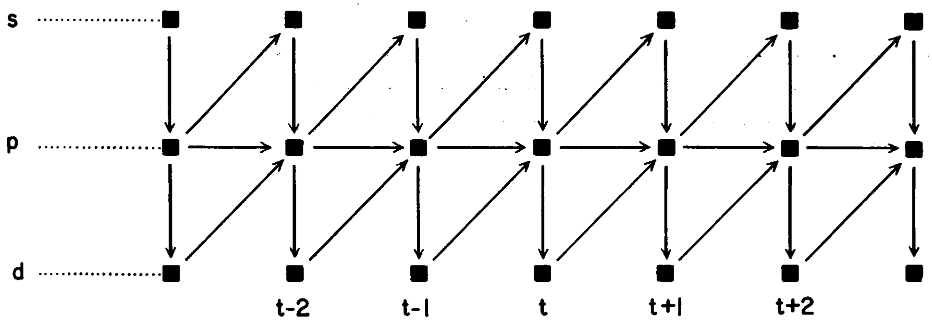

FIGURE 3

Arrow scheme for the pure casual chain system (3.8) to (3.10).

(ii) the residuals ν satisfy assumptions to the effect that the relations of the reduced form are unbiased predictors; hence

$$E(y|z) = C^{-1} Bz. \tag{3.18}$$

REMARK 3.2.2a. In the theory of ID systems the customary assumption is that the variables are jointly normally distributed, and that the current residuals are uncorrelated with the (relevant) exogenous and lagged endogenous variables. The residual noncorrelation assumptions will under general conditions be equivalent to assumption (3.18). (Compare remark 3.2.1.)

ILLUSTRATION 3.2.2. We quote the following ID system, a simple cobweb model subject to disturbance,

$$d_t(=q_t) = \alpha_1 - \beta_1 p_t + \nu_t \qquad \text{demand relation,} \tag{3.19}$$

$$s_t(=q_t) = \alpha_2 + \beta_2 p_{t-1} + \nu'_t \qquad \text{supply relation,} \tag{3.20}$$

$$d_t = s_t = q_t \qquad \text{relation of instantaneous equilibrium.} \tag{3.21}$$

See [78] for further discussion from the present points of view. Models of the same simple type were used in the early works on ID systems; see [19].

All variables are endogenous, and since there is no behavior relation for price p_t the model is not a PCC system. The two first relations give

$$c_{11} = 1, \qquad c_{12} = \beta_1, \qquad c_{21} = 1, \qquad c_{22} = 0, \tag{3.22}$$

showing that the matrix C is nonsingular. The reduced form is

$$q_t = \alpha_2 + \beta_2 p_{t-1} + \nu_t^*, \qquad p_t = \frac{\alpha_1 - \alpha_2}{\beta_1} - \frac{\beta_2}{\beta_1} p_{t-1} + \nu_t^{**}, \tag{3.23}$$

where $\nu_t^* = \nu'_t$ and ν_t^{**} are linear in ν_t and ν'_t.

According to 3.2.2(ii), the relations (3.23) are unbiased predictors,

$$E(p_t|p_{t-1}) = \alpha_2 + \beta_2 p_{t-1}, \qquad E(p_t|p_{t-1}) = \frac{\alpha_1 - \alpha_2}{\beta_1} - \frac{\beta_2}{\beta_1} p_{t-1} \tag{3.24}$$

and further unbiased predictors can be obtained from (3.24) by iterated substitutions. The supply relation (3.20) is an unbiased predictor since it so happens that it coincides with the first relation (3.23) of the reduced form, whereas the demand relation (3.19) in general is not an unbiased predictor,

$$E(q_t|p_t) \neq \alpha_1 - \beta_1 p_t. \tag{3.25}$$

For later reference we note the special case

(3.26)

$$\alpha_1 = \alpha_2 = 0, \qquad \beta_1 = \frac{0.6}{\rho}, \qquad \beta_2 = 0.6, \qquad r(\nu_t, \nu_t') = \frac{25\rho - 12}{4(25\rho^2 - 24\rho + 9)^{1/2}},$$

where ρ is arbitrarily fixed in the interval $0 < \rho \leqq 0.96$.

Relation (3.21) illustrates a typical situation in ID systems, namely that the endogenous variables are subject to one or more exact relationships, also called *linear constraints*. Such constraints in general will reduce the number of relevant endogenous variables, in systems (3.19) to (3.21) from three (d, s, p) to two (q, p). In the representation (3.16) it is assumed that such elimination of variables has already been performed. For some purposes it is useful to have a representation of the model before performing the elimination, and the following formula refers to this situation,

$$\begin{aligned} Gy &= Bz + \nu && m \text{ relations} \\ L_j(y, z) &= 0 && j = 1, \cdots, k; m + k = n. \end{aligned} \tag{3.27}$$

In current terminology, the first part of (3.27) is m behavioral relations, and the second is k linear constraints.

REMARK 3.2.2b. For ID systems, as we know from lemma 3.2 and table II, the behavior relations (3.27) of the primary form in general are not unbiased predictors. If we ask what kind of relations they are, a general answer is: given the exogenous and lagged endogenous variables z, suppose we wish to find values $\bar{y}$ for the current endogenous variables y such that the relations of the primary form, taken without residuals (compare the *ceteris paribus* clause A of 2.5.2) are identically satisfied by $\bar{y}$ and z; then $\bar{y}$ are obtainable from the reduced form as the conditional expectations of y for given z. That this is so is seen if we determine $\bar{y}$ from the reduced form, taken without residuals; then on transformation back to the primary form, this, with the residuals still omitted, will be satisfied by $\bar{y}$ and z. (For a more elaborate discussion see [51], where the interpretation is the same, except for the terminology.)

Illustration: (3.19) to (3.21) specified by (3.26). The reduced form gives us

$$\bar{p}_t = E(p_t|p_{t-1}) = -\rho p_{t-1}, \qquad \bar{q}_t = E(q_t|p_{t-1}) = 0.6\, p_{t-1} \tag{3.28}$$

while the primary form, taken without residuals, is required to give

$$\bar{q}_t = -\frac{0.6}{\rho}\bar{p}_t, \qquad \bar{q}_t = 0.6\, p_{t-1}, \tag{3.29}$$

and (3.29) is identically satisfied by $\bar{p}_t$ and $\bar{q}_t$ as given by (3.28).

DEFINITION 3.2.2. *An ID system is called just identified, over- or underidentified according as the coefficients of the primary form are uniquely, over- or underdetermined by the coefficients of the reduced form.*

While (3.19) to (3.21) is just identified, the usual situation in practice is that ID systems are overidentified (see theorem 3.2.4.).

3.2.3. *Conditional causal chain* (CCC) *systems.* CCC systems have the same general coefficient pattern as ID systems, that is, (3.16) or (3.27). To make for easy comparison with (3.27) we shall use the following representation for linear CCC systems in primary form,

$$y^* = \mathrm{H}'y + Bz + \omega, \qquad m \text{ relations}, \tag{3.30}$$

$$L_j(y, z) = 0, \qquad j = 1, \cdots, k; m + k = n, \tag{3.31}$$

where

(i) any component of the column vector y^*, say y_i^*, is a current endogenous variable, say y_{r_i}. The components $y_1^* = y_{r_1}, \cdots, y_m^* = y_{r_m}$ are not necessarily different variables.

(ii) The prime (′) of matrix H' indicates that

$$h_{ir_i} = 0, \qquad i = 1, \cdots, m, \tag{3.32}$$

that is, the term $h_{ir_i}y_{r_1}^*$ is missing in the ith relation in (3.30).

(iii) Relations (3.30) are unbiased predictors; hence

$$E(y^*|y', z) = H'y + Bz, \tag{3.33}$$

where the prime of the vector y' indicates that y_i^* does not enter among the variables y' in the conditional expectation $E(y_i^*|y', z)$.

REMARK 3.2.3. By lemma 2.1, assumption (3.33) implies that the residual ω_i in the ith relation of the system is uncorrelated with the explanatory variables y', z of the same relation.

ILLUSTRATION 3.2.3. On appropriate respecification of the stochastic assumptions, system (3.19) to (3.21) provides an illustration of the CCC approach. (See [78] for a discussion of this and related designs of CCC systems.)

The ensuing reduced form may again be written as in (3.23). According to 3.2.3(iii) it is now relations (3.19) and (3.20) that are unbiased predictors. Hence

$$E(q_t|p_t) = \alpha_1 - \beta_1 p_t, \qquad E(q_t|p_{t-1}) = \alpha_2 + \beta_2 p_{t-1}. \tag{3.34}$$

In the reduced form the first relation (3.23) is now an unbiased predictor since it so happens that it coincides with the supply relation (3.20) in the primary form of the model, whereas the second relation (3.23) in general is not an unbiased predictor,

$$E(p_t|p_{t-1}) \neq \frac{\alpha_1 - \alpha_2}{\beta_1} - \frac{\beta_2}{\beta_1} p_{t-1}. \tag{3.35}$$

It will be noted that characteristic differences between ID and CCC systems show up in (3.25) and the first part of (3.34) and also in the second part of (3.24) and (3.35).

It is illuminating to consider the special case

$$\alpha_1 = \alpha_2 = 0, \qquad \beta_1 = 0.8, \qquad \beta_2 = 0.6, \qquad r(\omega_t, \omega_t') = \rho, \tag{3.36}$$

where ω_1 and ω_t' denote the residuals in model (3.19) to (3.21) when specified as a CCC system. It can be shown that if p_t and q_t are normally distributed with zero mean and unit variance, a CCC system (3.19) to (3.21) specified by (3.36) defines the same stationary Gauss-Markov process as an ID system (3.19) to (3.21) specified by (3.26). Thus when the two models are used for generating artificial time series the ensuing two sets of series will be identically the same, [78].

