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'optimation 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 Edgworthienne pour représenter ou ajuster les coûts totaux à court terme.

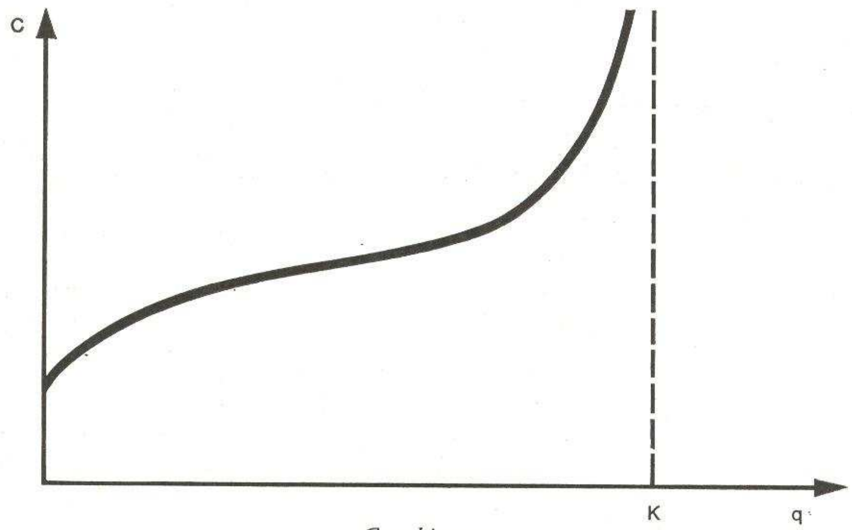
Une forme adéquate sera, par exemple, une sigmoïde inversée, telle la courbe de fréquence cumulée lognormale (Λ), les axes étant inversés (graphique 1).

L'expression analytique en serait

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

(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 Λ 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 Edgworthienne, 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

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

(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)

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

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

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

w^* étant le salaire réel.

On a en effet

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

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'optimation 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'efficience, 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

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

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

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

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

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

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

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

d'où

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

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

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

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 : $D = \prod_i D_i^{\alpha_i}$.

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

$$A\,x = 0 \qquad [12]$$

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'économètre 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 pous supposerons être une Cobb-Douglas

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

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

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

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

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

La matrice dimensionnelle est

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

(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)

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

d'où

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

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

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

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

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

et qui, dans ce cas particulier, implique

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

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 traiterons 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, $\bar{L}$, 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 $h(\chi, \rho) = 0$, 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)

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

(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

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

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

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

$$E_1 = a_3 + b_3 \, M \qquad [34]$$

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

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

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

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

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

$$I = a_5 + b_5 \, M_I \qquad [39]$$

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

$$M_C = a_6 + b_6 \, Q \qquad [41]$$

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

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

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

$$D = a_7 + b_7 \, Q \qquad [44]$$

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'optimation.

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

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

$n (n + 1)$ 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 :

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

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

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

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

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

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

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

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-independence » 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

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

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-Somermeyer (26)

$$p_i\, q_i = \frac{c_i\, (p_i/d)^{\alpha_i}\, d}{\sum_j c_j\, (p_j/d)^{\alpha_j}} \qquad [51]$$

est intéressant, puisqu'il ne souffre pas de ces limitations et que, parfaitement additif, il n'impose que l'estimation de $2\,n$ 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

$$X_{ij} = (A_i + B_i\, d_{ij})\, X_{i\,.} \qquad [52]$$

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

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

Par ailleurs

$$M_{.j} = \sum_i x_{ij} = \frac{\sum X_{i.}}{n} + \sum_i B_i\, X_{i\,.}\,(\sum_j \frac{d_{ij}}{n} - d_{ij}) \qquad [54]$$

soit, posant

$$\sum_j \frac{d_{ij}}{n} = \mu_i\,,\ \mu_i - d_{ij} = \delta_{ij}\,,\ \xi_i = \frac{\sum X_{i.}}{n} \qquad (27)$$

$$M_{.j} - \xi_i = \sum_i B_i\, X_{i\,.}\,.\,\delta_{ij}\,,\ \text{et posant } M_{.j} - \xi_i = V_j$$

on obtient en notation matricielle

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

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

Partitionnant, on obtient

$$\left[\frac{\mu^*}{\mu_n}\right] = \left[\begin{array}{c:c} \Delta^{*\prime} & \delta^* \\ \hline \delta^{**\prime} & 0 \end{array}\right] \left[\begin{array}{c:c} \hat{\mathrm{X}}^* & 0 \\ \hline 0 & 0 \end{array}\right] \left[\frac{b^*}{b_n}\right] \qquad [56]$$

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

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

et $\quad \mu_n = \delta^{**\prime} \hat{\mathrm{X}}^* b^*$

qui vérifie

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

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

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* $\underline{b}_n$ *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

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

ce qui revient à prouver que $\mathrm{I} \geqslant - n \Delta' \hat{b}$
ou, partitionnant

$$\left[\begin{array}{c:c} \mathrm{I}^* & i \\ \hline i' & 0 \end{array}\right] \geqslant n \left[\begin{array}{c:c} \Delta^{*\prime} & \delta^* \\ \hline \delta^{**\prime} & 0 \end{array}\right] \left[\begin{array}{c:c} \hat{b}^* & 0 \\ \hline 0 & b_n \end{array}\right]$$

soit les relations

$$\mathrm{I}^* \geqslant n \Delta^{*\prime} \hat{b}^*$$

$$i \geqslant n \hat{\delta}^* b^*$$

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

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

Sachant que

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

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

Ceci impliquerait la preuve de

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

où la division par $\hat{d}^{**}$ 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 $B_i \geqslant 0$ 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_i et δ_{ij}).

Pour $I^* \geqslant n\, \Delta^{*\prime}\ \hat{b}^*$, 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 :

$$A_1 = \tfrac{1}{2} + 2\ B_1$$

[53] $$A_2 = \tfrac{1}{2} + 2{,}5\ B_2$$

$$A_3 = \tfrac{1}{2} + 3{,}5\ B_3$$

$$30\ B_2 + 5\ B_3 = 5$$

[55] $$30\ B_1 = 5\ B_3$$

$$30\ B_1 + 30\ B_2 = 5$$

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 :

$$B_1 = 1/6 \qquad A_1 = 5/6$$

$$B_2 = 0 \qquad A_2 = 1/2$$

$$B_3 = 1 \qquad A_3 = 4$$

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_n, 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

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

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'efficience 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.

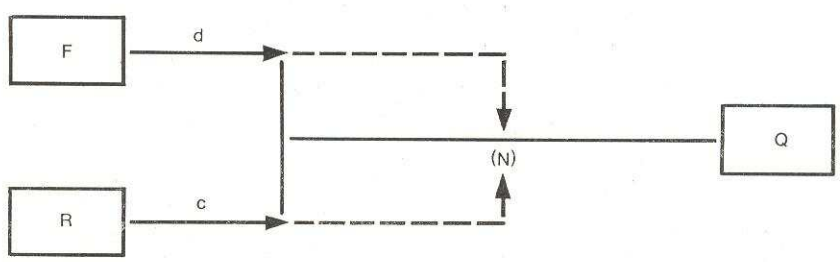
Quid si l'on essaie de raffiner l'analyse en introduisant les revenus disponibles, par voie de l'équation

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

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] cobinant 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 :

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

(31) Research Seminar in Quantitative Economics.

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

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

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

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

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 bourbier 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...