

**REVUE BELGE DE STATISTIQUE
ET DE RECHERCHE OPERATIONNELLE**

**Vol. 9 - N° 2
JANVIER 1969**

**BELGISCH TIJDSCHRIFT VOOR STATISTIEK
EN OPERATIONEEL ONDERZOEK**

**Vol. 9 - N° 2
JANUARI 1969**

REVUE BELGE DE STATISTIQUE ET DE RECHERCHE OPERATIONNELLE

VOL. 9 - N° 2 — JANVIER 1969

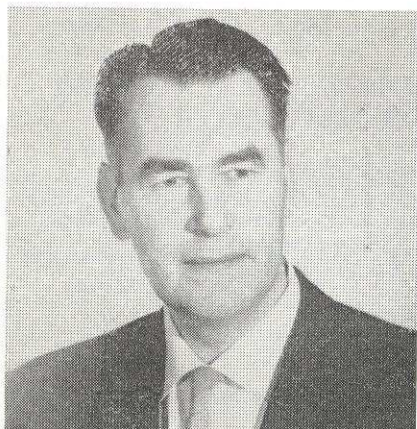
VOL. 9 - N° 2 JANUARI 1969

SOMMAIRE — INHOUD

In memoriam : Nicolas Basilewsky	1
H.C.S. THOM. — Application of climatological analysis to engineering design data	2
J. PAELINCK. — Sur quelques principes de construction de modèles économétriques	14
Publications reçues — Ontvangen publicaties	3 ^e couv.

BELGISCH TIJDSCHRIFT VOOR STATISTIEK
EN OPERATIONEEL ONDERZOEK

IN MEMORIAM



Le 31 décembre 1968, Nicolas BASILEWSKY, Président de la Société Belge pour l'Application des Méthodes Scientifiques de Gestion (SOGESCI), était victime d'un accident de roulage.

Il devait succomber à ses blessures le vendredi 10 janvier 1969.

Nicolas BASILEWSKY était Ingénieur civil des Mines, sorti de l'Université de Louvain.

Sa carrière, qui débute dans la construction mécanique, se poursuit dès 1945 dans le domaine de l'organisation à la filiale belge de la Société Bedaux. En 1947, il est engagé par le groupe français FOULD en qualité d'Ingénieur en Chef au service des Aciéries de Pompey dont il devient Directeur en 1953. Cinq ans plus tard, il est attaché comme Ingénieur-Conseil à la Direction Générale des Chantiers de l'Atlantique. Puis, en 1962, IBM France le prend à son service et enfin, en 1963, il passe à SIDMAR où il a la charge de Chef du Bureau des Méthodes, dépendant de la Direction Générale de cette Société.

Membre de la SOGESCI depuis 1963, il en devient Administrateur le 19 janvier 1966, puis Président le 21 février 1968.

Frappé par le destin alors qu'il poursuivait une carrière brillante qui n'avait cessé d'être ascendante, Nicolas BASILEWSKY laissera, dans la mémoire de tous ceux qui l'ont approché à la SOGESCI, le souvenir d'un collègue loyal, dont le caractère généreux et la grande intelligence l'inclinaient à dispenser, sans les mesurer, les conseils éclairés que sa grande compétence faisait rechercher.

Le Conseil d'Administration de la SOGESCI exprime à sa veuve et à son fils ses sentiments de condoléances très sincères.

APPLICATION OF CLIMATOLOGICAL ANALYSIS TO ENGINEERING DESIGN DATA (*)

by H.C.S. THOM

Silver Spring, Environmental Data Service - ESSA, U.S.A.

ABSTRACT

The general climatological prediction is defined, and its proper role in engineering design is discussed. The fundamental principles involved in the selection of the best meteorological variables for determining design values are discussed together with their analytical representation.

There are two general types of design values: quantiles on extreme values and quantiles on subextreme values. The extreme value quantile applies where the engineering system must withstand some recurring extreme. Examples are all forms of structures exposed to meteorological forces which could cause failure or destruction. The subextreme value quantile applies where the overloading of the engineering system does not cause destruction but only discomfort or inconvenience with perhaps some tolerable economic loss. Examples are heating and air conditioning systems and storm sewer and small drainage systems.

The proper analysis for extreme value design data comes from the statistical theory of extreme values. Examples of application to wind and snow loads are discussed. For subextreme value data either the recurrence of critical events within an individual time period may be limited or their spacing in time controlled. Both of these devices depend upon the fact that frequent repetition of the critical event is more undesirable than less frequent repetition. The statistical theory of the m^{th} value or the theory of the time interval distribution apply here. Examples of application to heating and air conditioning and to intense rainfall are discussed.

(*) Presented at the Symposium on Urban climates and Building climatology, Brussels, October 1968.

Introduction.

Weather enters into many of man's activities sometimes facilitating sometimes limiting his activity. The weather forecast with its severe limitation of time range makes it useful only for operational decisions being made a few days in advance at most. For the large area of planning decisions for years, even many years in the future, the statistical analysis of historical data resulting in the climatological prediction is required. This use in planning covers a wide variety of problems one of the most important being the application to engineering design discussed here.

From time immemorial builders knew something of the weather factor at least from experience and used this in their designs. It was not, however, until early in this century that the statistical analysis necessary to the development of climatological analysis began to be developed. The coming of the airplane, at about the same time which is so subject to weather factors, added greatly to the needs for climatological analysis. At first the methods were rather crude and heavily empirical or nonparametric. Later with the sophistication of engineering systems brought on by the second world war, climatological analysis began a rapid development with greater and greater use of parametric methods. Unfortunately, there are still relatively few who understand modern climatological analysis; so some work is still crude and even faulty. Notwithstanding there is today a good body of valid climatological analysis available which can give unbiased climatological prediction for the development of engineering design data almost anywhere in the world.

A weather design value may be defined as the magnitude of a meteorological variable which, when used in the design of an engineering system, nonmeteorological factors having been accounted for jointly or independently, will assure with a given probability that the system will adequately meet a set of prescribed design requirements. The choice of the meteorological value naturally depends on the assigned design conditions and on the prescribed value of the probability of being surpassed. The assignment of the design requirements and the probability level is clearly an engineering problem while the analysis to determine a set of probabilities containing the one chosen by the engineer or from which he can make a choice is the problem of the climatologist. The choice of the meteorological design variable and its relationship to the engineering system variables is the joint problem of the climatologist and engineer, for often the choice of meteorological element is a compromise and must be adapted to the problem. Thus,

the weather factor design problem consists of three parts: (1) choice of the design variable, (2) the solution of the relationship problem, and (3) the climatological analysis. It is the objective here to outline the general development of design values and to give the actual climatological analysis in certain special areas of application.

The design variable and its prediction.

The weather design variable may be defined as a variable quantity which is rationally related to the engineering variable of the design problem. Since it is also a climatological variable, it is a random variable with an associated distribution function which defines probability over a range of design requirements. The formal analysis may be illustrated by using functional notation applied to a hypothetical design variable [1]. The prescribed design requirements are completely specified by a load t_E which the engineer determines by engineering analysis. This results in the relationship

$$t_E = b(t_W) \quad (1)$$

where t_W is the weather design variable, t_E is the load variable, and b is the relationship function. t_W and b may be given by the engineer or obtained in collaboration with the climatologist, and t_W may be multidimensional. Conversely weather will also be associated with design values. This may be expressed by the inverse of (1) which is

$$t_W = b^{-1}(t_E) \quad (2)$$

By formal climatological analysis a distribution function on t_W is given by $F(t_W)$ [2]. Associated with t_E one must then admit a probability with which failure to meet the load can be tolerated. If $1 - p$ is the probability of failure then the relationship

$$p = F(t_W) \quad (3)$$

defines the weather design value t_W . Substitution from equation (2) defines the design load t_E

$$p = F[b^{-1}(t_E)] \quad (4)$$

t_E may also be thought of as a transformed variable and will then have a transformed distribution which may be expressed by a transformed distribution function

$$p = G(t_W) \quad (5)$$

Occasionally h will be a tractable function in which case G can be found explicitly. Taking the inverse of equation (4) yields

$$h^{-1}(t_E) = t_W = F^{-1}(p) \quad (6)$$

Solving for t_E gives

$$t_E = h[F^{-1}(p)] = G^{-1}(p) \quad (7)$$

which shows that the probability of being less than t_W is equal to the probability of being less than t_E which is prescribed by the engineer; hence, (6) defines an equiprobability transformation from t_W to t_E . The weather design value which is a p -quantile on t_W is determined by equation (6). This equation naturally offers the possibility of a selection of design values over any range of design probabilities p .

It is now the climatologist's task to choose a series of t_W in such a way as to form a random variable capable of giving rational probabilities. This in general will be a climatological series which is a stationary series of independent meteorological observations. In many instances the choice of weather design variable must be a compromise, for the observed meteorological elements do not always fit engineering relationships and, therefore, must be transformed or modified.