3.2.4. *Reversal of the direction of inference in* PCC *and* CCC *systems.*

(i) The direction of inference from PCC systems can be reversed in the sense of 2.4.3. Thus if the current endogenous variables are given, say $y = y^{(0)}$, the system (it does not matter if taken in primary or reduced form) makes a set of conditions that the exogenous and lagged endogenous variables z must satisfy in order that the expected value of y should equal $y^{(0)}$, conditions by which z may or may not be uniquely determined.

(ii) For CCC systems the corresponding problem will be dealt with in two installments since the reduced form relations in general are not unbiased predictors.

(a) To begin with, the situation is related to that in definition 3.2.2 in a conjugate fashion.

THEOREM 3.2.4. *In reversing the direction of inference from the behavior relations of a* CCC *system, the exogenous and lagged endogenous variables will be uniquely determined, over- or underdetermined by the current endogenous variables according as the corresponding* ID *system is just identified, over- or underidentified.*

(b) If we ask what interpretation can be given to the reduced form relations of a CCC system, the answer involves a reformation of remark 3.2.2b in terms of expected values on *ceteris paribus* clause B. "Given p_{t-1}, for what current price $\hat{p}_t$ will the expected demand equal the expected supply?" The answer is given by the reduced form, taken without residuals,

$$\hat{p}_t = -0.75\, p_{t-1}. \tag{3.37}$$

Since the behavior relations of the primary form are unbiased predictors, we obtain

$$E(d_t|\hat{p}_t) = -0.8\, \hat{p}_t = 0.6\, p_{t-1}, \qquad E(s_t|p_{t-1}) = 0.6\, p_{t-1}. \tag{3.38}$$

As required, the linear constraint (3.21) is identically satisfied by the conditional expectations (3.38).

The general problem is in essence the same as for PCC systems, but the formulation is somewhat more complicated than under (i): given the exogenous and lagged endogenous variables z, suppose we wish to find values $\hat{y}$ for the current endogenous variables y such that the corresponding expectations as given by the behavior relations are compatible with the linear constraints of the primary form; then $\hat{y}$ is obtainable from the reduced form, taken without residuals. That this is so is seen if we seek $\hat{y}$ so as to satisfy

$$\hat{y}^* = \mathrm{H}'\hat{y} + Bz, \qquad L_j(\hat{y}, z) = 0, \tag{3.39}$$

and when solving for $\hat{y}$ this system gives the reduced form, residuals omitted.

3.3. *On* PCC, CCC *and* ID *systems as stochastic processes.* By the introduction of stochastic residuals the theory of disturbance-free dynamic models becomes a branch of the theory of stochastic processes. The analytical tools that thereby are placed at our disposal are very useful for exploring the rationale of multirelation models. It is primarily some chapters on stationary, Markovian and Gaussian processes that come into play. For an orientation see [36], [66], [82]. The recent treatise of Robinson [45], more advanced, places emphasis on regression, prediction, and other aspects of central importance for the models under review. In the following references, the results under 3.3.1 have the nature of general existence theorems. Under 3.3.2 special processes serve as illustrations of general arguments.

3.3.1. *On the general scope of* PCC *systems.* The specification of PCC systems for unbiased prediction both in primary and reduced form involves two sets of restrictive conditions, (3.5) and (3.7). In CCC and ID systems there is only one set of restrictions. At first sight this may give the impression that PCC systems are very special and of narrow scope. The more general structure of CCC and ID systems is however only apparent. The theorems under (i) to (iii) establish PCC systems as a class of stochastic processes which in principle is of perfectly general scope.

For references under (i), see [66], [7], [75]. Under (ii), see [66], theorem 7, for predictive decomposition; for the corollary, see [73]. The theorem of predictive decomposition has been deepened and generalized in several respects. For the results under (iii), see [70] and [71]; the area is still in development [11], [22], and [42].

For students who are not familiar with predictive decomposition it is instructive to carry out its application to (3.8) to (3.10) in detail, noting in particular that in the substitutive deduction (3.12) to (3.15) of unbiased predictors with increasingly higher lags the variables must be substituted in the specified order, starting somewhere in the chain $s_t, p_t, d_t, s_{t-1}, \cdots$ and then proceeding in this order. (Compare lemma 2.4.4.)

(i) For Gaussian processes, stationary or evolutive, PCC systems constitute a straightforward generalization of processes of the autoregressive type, an extension from univariate to multivariate models.

(ii) The theorem known as predictive decomposition of stationary processes has the following corollary: Any given stationary Gaussian process can be approximated, to any prescribed accuracy, by a suitably chosen PCC system. That is, all moments of first and second order of the given process will be approximated by the corresponding moments of the process that is generated by the PCC system. This existence theorem remains valid if the given moments are generated not by a process but by a given set of observed time series.

(iii) The notion of PCC systems allows a straightforward generalization by not postulating (3.5) and (3.7) but instead that the relations of the primary and reduced forms should be mean square predictors. Any stationary process with finite variance can be approximated, to any prescribed accuracy in first and second order moments, by a suitable PCC system of this more general type.

Similarly, the notions of CCC and ID systems can be generalized by the use of mean square predictors.

3.3.2. *Similarities and differences between* CCC *and* ID *systems.* With reference to the specifications (3.27) and (3.30) and the simple cobweb models in illustrations 3.2.2 and 3.2.3 we note three points.

(i) Given an ID system it is always possible to construct a corresponding CCC system with the same coefficient pattern. In general the two systems will differ in the numerical coefficients and in the residuals. The ID system (3.19) to (3.21) is particularly illuminating, inasmuch as it coincides with the corresponding CCC system in illustration 3.2.2 except for the demand relation, where it differs in the coefficient β_1 and the residual ν_t.

By further specification of the ID and CCC systems of form (3.19) to (3.21) it is possible to arrange so that they define one and the same stochastic process and generate one and the same set of artificial time series. As indicated under (3.26) and (3.36) this can be achieved if we take

$$(3.40) \qquad \beta_1 = 0.8, \qquad \beta_1 = \frac{0.6}{\rho},$$

for the CCC and ID system, respectively. In view of the range of possible ρ values (3.40) shows that great numerical differences may occur between CCC and ID systems that refer to one and the same set of given time series.

(ii) The main point in our comparison between CCC and ID systems is that their behavior relations do not have the same operative significance, are not designed to answer the same type of question. This important fact, which belongs under table II, is brought in relief by (3.40). The conclusion is inescapable since the CCC and ID systems behind (3.40) generate the same time series, while they differ in just one of the coefficients, [78].

Subsection 3.2.4 and remark 3.2.2b provide a comparison between the operative significance of CCC versus ID systems. The salient point is that *ceteris paribus* clause A enters in the interpretation of ID systems in the very *posing* of the question of the operative use of the behavior relations. To paraphrase, the passage back and forth from the primary to the reduced form has led the ID approach into a "pitfall of stochastization" in the sense of 2.5.3(iii). The inescapable conclusion is that on the more general clause B the behavior relations are unbiased predictors in CCC systems, whereas they are not so in ID systems. Thus on clause B in the case of logarithmic variables the price elasticity of demand is 0.8 in the CCC system behind the first part of (3.40), whereas the coefficient $(0.6)/\rho$ of the ID system behind the second part does not allow this interpretation.

It lies near at hand to conclude from the foregoing that the ID approach from its very beginning has run into a blind alley, inasmuch as the model is treated as disturbance-free in the transformation back and forth from primary to reduced form. The situation is related to, but not the same as, the unqualified reversal of regression relations, and the numerical differences at issue are of the same order of magnitude, as illustrated by (3.40) and (3.43) to (3.44). The reader who is convinced by this argument may think that the entire ID approach is victim to a pitfall of stochastization and should therefore be abandoned. There is however also the question whether ID systems, granted that they are of no operative use in primary form, may constitute a sound approach when it comes to the reduced form. We shall come back to this question in subsection 5.3.2.

(iii) Subsection 3.2.4 and remark 3.2.2b further show that on *ceteris paribus* clause B the reduced form does not have the same operative significance in CCC and ID systems. In CCC systems the reduced form provides certain conditional relations for the current endogenous variables, and these conditions have an operative meaning on clause B. It will be noted that the conditional relations may be exploited for a causal interpretation of the linear constraints. Thus in illustration 3.2.3 the relation of instantaneous equilibrium (3.21) may be interpreted as a price mechanism that works to clear the market at period t, the clearance referring to expected demand and supply for known price at period $t - 1$.

3.4. *An empirical illustration:* PCC, CCC, *and* ID *systems fitted to the same statistical data.* We quote the demand relations of four multirelation models (each of which is a system of three relations) that have been designed for the watermelon market in the United States from 1930 to 1951. Notations: X is demand, P price, Y income, N population, F freight rate, L cost of living. The variables are logarithmic, and the ratios are logarithms of the nonlogarithmic ratios. (See the quoted papers for accounts of the complete models and for comparative comments.)

Pure causal chain [77]

$$\frac{X}{N} = -0.206\,\frac{P}{L} + 0.430\,\frac{Y}{NL} - 1.088 + \upsilon. \tag{3.41}$$

Conditional casual chain [78]

$$\frac{X}{N} = -0.315\,\frac{P}{L} + 0.509\,\frac{Y}{NL} - 1.077 + \omega. \tag{3.42}$$

Conditional causal chain, second variant [78],

$$\frac{X}{N} = -0.417P + 0.776\,\frac{Y}{N} - 0.563F - 0.882 + \omega. \tag{3.43}$$

Interdependent system [52],

$$\frac{X}{N} = -0.901P + 1.378\,\frac{Y}{N} - 0.614F - 0.126 + \nu. \tag{3.44}$$

Great differences show up in the numerical coefficients. This is not surprising after the preceding review, having seen that the behavior relations do not have the same operative meaning in PCC and CCC systems on the one hand and ID systems on the other. On the logic of the different approaches, (3.41) to (3.43) are estimates of unbiased predictors, whereas (3.44) is not. Thus in the PCC relation (3.41) the coefficient 0.206 is the price elasticity of watermelon demand in the stochastic sense of *ceteris paribus* clause B of 2.5.2 (see also illustration 2.5.2a). The same interpretation applies, in principle, to the coefficients in the CCC relations (3.42) and (3.43). The coefficient 0.901 in the ID relation (3.44), however, is not a demand elasticity in the same stochastic sense.

