

1989-07

Les limites des modèles économétriques dans la prévision, pp. 1-29,

NIBITEGEKA Emmanuel

UB, Cahiers du CURDES

NIBITEGEKA Emmanuel, Les limites des modèles économétriques dans la prévision, pp. 1-29,
Cahiers du CURDES n° 7, Juin 1989.

<https://repository.ub.edu.bi/handle/123456789/626>

Téléchargé depuis le dépôt institutionnel officiel de l'Université du Burundi

LES LIMITES DES MODELES ECONOMETRIQUES DANS LA PREVISION (1).

Par Emmanuel NIBITEGEKA.

0. INTRODUCTION

La planification a été longtemps perçue comme une collection de voeux pieux que le hasard effectuerait tout seul ou un simple inventaire d'actions faisables ou souhaitables. Puis les techniques quantitatives ont fait leur apparition, apportant ainsi des "supports techniques" au planificateur. Parmi les supports, il y a l'économétrie qui lui a permis de quantifier l'économie et de prédire, à l'aide de modèles, l'évaluation de variables telles que le produit national brut, le revenu national, etc. Il était alors possible, avec les prévisions disponibles, de "prescrire" une politique adéquate.

Tout d'abord, quels sont les objectifs de l'Econométrie? Pourquoi désire-t-on obtenir la grandeur des paramètres économiques? Comment l'Econométrie peut-elle contribuer à l'analyse économique?

Sur ces questions pertinentes, nous pouvons répondre que:

- Tout d'abord, toute décision économique est basée sur certaines connaissances du cours des événements futurs. Un homme d'affaires désire savoir combien il peut vendre de sa marchandise et à quel prix? Un pays donné peut suivre une politique qui résulte du plein emploi, de sa force de travail et une utilisation maximale du capital. Au niveau de la firme, de l'économie nationale, la décision à prendre est dictée par certaines présomptions sur le futur. L'économètre peut aider à comprendre le futur à l'aide de modèles systématiques. C'est pourquoi la prédiction et la prévision ont souvent été citées comme la première contribution de l'économétrie.

- Une autre non moins importante contribution des modèles économétriques réside dans les alternatives qu'on peut envisager dans les processus de prise de décision. Par exemple, il est intéressant de connaître les effets sur le revenu national si l'on adopte des politiques fiscales alternatives. La technique de simulation permet à l'économètre d'utiliser un modèle approprié pour analyser ces effets alternatifs. Les pouvoirs publics - qui prennent les décisions - ont alors l'option de baser leurs décisions sur une telle information.

- Même si l'économètre ne serait pas intéressé à prédire ou à simuler, un autre intérêt réside dans l'analyse et la compréhension de la structure de l'économie. Celle-ci est en effet le premier objectif dans une investigation économétrique. Par exemple, l'on peut être intéressé de savoir ce qui se passerait dans le secteur investissement si l'on agit sur le secteur monétaire ou financier. Ou ce qui peut arriver dans le secteur consommation lorsqu'il y a une baisse des taxes. Non seulement, nous souhaitons savoir ce qui arriverait mais également l'intensité de l'impact de cet événement. L'économétrie est indiquée pour répondre à de telles questions.

Bien sûr, la variété des objectifs d'une étude économétrique n'implique pas qu'il n'y a aucune relation entre eux: ils sont interconnectés.