The relationship problem

Meteorological variables are purely physical variables which really apply to nothing but weather. In order, therefore, to use them in non-meteorological applications a means of conversion to the applied variable must be provided so that the climatological prediction can be transformed to the applied variable. This is done as illustrated above by establishing a functional relationship between an appropriate meteorological variable and the applied variable. The problem of finding such a relationship is called the relationship problem. It is often solved jointly by the engineer and the climatologist.

The solution to the relationship problem is always some form of functional relationship between the meteorological variable and the application variable. In some instances the function is known from fundamental principles or can be derived from them; in others varying amounts of fundamental principles and empiricism must be employed. In all situations the problem has the formal structure set out below:

Assuming that the variable of the application is given, the first step is the choice of the proper meteorological variable. This must be selected from the elements being observed or combinations of them. Not only must these elements be a part of the observation program, but they must have been observed for a sufficient period of record so that a satisfactory climatological analysis may be made. Let the application variable be y and the meteorological variable x . Then the required relationship may be expressed by the equation

$$y = b(x), \quad (8)$$

where x may be a set of variables. This relates the application variable y to the meteorological variable x , and serves to transform the prediction on the climatological variable to one on the applied variable.

The solution of the relationship problem must have the close collaboration of the climatologist and the engineer. The choice of the variable y is, of course, the engineer's task, but because of the limited availability of meteorological elements, and the varied possibilities of developing a climatological prediction equation, the climatologist must have a hand in choosing the variable x . In addition, the two together must provide a suitable series of data on y and x concurrently for estimating the function $b(x)$ as a regression unless this relationship is already available from purely theoretical analysis. This series need not be a climatological series, i.e., one value per year, but for complete validity in carrying out the regression analysis, variable x must form a series in which the successive values in time are independent of each other. This may usually be accomplished by spacing the values of x sufficiently in time so that successive values are independent. Advantage may be taken of all data by averaging the several regressions based on the lagged data. The climatological prediction, however, must be made only on the basis of a climatological series, i.e., a stationary series of independent values.

Extreme value design data

A set of extreme values will constitute the weather design variable wherever the engineering system cannot otherwise be protected against the occurrence of an extreme [3]. This set will always consist of a series of extremes for successive periods of one year or more because weather has an annual fluctuation. The reasoning back of this is as follows: If the period were less than one year, it would be impossible to form a proper climato-

logical series of extremes for the series for one part of the year must always be different from the series for the other part of the year so the composite series would not form a proper random variable. On the other hand series of extremes for periods longer than a year would form a proper series but the distributions of such series can be shown to be identical with the annual series; hence, the result would be an undesirable reduction in sample size where lengthy homogeneous series are already seldom available. Therefore, the series of annual extremes is the best climatological series for establishing extreme value design data. The following discussions of climatological analysis for wind and snow load are examples of extreme value design data.

Wind load design data

Since dangerous loads on structures are often the result of the shorter bursts of wind such as those of the same scale as the structure, and these are of only a few seconds duration, it has always appeared desirable to use the regularly observed wind speed with the shortest duration. This is one mile or one minute in the U.S. and sometimes less and sometimes more elsewhere. To go from the observation to the shorter duration speed a gust conversion factor is applied. While there are some weaknesses to this procedure, it has appeared to be the best procedure available at present.

In the U.S. it has been found that fitting the Fréchet extreme value distribution to the annual extreme winds provides a satisfactory basis for obtaining design loads. This distribution function may be expressed as

$$F(v) = \exp [-(v/\beta)^\gamma] \quad (9)$$

where F gives the probability of the speed being less than v and β and γ are statistical parameters. Maps of several quantiles from the distribution have been provided for the use of engineers [4].

Since the wind speed is readily converted to pressure, the relationship problem involves mainly the application of a gust factor and the distribution of the forces over the structure. These have been largely handled by engineers with some assistance from meteorologists.

Since the earth is very inadequately covered with observations of extreme wind speeds, the writer recently developed a procedure for obtaining approximate design wind speeds from much more widely available or readily estimated data [5]. The method is based on two principles: (1) that an extreme wind speed is composed of a general level of wind on which a

wind speed due to a violent meteorological disturbance is superimposed and, (2) that based on a large number of observations the extreme wind distributions may be divided into two classes — one for extratropical storms and thunderstorms, the other for tropical cyclones.

For extratropical cyclones and thunderstorms a study of a large number of sample extreme wind records showed that the shape parameter γ tended to 9.0 as the number of years of record became large. The scale parameter β was determined by the level of the wind which was assumed to be the highest mean monthly speed $\bar{u}$ for the year. The estimate β^* was related to $\bar{u}$ by the relationship

$$\beta^* = (347.5 \bar{u} + 364.5)^{1/2} - 19.1 . \quad (10)$$

Thus, the extreme wind distribution for extratropical storms and thunderstorms becomes

$$F_E(v) = \exp [-(v/\beta^*)^{-9.0}] . \quad (11)$$

Tests of this relationship over a wide range of climatic conditions of small sample sizes gave results within a few miles per hour of the empirical estimate.

For tropical storms it was found for the longest and best records that γ averaged near 4.5. The scale parameter, as was to be expected, followed the same relationship as for extratropical storms, i.e., equation (10). The extreme value distribution may then be expressed by

$$F_T(v) = \exp [-(v/\beta^*)^{-4.5}] . \quad (12)$$

Equations (11) and (12) may be readily inverted by taking logarithms to obtain design values of v .

In the tropical storm area of the Atlantic Ocean and in part of the typhoon area of the Pacific extratropical storms also occur. This results in a mixture of the two distributions (11) and (12) giving

$$F_{ET}(v) = (1 - p_T) F_E(v) + p_T F_T(v) . \quad (13)$$

Here p_T is the probability of an annual extreme wind speed from a tropical cyclone. The mixture parameter p_T was estimated by a relationship with the mean annual frequency of tropical cyclone passages per five degree square. Since the frequency of tropical storm passages is about three times as large

in the maximum square east of the Philippine Islands as it is in the maximum square east of Miami, the relationship for p_T had to be different in the two tropical storm areas. For the Atlantic Ocean the relationship was found to be

$$p_T = 1/[1 + 99 \exp (-3.0 f)] \quad (14)$$

and for the Pacific

$$p_T = 1/[1 + 99 \exp (-4.4 f)] \quad (15)$$

where f is the mean annual frequency of tropical storm passages for the five degree square under consideration. Equations (10), (11), (12), (14) or (15) provide the required substitution in equation (13) to obtain approximations to the extreme wind distribution for all areas of the earth. The result cannot be inverted directly to obtain design speed but must be inverted empirically by plotting several values of F_{ET} .

Snow load design data

The variable which appeared to have the best prospects of measuring the snow load on structures is the maximum annual weight of the snow pack on the ground expressed in depth of water. In the U.S. this observation is called the water equivalent of the snow pack. It is superior to the depth of the snow pack with an assumed density since the density has been found to vary widely and in fact, have its own distribution [6]. Except for very large accumulations the density of the snow pack even appears to be uncorrelated with the depth.

The annual extremes of water equivalent of the snow pack greater than zero were found to follow a logarithmic normal distribution given by

$$G(\ln x) = N [\ln x; \alpha_1(\ln x), \sigma(\ln x)] \quad (16)$$

where x is the water equivalent, $\alpha_1(\ln x)$ is the mean of $\ln x$, and $\sigma(\ln x)$ is the standard deviation of $\ln x$. This distribution, of course, holds only where a snow pack exists every year. In most southern areas a snow pack does not exist every year although when it does occur dangerous snow loads may be produced. This makes the universal water equivalent distribution a mixed distribution given by

$$G(\ln x) = (1 - p) + p N [\ln x; \alpha_1(\ln x), \sigma(\ln x)] \quad (17)$$

Here p is the probability of a snow pack greater than zero and $1 - p$ is, of course, the probability of no snow pack.

The design water equivalent on the ground is then found by inverting (17) and substituting $z = \ln x$ and the statistical estimates $\bar{z}$, $s(z)$ and p yielding the G^{th} quantile

$$z(G) = s(z) N^{-1} [G(z) - q]/p + \bar{z} \quad (18)$$

Design snow loads on the ground are given by multiplying $z(G)$ by a suitable constant k to convert depth of water to pressure.

The relationship problem is complex, but the best solutions to it are given in the National Building Code of Canada [7]. Here coefficients are given to convert snow load on the ground to snow load on roofs for a wide variety of structures.

Subextreme value design data

There are a number of areas where it is not economically feasible or desirable to design for the extreme value. Here the objective is either to space a critical event so that its recurrence in time is not intolerable or to limit the risk in a way which prevents excessive recurrence in a single season. The first may be accomplished by use of a time interval distribution and the second by use of an m^{th} value distribution. Occasionally a situation arises where the m^{th} values cannot be readily made available. Then roughly similar results can be attained by treating the problem by an extreme value distribution on a subextreme statistic such as the daily average.