4. Unbiased predictors and least squares regression

4.1. *Introductory. Prediction versus coefficient estimation.* Once an explanatory model is specified with regard to its intended operative use for predictive and other purposes, the statistical estimation of its coefficients and other parameters is technical matter and therefore, in principle, a noncontroversial issue. In accordance herewith, different estimation techniques are available for PCC and CCC systems on the one hand and ID systems on the other. (See references at the beginning of section 3.) Speaking generally, the techniques give consistent parameter estimates when applied to the models for which they are designed.

This section is a brief excursus on least squares estimation as the principal estimation technique for unbiased predictors. Lemma 4.2 shows that the approach is (a) unbiased under very general conditions, and (b) distribution-free with regard to the residuals. The ensuing comments emphasize the flexibility and wide scope of the approach. Its efficiency is discussed with reference to two aspects of the estimation problem for a linear predictor (2.3): (i) prediction of y in terms of the explanatory variables x_i; (ii) estimation of the coefficients β_i.

The approach of least squares regression is primarily designed for problem (i), and at the same time the procedure gives asymptotically consistent estimates under (ii). In view of the robustness of the approach it would seem that the classic principle of minimum residual variance under (i) acts like a balance wheel to keep the coefficient estimates under (ii) consistent and not far from optimal efficiency.

4.2. *A companion to lemma* 2.2.1.

LEMMA 4.2. *Considering the unbiased linear predictor* (2.3) *with* (2.4), *let a corresponding set of observations constitute the sample S. Suppose that the observed means and second order moments of* $y, x_1, \cdots, x_h$ *tend to the corresponding theoretical means and second order moments as the sample size increases indefinitely*; *in symbols,*

$$\lim m_i = \mu_i, \qquad \lim m_{ik} = \mu_{ik}, \qquad i, k = 0, 1, \cdots, h, \tag{4.1}$$

where the subscript $i = 0$ *refers to* y. *Further, suppose that the theoretical distribution of* $x_1, \cdots, x_h$ *is not confined to a linear subspace* (*that is, the positive semidefinite matrix* $[\mu_{ik}]$, $i, k = 1, \cdots, h$, *should be definite*). *Let*

(4.2) $$y = b_0 + b_1x_1 + \cdots + b_hx_h + u$$

be the least squares regression of y on $x_1, \cdots, x_h$ as observed in S. Then the observed coefficients are consistent in the sense of large samples, that is,

(4.3) $$\lim b_i = \beta_i, \qquad i = 0, 1, \cdots, h.$$

See [78] to [80]. Relations (4.1) and (4.3) make a general frame within which more accurate results can be obtained by suitable specialization. Thus under supplementary and not very restrictive conditions the coefficients b_1 will be asymptotically unbiased, have standard errors that tend to zero, and in the limit be normally distributed; see subsection 4.3.1. For a case where (4.3) is established in the sense of stochastic convergence, see Hurwicz [27].

As always in estimation theory, S is to be regarded as generated from the theoretical model by way of a random sample from a joint probability distribution or from a stochastic process, and the ensuing large sample limits in (4.3) are valid in the same sense as in (4.1), stochastic convergence or convergence with probability one as the case may be. Thus qualified, lemma 4.2 is an immediate consequence of the facts that the observed coefficients b_i are rational functions of the $h + 1$ observed means and the $h(h + 1)/2$ second order moments, and that the theoretical coefficients β_i can be expressed as the same functions of the theoretical means and second order moments.

4.3. *On the wide applicability of least squares regression.* We shall refer to an array of situations where assumption (4.1) is fulfilled and least squares regression, accordingly, gives coefficient estimates that are asymptotically consistent. The survey does not aim at completeness; its purpose is to demonstrate the flexibility and robustness of the regression approach by a varied set of illustrations.

4.3.1. *Theoretical aspects.*

(i) *Nonstochastic versus stochastic variables as explanatory variables.* The treatment of these two general situations belongs to the fundamentals of regression theory [10]. They are also known as specification by fixed numbers versus random variables, or Gauss-Fisher versus Galton-Yule specification [82].

(ii) *Exogenous versus lagged endogenous variables as explanatory variables.* The treatment under (i) usually refers to the case when the explanatory variables are exogenous. The introduction of lagged endogenous variables carries over to the theory of stochastic processes. In early applications of stochastic processes, more or less intuitive procedures in line with lemma 4.2 are abundant (for example, see [66], section 29, with discussion of pioneering work by Yule [88] and Walker [62]). If nothing more than large sample consistency is aimed at, the principle of ergodicity makes a general argument that assumption (4.1) is satisfied ([78], compare also [66], section 12, with reference to Khinchin [28]). Mann and Wald [40], applying maximum likelihood methods, were first to obtain standard errors of the coefficient estimates, results that in recent years have been extended in many directions [63], [64], [22].

(iii) *Uni- versus multirelation models.* Lemma 4.2 applies to an unbiased predictor, whether this is a unirelation model or forms part of a system of relations. Specifically, the results under (ii) obtained by the use of the ergodicity principle extend to multirelation models [78].

Hence for FCC and CCC systems, least squares regression as applied to the primary form will under general conditions give consistent estimates for its coefficients. Coefficient estimates for the reduced form can then be obtained by formal transformation. For PCC systems this technique does not make explicit use of the hypothesis (3.7) that the reduced form relations are unbiased estimators. This information may instead be exploited to test the coefficient estimates with regard to the validity of (3.7), that is, to explore whether (3.7) can be accepted as an approximation.

(iv) *Least squares regression and analysis of variance.* The estimation aspects of analysis of variance are covered by the regression approach (2.3), inasmuch as one or more of the explanatory variables x_i may be dummy $0 - 1$ variables that indicate control group versus treatment group, block effects, and so forth. This is a typical case for the Gauss-Fisher approach of nonstochastic explanatory variables. On the close relation between regression analysis and analysis of variance, see, for example, [65].

(v) *Conditional regression.* In this case one or more coefficients are known by ancillary information. The device is simply to introduce these values into the relation and estimate the remaining coefficients on the least squares principle [67], [68], [82].

(vi) *Mutually independent versus interdependent residuals.* The classic results under (i) require that the residuals are normal and mutually independent. Then, as is well known, the coefficient estimates are of optimal efficiency. It is only recently that the empirical assessment of explanatory relations has been explored with regard to least squares regression as an estimation technique that with regard to the residuals is distribution free, and in particular parameter free with regard to residual intercorrelation. The first results in this direction refer to time series relations with autocorrelated residuals, unbiasedness, and generalized standard errors being established in the case when all explanatory variables are exogenous [72]. Corresponding results have been obtained by Zellner [89] for the case of multirelation models. Still more generally, Lyttkens [37] has derived standard errors for the case when the explanatory variables may include lagged endogenous variables. The question about the efficiency of this type of estimates lies deeper and has been explored by the use of spectral analysis [20], [22], [63]. Under fairly general conditions the estimates of optimal efficiency have standard errors that for large samples are proportional to those of the least squares estimates. In some cases the proportional factor tends to the unit, showing that the least squares coefficient estimates are of optimal efficiency in the large sample sense. Notably, this is so in the search for hidden periodicities and in the fitting of polynomial trends.

(vii) *Homoscedastic versus heteroscedastic residuals.* The classic results under (i) further require that the residuals have the same variance for all values of the explanatory variables. Theil [53] has established unbiasedness and given standard errors for the least squares coefficient estimates in the case when the residuals do not have constant variance.

4.3.2. *Least squares regression in applied work.* The wide range of theoretical assumptions under 4.3.1(i) to (vii) meet the requirements of a variety of situations in applied work. We note:

(i) *Experimental versus nonexperimental situations.* There is a close affinity between this dualism and that under 4.3.1(i). In controlled experiments an explanatory variable—a typical case being the stimulus levels that the experimenter varies systematically—is usually treated as nonstochastic, in nonexperimental models usually as a random variable. The special devices under 4.3.1(ii), (v), (vi) mainly come to use in the treatment of nonexperimental data, while those under 4.3.1(iii), (vii) are of relevance both in experimental and nonexperimental situations.

(ii) *Cross section versus time series data.* This distinction is primarily of relevance in nonexperimental situations. The estimation technique for time series data belong under the theory of stochastic processes, and there are the important case distinctions under 4.3.1(ii) and (vi).

(iii) *Causal analysis versus model adaptation for strategic purposes.* In PCC systems the relations are *autonomous* in the sense that the behavior relation of one or more decision taking units may be altered without this necessarily requiring any change in the other behavior relations. Hence if the system has been constructed by way of causal analysis of actual observations a behavior relation may be replaced by a *strategic relation,* that is, a relation that refers to the future and provides a planned scheme of behavior that serves some specified purpose, for example, profit maximization. In the system so modified the relations that have not been changed may still be estimated by least squares regression, whereas the strategic relation wholly or partly is a theoretical construct. The mixed system may be used for exploring how the postulated strategy would work in practice, by calculating the expected future development from suitable sets of initial conditions.

ILLUSTRATION 4.3.2. In the PCC system (3.8) to (3.10), the price mechanism (3.10), which is designed for a market with free competition, may be replaced by some specified monopolistic price strategy. Such a model seems to be of relevance for the present agricultural situation in several countries. By marketing organizations and/or political pressure the price mechanism for agricultural products has shifted in monopolistic direction. At the same time, owing to the large number of autonomous consumers and producers, the demand and supply functions have not been much affected, with the result that market price is higher than the equilibrium price, as shows up in the growing stocks of agricultural products.

This illustration is not given because of the subject matter conclusion, which is quite plain. The emphasis is on the negative comment that although the argument is simple it is difficult to cover by a model of closed form. In situations like this the approach of strategic relations comes to the fore as a theoretical supplement to models where all relations are jointly theoretical-empirical constructs.