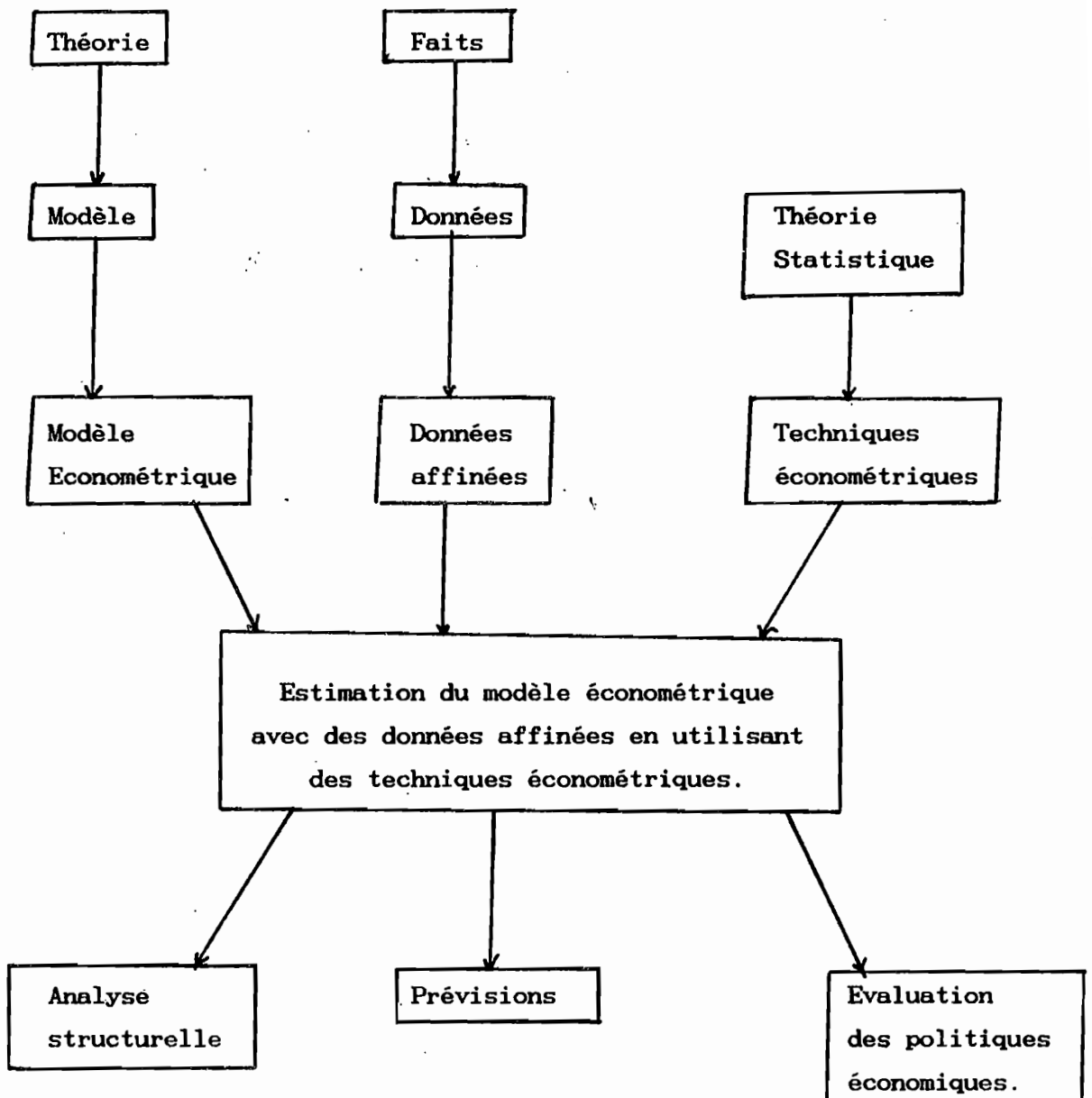
Mais cela se révèle parfois si théorique que les politiques s'avèrent irréelles, inadaptées..... Puisque l'économiste planificateur mettait tant de foi en leur pouvoir de prédiction, il est temps aujourd'hui de jeter un regard plus approfondi sur les modèles économétriques. Le sujet du travail que nous avons intitulé "Les Limites des Modèles économétriques dans la prévision" va essayer de cerner les principaux obstacles que rencontrent ces modèles dans la prévision.

Pour cela, un rappel sur la méthodologie générale d'une analyse économétrique constituera la première partie du travail, qui sera suivie par les principales difficultés pour le planificateur d'obtenir une bonne prévision et ainsi prescrire une politique économique appropriée.

I. RAPPEL SUR LA METHODOLOGIE GENERALE D'UNE ANALYSE ECONOMETRIQUE

L'analyse économétrique se base essentiellement sur deux éléments à savoir les faits d'une part et la théorie de l'autre.

Schéma n° 1 (2).



Ainsi, de la théorie on tire un modèle qui peut se définir comme toute représentation d'un phénomène réel, tel un système économique ou un processus. Pour les économistes, cette définition est plus stricte: le modèle est une simplification qui représente le système économique réel par un système d'équations simples ou simultanées.