Intense rainfall design data

In the design of storm drainage systems for buildings and groups of buildings it is often not economical to design against overflows which cause only moderate inconvenience or insignificant damage. In the past the so-called frequency or mean recurrence interval of depth for a given storm duration has been employed to design such drainage systems. Since the time interval between events is averaged, such design values give little information on the spacing of critical rainfall amounts. The spacing appears to be the important factor since the system would be considered a failure if events were spaced too close together. Thus, it appears desirable to fix the rainfall amount for the given duration and give the probability that a rainfall greater than this amount will recur in a certain time interval.

The discrete events of high intensity rainfall follow a Poisson time interval distribution [8] which may be expressed by

$$dp = \alpha \exp(-\alpha t) dt \quad (19)$$

where p is probability, t is time, and α is the mean frequency of occurrence per year. The distribution function is found by integration of (19) giving

$$F(T) = 1 - \exp(-\alpha T) \quad (20)$$

where T is the time interval in years and $F(T)$ is the probability of a spacing less than T years. The mean recurrence interval is the expected value of t which is $1/\alpha$. Design T 's are easily obtained by inverting (20) by taking logarithms.

In some areas estimates of the mean recurrence interval $1/\alpha$ are available for various durations from which the time distribution function is immediately available from equation (20).

The relationship problems in this instance need not be of concern to the climatologist since they have already been fully analyzed by engineers.

Air conditioning design data

In this area mean frequency design data have occasionally been employed which fail completely to reflect the risk of failure of the system. As in other subextreme value problems failure of the system is temporary and an inconvenience but not critical. What is desired is a design value which balances the risk with the potential discomfort and the cost of installation.

The deficiency of the mean frequency design data lay in its failure to reflect within seasonal risk. Thus, the mean frequency data gives only the risk of failure on the average and ignores the possibility of a run of high meteorological values causing system failure in a single season. Clearly the user of the system will be most concerned about successive failures to carry the load; hence, it is necessary to limit the within season risk as well as the between season risk. This can best be accomplished by using the m^{th} value distribution [9].

In the U.S. the air conditioning season is often assumed to be the four months June through September. This period contains 122 days; thus, the highest value in this period has a probability of about 0.0125 of occurring, the second highest 0.025, and the m^{th} highest 0.0125 m . Hence, if the m^{th} value distribution, of which the Type I extreme value distribution

is a special case for $m = 1$, is fitted to an annual series of m^{th} values for the air conditioning season, the within year risk is $m/122$ and the between year risk is given by the distribution function. The complement is expressed by the integral

$$H(t) = \frac{m^m}{\beta_m \Gamma(m)} \int_t^\infty \exp \left[m \left(\frac{\alpha - t}{\beta} \right) - m \exp \left(\frac{\alpha - t}{\beta} \right) \right] dt \quad (21)$$

where t is the design value, m is the rank of the variable from the highest value, α is a location parameter and β is a scale parameter. Tables for obtaining $H(t; m)$ are available. The distribution is applied to both the wet bulb and dry bulb temperatures although the dry bulb load is small compared to the wet bulb load. Design values are obtained from (21) by empirical inversion as for equation (13).

The conversion to the applied variables is made through well known equations for heat transmission and sensible and latent heat loads. Radiation is always an appreciable part of the sensible heat load and is usually taken as that produced by the clear day value for highest sun.

Heating design data

The heating design variable problem is somewhat different from the air conditioning problem because the highest loads often occur during hours when the heating system is operating at comparatively low output or there is little concern for the system meeting the load. Thus, the extreme value is not an appropriate weather design variable nor is an m^{th} highest value useful since it does not reflect diurnal variation which is the problem here. The most desirable variable would be that for a chosen hour but then the choice of hour would be difficult. Finally the daily average was chosen as the variable since it also occurs at some hour, and more important, it was highly correlated with heating load [1]. The climatological series chosen was that of annual minimum of the daily average temperature.

The appropriate distribution is the Type I extreme value distribution for minimum values. This is expressed by the distribution function

$$F(t) = \exp \left[- \exp \left(\frac{t - \alpha}{\beta} \right) \right] \quad (22)$$

where t is the weather design variable, α is the location parameter, β is the scale parameter, and $F(t)$ is the probability of meeting heat load associated with the minimum daily average t . Design values may be obtained by simple inversion of (22).

The relationship problem has been solved by engineers using heat transmission coefficients and quantities of infiltration and ventilation by outside air.

REFERENCES

- [1] Thom, H.C.S., « Revised Winter Outside Design Temperature », *Trans. ASHAE*, Vol. 63, 1957, pp. 111-112.
- [2] Thom, H.C.S., « Some Methods of Climatological Analysis », *WMO Tech. Note No. 81*, 1966, Geneva.
- [3] Thom, H.C.S., « Frequency of Maximum Wind Speeds », *Proceedings, ASCE*, Vol. 80, 1954, p. 2.
- [4] Thom, H.C.S., « New Distribution of Extreme Winds in the United States », *Journal Structural Div. Proceedings ASCE*, July 1968, pp. 1787-1801.
- [5] Thom, H.C.S., « Toward a Universal Climatological Extreme Wind Distribution », Paper No. 24, *Proceedings Seminar on Wind Forces on Buildings and Structures*, Sept. 1968, pp. 669-683.
- [6] Thom, H.C.S., « Distribution of Maximum Annual Water Equivalent of Snow on the Ground », *Monthly Weather Review*, Vol. 94, No. 4, April, 1966, pp. 265-271.
- [7] National Building Code of Canada, NRC No. 3188, National Research Council, Ottawa, 1953.
- [8] Thom, H.C.S., « A Time Interval Distribution for Rainfall », *Journal Hydraulics Div. Proceedings, ASCE*, July 1959, pp. 83-91.
- [9] Thom, H.C.S., « A Rational Method of Determining Summer Weather Design Data », *Trans. ASHRAE*, Vol. 66, 1960, pp. 248-263.

SUR QUELQUES PRINCIPES DE CONSTRUCTION DE MODELES ECONOMETRIQUES (*)

par J. PAELINCK

Faculté des Sciences Economiques et Sociales, Namur.

*« La République des Benauges
est une démocratie spécifique,
laïque et statistique ».
Robert Escarpit, Honorius, Pape.*

I. Introduction.

La construction de modèles économétriques de croissance, d'exploration, d'optimisation et de contrôle devient une activité de plus en plus courante.

Nous y avons versé aussi occasionnellement; les pages qui suivent sont le fruit à la fois de notre expérience et de nos lectures. Elles n'ont rien d'original et visent uniquement à attirer l'attention sur certaines techniques ou « sous-routines » qui peuvent constituer des auxiliaires précieux ou des garde-fous indispensables au cours du processus de modélisation de l'économie.

Nous en développons cinq ci-après.

II. Théorie économique.

Keynes a défini l'économétrie comme une espèce d'alchimie statistique (1); en quoi il se trompe, car l'économétrie résulte en réalité d'une synthèse ternaire de la théorie économique, de l'appareil mathématique — au sens le plus large — et de la statistique mathématique.

A ce titre, la théorie économique — coulée en forme mathématique — fournit un certain nombre d'hypothèses sur les réactions des sujets économiques.

(*) Conférence donnée à la Société Belge de Statistique et de Recherche Opérationnelle, le 28 mars 1968.

(1) « Econometrics, this brand of statistical alchemy ... ».

Ces hypothèses — théorèmes opératoires, à vérifier dans les faits — peuvent nous guider dans la formulation de certaines relations ou dans l'estimation de certaines variables ou paramètres.

Les auteurs les appliquent avec un bonheur inégal, comme le montreront les exemples qui suivent.

A. Forme de la courbe du coût total.

Rappelons qu'il s'agit déjà d'une construction théorique de second ordre, dérivée des fonctions de production et d'une hypothèse sur la minimisation des coûts.

De nombreuses tentatives de mesure ont été faites (Dean, Johnson, Lyle); le plus souvent, une droite est retenue comme ajustement final.

Les explications de ce phénomène sont les suivantes. Souvent, le domaine de variation des données ne couvre pas les portions décroissantes ou croissantes de la courbe des coûts marginaux, mais seulement une zone autour d'un point d'inflexion de la courbe totale, zone approchable par une droite. En d'autres mots, la courbe du coût marginal a un minimum *aplati*, ajustable par une droite parallèle à l'axe des quantités.

De plus, les courbes empiriques sont le résultat d'une recherche de flexibilité et de possibilité d'adaptation dans le cycle conjoncturel; la firme recherche une technique pour laquelle elle est raisonnablement efficiente dans un domaine relativement large de variation des quantités produites et écoulées; au-delà de ce domaine, les coûts moyen et marginal peuvent croître assez rapidement.

L'analyse statistique sera donc souvent impuissante à détecter ces caractéristiques. Les arguments techniques empruntés par la science économique pour décrire un établissement, ou la constatation de la multiplicité d'établissements dans un marché concurrentiel, incitent à choisir souvent (2) la courbe Edgeworthienne pour représenter ou ajuster les coûts totaux à court terme.