5. Why econometrics?

It is a striking feature of the area under review that pioneering developments have been made in econometric setting (as in 1.1 we refer to the two key issues of "choice of regression" and "interdependent systems") whereas the ensuing models are of much wider scope. Their field of potential applications extends all over the socioeconomic sciences, and, still wider, over the entire realm of nonexperimental model building. In 5.1 we shall turn to these matters by way of two slightly provocative questions. Then in 5.2 we shall adduce some comments on the place of PCC, CCC and ID systems in scientific model construction in general, and in 5.3 sum up towards a comparative appraisal of the models.

5.1. *Stochastization in retrospect.* First, how is it that econometrics, this modern offspring of "the dismal science," has been a melting pot for developments of such fundamental nature and far-reaching scope?

Second, how is it that there has been so much obscurity around the key problems about choice of regression and interdependent systems, and how is it that the debate has been such a long-drawn affair?

The first question is rhetorical. Economics is much more developed than other behavioral sciences, whether we look at the current flow and accumulated stock of statistical data, or at the advanced nature and the variety of theoretical approaches, not to speak of the institutional, personal and material resources. Hence it is no wonder that econometrics makes the playground *hors concours* for innovations and spearheads advancements in the behavioral sciences.

As to the second question, several features have combined to obscure the issues, and the debate has only gradually dug down to the root of the troubles. For one thing, model construction was well developed in economics long before the first attempts were made to exploit statistical data and techniques for the purpose, and so the stochastization has been hampered by forces of inertia. For another thing, a satisfactory treatment of stochastic models was hardly possible before 1933 when Kolmogorov [32] strengthened the mathematical foundations of probability theory. Third, the stochastization has led to genuine innovations in the large arsenal of scientific models, innovations that have repercussions at all stages of the model construction: the specification of basic hypotheses, the statistical methods for parameter estimation and hypothesis testing, and the various procedures for operative use of the model. At the same time the stochastization has impacts on the subject matter theory that gives substance to the model construction, and last, but not least in being bewildering, the theory of knowledge comes to the fore in the causal aspects of the stochastization.

Tables I and II provide a unifying view of the situation, and locate the main stumbling blocks to *the stochastic treatment of models that in disturbance-free form serve two different purposes.* In the choice of regression the operative procedures at issue are direct and reverse prediction; in multirelation models they are prediction from the model in primary and reduced form. Hence there is rather just one main stumbling block in the picture: the pitfall of stochastization to avoid both in "choice of regression" and "interdependent systems" lies in a symmetric treatment of explanatory variables and variables to be explained. To obtain logical consistency in the model construction the stochastization forces us to treat the explanatory relations as *directed* relations. The simplest way of doing this is to specify the relations as unbiased predictors.

Let us recapitulate very briefly how the econometric developments have brought clarification and progress in the general area of stochastic model construction for explanatory purposes in nonexperimental situations.

As to the choice of regression in analyzing two variables x, y the formal symmetry between the regression of y upon x and of x upon y was emphasized from the very beginning of correlation theory. In demand analysis the argument about symmetry and the ensuing indeterminacy of demand relations was voiced by Mackeprang [39] as early as 1906. The argument was vigorously stressed by Frisch [18]. In his standard work on demand analysis Schultz [48] as late as 1938 has no general answer to the problem, and therefore calculates two sets of demand relations, one being the (multiple) regression of demand on price (and other variables), the second the regression of price on demand.

Benini [5] in 1907 is the first on record to have estimated demand relations consistently by what today is the current technique, namely the regression of demand on price. Benini used the asymmetric approach without discussing its rationale. Tinbergen [58] gave the problem a more articulate treatment in the late 1930's, the asymmetry being based on a cause-effect interpretation of demand relations and other explanatory relations. More recently the casual argument has been further strengthened, under reference to the operative use of explanatory relations [68], [82], [76], and with specification in terms of unbiased predictors [78], [79], [80].

In the 1920's and 1930's part of the obscurity around the choice of regression was due to a failure to distinguish between demand and supply as different observables. The classic paper of E. J. Working [83] sheds light on the situation, without however making full distinction between observed demand and supply. A source of confusion in this connection is the cobweb theory for demand-supply balance. As illustrated by (3.1), the cobweb approach is quasi-dynamic and thereby of historical importance as a forerunner to completely dynamic systems, but at the same time the approach made for undue persistency of the short cut of dealing with demand and supply as the same observable.

As to multirelation models, the 1930's brought several important developments. Yule [88] in 1927 introduced the autoregressive stochastic processes, a unirelation model that is dynamic and stochastic. Yule's fundamental innovation entered econometrics with Frisch [17], and was further developed by Tinbergen in his work with FCC systems [57] to [60]. Tinbergen's treatment of the rationale for his completely dynamic approach was to some extent intuitive. For one thing, this was so in his use of least squares regression for the parameter estimation, for another in the substitutive inference from the behavior relations, whereas their use for direct inference was supported by a causal interpretation. (In retrospect, an intuition beyond praise!)

The ID approach was introduced as a general device for the stochastization of disturbance-free models. The basic illustrations in the early works included cobweb models and systems of Keynesian type [23], [33], [19]. Haavelmo in 1943 opened up the approach with a wholesale dismissal of least squares regression for the estimation of multirelation models [23]. PCC and ID systems were soon taken up for comparative analysis [7], [71]; the difference in approach was emphasized, and least squares regression was shown to be appropriate for the estimation of a general class of PCC systems. In the next round the general scope of PCC systems was established [70], [73], [75].

Perhaps it is in the casual aspects that the intrinsic differences in approach have shown up at their sharpest. In PCC systems the behavior relations allow a straightforward causal interpretation, with the (current) endogenous variables for effects and the corresponding explanatory variables as causal factors, [82], [74], [77]. In the ID approach, owing to the symmetric treatment of the current endogenous variables, a causal interpretation of the behavior relations is available only in a highly attenuated sense, [49], [51], [77].

The differences in the causal interpretation, although highly relevant, do not go to the bottom of the problem [78]. As we know from table II, the parting of the ways between PCC and ID systems is at the basic level of how to design the models for stochastic inference. The ensuing implications with regard to the rationale of the model construction are in agreement with the conclusions based on causal arguments, and at the same time the conclusions are carried further. Three points will be noted. (a) In designing the models in terms of unbiased predictors, CCC systems enter as an alternative to ID systems. For one thing, this dualism makes for a more balanced comparison since the operative use of PCC systems combines the performances of CCC and ID systems. (b) The high parameter values obtained in the ID approach have the nature of a bias, a pitfall of stochastization due to a symmetric treatment of variables in the transformation from the primary to the reduced form. Model (3.19) to (3.21) is a typical case: the representation in the second part of (3.24) involves a reversal of relation (3.19), and as we know from table I the reversal makes it impossible to specify both the second part of (3.24) and (3.19) as unbiased predictors. As illustrated by (3.40) and (3.43) to (3.44) the ensuing numerical differences are often quite striking. (c) The *identification* problem, this cumbersome feature of the ID approach, does not arise in PCC and CCC systems. It is a normal and frequent situation that a multirelation model involves more explanatory variables than variables to be explained. The "conjugate" theorem 3.2.4 shows that ID systems then are overidentified. This awkward situation, I think, shows better than anything else that the ID approach from the very beginning suffers from a fundamental construction error. If the alternative approach of CCC systems has not been discussed earlier it is perhaps partly because it does not provide a basis for predictive inference from the reduced form, partly because the mutual independence between lagged residuals involves a nontrivial technicality (the residuals in (3.19) are generated as moving averages of mutually independent random variables, averages where all autocorrelations come out as zero; see [78]).

5.2. *On the place of* PCC, CCC, *and* ID *systems in scientific model construction in general.*

(i) The models under review are designed as *explanatory* approaches in *nonexperimental* situations.

With regard to ends and means we distinguish between four broad sectors of applied statistics, according as the models are designed for descriptive versus explanatory purposes, on the basis of controlled experiments versus nonexperimental data [76]. Explanatory models involve *directed* relationships by way of explicit or implicit systems; that is, the variables (or the variation in the variables) to be explained are expressed in terms of (the variations in) the explanatory variables.

(ii) The stochastization of multirelation models has led to genuine innovations in the general area of scientific model construction. Speaking generally, the innovations bring in relief that a stochastic model is less flexible for purposes of predictive inference than a corresponding disturbance-free model. This is in particular so when it comes to the dualism between cross section inference and time series inference.

It is instructive to compare with predictive inference from differential equation system. Let us consider one of first order,

$$\frac{dy}{dt} = f(x, y), \tag{5.1}$$

where y is a vector of endogenous variables and x one of exogenous variables. An equivalent form of the system is

$$y_t = y_{t-at} + f(y_{t-dt}, x_{t-dt})\, dt, \tag{5.2}$$

which gives the endogenous variables at t explicitly in terms of lagged variables (at $t - dt$, and for systems of higher order at $t - 2dt$, $t - 3dt$, $\cdots$). The point of this comment is that system (5.2) involves no cross section relation between the endogenous variables at time t. In this respect, quite apart from the stochastization, the models under review are innovations in dynamic model construction, and in particular so PCC systems.

The problems at issue thus can be avoided, in principle, by dealing with time as a continuous variable [47], [44]. If however the empirical data force us to deal with time as a discrete variable, as they usually do in econometrics, this device does not really meet the situation—the problems come back in equivalent form by way of aggregation problems; see [7], [54]. For a treatment that emphasizes the causal aspects of aggregation see [68], [82].

Next, some general comments on stochastic models with regard to cross section and/or time series prediction.

(a) PCC systems, as we know from table II, are designed both for cross section prediction (primary form) and time series prediction (reduced form). See also figure 2.

(b) If we are only interested in prediction in a literal sense, from the past and present into the future, not in conditional cross section prediction (see 2.6), there is no point in designing the model for inference from the primary form. In such case, in other words, the cross section interrelations are a source of information that can, at best, serve auxiliary purposes. This seems to be a main argument in support of the ID approach [34], [26].