Mais, pourquoi un modèle économétrique? En fait, celui-ci aide à quantifier l'économie en mesurant les relations économiques pour les comparer avec les différentes théories traitant d'un même phénomène. En second lieu, on peut citer les prévisions: il s'agit d'utiliser le modèle estimé pour prédire les grandeurs des variables endogènes en dehors de la période d'observation. Ceci aboutira à l'évaluation de politiques économiques. Il va de soi que les trois objectifs sont liés les uns aux autres et intéressent de très près les planificateurs.

1.1. Méthodologie d'une analyse économétrique.

Pour cela, nous utiliserons le tableau synoptique ci-dessous qui résume parfaitement la méthodologie générale d'une étude économétrique.

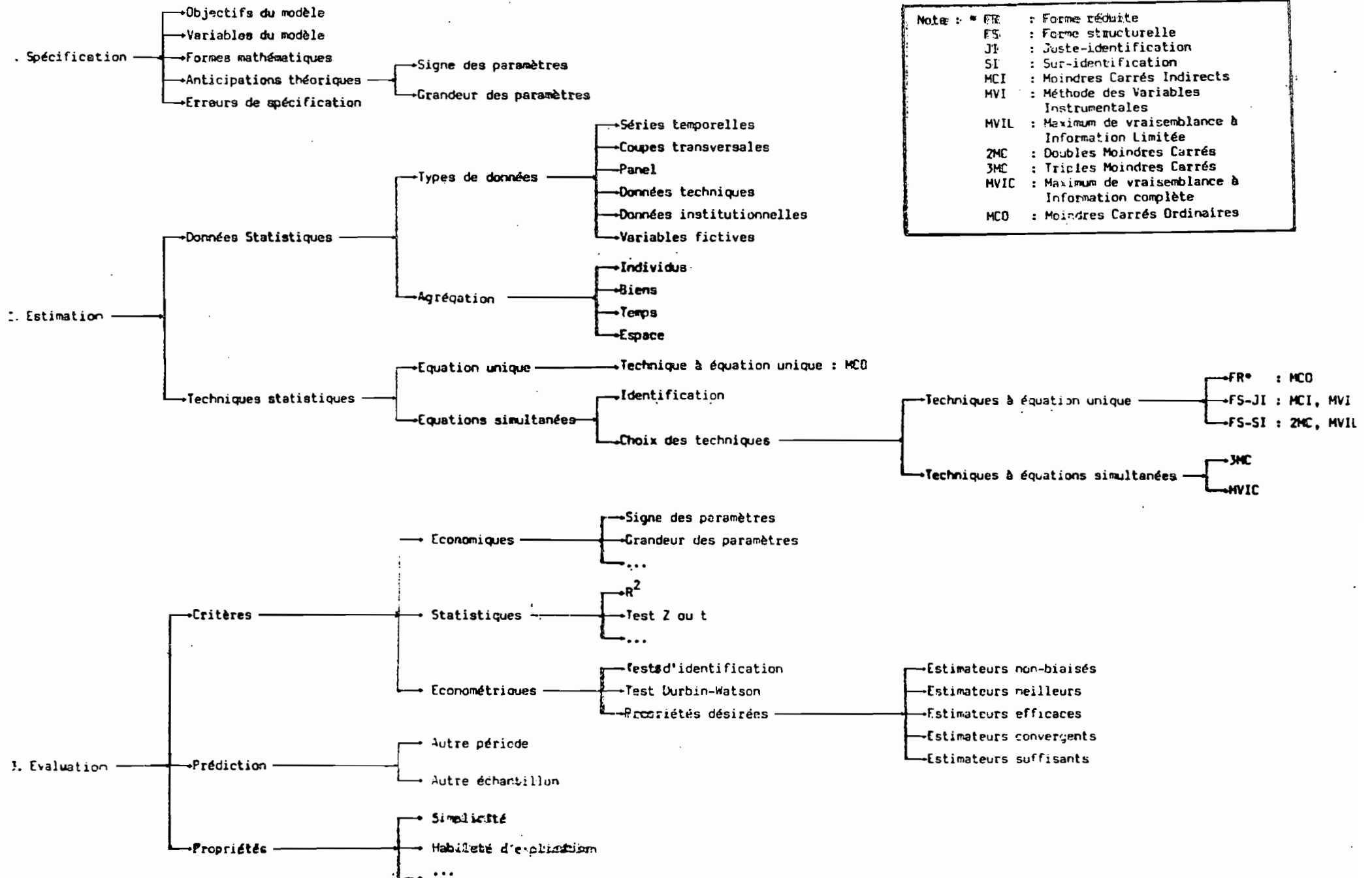
Schéma n° 2: Méthodologie d'une analyse économétrique (3).