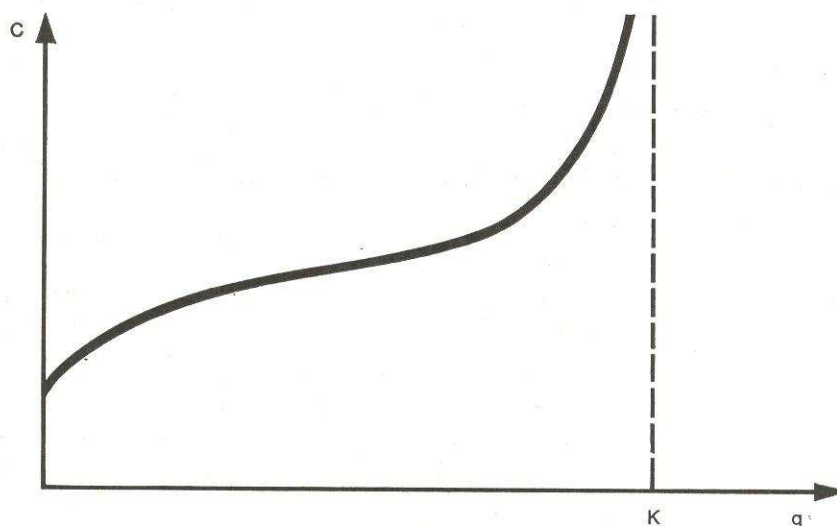
Une forme adéquate sera, par exemple, une sigmoïde inversée, telle la courbe de fréquence cumulée lognormale (Λ), les axes étant inversés (graphique 1).

L'expression analytique en serait

$$\frac{q}{\kappa} = \frac{1}{C \sqrt{2\pi}\sigma} \int_0^{C^*} e^{-\frac{[\log(C-a) - \mu]^2}{2\sigma^2}} dC \quad [1]$$

(2) Pas toujours !

courbe lognormale tétra-paramétrique (paramètres μ , σ^2 , a , κ); a représente les coûts fixes, κ la capacité de production maximale, μ et σ^2 les paramètres caractéristiques des distributions normales.



Graphique 1

Une application en a été faite par Klein (3) au cas des centrales électriques thermiques au charbon du nord-est des Etats-Unis. Si pareille courbe est estimée sur coupes instantanées (« cross-section »), elle sera représentative de la courbe de coût individuelle; les propriétés de reproduction de Δ permettent de l'agréger facilement sur l'ensemble des firmes si les coûts sont distribués lognormalement (4).

Si les considérations de théorie économique — appliquées au domaine d'investigation particulier du chercheur — ne rejettent pas l'hypothèse d'une courbe Edgeworthienne, l'extrapolation des coûts en dehors du domaine d'ajustement est légitime, mais, répétons-le, cette conclusion est un *cas d'espèce*.

B. Elasticité de substitution d'une fonction de production.

Rappelons que l'élasticité de substitution d'une fonction à deux arguments — par exemple K , le capital, et L , le travail — est définie comme

$$\sigma = \frac{d \log (K/L)}{d \log r} \quad [2]$$

(3) L. Klein, *An Introduction to Econometrics*, Prentice-Hall, 1962, pp. 121-125.
 (4) J. Aitchison and J.A.C. Brown, *The lognormal distribution*, C.U.P., 1957.

où K/L représente « l'intensité capitalistique » (rapport capital-travail) et r le taux marginal de substitution — dK/dL .

Elle est inversement proportionnelle à la dérivée seconde de la courbe $K = g(L)$, et, comme « coefficient inverse de courbure », mesure les facilités de substitution entre facteurs de production.

Dans la fonction classique à élasticité de substitution constante (version SMAC (5) de la CES)

$$P = \gamma [\delta K^{-\rho} + (1 - \delta) L^{-\rho}]^{-1/\rho} \quad [3]$$

le paramètre ρ est relié à σ par la relation $\rho = \sigma^{-1} - 1$ (6).

L'estimation de ρ est *a priori* assez complexe; aussi les auteurs SMAC proposent-ils d'utiliser comme estimateur de ρ l'expression $s^{-1} - 1$, où s est le paramètre de pente dans

$$\log \frac{P}{L} = s \log w^* + C^{te} \quad [4]$$

w^* étant le salaire réel.

On a en effet

$$s = \frac{d \log (P/L)}{d \log w^*} = \frac{d (P/L) L P'_L}{d P'_L P} = \frac{P'_K d K \cdot P'_L}{P'_{LK} d K \cdot P} = \sigma \quad [5]$$

dans le cas de fonctions homogènes de degré 1, ce qui est vérifié par [3].

Le pas décisif dans le raisonnement est l'assimilation de w^* et P'_L (7); cette hypothèse microéconomique classique n'est justifiée que sous des conditions d'optimisation et de structure de marché (concurrence pure et parfaite) strictes. Lorsque l'on ajuste la fonction [4] par « cross-section » internationale — comme l'ont fait les auteurs SMAC — il n'est pas du tout certain que ces conditions soient satisfaites et que, de plus, les secteurs se trouvent dans chaque pays au même stade d'adaptation vers l'équilibre.

(5) Du nom des quatre auteurs de l'article : Arrow, K.J., Chenery, H.B., Minhas, B.S., Solow, R.M., « Capital-Labor Substitution and Economic Efficiency », Review of Economics and Statistics, XLIII, 1961, N° 3.

(6) γ est un paramètre d'efficacité, δ un paramètre de distribution des revenus entre facteurs; ils est d'ailleurs dimensionnellement non homogène par rapport à $(1 - \delta)$, comme on s'en apercevra d'après les principes de la Section III.

(7) P'_L est la dérivée de [3] par rapport à L , soit la productivité marginale physique du travail; de même, P'_{LK} est la dérivée seconde croisée.

Ceci explique (8) les résultats pour le moins bizarres obtenus; le tableau 1 ci-après en sélectionne certains en les arrondissant.

TABLEAU 1

Secteurs	σ
— Verre	1,00
— Bois	0,90
— Ciment	0,90
— Fabrications métalliques	0,90
— Chimie	0,85
— Sidérurgie	0,80
— Textiles	0,80

Le coefficient 0,90 semble plus vraisemblable pour l'industrie du bois que pour les fabrications métalliques et le ciment; de même, le coefficient 0,80 semble moins vraisemblable pour les textiles que pour la sidérurgie (ou même le coefficient de 0,85 pour la chimie). Ces observations ne répondent en effet pas aux connaissances techniques que nous avons des possibilités respectives des substitutions de facteurs dans ces secteurs.

Aussi préférons-nous à cette méthode d'estimation les approches plus proprement statistiques proposées récemment (9).

C. Estimation de l'emploi lié à un programme d'investissement.

Les difficultés dans lesquelles se débat encore la théorie des fonctions de production en régime de croissance ont fait rechercher des modèles simplifiés pour l'estimation conjointe de l'emploi et de l'investissement dans des modèles d'exploration.

Graham Pyatt (10) développe ainsi le sous-modèle suivant.

(8) En partie; pour une critique statistique, voir V.R. Fuchs, Capital-Labor Substitution: A Note, Review of Economics and Statistics, 1963, N° 4, pp. 436-438.

(9) R. Thoss, Estimation des Fonctions de Production CES: une comparaison, Institut für Statistik und Oekonometrie an der Universität Mannheim, Faculté des Sciences Economiques et Sociales de Namur, Travaux d'économie mathématique et de recherche opérationnelle; traductions en vue de diffusion et de discussion interuniversitaires, Série Française, Document N° 1/68.

(10) In: Economic Growth and Manpower, BACIE Spring Conference, 1963, p. 37.

Soit ΔY l'augmentation des capacités d'une industrie déterminée pendant une période, X l'investissement brut, S la mise au rebut. On aura

$$\Delta Y = X - S \quad [6]$$

Soit également ΔL l'augmentation d'emploi, N l'emploi correspondant à X , R les réductions d'emploi correspondant à S . On aura

$$\Delta L = N - R \quad [7]$$

Soit p le prix de l'output, égal au coût du travail et du capital par unité d'output, et soit w le taux des salaires. Multipliant [6] par p , [7] par w , et soustrayant, on obtient

$$p \Delta Y - w \Delta L = (p X - w N) - (p S - w R) = p X - w N \quad [8]$$

si $p S = w R$, en d'autres mots si le matériel investi est mis au rebut dès qu'il cesse de gagner un revenu au capital.

Si nous admettons cette hypothèse, les projections d'emploi liées à un programme d'investissement peuvent être effectuées comme suit.

Soit r le taux de rendement brut du nouveau capital investi pendant la première année

$$r = \frac{p X - w N}{I} = \frac{p \Delta Y - w \Delta L}{I} \quad [9]$$

d'où

$$\Delta L = \frac{p \Delta Y - r I}{w} \quad [10]$$

en supposant r constant dans une projection à prix relatifs constants.