(c) When the distances in time between different cross sections are infinitesimal, as in (5.2), it does not bring any advantage to generalize the system by introducing relations between the variables at t. The situation is different, however, when the distances are genuine time spans, such as months, quarters, or years, as they usually are in the statistical data of economics. The introduction of cross section relations then makes it possible to squeeze more information from the observed data. One way to put it is that in differential equations it is assumed that the variables influence each other at the beginning of each (infinitesimal) period, the influences within the period being ignored, whereas the PCC and CCC systems allow for influences also within a period. In this respect PCC and CCC systems involve an approximation, inasmuch as it is ignored that cross section influences require time. The rationale of this type of approximation is that the influences within a cross section may be rapid relative to the influences between variables that refer to consecutive time points. For example, the assumption behind the cross section relation (3.8) is that the consumers in their purchases react immediately upon the existing price, whereas the feedback of demand on price takes place with a lag, price being adjusted according to changes in stocks [7], [71], [77].

(iii) Table II shows that PCC, CCC, and ID systems are a basic triad of models, each covering one of the three possibilities that exist with regard to unbiased prediction from the primary and/or the reduced form: prediction from both forms (PCC systems), only from the primary form (CCC systems), and only from the reduced form (ID systems).

Tables I and II have the irrevocable nature of theorems in probability theory, and this is in particular so for the parting of the ways shown in table II. The theorems make for a systematic coordination between the operative use of explanatory models and their stochastic design. Specifically, the design in terms of unbiased predictors involves a logical sharpening of the construction principles for explanatory models in nonexperimental situations.

(iv) The stochastization of explanatory models by way of unbiased predictors places the model builder under the onus of specifying the operative use of the model in terms of *directed* relations, or, to paraphrase, he must specify whether and in what direction the relations of the model, primary or derived, are designed for purposes of prediction.

There are several different situations in which directed relations and unbiased predictors enter in explanatory models. We note

(a) *Stimulus-response relations in controlled experiments.* Here the direction of the relation is indicated by the very design of the experiment.

(b) *Behavior relations*. Here the direction may be indicated by a causal interpretation, the behavior to be explained being assumed to be influenced by a catalogue of explanatory factors. A definition of causal relationships that covers the hypotheses both under (a) and (b) is that *causal relations are stimulus-response relationships in real or fictitious experiments*. (For a discussion of causal concepts from the points of view of general theory of knowledge, see [81]; see also [82], [74], [77].)

(c) *Lagged relationships*. When a current variable is explained in terms of lagged variables, the direction of the relation coincides with the flow of the time variable.

One way to put the problems under review is that it may be desirable to make operative use of relations both under (b) and (c), and that the requisite "directionizing" may be in irrevocable conflict.

(v) To "directionize" explanatory models makes a general development program for subject matter work in the realm of nonexperimental sciences. As regards economic theory, the present state of things leaves much to be desired in this respect. A purely formal treatment of the models dominates the picture even at the highest levels of economic literature, as witnessed by Tinbergen's recent book on economic policy [61] or Allen's advanced exposition of mathematical economics [1]. At the same time the verbal parts of economic theory often provide material that suffices to directionize the model, although such material at present usually is left aside as not being essential for the model construction. The ensuing stochastization may or may not necessitate formal changes in the model. An approach where it lies close at hand to "directionize" is the input-output models of W. Leontief [35]. Here the stochastization gives a CCC system if we take the final demands as the variables to be explained and the total inputs as explanatory variables [80].

(vi) *Errors in equations versus errors in variables*. Perhaps the best known device in the approach of errors in variables is Frisch's bunch map analysis [18]. For a recent review of the area see [12]. For treatments that stress the casual aspects of the situation, see [19], [82].

The crucial question about symmetric treatment of variables here comes up in yet a different context. PCC, CCC, and ID systems belong under "errors in equations"; more precisely, the residual disturbance of an explanatory relation serves to cover the effect of influencing variables that have not been taken into explicit account in the relation. In models that involve errors in variables the typical hypothesis is that the variables, if their measurement errors were removed, satisfy a disturbance-free relationship, that is, a relation that allows a symmetric treatment of the variables. A few comments:

(a) From a formal point of view it makes no difference whether the residual v of an explanatory relation (2.3) is specified as (1) a measurement error in the variables y to be explained, or (b) as an error in the relation, an error due to causal factors not explicitly taken into account. Specification (2) is however much more general, as is clear from the fact that on assumption (1) the relation between $y - v$ and $x_1, \cdots, x_h$ may be operated with in the same way as a disturbance-free relation, but not so on assumption (2).

The hypothesis of errors in equations versus errors in variables are complementary approaches, not competitive. To paraphrase, the models under review are designed for situations where measurement errors are negligible relative to the effect of disturbance variables. In situations where both types of errors are of relevance, a synthesis of the two approaches is called for.

(b) The approach of errors in variables has a clear-cut *raison d'être* in situations where "the disturbance-free relation between the error-free variables" constitutes a real phenomenon whose exact measurement in principle is a matter of observation technique, and where "the variables" are coordinates of measurement in the space in which the phenomenon is observed. The coordinate axes can then be placed arbitrarily and the coordinates should therefore allow a symmetric treatment. Cases in point are (1) the position of a steel rod in a plane; (2) the connecting line between the triangulation points in a system of geodetic measurements; (3) astronomical data on the orbit of a planet or a sputnik. Such situations not only require a symmetric treatment of the variable coordinates; estimation methods are called for that are invariant to linear transformations in the variables-coordinates.

(c) Behavior relations do not belong under (b). First, when representing the relation in a graph the coordinates are the different scales of measurement of the variables under analysis, so there is no arbitrariness in the choice of coordinates. Second, there is no reason to believe that if the variables could be freed from technical measurement errors the ensuing error-free variables would satisfy a disturbance-free relation.

The literature on ID systems is sometimes ambivalent on the fundamental question of whether the relations in primary form are to be interpreted as subject to measurement errors in the variables to be explained or to errors in the relations. The review in subsection 3.2 deals with multirelation models as subject to errors in equations. This approach is adopted because it is so much more general, and in the econometric area so much more natural. There are in particular two weak points in the approach of measurement errors. What is the rationale for assuming that the current endogenous variables, if freed from measurement errors, are subject to a disturbance-free relation? Second, what is the rationale for assuming that there are measurement errors in those endogenous variables that are explained by behavioral relations, but not in the remaining endogenous variables, and not in the exogenous variables?

(vii) *Sewall Wright's approach of path coefficients.* The econometric models under review have interesting parallels with the models of genetic linkage which were developed by S. Wright in the 1920's and 1930's. The fundamental references are [85] and [86]; see also [87] for references to recent work by J. Tukey (1954) and M. E. Turner and C. D. Stevens (1959). Both approaches have primarily been designed for a special field of applications and both are, in point of principle, of general scope in nonexperimental situations. Wright's models specify the flow of causation in a network of consecutive generations; hence they are of more general form than the econometric models with their emphasis on stationary processes in a grid of equidistant time points. The key device in Wright's approach is to assess genetic linkage by iterated substitutions in a chain of regression relations, an elimination procedure that is analogous to the substitutions that in PCC systems lead from the primary form to the reduced form. In situations where the compound path coefficients are overidentified there is a parting of the ways between "errors in equations" and "errors in variables." Here as always Wright makes use of the approach of "errors in equations," and obtains unique estimates for the overidentified coefficients on the basis of supplementary assumptions. Turner and Stevens (1959) deal with the overidentified case by an adaptation of the approach of "errors in variables," under reference to the statistical techniques of ID systems.

Lemma 2.4.4 on substitutive inference lends itself for exploring the rationale of the approach of path coefficients. The genetic models will be adaptations of (3.4), (3.16), (3.27), (3.30), and (3.31); to this end we must reinterpret t as referring to consecutive generations, and each variable y_t or z_t will in general be subject to several observations. In a broad class of models the ensuing results are in agreement with Wright's treatment of compound path coefficients. As likewise might be expected it turns out that the approach of errors in variables give parameter estimates that may be very different from those obtained by Wright's approach. This last result is parallel to the situation in ID versus CCC systems, an analogy which however is of limited relevance, for according to genetic theory the linkage is between nonobserved genotypes, not between observed phenotypes, and therefore the approach of errors in variables is less farfetched than in the econometric models.

5.3. *Toward a comparative appraisal of* PCC, CCC, *and* ID *systems*. In section 3.2 the three approaches are specified as stochastic model frames of a general nature. In applied work the frames are filled in by subject matter content, and the parameters are assessed on an empirical basis. The following comparison focuses on the general framework as such. A comparison and appraisal of applied work falls outside the scope of our review. Empirical results are referred to only for illustration and to stress the need for more exploratory research.

5.3.1. PCC *systems*. The argument summed up in table II and section 5.2 speaks strongly in favor of the PCC approach. PCC systems are designed for unbiased prediction both in primary and reduced form, just as with disturbance-free multirelation models, only that the predictions give expected instead of exact values. Their twofold use does not involve any restriction in scope; in point of principle, they are applicable to any set of observed data, [70], [73], [75].

From the point of view of static versus dynamic analysis PCC systems are designed as completely dynamic models. The reduced form provides kinematic inference in showing how the system develops from one period to the next, while the behavior relations of the primary form specify the driving forces of the system, thereby supplementing the kinematic inference so as to make full-fledged dynamics [16], [58], [82], [78], [81].

5.3.2. CCC *versus* ID *systems.* These models may perhaps best be interpreted as mixed static-dynamic approaches, which in principle cover the entire range between purely static models at the one extreme and purely dynamic systems at the other. Static elements in the form of equilibrium conditions and other constraints being introduced in the primary form, this makes a formal generalization relative to PCC systems, but the generalization is bought at the price of a drastic cut in aspiration levels. It is no longer possible to make unbiased prediction both from the primary form and the reduced form. CCC systems are designed for prediction from the primary form, ID systems from the reduced form. What can be said about the choice between the two approaches?