(voir verso)

Selon ce schéma, on rencontre dans toute recherche économétrique trois éléments que sont:

1. La spécification qui se définit comme l'ensemble des activités qui donnent aux relations économiques ou aux théories une forme permettant leur vérification.
2. L'estimation de calcul numérique des différents paramètres - ou coefficients - du modèle.
3. L'évaluation qui consiste à l'appréciation des résultats numériques obtenus et des qualités du modèle.

SCHEMA NO 2. : METHODOLOGIE D'UNE ANALYSE ECONOMETRIQUE



Note : * FR : Forme réduite
 FS : Forme structurelle
 JI : Juste-identification
 SI : Sur-identification
 MCI : Moindres Carrés Indirects
 MVI : Méthode des Variables Instrumentales
 MVIL : Maximum de vraisemblance à Information Limitée
 ZMC : Doubles Moindres Carrés
 JMC : Triples Moindres Carrés
 MVIC : Maximum de vraisemblance à Information complète
 MCO : Moindres Carrés Ordinaires

1.1.1. La spécification.

Toute étude, quelque soit sa nature, commence par la détermination d'objectifs. Pour prendre un exemple, supposons que nous sommes intéressés à connaître le comportement des consommateurs de haricots au Burundi. Ceci revient à analyser la fonction de demande du produit. Pour cela nous avons besoin de variables. On va donc procéder à un inventaire de variables pouvant influencer cette demande. On a ici, comme variable endogène, c'est-à-dire expliquée par le modèle, la demande de haricots (4): Q^d_t ($t = 1, \dots, n$). Puis, parmi les variables exogènes - qui expliquent le modèle - nous pouvons citer le prix du haricot, (P_t), le prix des autres produits vivriers (PA_t), le revenu des consommateurs (Y_t), ...etc.

Dans cette liste, doit figurer le terme stochastique représenté par la variable u_t . En effet, quelque soit la précision avec laquelle on détermine les variables explicatives - indépendantes ou exogènes - celles-ci ne constituent jamais une relation de type déterministique (relation précise) avec la variable expliquée. En d'autres termes, dans notre cas précis où nous avons: $Q^d_t = f(P_t, PA_t, Y_t)$ (i) le fait de connaître le niveau de P_t , de PA_t et de Y_t ne nous permet pas de déterminer avec exactitude le niveau de Q^d_t parce que celle-ci est certes influencée par ces variables, mais pas uniquement celles-là. Ainsi, par le fait qu'en économie les relations sont stochastiques (c'est-à-dire que les influences sont tellement multiples qu'on ne les répertorie pas toutes), il est nécessaire d'ajouter un terme pour prendre en considération toutes les autres influences non prises en compte par les variables explicatives. Ainsi nous avons: $Q^d_t = f(P_t, PA_t, Y_t, u_t)$ (ii).

Une fois nos variables déterminées, nous pouvons passer à la forme mathématique du modèle qui peut être linéaire ou non-linéaire. Comme la théorie économique ne nous suggère rien, on prend l'hypothèse la plus simple, c'est-à-dire l'hypothèse linéaire. Ainsi (2) devient.

$$Q^d_t = \beta_0 + \beta_1 P_t + \beta_2 PA_t + \beta_3 Y_t + u_t \quad (iii) \quad (5)$$

Le point suivant de la spécification concerne les anticipations théoriques qui portent essentiellement sur la grandeur et le signe des paramètres. Pour cela, il faut savoir ce que représente les coefficients $\beta_0, \beta_1, \beta_2, \beta_3$:

β_0 est la constante qui nous donne la valeur de Q_t^d lorsque toutes les variables sont nulles. Le coefficient β_1 représente la modification de Q_t^d lorsque P_t change. En termes mathématiques, cela s'exprime par la dérivée première.