Le modèle est astucieux, mais son emploi dépend des conditions en vigueur dans l'industrie étudiée. Certaines industries gardent un matériel qui, par une température conjoncturelle ordinaire, ne serait pas rentable, mais le devient à prix croissants (cas de certaines centrales électriques en Belgique); d'autres, se servant de certains critères d'investissement (11), se débarrassent de leur matériel avant qu'il ait cessé de leur gagner un profit.

Dans des conditions pareilles, des sur- et sous-estimations de ΔL sont possibles. Les conditions empiriques de validité des hypothèses de comportement des entreprises sont donc à vérifier soigneusement.

(11) Type analyse de « challenger » (Joël Dean, MAPI).

D. Conclusion.

On a vu que la théorie économique, utilisée en économétrie, pouvait être la meilleure et la pire des choses. Son emploi judicieux guide la tâche du constructeur de modèles et souvent même la facilite. Elle lui tend cependant des pièges auxquels — par paresse d'esprit — il n'échappe pas toujours.

Comme quoi la pratique économétrique suppose de bons réflexes.

III. Analyse dimensionnelle (12).

L'analyse dimensionnelle est une branche de l'algèbre qui étudie des grandeurs sous l'angle de leurs dimensions. Ces dernières sont définies comme des ensembles de grandeurs considérées comme additives.

Classique en physique, l'analyse dimensionnelle a été récemment introduite en économie, en attendant de passer le test de l'économétrie.

De simples considérations dimensionnelles permettent cependant déjà de contrôler l'exactitude de certains calculs. Ainsi le coefficient moyen de capital a-t-il la dimension du temps, comme l'atteste la dérivation suivante

$$x_M = \frac{K}{P} \in \left[\frac{M}{M T^{-1}} \right] = [T] \quad [11]$$

Il a, par conséquent, une unité de mesure et sa valeur numérique changera avec l'unité choisie; ainsi un coefficient calculé sur base de données de production annuelles devra être multiplié par quatre si on l'utilise dans un modèle trimestriel (13).

L'analyse dimensionnelle proprement dite repose essentiellement sur le théorème de *Buckingham*. Celui-ci établit qu'une équation est dimensionnellement homogène (14) si et seulement si elle peut être réduite à un ensemble complet de produits non dimensionnels *indépendants*.

(12) Voir l'ouvrage récent de F.J. de Jong, *Dimensional Analysis for Economists*, Rotterdam, North Holland Publishing Cy, 1967.

(13) Pour un autre exemple simple, voir F.J. de Jong, *L'Analyse Monétaire*, élaborée par la Nederlandse Bank, Bull. d'Infor. et de Docum. de la B.N.B., 1965, Vol. 1, N° 4-5.

(14) C'est-à-dire que le membre de gauche et de droite ont la même dimension, ce qui implique que l'équation est invariante pour un changement d'unité de mesure dans les dimensions fondamentales. Celles-ci sont M et T dans l'expression [11], et la dimension de x_M , dimension composée, en a été dérivée selon une formule dimensionnelle spécifique : $D = \prod_i D_i^{a_i}$.

C'est surtout par son corollaire que ce théorème est intéressant; il permet d'orienter la recherche de la forme fonctionnelle correcte d'une équation par la résolution d'un système d'équations linéaires homogène

$$A x = 0 \quad [12]$$

où A est la matrice des exposants des dimensions fondamentales entrant dans la dimension éventuellement dérivée de chaque variable du problème. Ceci permet à l'économetre de dégager la forme fonctionnelle *a priori* inconnue d'une relation entre différentes variables; nous verrons cependant que, là encore, la théorie économique joue son rôle.

Supposons par exemple que, dans le courant de la construction d'un modèle économétrique, l'on éprouve le besoin d'insérer une équation de demande de travail. Le modèle comporte déjà une fonction de production que nous supposons être une Cobb-Douglas

$$P = c L^{\alpha} K^{1-\alpha} \quad [13]$$

Par ailleurs, la connaissance de la théorie économique permet de poser comme hypothèse de travail que la demande de travail est fonction du capital installé, du prix du produit, du salaire nominal et du progrès technique, soit

$$F(L; K, p, w, c) = 0 \quad [14]$$

Choissant les dimensions fondamentales R_L , R_K , R_P , M et T (15), soit quatre dimensions « stocks » et la dimension « temps », on dérive

$$\left\{ \begin{array}{l} L \in [R_L T^{-1}] \\ K \in [R_K] \\ p \in [M R_P^{-1}] \\ w \in [M R_L^{-1}] \\ c \in [R_P R_T^{-\alpha} R_K^{\alpha-1} T^{\alpha-1}] \end{array} \right. \quad (16) \quad [15]$$

La matrice dimensionnelle est

$$\begin{array}{c} \begin{array}{ccccc} & L & K & p & w & c \\ \begin{array}{c} M \\ R_P \\ R_T \\ R_K \\ T \end{array} & \left[\begin{array}{ccccc} 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & -1 & 0 & 1 \\ 1 & 0 & 0 & -1 & -\alpha \\ 0 & 1 & 0 & 0 & \alpha-1 \\ -1 & 0 & 0 & 0 & \alpha-1 \end{array} \right] \end{array} \end{array} \quad [16]$$

(15) Le principe du choix de ces dimensions ne sera pas abordé ici.

(16) Depuis la fonction [13].

et la résolution du système [12] donne les exposants du seul produit non dimensionnel (17)

$$\pi = L K^{-1} p^{(a-1)^{-1}} w^{(1-a)^{-1}} c^{(a-1)^{-1}} \quad (18) \quad [17]$$

d'où

$$L = \pi K \left(\frac{p c}{w} \right)^{(1-a)^{-1}} \quad [18]$$

L'équation de demande de travail est donc déterminée à une constante près. L'analyse économique permet d'en dégager la valeur; si nous admettons en effet une hypothèse de minimisation de coûts dans un marché de concurrence pure et parfaite, le travail sera utilisé de telle façon que sa productivité marginale en valeur soit égale à son salaire nominal, soit

$$p \frac{\partial P}{\partial L} = w \quad [19]$$

ce qui, combiné avec [13], permet de dériver

$$L = \alpha^{(1-a)^{-1}} K \left(\frac{p c}{w} \right)^{(1-a)^{-1}} \quad [20]$$

et qui, dans ce cas particulier, implique

$$\pi = \alpha^{(1-a)^{-1}} \quad [21]$$

La comparaison de [18] et [20] permet d'apprécier la contribution de l'analyse dimensionnelle; elle a donné, avec un minimum d'hypothèses, une forme fonctionnelle égale à la forme dérivée de la théorie économique complète et, de plus, réduit le problème économétrique à la seule estimation du paramètre non dimensionnel π . Dans le cas où l'on ne désire pas — pour des raisons de fait — s'encombrer de trop d'hypothèses, l'analyse dimensionnelle fournit une solution aisée et élégante.

Ce sera encore le cas si aucune théorie valable n'existe; on essaiera alors, par analyse dimensionnelle, de trouver une relation fonctionnelle entre certaines variables sélectionnées par induction économique. Si aucune solution n'existe, on sera de toute façon sur la voie des variables manquantes; le choix de celles-ci est donc facilité.

Toutes les potentialités de l'analyse dimensionnelle en économétrie n'ont pas encore été exploitées; l'on aurait à étudier dans quelle mesure

(17) A est ici de rang 4.

(18) π est la valeur de ce produit.

elle est efficace dans la suppression de difficultés typiquement économétriques (multicollinéarité, autocorrélation).

Nous sommes d'avis qu'elle deviendra un auxiliaire de plus en plus important dans la construction de modèles économétriques.

IV. Comptage d'équations.

Rappelons qu'en termes mathématiques il s'agit de rechercher la compatibilité d'un système d'équations avec une solution. La simple égalité du nombre de variables endogènes (inconnues) et du nombre d'équations n'est d'ailleurs pas une condition suffisante pour l'existence d'une solution; des cas d'inexistence ou d'existence de solutions multiples peuvent se présenter dans des systèmes non linéaires.

Le modèle à un ou plusieurs degrés de liberté (modèle « ouvert ») est bien connu; nous le rappellerons ci-après pour donner quelques précisions. Le modèle à degrés de liberté « négatifs » est moins courant; nous le traitons plus loin.

A. Modèles ouverts.

Comparons les modèles de Domar et de Harrod (19).

TABLEAU 2

Domar		Harrod	
$I = I_0 e^{rt}$	[23]	---	
$\dot{Y} = \dot{I}/s$	[24]	$S = s Y$	[24']
$\dot{P} = I/v$	[25]	$I = v \dot{Y}$	[25']
$\dot{Y} = \chi \dot{P}$	[26]	$I = \chi S$	[26']
$\bar{L} = \bar{L}_0 e^{ct}$	[27]	idem	
$L = k Y^b$	[28]	idem	
$L = \rho \bar{L}$	[29]	idem	

(19) Cfr W. van Ryckeghem, De aanwending van macro-economische groeimodellen in alternatieve economische situaties, *Revue Belge de Statistique et de Recherche Opérationnelle*, 1964, vol. 4, N° 4, pp. 3-13.