(i) ID systems aim at unbiased prediction from the reduced form, and renounce from unbiased prediction on the basis of the primary form. The aim is a natural extension of the quasi-dynamic approach of the disturbance-free models of the cobweb theory around 1930, but in setting up this aim for the stochasticized model the ID systems have run into a dilemma. When the reduced form is specified as unbiased predictors, then as we know from table II the primary form cannot in general be used for unbiased prediction. But when the behavior relations of the primary form are not unbiased predictors, will then the ensuing reduced form have any meaning at all?

As pointed out by several commentators, the reduced form of ID systems has no obvious rationale. Thereby the entire approach is questioned. To begin with, there is the question of whether the relations of the reduced form have the appropriate catalogues of explanatory variables. The best that can be said, I think, is that if each behavior relation of the primary form has a realistic catalogue of explanatory variables, the catalogues of the reduced form should also be realistic. This argument in favor of the approach is valid in disturbance-free models, but it is open to doubt in a stochasticized model if the primary form, as happens with ID systems, does not constitute an unbiased predictor. The rationale of the approach is further blurred if the ID system is *overidentified,* which is the usual situation in practice. In this case there are more relations than unknowns to determine the coefficients of the primary form when transforming back from the reduced form. For overidentified systems the estimation techniques of Theil [55] and Basmann [3] make combined use of the primary form and the reduced form. In consequence, the ensuing parameter estimates give a system where the relations neither in primary form nor in reduced form are unbiased predictors.

(ii) CCC systems are designed for situations where a purely dynamic approach for some reason or other is not feasible or desirable. They serve what may be regarded as a minimum program for the operative use of the models under review, namely, unbiased prediction from the behavior relations. The relations are knit together with loose ties; the emphasis is more upon the separate relations than upon the relations as forming a system. Depending upon the design, CCC systems in primary form may yield both cross section inference and time series inference. In this connection arise estimation problems of new types that it would carry too far to take up in this review.

5.3.3. *The empirical outlook.* On the applied side much more experience is needed about the reach and limitation of the models under review, but the outlook on the existing material is in accordance with the theoretical argument summed up in 5.3.1 and 5.3.2. The applied work with ID systems has given poor predictions and is on the whole disappointing, to quote from a recent authoritative address of the initiator of the approach [24]. For PCC and CCC systems the need for more empirical work is particularly urgent. After Tinbergen's pioneering applications of PCC systems in the 1930's no further applied work with this type of model was reported for a long while, the scene being dominated by the ID approach. The applications of causal chain systems recently reported are promising, and give increased weight to the need for continued work in this direction. (At least a partial dismissal of the ID approach is embodied in the first macroeconomic model to be in continuous use as a forecasting instrument, see C. Clark's contribution to the discussion on "The present position of econometrics" [46], pp. 287–288. See also [9] and [77].)

6. Can causal relations and their direction be assessed by purely statistical devices?

The question may sound heretic. It is a contention of old standing that the specification of causal hypotheses belongs to subject matter theory, and is not a matter of statistical technique. The traditional argument refers on the one hand to the asymmetry between cause and effect in causal hypotheses, on the other to the symmetry between the variables in scatter diagrams, in correlation coefficients, in pairs of regression lines. Presence of a flaw in this argument was initially revealed by Holbrook Working [84] in an empirical statistical study of the causal connection between two variables.

The flaw of the argument is that while there is perfect statistical symmetry in correlation coefficients and the like when only two variables x, y enter the picture, the introduction of a third variable z may give rise to asymmetries between x and y relative to z, asymmetries that under favorable circumstances may give a clue to causal inference. In subsections 6.1.1 and 6.1.2 we shall consider two situations of this type that arise in uni- and multirelation models, respectively. Section 6.3 gives an empirical illustration, and comments upon the rationale of the approach.

We note in advance that the formal aspects of the following arguments are quite simple. The intricacies and the open questions lie in what causal conclusions can be drawn, questions that partly cross the border to philosophy and the theory of knowledge.

6.1. *Two theorems with bearing upon causal directions.* We shall discern between three hypotheses,

$$x \rightarrow y,\ x \leftarrow y,\ \text{neither } x \rightarrow y \text{ nor } x \leftarrow y, \tag{6.1}$$

where the two first read: x influences y; y influences x. Cases under the third are, for example, that x and y influence each other mutually, or that x and y do not influence each other but are both influenced by a third variable z; in symbols,

$$x \leftrightarrow y,\ x \leftarrow z \rightarrow y. \tag{6.2}$$

Our theorems deal with situations where the three cases (6.1) give rise to statistical asymmetries. Hypotheses (6.1) and (6.2) being a matter of model specification, we shall disregard sampling aspects and formulate the theorems in population parlance so as to apply to large samples.

6.1.1. *A unirelation theorem.* The idea behind the following simple theorem is that if the regression of y on x turns out to have a stable slope when the observed sample is broken up in subsamples according to some stratifying variable z, while the regression of x on y has varying slope in the subsamples, this is an indication in favor of the first hypothesis in (6.1).

THEOREM 6.1.1. *Let*

$$y = \alpha + \beta x + v \tag{6.3}$$

be an unbiased predictor in a population P. Let

$$y = \alpha_i + \beta x + v_i, \qquad i = 1, \cdots, k, \tag{6.4}$$

be unbiased predictors in subpopulations P_i obtained by stratifying P after some variable z. Let

$$x = \alpha_i^* + \beta_i^* y + v_i^*, \qquad i = 1, \cdots, k, \tag{6.5}$$

be the least squares regression of x on y in the subpopulations P_i. Finally, let $\rho_i = \rho_i(x, y)$ denote the correlation coefficient of x, y in P_i. Then in order that we should have

$$\beta_1^* = \cdots = \beta_k^* \tag{6.6}$$

it is necessary and sufficient that

$$\rho_1 = \cdots = \rho_k. \tag{6.7}$$

The proof is immediate, for in obvious notation we have

$$\beta = \frac{\sigma_i(y)}{\sigma_i(x)} \rho_i, \qquad \beta_i^* = \frac{\sigma_i(x)}{\sigma_i(y)} \rho_i, \tag{6.8}$$

which gives

$$\beta_i^* = \frac{\rho_i^2}{\beta}, \qquad i = 1, \cdots, k. \tag{6.9}$$

Theorem 6.1.1 is an adaptation of a device due to H. Working [84]. Considering two causal hypotheses about price features of wheat, namely $y \rightarrow x_1$ and $y \rightarrow x_2$, where x_1 and x_2 are current prices for delivery of wheat in May (old crop) and

July (new crop) respectively, and $y = x_2 - x_1$ is the price spread, Working formed the regressions of x_1 upon y and of x_2 upon y on the basis of weekly data, stratified into monthly subsamples, and discriminated between the two hypotheses by comparing the stability of the two regressions in the subsamples. In the first part of his study Working discussed the results of the same device as applied to two hypotheses of type $x_2 - x_1 \rightarrow x_1$ and $x_1 \rightarrow x_2 - x_1$.

The fact that Working's device was presented more than 25 years ago and since then has been unnoticed in spite of its eminent potential importance is a striking illustration of how causal arguments have been pushed to the side in the 1940's and 1950's. Independently of Working's approach, statistical methods for the analysis of causal patterns have come to the fore in recent years (see [13]). I am greatly indebted to Dr. Working for showing me his device in 1957 and for the ensuing exchange of thoughts by correspondence and discussions.

Theorem 6.1.1 can be varied and extended in several ways. For example, we may replace models (6.3) and (6.4) by

$$y = \alpha + \beta x + \gamma z + v, \qquad y = \alpha_i + \beta x + \gamma_i z + v_i, \qquad i = 1, \cdots, k, \tag{6.10}$$

assuming

$$\sigma(v_1) = \cdots = \sigma(v_k) \tag{6.11}$$

and letting

$$x = \alpha_i^* + \beta_i^* y + \gamma_i^* + v_i^*, \qquad i = 1, \cdots, k, \tag{6.12}$$

denote the least squares regression of x on y and z in $P_1, \cdots, P_k$. Then $\beta_1^* = \cdots = \beta_i^*$ if and only if

$$\gamma_1 = \cdots = \gamma_k. \tag{6.13}$$

6.1.2. *A theorem on PCC systems.* The theorem refers to systems of a very simple type where there are only two relations, and, accordingly, only two possibilities (6.1) for the causal ordering of the current endogenous variables. The difference in causal ordering implies the statistical asymmetry (6.16) relative to the third variable in play.

THEOREM 6.1.2. *Let two PCC systems be defined in accordance with* 3.2.1,

$$x = \alpha_0 + \beta z + v, \qquad y = \alpha_1 + \gamma x + \omega, \tag{6.14}$$

$$y = \alpha_0 + \beta z + v, \qquad x = \alpha_1 + \gamma y + \omega, \tag{6.15}$$

thus letting x, y denote current endogenous variables and z an exogenous or lagged endogenous variable. Then in the model (6.14) *the correlation between x, z is numerically greater than the correlation between y, z; and conversely in the model* (6.15). *In symbols,*

$$|\rho(x, z)| > |\rho(y, z)|, \qquad |\rho(x, z)| < |\rho(y, z)|. \tag{6.16}$$

From the construction principles of PCC systems (see section 3.2.1) we know that both relations of system (6.14) are unbiased predictors, and that the relations form a chain where each relation specifies the relevant variables. Hence system (6.14) implies $E(x|z) = \alpha_0 + \beta z$ and $E(y|x, z) = \alpha_1 + \gamma x$, and similarly for system (6.15). These relations make it possible to apply the substitution lemma 2.4.4.

To prove the first part of (6.16), we may without loss of generality take the variables to have zero mean and unit variance. Then, on substitution in system (6.14), we obtain

$$y = \beta\gamma z + \omega + \gamma v \tag{6.17}$$

with $E(y|z) = \beta\gamma z$, and

$$\beta = \rho(x, z), \qquad \gamma = \rho(x, y), \qquad \beta\gamma = \rho(y, z), \tag{6.18}$$

which implies the first part of (6.16). For the second part the proof is similar.