Ainsi, $\beta_1 = \frac{\partial Q_t^d}{\partial P_t}$ (iv) et celle-ci représente en fait une partie de l'élasticité-prix moyenne. Celle-ci se définit comme:

$$\eta_{p,q}^d = \frac{\partial Q_t^d}{\partial P_t} \cdot \frac{\bar{P}}{\bar{Q}} \quad (v)$$

Il en va de même pour les autres coefficients. Quant à leur signe et grandeur, on peut s'attendre à un élevé et positif, tandis que est faible si l'on considère que le haricot fait partie des produits de première nécessité. Il aura un signe négatif parce qu'une augmentation du prix du haricot entraîne une baisse de sa demande.

Il n'est pas inutile d'attirer l'attention sur les principales erreurs de spécification parce qu'une mauvaise spécification entraîne des résultats rendant erronée toute investigation basée sur eux. Les principales sont:

- l'omission de variables-clés.
- l'omission d'équations-clés.
- une forme mathématique erronée.
- la faiblesse de la théorie économique.

Nous pouvons maintenant passer à l'estimation qui, rappelons-le, est le calcul numérique des coefficients. Cela implique l'existence de données statistiques et des techniques adéquates.

1.1.2.2. L'estimation.

a). Données statistiques.

Si la récolte de données n'est pas le travail de l'économètre, ce dernier doit, dans le meilleur des cas, choisir celles qu'il veut utiliser. Elles seront, suivant les objectifs poursuivis, et la nature du sujet traité, de différents types:

On rencontre:

- des données de type séries temporelles,
- des données coupe transversale,
- des données "panel" ou données longitudinales,
- des données techniques,
- des données institutionnelles,
- des variables fictives.

Celles-ci, pour être utilisées, sont souvent agrégées. Cette opération peut concerner tant les individus que les biens, elle peut se faire dans le temps et l'espace.

Face à ces données agrégées, il ne reste plus qu'à les confronter avec des techniques appropriées pour dégager le modèle final.

b) Techniques statistiques.

S'il est vrai que les modèles à équation unique s'estiment à l'aide de la plus simple des techniques à équation unique, à savoir la méthode des moindres carrées ordinaires, il est aussi vrai que les modèles à équations simultanées peuvent être estimés aussi bien par des techniques à équation unique que par des techniques à équations simultanées. Pour ce dernier cas, parmi les critères permettant de choisir entre les différentes techniques d'estimation, il y a la forme (6) - structurelle ou réduite - et le degré d'identification (7).

Le schéma suivant résume les techniques pouvant être utilisées.

Schéma n° 3.

	Juste identifié	Sur-identifié	Non-identifié
Forme réduite (1)	MCO (2)		
Forme structurelle	MCI, MVI (2)	2MC, MVL (2)	Aucun

(1) Un modèle dans sa forme réduite est toujours identifié.

(2) Voir schéma n° 2.

Le classement du schéma n°3 n'est pas impératif. Il s'agit simplement de celui qui semble le plus approprié.

Il y a aussi un certain nombre d'autres critères qui peuvent influencer le choix de techniques comme:

- les propriétés désirées.
- la simplicité de la méthode.
- le coût en temps et en argent.

Nous pouvons passer maintenant au dernier point de cette partie, à savoir l'évaluation.

1.1.3. L'évaluation.

Une fois le modèle estimé, il est nécessaire de tester la validité des résultats obtenus, et si celle-ci s'avérait juste, procéder à des investigations - projections, politique économique.

Pour tester cette validité, il existe des critères économiques comme:

- le signe et la grandeur des paramètres;
- statistiques comme le coefficient de détermination au carré (R^2);
- économétriques comme le test d'identification, le test de Durbin-Watson, les tests des estimateurs (non biaisés, les meilleurs, efficaces, convergents, suffisants).

Alors, il est possible de pouvoir prévoir et prédire sur une période. C'est en cela que consiste l'évaluation. Après ces considérations d'ordre théorique, voyons maintenant les difficultés auxquelles est confronté le planificateur dans la construction de son modèle et dans son exploitation.

II. LES LIMITES DES MODELES ECONOMETRIQUES.

Un bon modèle doit avoir certaines propriétés additionnelles même s'il satisfait les critères économiques, statistiques et économétriques.

- La plausibilité théorique c'est-à-dire qu'un modèle doit être compatible avec la théorie économique.
- La simplicité: un bon modèle doit être simple, c'est-à-dire compréhensible pour tout le monde. Un modèle qui ne peut être compris que par son auteur devient inutile.