L'équation [23] définit l'évolution des investissements, l'équation [24] l'effet de revenu des investissements, l'équation [25] leur effet de capacité; l'équation [26] définit le degré d'équilibre monétaire. L'équation [27] retrace l'évolution de l'offre de travail, la fonction [28] est celle de la demande de travail; enfin, l'équation [29] définit le degré d'équilibre sur le marché du travail.

Le modèle de Harrod est caractérisé par l'absence de l'équation [23]; l'équation [24'] définit l'offre d'épargne, l'équation [25'] sa demande; [26'] est une autre formulation de l'équilibre monétaire.

Le modèle de Domar est « fermé » dans les variables endogènes Y , P , I , L , χ et ρ ; il interdit une politique *instrumentale*, mais permet de définir une politique *paramétrique*, par exemple en termes de r , le taux de croissance des investissements; le principe général Tinbergenien reste cependant vrai : à lui seul, le paramètre r est impuissant à contrôler à la fois deux objectifs de politique économique, par exemple les degrés d'équilibre χ et ρ . Il existe dès lors une fonction $b(\chi, \rho) = 0$, paramétrique implicite en r . Le double contrôle devient possible, par exemple par l'addition de la variable de contrôle s ; dans ce cas, en principe, tout couple (χ, ρ) est contrôlable.

Le modèle de Harrod, possédant un degré de liberté, permet un contrôle « mixte » du couple (χ, ρ) , par exemple par I et s . On voit cependant combien les distinctions sont formelles; en pratique, I peut être considéré comme exogène dans l'équation [23] du modèle de Domar. Si nous esquissons ces distinctions, c'est pour insister sur le fait qu'il ne faut pas limiter la recherche des instruments de contrôle aux seules variables; des paramètres du modèle peuvent jouer ce rôle. Ceci peut être utile dans le « design » de certains modèles utilisés à des fins de politique économique.

B. Modèles surdéterminés.

Des enseignements intéressants peuvent être tirés de modèles surdéterminés, ainsi que le montre le modèle développé ci-dessous. Il s'agit du modèle des Nations-Unies (20) relatif aux conditions de développement du Tiers-Monde; il compte quatre groupes d'équations.

1. Groupe 1 : équation technique (production)

$$Q(t) = f\left(\sum_{\tau=1}^t I_{\tau}\right) \quad [30]$$

(20) Nations-Unies, Studies in long term economic projections for the world economy; Aggregative models, Doc. E/3842/51/ECA/80.

2. Groupe 2 : équations d'exportation

$$X = X_A + X_B \quad [31]$$

$$X_A = a_1 (Q_A)^{b_1} \quad [32]$$

$$X_B = a_2 c^{b_2 t} \quad [33]$$

$$E_1 = a_3 + b_3 M \quad [34]$$

$$E_2 = a_4 + b_4 (p_x X) \quad [35]$$

$$C = F(t, P) \quad [36]$$

$$Z = \frac{p_x}{p_m} X + \frac{E_1}{p_m} + \frac{E_2}{p_m} + \frac{C}{p_m} \quad [37]$$

Les équations [32] et [33] sont les fonctions d'exportation vers les pays développés et les pays socialistes, respectivement. Les équations [34] et [35] ont trait aux exportations nettes à prix courants des services et aux revenus nets de facteurs. [36] définit les transferts de capitaux (P représente les variables politiques qui entrent dans leur détermination). Enfin, l'équation [37] définit la capacité d'exportation.

3. Groupe 3 : équations d'importation

$$M = Z \quad [38]$$

$$I = a_5 + b_5 M_I \quad [39]$$

$$M_I = M - M_C \quad [40]$$

$$M_C = a_6 + b_6 Q \quad [41]$$

Ces équations sont d'interprétation évidente, si l'on précise que M_I et M_C sont les importations de biens d'investissement et de consommation, respectivement.

4. Groupe 4 : équilibre financier

$$S = I \quad [42]$$

$$S \equiv C + D \quad [43]$$

$$D = a_7 + b_7 Q \quad [44]$$

Notons que D représente l'épargne domestique.

(α est fonction des niveaux d'équilibre de Z et C); or, il est peu vraisemblable que [45], le niveau d'épargne domestique requis ex-post, soit compatible avec l'épargne domestique formée ex-ante (équation [44]). Le modèle offre alors deux possibilités d'interprétation :

a) via un sous-modèle à limitation des réserves de change extérieur; dans ce cas, l'épargne domestique doit s'ajuster au taux de croissance implicite dans le modèle, et l'équation [44] devient redondante. La différence entre D calculé par [45] et [44] donne le « savings-gap »;

b) via un sous-modèle qui élimine l'équation [36]; dans ce cas, C , les transferts de capitaux, devront s'adapter au taux de croissance implicite du modèle; la différence entre C calculé via [45] — rappelons que $\alpha = g(C)$ — et via [36] donne le « foreign-exchange gap ». L'équation [39] est là pour montrer combien ce frein à la croissance du tiers-monde peut être puissant (cfr aussi [38]).

Ceci est, en résumé, la mécanique des « two-gap models » (modèles à double déficit). L'idée peut être généralisée à d'autres situations où des écarts entre variables ex-ante et ex-post sont susceptibles de se produire. La construction d'un modèle surdéterminé permet aussi de déceler l'écart le plus « actif » et de mesurer son effet sur la solution, un peu à la façon d'une variable duale dans des modèles d'optimisation.

V. Estimation de paramètres.

Les modèles, surtout une fois qu'ils sont sectorialisés, comprennent un nombre considérable de paramètres; le cas de 40.000 paramètres n'est pas rare (21).

La tâche d'estimer pareil nombre de coefficients serait titanesque, si des procédures limitatives des degrés de liberté du problème n'existaient pas.

Ci-dessous, nous donnerons deux exemples de leur *modus operandi*.

A. Elasticités des fonctions de demande.

Supposons que l'on distingue n catégories de consommation des ménages (alimentation, boissons, tabac, habillement, etc.); le problème consiste à estimer

(21) Cfr J.S. Cramer et L. Mennes, Estimation in medium term econometric models: the expert's practice, in *Modelli econometrici per la Programmazione*, Firenze, Scuola di Statistica, 1965, pp. 351-358.

Il y a en tout 15 équations en 14 variables endogènes ($Q, F, X, X_A, X_B, E_1, E_2, C, M, M_I, M_C, S, D$). Une interprétation intéressante est la suivante.

Considérons les équations [38] à [43]; elles permettent d'exprimer D comme une fonction linéaire de Q , par exemple

$$D = \alpha + \beta Q \quad [45]$$

$n(n+1)$ élasticités de la demande (n^2 élasticités-prix directes et croisées, n élasticités-dépense).

Certaines relations permettent de réduire le nombre de degrés de liberté au départ. Ainsi en est-il :

— de la relation d'agrégation d'Engel, retraçant un lien linéaire entre élasticités-dépense :

$$\sum_i a_i E_i = 1 \quad (22) \quad [46]$$

— de la relation d'agrégation de Cournot, retraçant un lien entre des élasticités-prix :

$$\sum_i a_i E_{ij} + a_j = 0 \quad [47]$$

— de la relation d'agrégation d'Euler, reliant élasticités-prix et dépense d'une fonction de demande :

$$\sum_j E_{ij} + E_i = 0 \quad [48]$$

— de la condition de Slutsky, relative à la symétrie de la matrice de substitution :

$$\sigma_{ij} = \sigma_{ji} \quad [49]$$

Ces relations fournissent un nombre de contraintes de l'ordre de $\frac{n(n+2)}{2}$ (23). D'autres contraintes, déduites des conditions du second ordre et relatives au Hessien bordé et ses mineurs principaux, pourraient encore limiter les degrés de liberté.

Ces conditions ne fournissent en principe pas de forme fonctionnelle pour les équations de la demande, et sont insuffisantes pour l'estimation de l'ensemble des paramètres souhaités.

(22) Les a_i représentent les parts budgétaires.

(23) Elles ne sont pas toutes indépendantes les unes des autres.

Des hypothèses spécifiques doivent être introduites. Elles ont trait, par exemple, aux caractéristiques du système d'utilités du consommateur; ainsi l'hypothèse de « want-indépendance » de Frisch permet de limiter le problème d'estimation à n paramètres (24).