6.2. *Empirical illustration. Comments.*

6.2.1. The following illustration has the limited purpose of serving as background for the subsequent comments.

Illustration of theorem 6.1.1. The calculations reported in table III are based on a fraction of the sample survey of family expenditures in Sweden 1913–1914. The data refer to 141 families, each with husband, wife, and two children. The notations are as in (6.3) and (6.4), with a shift to Roman letters to denote sample values. The variables of the regressions are x = log *expenditure on clothing*, and y = log *income*. The stratification is regional, the subsamples referring to the three cities specified in the first column.

TABLE III

REGRESSION VARIABILITY OF FAMILY EXPENDITURE

City	n	b_i	b_i^*	r_i	$1/b_i^*$
Uppsala	45	1.103	0.464	0.691	2.155
Eskilstuna	44	1.110	0.339	0.601	2.950
Hälsingborg	52	1.267	0.258	0.572	3.868
Aggregate	141	1.152	0.327	0.704	3.058

On the logic of the approach we want to compare the variability of b_1, b_2, b_3 relative to b_1^*, b_2^*, b_3^*. A simple device for the purpose is to compare the Charlier coefficients of variability. In customary symbols, this gives

$$\frac{s(b_i)}{\bar{b}_i} = 9.3\%, \qquad \frac{sb_i^*}{\bar{b}_i^*} = 23.4\%. \tag{6.19}$$

The percentages differ in the expected direction, for it is a basic hypothesis in economic theory that consumer demand is causally influenced by consumer income, not—or not so much—the other way around, and according to theorem 6.1.1 this should make the variability smaller in the column of b_i than in the column of b_i^*.

6.2.2. *Comments.* We have seen that the causal hypotheses in 6.1.1 and 6.1.2 give rise to statistical asymmetries, namely the stability versus instability of the regression slopes in (6.4) and (6.5) and the two inequalities in (6.16). Hence the asymmetries can serve as statistical tests for discriminating between the hypotheses at issue.

Thus far everything is in line with current procedures of hypothesis testing. The causal directions are part of the model, with implications that are liable to empirical tests just as any other part of the hypothetical setup. The question we have posed, however, goes farther: the question is whether the statistical asymmetries can serve as independent vehicles in the search for causal relations, and in particular for the assessment of causal directions. This aspect of the matter was keenly debated at the Berkeley Symposium. Specifically, I am indebted to Professor J. Tukey for his constructive comments. Insofar as any divergencies came up in the debate, it is my understanding that these partly lie on the plane of causal terminology, partly refer to questions which essentially depend upon future experience about the reach and limitation of the purely statistical devices and which therefore at the present stage must be kept open.

(i) The devices at issue are subject to several limitations. For one thing, they require that the correlations in play must not be too small, nor too large. This is clearly so for the correlations between x, y in both theorems. Further we note that the stratifying variable z in (6.4) and (6.5) must not be too strongly correlated with x or y. If strongly correlated with x, in fact, the stratification will be nearly equivalent with stratification after x, and so the regression (6.4) will automatically be more stable than (6.5), and conversely if z is strongly correlated with y. In neither case can the regression variability give a basis for conclusions about causal direction.

(ii) The approaches of theorems 6.1.1 and 6.1.2 refer to situations with errors in equations, not to situations with errors in variables. In models with errors in variables, in fact, the underlying hypothesis is that the error-free variables satisfy a disturbance-free equation, and this hypothesis leads to quite other considerations. For example, if relation (6.4) were disturbance free (and the variables error free) it would be the inverse of (6.5), and so the two regressions would both be either stable or unstable in the subsamples. Similarly, if models (6.14) and (6.15) were disturbance and error free, (6.16) would be equalities of inequalities.

(iii) On the logic of the approach in subsection 6.1.1 the stratification after z explores whether predictors like (6.3) or the left of (6.10) with constant β are applicable in the subsamples, that is, applicable under circumstances that differ along with the stratification variable. Hence by repeating the procedure by the use of other stratification variables, say $z_1, z_2, \cdots$, we may hope to extend gradually the field of operative use of the predictors. This makes sense when specifying (6.3) and the left of (6.10) as unbiased predictors, quite irrespective of the causal interpretation of the first of (6.1). This being so, is it possible, or even meaningful, to interpret such a statistical analysis and the ensuing conclusions about variability as referring to the causal hypotheses in the first or second of (6.1)? The following comments on this question refer to statistical approaches in general, not only to the situation in subsection 6.1.1.

(a) Certainly it is meaningful to pose causal hypotheses like the first or the second of (6.1). This is everyday routine in experimental laboratories, although their cause-effect hypotheses often go under the name of stimulus-response relations [77], [81]. The terminology is irrelevant. If however the causal terms were cut out, the terminology about hypotheses like (6.1) would depend upon whether or not they could be tested and verified by way of controlled experiments. And it would be very confusing indeed if the terminology of a hypothesis were to depend upon current procedures for its verification—the terminology would then have to be changed as the research front moves between nonexperimental and experimental positions.

(b) It is my understanding that the discrimination between causal hypotheses (6.1) is not a matter for hypothesis testing by way of significance tests of the conventional stochastic type. Such tests might under favorable circumstances be applied to the statistical asymmetries in play, such as the coefficients of variability (6.19), but in point of principle they make no basis for inference about the causal hypothesis as such.

Even if we limit our attention to the statistical aspects of the asymmetries, the stochastic tests of significance are of restricted relevance in the present situation, since we are dealing with nonexperimental data. The stochastic tests of significance are designed to cover sampling errors, but in model building on the basis of nonexperimental data specification errors often dominate over sampling errors, and the specification errors in point of principle are not covered by the significance tests [76].

(c) As always, scientific model building is governed by Mach's principle of economy of thought [38]. The best we can hope for is that the statistical techniques discussed may sometimes serve as a vehicle for such economy. In (6.3), for example, the causal model in the first of (6.1) makes use of only one parameter to measure the influence of x upon y, whereas in (6.4) the reverse model in the second of (6.1) makes use of one parameter for each of the substrata. Hence in situations where the statistical asymmetry (6.3) and (6.4) shows up in the observations, the causal hypothesis in the first of (6.1) is supported by the Machian principle.

I do not wish to stress the reference to Mach's principle. Whether or not the approach of statistical asymmetries will prove useful for establishing the direction of causal relationships can only be shown by experience. The situation is very different from the routine technique of experimental statistics. As I see it, the search for causal relations by the approach of statistical asymmetries is like following the scent in a hunt. The flows of causation reveal themselves in various statistical asymmetries that under favorable circumstances form a scent that may be pursued by statistical methods and lead to causal conclusions. And it might even be possible that such clues will provide material for fresh causal hypotheses and conclusions, hypotheses that under still more favorable circumstances might be taken up and tested by controlled experiments.

REFERENCES

[1] R. G. D. Allen, *Mathematical Economics*, London, Macmillan, 1956.

[2] R. J. Ball, A. Hazelwood, and L. R. Klein, "Econometric forecasts for 1959," *Bull. Oxford Univ. Inst. Statist.*, Vol. 21 (1959), pp. 3–16.

[3] R. L. Basmann, "On the asymptotic distribution of generalized linear estimators," *Econometrica*, Vol. 28 (1960), pp. 97–107.

[4] W. J. Baumol, *Economic Dynamics. An Introduction*, New York, Macmillian, 1951, 1959 (2nd ed.).

[5] R. Benini, "Sull'uso delle formule empiriche nell'economia applicata," *Giorn. Economisti*, Vol. 35 (1907), pp. 1053–1063.

[6] R. Bentzel and B. Hansen, "On recursiveness and interdependency in economic models," *Rev. Econ. Studies*, Vol. 22 (1954), pp. 153–168.

[7] E. Bentzel and H. Wold, "On statistical demand analysis from the viewpoint of simultaneous equations," *Skand. Aktuarietidskr.*, Vol. 29 (1946), pp. 95–114.

[8] C. Christ, "A test of an econometric model for the United States 1921–1947," *Conference on Business Cycles*, New York, National Bureau of Economic Research, 1951, pp. 35–107.

[9] C. Clark, "A system of equations explaining the United States trade cycle 1921–1941," *Econometrica*, Vol. 77 (1949), pp. 93–124.

[10] H. Cramér, *Mathematical Methods of Statistics*, Uppsala, Almqvist & Wiksell, 1945; Princeton, Princeton University Press, 1946.

[11] ———, "On some classes of nonstationary stochastic processes," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1961.

[12] J. Durbin, "Errors in variables," *Rev. Inst. Internat. Statist.*, Vol. 22 (1954), pp. 23–32.

[13] G. Eklund, *Studies of Selection Bias in Applied Statistics*, Uppsala, Almqvist & Wiksell, 1959.

[14] M. Ezekiel, "The cobweb theorem," *Quart. J. Econ.*, Vol. 52 (1938), pp. 255–280.

[15] P. Faxér and H. Wold, "On the specification error in regression analysis," *Ann. Math. Statist.*, Vol. 28 (1957), pp. 265–267.

[16] R. Frisch, "Static and dynamic approaches in economic theory," (Norwegian.) *Nationaløkon. Tidskr.*, Vol. 41 (1929), pp. 321–379.

[17] ———, "Propagation problems and impulse problems in dynamic economics," *Economic Essays in Honour of Gustav Cassel*, London, Allen and Unwin, 1933, pp. 171–205.

[18] ———, *Statistical Confluence Analysis by Means of Complete Regression Systems*, Publ. No. 5, Univ. Inst. of Economics, Oslo , 1934.

[19] M. A. Girshick and T. Haavelmo, "Statistical analysis of the demand for food," *Econometrica*, Vol. 15 (1947), pp. 79–110. (Reprinted in revised form in [26].)

[20] U. Grenander, "On the estimation of regression coefficients in the case of an autocorrelated disturbance," *Ann. Math. Statist.*, Vol. 25 (1954), pp. 252–272.

[21] ———, (ed), *Probability and Statistics. The Harald Crámer Volume*, Stockholm, Almqvist & Wiksell, and New York, Wiley, 1960.

[22] U. Grenander and M. Rosenblatt, *Statistical Analysis of Stationary Time Series*, New York, Wiley, 1956.