- L'habilité d'explication: un modèle doit être à même de traduire fidèlement les phénomènes économiques qu'il est sensé représenter. En d'autres termes, il doit être cohérent avec le comportement observé et être à même de les prédire.
- Le pouvoir de prédiction: un bon modèle doit pouvoir aider à prévoir. Pour voir cela, on teste le modèle soit avec un autre échantillon - d'autres données statistiques - soit en dehors de la période d'observation qui a servi de base à l'établissement du modèle.

Jusqu'ici, nous n'avons considéré que le bon côté des choses. En effet, le modèle d'un phénomène économique est en fait une "mutilation de la réalité" car certaines variables sont privilégiées, d'autres sont volontairement privées de mouvement, d'autres sont tout simplement omises.

Dans la mesure où une analyse économique fait appel à un modèle qui a été "validé" par des simulations systématiques, et plus encore par des prévisions couvrant une période suffisamment longue et mouvementée, nous pouvons dire que cette analyse est basée sur une représentation "réaliste", c'est-à-dire suffisamment fidèle des mécanismes économiques.

C'est oublier que l'économètre est confronté à de nombreuses difficultés, les unes étant inhérentes au modèle lui-même, les autres étant imputables à la théorie et aux méthodes d'estimation.

C'est dans cette partie que nous analyserons les limites des modèles économiques. Pour cela, nous avons distingué deux parties, la première traitant des difficultés liées aux données; la seconde traitant des problèmes techniques.

2.1. Les difficultés liées aux données.

2.1.1. La base des données.

L'objectif des techniques économétriques est d'obtenir des équations et partant des coefficients non-biaisés, consistants et dont l'estimation est optimale. Si les données sont sérieusement déficientes, incomplètes, tous les efforts faits pour adopter les méthodes d'estimation sont d'une utilité moindre. L'exemple typique est celui des pays en développement. Leurs données présentent trois principaux problèmes à l'économètre:

1) Les séries macroéconomiques couvrent une période assez courte; quelques années. Par conséquent, du point de vue du planificateur, le modèle reste opérant sur une marge très étroite; alors qu'il lui faudrait des données étalées sur une longue période, qui pourraient refléter la tendance générale.

2) Les chiffres fluctuent considérablement et très souvent et on peut difficilement y attacher beaucoup d'attention. On les estime alors un peu arbitrairement; ce qui ne devrait pas être le cas. Les données du commerce international et une portion du national constituent le seul ensemble de chiffres de ces pays en lesquels les économistes et les planificateurs peuvent placer un peu plus de confiance. Même si l'on dispose de plus d'informations, et qu'on les examine systématiquement et à fonds, on retrouve diverses et substantielles divergences.

Un exemple de chez nous: les données macroéconomiques du Burundi que publie le Ministère du Plan sont différentes de celles publiées par la Banque de la République du Burundi. On se demande alors lesquelles utiliser.

Considérons par exemple le fait que beaucoup de modèles en usage basent leurs estimations sur des données à prix constants. Ceci est fait généralement pour séparer le revenu national ou le produit national des effets des prix.

Ces données sont généralement celles de la comptabilité nationale et les déflateurs utilisés ne sont pas le plus souvent les plus appropriés pour avoir des séries à prix constant.

Ainsi, on se réfère souvent à des tests statistiques - R^2 , t - pour conclure hâtivement sans examiner la manière dont les données utilisées ont été estimées.

3) La troisième source de difficultés quant à la base de données vient des comptes nationaux. Ceux-ci sont utilisés par les économètres alors qu'ils ont été estimés souvent récemment et qu'ils subissent des transformations continues en les améliorant.

Nous avons l'exemple de nos données macroéconomiques qui sont en révision et adaptation annuelles. Ce qui fait que tout modèle construit est "daté" dans le sens où ils sont fixés aux séries de telle période. Les révisions des comptes nationaux ne sont pas mauvaises en soi; ils doivent être revus chaque fois qu'une information supplémentaire est assez importante pour la tenir en considération. Ainsi affectent-elles (les révisions) non seulement les coefficients mais aussi les conclusions sur lesquelles le modèle aboutit. Cela a pour conséquence de tout remettre en question pour le planificateur. Dans l'élaboration de plans quinquennaux par exemple, l'on est obligé de revoir les prévisions pour tel projet qui ont été faites sur base du modèle en question, à telle date donnée.