Les limitations du système ainsi résultant, et des objections empiriques (25), nous ont fait préférer le système linéaire de dépense de Stone-Samuelson, qui, de plus, est très performant pour les projections; il impose l'estimation de $2n - 1$ paramètres dans le système

$$p_i q_i = p_i \bar{q}_i + b_i (d - \sum_j p_j \bar{q}_j) \quad [50]$$

où $\bar{q}_i$ représente des quantités minimales achetées quel que soit le prix en vigueur, et $(d - \sum_j p_j \bar{q}_j)$ le revenu excédentaire après satisfaction des dépenses minimales. Le modèle souffre cependant de limitations (substituabilité et inélasticité-prix généralisées).

A ce titre, le modèle de Leser-Hilhorst-Sommerey (26)

$$p_i q_i = \frac{c_i (p_i/d)^{a_i} d}{\sum_j c_j (p_j/d)^{a_j}} \quad [51]$$

est intéressant, puisqu'il ne souffre pas de ces limitations et que, parfaitement additif, il n'impose que l'estimation de $2n$ paramètres indépendants; ses performances prévisionnelles sont cependant encore à l'étude.

Enfin, d'autres systèmes, fonctionnels ou non, sont à l'étude; un immense domaine d'analyse reste encore à défricher, tant du point de vue des implications théoriques que de celui de l'opérationalité empirique.

B. Modèles interrégionaux.

L'analyse et la programmation multirégionale supposent connues une ou plusieurs matrices de flux interrégionaux. Les statistiques font le plus souvent défaut.

(24) R. Frisch, A Complete Scheme for Computing All Direct and Cross Demand Elasticities in a Model with many Sectors, *Econometrica*, 1959, pp. 177-196.

(25) Voir J. Paelinck, « Fonctions de Consommation pour la Belgique 1948-1959, Contribution à l'étude des systèmes linéaires de dépenses », Travaux de la Faculté des Sciences Economiques et Sociales de Namur, 1964; et « Sensibilité de projections à moyen terme de fonctions de consommation à différents systèmes d'ajustement statistique », *Economie Appliquée*, Paris, 1966.

(26) Somermeyer, W.H. (met medewerking van Hilhorst, J.G.M. en Wit J.W.W.A.), een methode voor de schatting van prijs- en inkomenselasticiteiten uit tijdreeksen en haar toepassing op consumptieve uitgaven in Nederland 1949-1959, *Statistische en Econometrische Onderzoekingen*, 1961, N° 4, pp. 205-228.

Nous développerons ci-dessous un petit modèle qui montre comment l'introduction d'hypothèses sur des facteurs de répartition et sur la forme fonctionnelle peuvent réduire considérablement le nombre de degrés de liberté, qui est de $n(n-1)$.

Soit X_{ij} un flux quelconque, X_i un total de ligne, M_j un total de colonne, $X_{..}$ le grand total, les totaux marginaux étant supposés connus.

1. *Etablissement de la formule générale.*

Si l'on pose que

$$X_{ij} = (A_i + B_i d_{ij}) X_i. \quad [52]$$

où d_{ij} représente la distance séparant les régions i et j , on obtient

$$\sum_j X_{ij} = X_i = (n A_i + B_i \sum_j d_{ij}) X_i.$$

d'où

$$A_i = \frac{1}{n} + B_i \sum_j \frac{d_{ij}}{n} \quad [53]$$

Par ailleurs

$$M_j = \sum_i X_{ij} = \frac{\sum_i X_i}{n} + \sum_i B_i X_i \left(\sum_j \frac{d_{ij}}{n} - d_{ij} \right) \quad [54]$$

soit, posant

$$\sum_j \frac{d_{ij}}{n} = \mu_i, \quad \mu_i - d_{ij} = \delta_{ij}, \quad \xi_i = \frac{\sum_i X_i}{n} \quad (27)$$

$$M_j - \xi_i = \sum_i B_i X_i \cdot \delta_{ij}, \text{ et posant } M_j - \xi_i = V_j$$

on obtient en notation matricielle

$$V = \Delta' \hat{X} b \quad [55]$$

où

V = vecteur des V_j

Δ' = matrice transposée des $[\delta_{ij}]$

$\hat{X}$ = vecteur diagonalisé des X_i .

b = vecteur des B_i .

(27) On vérifie $\sum_i \xi_i = X_{..}$, le grand total.

On vérifie $i' V = i' \Delta' \hat{X} b = 0$, vu que $i' \Delta' = 0$.

Partitionnant, on obtient

$$\begin{bmatrix} \mu^* \\ \mu_n \end{bmatrix} = \begin{bmatrix} \Delta^{*'} & \delta^* \\ \delta^{**'} & 0 \end{bmatrix} \begin{bmatrix} \hat{X}^* & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} b^* \\ b_n \end{bmatrix} \quad [56]$$

d'où $\mu^* = \Delta^{*'} \hat{X}^* b^* + \delta^* (x_n b_n)$

soit $b^* = \hat{X}^{*-1} \Delta'^{-1} [\mu^* - \delta^* (x_n b_n)]$

et $\mu_n = \delta^{**'} \hat{X}^* b^*$

qui vérifie

$$\begin{aligned} \mu_n &= \delta^{**'} \hat{X}^* b^* \\ &= \delta^{**'} \hat{X}^* \hat{X}^{*-1} [\mu^* - \delta^* (x_n b_n)] \end{aligned}$$

et vu que $\delta^{**'} = -i' \Delta'^*$

on obtient

$$\begin{aligned} \mu_n &= -i' \Delta'^* \Delta^{*-1} [\mu^* - \delta^* (x_n b_n)] \\ &= -i' \mu + i' \delta^* x_n b_n = -i' \mu, \end{aligned}$$

comme il se doit, et vu que $i' \delta^* = 0$, en vertu de la relation $i' \Delta' = 0$.

2. Le choix de b_n se fera de sorte à rendre positifs les X_{nj} .

Si I est la matrice ne comprenant que des valeurs unitaires, sauf dans la diagonale principale, il y a lieu de vérifier que

$$[X_{ij}]' = [X_{ji}] = n^{-1} I \hat{X} + \Delta' \hat{X} \hat{b} > 0$$

ce qui revient à prouver que $I \geq -n \Delta' \hat{b}$

ou, partitionnant

$$\begin{bmatrix} I^* & i \\ i' & 0 \end{bmatrix} \geq n \begin{bmatrix} \Delta^{*'} & \delta^* \\ \delta^{**'} & 0 \end{bmatrix} \begin{bmatrix} \hat{b}^* & 0 \\ 0 & b_n \end{bmatrix}$$

soit les relations

$$I^* \geq n \Delta^{*'} \hat{b}^*$$

$$i \geq n \delta^* b^*$$

étant connu que $i' \geq (n b_n) \delta^{**'}$.

Pour des b tous positifs, nous ne pouvons qu'esquisser une preuve au sujet de $i \geq n \delta^* \hat{b}^*$.

Sachant que

$i \geq (n b_n) \delta^{**}$, l'on prouverait $i \geq n \delta^* b^*$, si l'on pose

$$(b_n) \delta^{**} \geq \hat{\delta}^* b^* = \hat{b}^* \delta^*.$$

Ceci impliquerait la preuve de

$$\hat{\delta}^{**} \hat{\delta}^* \leq \hat{b}^{*-1} b_n$$

où la division par $\hat{a}^{**}$ laisse le sens de l'inégalité intact si l'on suppose que la région n est la moins éloignée de toutes les autres.

Cette inégalité est satisfaite pour tous les $B_1 \geq 0$ dans le cas de $\delta_{ij}^* < 0$. Les autres cas ($\delta_{ij}^* > 0$) ne posent pas de problème, puisque dans ce cas X_{ij} est automatiquement positif (cfr. l'expression de X_{ij} en fonction de B_1 et δ_{ij}).

Pour $I^* \geq n \Delta^* \hat{b}^*$, on répéterait la preuve colonne par colonne.

3. Un exemple montrera le fonctionnement du modèle.

Soit la matrice de distances et les marges $X_{i.}$ et $M_{.j}$ suivantes :

	A	B	C	$X_{i.}$
A	0	1	3	30
B	1	0	4	20
C	3	4	0	10
$M_{.j}$	20	20	20	60

Les équations [53] et [55] sont respectivement :

$$\begin{aligned}
 [53] \quad & A_1 = \frac{1}{2} + 2 B_1 \\
 & A_2 = \frac{1}{2} + 2,5 B_2 \\
 & A_3 = \frac{1}{2} + 3,5 B_3 \\
 & 30 B_2 + 5 B_3 = 5
 \end{aligned}$$

$$[55] \quad \begin{aligned} 30 B_1 &= 5 B_3 \\ 30 B_1 + 30 B_2 &= 5 \end{aligned}$$

où l'on voit que les équations [55] ne sont pas indépendantes.