[23] T. Haavelmo, "The statistical implications of a system of simultaneous equations," *Econometrica*, Vol. 11 (1943), pp. 1–12.

[24] ———, "The role of the econometrician in the advancement of economic theory," *Econometrica*, Vol. 26 (1958), pp. 351–357.

[25] C. Hildreth and F. G. Jarrett, *A Statistical Study of Livestock Production and Marketing*, New York, Wiley, 1955.

[26] W. C. Hood and T. C. Koopmans (eds.), *Studies in Econometric Method*, New York, Wiley, 1953.

[27] L. HURWICZ, "Prediction and least squares," *Statistical Inference in Dynamic Economic Models*, New York, Wiley, 1950, pp. 266–300.
[28] A. KHINCHIN, "Korrelationstheorie der stationaren stochastischen Prozesse," *Math. Ann.*, Vol. 109 (1934), pp. 604–615.
[29] L. R. KLEIN, *Economic Fluctuations in the United States, 1921–1941*, New York, Wiley, 1950.
[30] ———, *A Textbook of Econometrics*, Evanston, Ill., Row, Peterson, 1953.
[31] L. R. KLEIN and A. GOLDBERGER, *An Econometric Model of the United States 1929–1952*, Amsterdam, North-Holland, 1955.
[32] A. KOLMOGOROV, "Grundbegriffe der Wahrscheinlichkeitsrechnung," *Ergebnisse der Mathematik und ihrer Grenzgebiete*, Vol. 2, No. 3, 1933.
[33] T. C. KOOPMANS, "Statistical estimation of simultaneous economic relationships," *J. Amer. Statist. Assoc.*, Vol. 40 (1945), pp. 448–466.
[34] T. C. KOOPMANS (ed.), *Statistical Inference in Dynamic Economic Models*, New York, Wiley, 1950.
[35] W. W. LEONTIEF, *The Structure of American Economy, 1919–39*, Cambridge, Mass., Harvard University Press, 1941.
[36] M. LOÈVE, "Fonctions aléatoires de second ordre," *Rev. Sci.*, Vol. 84 (1946),pp. 195–206.
[37] E. LYTTKENS, *Regression Analysis with Endogenous Variables and Autocorrelated Residuals*, to be published.
[38] E. MACH, *Die Mechanik in ihrer Entwicklung, Historisch-kritisch*, 1883, Leipzig, Barth, 1933 (8th ed.).
[39] E. P. MACKEPRANG, *Pristeorier*, Copenhagen, Bagge, 1906.
[40] H. B. MANN and A. WALD, "On the statistical treatment of linear stochastic difference equations," *Econometrica*, Vol. 11 (1943), pp. 173–220.
[41] A. MARSHALL, *Principles of Economics*, 1890, London, Macmillan, 1946 (8th ed.).
[42] P. MASANI, "Cramér's theorem on monotone matrix-valued functions and the Wold decomposition," *Probability and Statistics. The Harald Cramér Volume*, Stockholm, Almqvist & Wiksell, and New York, Wiley, 1960, pp. 175–189.
[43] D. PATINKIN, *Money, Interest and Prices*, Evanston, Ill., Row, Peterson, 1956.
[44] A. PHILLIPS, "The estimation of parameters in systems of stochastic differential equations," *Biometrika*, Vol. 46 (1959), pp. 67–76.
[45] E. A. ROBINSON, *Infinitely Many Variates*, London, Griffin, 1959.
[46] Royal Statistical Society, "The present position of econometrics, a discussion," *J. Roy. Statist. Soc., Ser. A*, Vol. 123 (1960), pp. 274–296.
[47] P. A. SAMUELSON, *Foundations of Economic Analysis*, Cambridge, Mass., Harvard University Press, 1947.
[48] H. SCHULTZ, *The Theory and Measurement of Demand*, Chicago, University of Chicago Press, 1938.
[49] H. SIMON, "Causal ordering and identifiability," *Studies in Econometric Methods*, New York, Wiley, 1953, pp. 49–74.
[50] H. STROTZ, "Interdependence as a specification error," *Econometrica*, Vol. 28 (1960), pp. 428–442.
[51] H. STROTZ and H. WOLD, "Recursive vs. nonrecursive systems: An attempt at synthesis," *Econometrica*, Vol. 28 (1960), pp. 417–427.
[52] D. SUITS, "An econometric model of the watermelon market," *J. Farm Econ.*, Vol. 37 (1955), pp. 237–351.
[53] H. THEIL, "Estimates and their sampling variance of parameters of certain heteroscedastic distributions," *Rev. Internat. Statist. Inst.*, Vol. 19 (1951), pp. 141–147.
[54] ———, *Linear Aggregation of Economic Relations*, Amsterdam, North-Holland, 1954.
[55] ———, *Economic Forecasts and Policy*, Amsterdam, North-Holland, 1958.
[56] J. TINBERGEN, "Bestimmung und Deutung von Angebotskurven. Ein Beispiel," *Z. Nationalökonomie*, Vol. 1 (1930), pp. 669–679.

[57] ———, *An Econometric Approach to Business Cycle Problems*, Paris, Hermann, 1937.
[58] ———, *Statistical Testing of Business Cycle Theories, I: A Method and its Application to Investment Activity. II: Business Cycles in the United States of America, 1919–1932.* Geneva, League of Nations, 1939.
[59] ———, "Fconometric business cycle research," *Rev. Econ. Studies*, Vol. 7 (1940), pp. 73–90.
[60] ———, *Eusiness Cycles in the United Kingdom, 1870–1914*, Amsterdam, North-Holland, 1951, 1956 (2nd ed.).
[61] ———, *Economic Policy: Principles and Design*, Amsterdam, North-Holland, 1956.
[62] G. WALKER, "On periodicity in series of related terms," *Proc. Roy. Soc. London, Ser. A.*, Vol. 131 (1931), pp. 518–532.
[63] P. WHITTLE, "The simultaneous estimation of a time series' harmonic components and covariance structure," *Trabajos Estadist.*, Vol. 3 (1952), pp. 43–57.
[64] ———, "The analysis of multiple stationary time series," *J. Roy. Statist. Soc., Ser. B.*, Vol. 15 (1953), pp. 125–139.
[65] S. WILKS, *Mathematical Statistics*, Princeton, Princeton University Press, 1944.
[66] H. WOLD, *A Study in the Analysis of Stationary Time Series*, Uppsala, Almqvist & Wiksell, 1938, 1954 (2nd ed.).
[67] ———, "Eine schwedische Untersuchung über die Elastizität der Nachfrage," *Metron*, Vol. 14 (1942), pp. 263–285.
[68] ———, "A synthesis of pure demand analysis, I-III," *Skand. Aktuarietidskr.*, Vol. 26 (1943), pp. 85–118 and 220–263; Vol. 27 (1944), pp. 69–120.
[69] ———, "A theorem on regression coefficients obtained from successively extended sets of variables," *Skand. Aktuarietidskr.*, Vol. 28 (1945), pp. 181–200.
[70] ———, "On prediction in stationary time-series," *Ann. Math. Statist.*, Vol. 19 (1948), pp. 558–567.
[71] ———, "Statistical estimation of economic relationships," *Econometrica*, Vol. 17 (1949), Supplement, pp. 1–22.
[72] ———, "On least square regression with autocorrelated variables and residuals," *Bull. Internat. Inst. Statist.*, Vol. 32, II (1950), pp. 277–289; Oral discussion: *Bulletin*, Vol. 32, I (1952), pp. 145–147.
[73] ———, "Time series and demand analysis," *Sankhyā*, Vol. 11 (1951), pp. 205–216.
[74] ———, "Causality and econometrics," *Econometrica*, Vol. 22 (1954), pp. 162–177.
[75] ———, "Possibilités et limitations des systèmes à chaîne causale," *Cahiers Sém. Econ. R. Roy.*, No. 3 (1955), pp. 81–101.
[76] ———, "Causal inference from observational data. A review of ends and means," *J. Roy. Statist. Soc., Ser. A.*, Vol. 119 (1956), pp. 28–61.
[77] ———, "A case study of interdependent versus causal chain systems," *Rev. Internat. Statist. Inst.*, Vol. 26 (1959), pp. 5–25.
[78] ———, "Ends and means on econometric model building. Basic considerations reviewed," *Probability and Statistics. The Harald Crâmer Volume*, Stockholm, Almqvist & Wiksell, and New York, Wiley, 1960, pp. 355–434.
[79] ———, "A generalization of causal chain models," *Econometrica*, Vol. 28 (1960), pp. 443–463.
[80] ———, "Construction principles of simultaneous equations models in econometrics," *Bull. Internat. Statist. Inst.* (Tokyo session), in press.
[81] ———, "On the definition and meaning of casual concepts," submitted to *Philos. Sci.*
[82] H. WOLD and L. JURÉEN, *Demand Analysis. A Study in Econometrics*, Stockholm, Almqvist & Wiksell, 1952; New York, Wiley, 1953.
[83] E. J. WORKING, "What do statistical demand curves show?" *Quart. J. Econ.*, Vol. 41 (1927), pp. 212–235.
[84] H. WORKING, "Price relations between July and September wheat futures at Chicago since 1885," *Wheat Studies*, Vol. 9 (1933), pp. 187–238; "Price relations between May and

newcrop wheat futures at Chicago since 1885," *Wheat Studies*, Vol. 10 (1934), pp. 183–230.

[85] S. WRIGHT, "Correlation and causation," *J. Agric. Res.*, Vol. 20 (1921), pp. 557–585.

[86] ———, "The method of path coefficients," *Ann. Math. Statist.*, Vol. 5 (1934), pp. 161–215.

[87] ———, "Path coefficients and path regression: Alternative or complementary concepts?" *Biometrics*, Vol. 16 (1960), pp. 189–202.

[88] G. U. YULE, "On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers," *Philos. Trans. Roy. Soc. London*, Vol. 226 (1927), pp. 267–298.

[89] A. ZELLNER, "Econometric estimation with temporally dependent disturbance terms," to appear in the *Internat. Econ. Rev.*