2.1.2. La signification des coefficients

Une deuxième difficulté, pour le constructeur d'un modèle, vient de la signification des coefficients et des erreurs de spécification que nous avons abordé dans la première partie. Les données doivent être mesurées sans erreurs pour un bon modèle.

Mais, même si elles ne sont pas biaisées et si elles couvrent une période assez longue, la signification des coefficients des équations de régression peut être très limitée.

Ils ne sont pas seulement considérés comme émergeant d'un système particulier d'équations simultanées - ce qu'ils sont en réalité - mais sont identifiés comme étant les propensions marginales ou les élasticités auxquelles les économistes sont habitués et qui ont toujours eu une certaine harmonie de causalité autour d'eux. Ainsi, en plus de la possibilité de prévoir ou de prédire, les modèles sont utilisés pour des contrôles éventuels grâce à ces coefficients: Exemple: Recommandation de réformes économiques sur base de certains coefficients. Les problèmes suivant sont liés à leur signification.

a) La prévalence de la multicollinéarité.

Avec cette première tare, l'abus d'une analyse des équations de régression est alors inévitable. Cette forme est universelle et un grand nombre d'exemples particuliers peuvent être donnés. Nous nous contenterons d'un pour illustrer ce problème (8). Soit une équation:

$$C_t = - 106.04 + 1.006 Y_{1t} + 0.655 Y_{2t}$$

(0.12) (3.574) (6.625)

Où $\bar{R}^2 = 0.93$ (coefficient de détermination au carré corrigé)

C_t = la consommation privée à la période t.

Y_{1t} = les revenus salariaux à la même période.

Y_{2t} = les revenus non salariaux, à la période t également.

Les chiffres entre paranthèses sont les valeurs t des coefficients.

A y regarder de plus près, nous aboutissons à la conclusion qu'il existe du point de vue marché, deux types de consommation entre des revenus salariaux, et les non-salariaux .

Tout changement dans la distribution des revenus doit avoir des effets significatifs sur le niveau de la consommation totale. Les coefficients semblent "raisonnables" seulement parce que nous sommes accoutumés à

croire que la propension à consommer en dehors des revenus salariaux est plus élevée qu'en dehors des revenus non-salariaux. On associe souvent les premiers avec les pauvres (qui ne vivent que de leur salaire) et les seconds avec les couches riches de la population (qui n'ont pas besoin de salaire pour consommer).

En effet, dans ce modèle, beaucoup de revenus tirés du secteur agricole sont traités comme non-salariaux; une évidence qui devrait permettre de regrouper les revenus des groupes salariés et non-salariés.

Seulement, l'important est de savoir que les deux variables indépendantes sont hautement corrélées (9), une situation qui remet en question la signification des coefficients de régression. Il est par conséquent fortuit que le coefficient des revenus non salariaux est plus petit que celui des revenus salariaux.

C'est pour ces raisons qu'il n'est jamais justifié d'utiliser ces coefficients et les équations pour une évaluation ou une recommandation de politique économique.

Il est donc vrai que les variables exogènes peuvent fausser la signification des coefficients lorsqu'elles sont hautement corrélées. On a alors la prévalence de la multicollinéarité.

Lorsque ce dernier cas arrive, les coefficients des variables indépendantes sont de plus en plus arbitraires. Ils ne peuvent donc plus être considérés comme les dérivés partielles dont on parlait dans la première partie.

Prenons une équation à titre d'exemple:

$$Y = a + b_1 X_1 + b_2 X_2 + b_3 X_3.$$

Si X_2 et X_3 sont hautement corrélés, les coefficients b_2 n'indiquent pas l'accroissement de Y lorsque X s'accroît d'une unité "toutes choses étant

égales par ailleurs". Comme X_3 varie avec X_2 , toutes choses ne sont plus égales par ailleurs. Si l'on veut calculer l'accroissement de Y lorsque X_2 s'accroît d'une unité, X_3 prend des valeurs qui altèrent les valeurs de X_2 . Le coefficient le plus approprié serait un \hat{b}_2 ; coefficient d'une équation simple:

$$Y = a + \hat{b}_2 X_2 \text{ qui prendrait en compte l'influence de } X_2.$$

De ce fait, à part le cas où les variables explicatives sont complètement indépendantes l'une de l'autre, le seul cas d'interprétation de b_2 et b_3 serait que X_2 et X_3 soient parfaitement corrélés; et un coefficient unique deviendrait la somme des deux. Celle-ci serait plus significative que les deux pris séparément.