Si l'on suppose le flux $(C B) = 0$, l'on résout facilement le système et l'on obtient les valeurs suivantes pour B_i et A_i :

$$\begin{aligned} B_1 &= 1/6 & A_1 &= 5/6 \\ B_2 &= 0 & A_2 &= 1/2 \\ B_3 &= 1 & A_3 &= 4 \end{aligned}$$

et la matrice des flux se présente comme suit :

	A	B	C	$X_{i.}$
A	0	20	10	30
B	10	0	10	20
C	10	0	0	10
$M_{.j}$	20	20	20	

4. Un certain nombre de problèmes sont soulevés par ce modèle.

a) Tout d'abord, dans la logique du système, $B_i > 0$ (28). L'on peut toujours obtenir B_i , l'estimation indépendante, positive; si un ou plusieurs des autres B_i se révélaient négatifs, l'un ou l'autre des deux phénomènes suivants se seraient produits :

- le modèle est exact, mais les données utilisées (totaux marginaux) ne sont pas correctes;
- le modèle n'est pas adapté au cas analysé.

Seule son application peut permettre d'approfondir le cas aberrant cité.

b) Il faut que le flux $X_{ij} > 0$. Le point 2 ci-dessus fournit une esquisse de preuve qu'il en est ainsi si $B_i > 0$.

c) L'on a insisté sur le fait que le modèle aidait à surmonter la difficulté d'obtention des statistiques de flux interrégionaux; il est, par essence, un modèle *résiduel*, applicable à un problème *tronqué*.

(28) Si $B_i > 0$, alors $A_i > 0$; cfr [53].

L'on commencera par estimer statistiquement le maximum de flux possibles; ce n'est qu'au tableau *réduit*, aux marges afférant aux seuls flux encore inconnus, que s'appliquera le modèle.

Cette méthode de *troncature* dans l'estimation de paramètres statistiques a donné d'excellents résultats dans d'autres cas (29).

d) Les d_{ij} enfin sont des distances *économiques* et non purement physiques. Ils peuvent varier de produit à produit, et tiendront compte, dans chaque cas, des différentes modalités de transport, de leur importance relative, des tarifs pratiqués, des difficultés du parcours, etc. Un soin tout particulier devra être apporté à leur estimation.

5. *En conclusion*, on a réduit le problème non tronqué à l'estimation d'un paramètre sur $n(n-1)$, soit un degré d'information nécessaire de

$$[n(n-1) - (2n-1)]^{-1} = [n^2 - 3n + 1]^{-1} \quad [57]$$

Dans le cas où on l'appliquerait aux provinces belges, ce degré d'information serait de $1/55 < 2\%$.

Le modèle est actuellement soumis au test et comparé à l'efficacité d'autres modèles, celui de programmation linéaire, par exemple; ce dernier est cependant incapable de fournir des flux symétriques.

VI. Structures particulières.

Un modèle est, en général, construit en vue d'applications futures; mais qu'il s'agisse de *prévisions*, de *politique économique à court terme* ou de *projections à moyen terme*, son pouvoir d'exploration sera fonction de sa *correction analytique*, surtout dans les deux derniers cas.

L'investigation des structures sous-jacentes aux équations permet de corriger certaines formulations douteuses ou de suggérer des formulations adéquates. Deux exemples illustreront cette affirmation.

A. Equation d'achat de biens durables (30).

Afin d'accroître la précision d'une prévision à court terme, l'on peut songer à introduire dans le modèle les résultats d'enquêtes (panels de consom-

(29) Cfr. J. Paelinck et J. Waelbroeck, « La méthode RAS de Cambridge, Application à l'étude de l'évolution des coefficients techniques du tableau inter-industriel belge », *Economie Appliquée*, 1963, N° 1.

(30) D.B. Suits, *The Theory and Applications of Econometric Models*, Athens, Center for Economic Research, Training Seminar Series, N° 3, 1963.

mateurs), ainsi qu'il a été fait dans le modèle R.S.Q.E. (31) de Michigan, qui incorporait les résultats du Survey Research Center.

L'incorporation, par cette procédure, d'un « baromètre » des intentions d'achat ne fournit cependant pas les causes explicatives des modifications d'intentions.

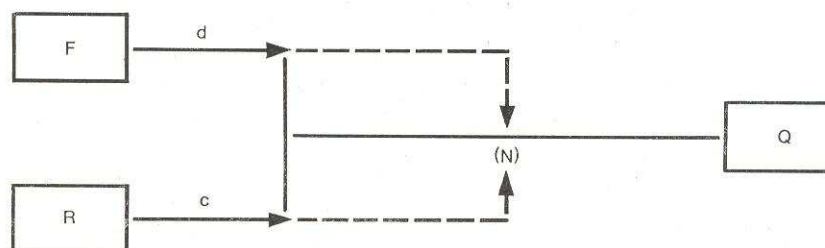
Quid si l'on essaie de raffiner l'analyse en introduisant les revenus disponibles, par voie de l'équation

$$Q = a + bN + cR \quad [58]$$

où N est la proportion de gens qui répondent affirmativement dans le cadre du Survey, et R le revenu disponible ?

L'erreur commise serait que N et R ne sont pas indépendants; c , en effet, ne représente qu'un coefficient marginal après élimination de l'effet bN .

La structure causale implicite dans [58] peut être schématisée comme suit :



Graphique 2

L'effet c est limité à la *modification partielle* des intentions originales par R , l'équation [58] combinant deux niveaux d'information :

- sur les stimuli : $e.a.R$;
- sur les réponses : N .

Analytiquement, ces variables ne peuvent être confondues, et une formulation correcte serait :

$$Q = a + cR + dF \quad [59]$$

(31) Research Seminar in Quantitative Economics.

la valeur 1 à λ , ce qui résulte dans la condition classique de « clearing » du marché

$$y_1 = y_2 \quad [62]$$

Si le prix est contrôlé, λ prend la valeur 0, et on obtient

$$y_3 = x_3 \quad [63]$$

La détermination de la valeur de λ se fait en comparant le prix calculé par le modèle et le prix réellement observé. Ainsi, par exemple, commencera-t-on par fixer $\lambda = 1$ en supposant un fonctionnement spontané du marché; on observe l'évolution des prix, pour finir par constater des écarts systématiques entre y_3 (observé) et y_3 (calculé), écarts que l'on suppose être dus à l'introduction d'un contrôle de prix. On corrige la trajectoire calculée par le modèle en posant $\lambda = 0$, jusqu'à l'observation de nouveaux écarts, supposés dus à l'abolition du contrôle des prix.

On voit donc comment la variable binaire λ peut formaliser des faits supposés en principe non modélisables.

VII. Conclusions.

Il est difficile de conclure sur ce qui précède.

Les quelques principes énoncés ne sont, ni des conditions nécessaires, ni surtout des conditions suffisantes, pour la construction de « bons » modèles économétriques.

Il est cependant raisonnable de les avoir « in the back of one's mind », même si l'on s'est déjà débattu dans le borbier statistico-mathématique qui peut transformer nos connaissances économiques en modèles opératoires.

Peut-être le débutant y aura-t-il trouvé quelque appui dans l'escalade de ce qui n'est certainement pas une colline parétienne des plaisirs; l'homme de métier pardonnera ce badinage qui a occupé agréablement une journée de vacances...

PUBLICATIONS REÇUES**ONTVANGEN PUBLICATIES**

- 1) Notas de logica mathematica, Nos 4, 5, 12, 13 à 22.
Notas de algebra y analisis, No 1.
Actas Jordanas 1958; Monografias mathematica 1 à 3. Instituto de Matematica. Universidad Nacional del Sur — Bahia Blanca (Argentina).
- 2) Boletin de la Sociedad Argentina de Investigacion Operativa, Buenos-Aires, No 18, 4ème trimestre 1967, 1º - 2º, 1968.
- 3) Annales de Sciences Economiques Appliquées, No 4, octobre 1968.
- 4) Journal of the Operations Research Society of Japan, Vol. 10, Nr 3 & 4, june 1968.
- 5) Trabajos de Estadistica y de Investigacion Operativa, Madrid, Vol. XIX, Cuadernos I y II.
- 6) Revue Française d'Informatique et de Recherche Opérationnelle, AFIRO, Paris, No 10, juillet 1968.
- 7) Cuadernos de Estadistica Aplicada e Investigacion Operativa, Barcelone, Vol. V, Fasc. 3, 1968.
- 8) Operationele Research Mededelingen, s' Gravenhage, Nr 9, september 1968.
- 9) Oost- Vlaanderen groeit, Nr 2, juni 1968.
- 10) Systèmes d'informatique Magazine (Bull-G.E.), No 2.
- 11) Revue IBN (Institut Belge de Normalisation) Nos 9, 10 et 11, 1968.
BIN Revue (Belgisch Instituut voor Normalisatie) Nrs 9, 10 en 11, 1968.
- 12) Revue Suisse pour l'Organisation Industrielle (Institut d'Organisation Industrielle de l'Ecole Polytechnique de Zurich), Nos 9 à 12, 1968.
- 13) Revue de Statistique appliquée, Vol. 16, No 3, 1968.
- 14) Operational Research Quarterly, London, Vol. 19, No 4, 1968.
- 15) Revue de l'Institut international de Statistique, La Haye, Vol. 36, No 3, 1968.