S'il est vrai que la multicollinéarité fausse toute interprétation et toute utilisation des coefficients, il est une autre source de limites: les variables sont mal spécifiées. On aboutit alors à des erreurs de spécification dans la mesure et le choix des variables explicatives.

b) Les erreurs de spécification et le choix des variables explicatives.

Dans la première partie, nous avons parlé de certaines erreurs de spécification; elles proviennent soit des omissions de certaines variables, soit du choix de la forme mathématique, etc.

Lorsqu'on omet d'inclure dans les équations certaines variables exogènes, il est erroné de considérer les coefficients comme des dérivés partielles. Exemple: soit Y, X, Z , trois ensemble de variables formant des matrices $(n \times m)$ et les valeurs b, β et θ des vecteurs tels que $Y = X\beta$ (i).

La méthode des moindres carrés estime β par b par:

$$b = (X'X)^{-1} X'Y \text{ (ii)}$$

$$b = (X'X)^{-1} X' (X)\beta \text{ (iii)}$$

où

$$(X'X)^{-1}(X'X) = I \Rightarrow b = \beta \text{ (iv)}$$

Ceci veut dire que b est un estimateur non - biaisé de β
 Si par contre, le modèle correct était:

$$Y = X\beta + Z\theta \quad (v)$$

Alors que le modèle utilisé dans l'estimation est $Y = X\beta$ parce que l'écomètre a omis d'inclure Z , on a au lieu de:

$$\begin{aligned} b &= (X'X)^{-1} X'Y, \\ b &= (X'X)^{-1} X' (X\beta) + (X'X)^{-1} X' (Z\theta) \quad (vi) \end{aligned}$$

Si l'on substitue Y dans (v)

où

$$b = \beta + \delta\theta \quad \text{où } \delta \text{ considère } X \text{ comme une fonction de } Z \text{ (} X = \delta Z \text{)}$$

Ainsi "b" peut être un estimateur sur ou sous-estimé de β

Comme la variable Z n'a pas été incluse dans le modèle, il est pratiquement impossible de déterminer a priori le changement et dans certains cas la direction du biais dans l'estimateur de Y .

Ainsi, l'omission de variables qui ont certains liens avec Y a pour conséquence que non seulement nous n'aboutissons pas à une estimation non biaisée, mais que les estimateurs des coefficients liant tous les autres X_i à Y peuvent être "viciés".

Par contre, si les Z sont incluses dans le modèle alors qu'elles n'ont aucun lien avec Y , elles ne vont pas biaiser les estimateurs de

Nous abordons maintenant le troisième point qui est la forme des équations.

2.1.3. La forme mathématique des équations

Une autre voie par laquelle un modèle peut ne pas satisfaire, c'est celle où le constructeur peut mal spécifier les relations entre les variables endogènes et l'ensemble des variables indépendantes. Ainsi, il lui appartient de choisir entre des relations linéaires et des relations quadratiques (non-linéaires). Ceci est un cas particulier des erreurs de spécification relevées plus haut.

Dans la pratique, beaucoup de modèles sont des équations linéaires. Celles-ci et d'autres comme log-linéaires - ont souvent des coefficients de détermination au carré élevés. On les préfère également pour leur linéarité facilement manipulable: elles sont faciles à régresser.

En réalité, la forme d'un modèle a des implications importantes sur les relations économiques entre les variables et partant les conclusions qu'on peut tirer.

Soit une équation $Y = a + bX$.

où Y peut être la consommation
et X le revenu.

b est la propension marginale à consommer et cela suppose que celle-ci restera constante durant toute la période. Mais personne ne peut l'affirmer. En effet, nous pouvons considérer plusieurs cas possibles:

- Il peut être raisonnable de prévoir que l'intercept et le coefficient b resteront inchangés. Ainsi l'équation deviendrait $Y = bX$.
- On peut supposer par ailleurs que les deux varieront mais que l'élasticité - revenu restera inchangée - on utilisera alors la forme logarithmique. Soit: $\log Y = \log a + b \log X$.

- Ou encore que l'élasticité décroîtra au fur et à mesure que Y augmente. On utilisera alors la relation semi-logarithmique: $Y = a + b \log X$.

- Ou encore que l'élasticité décroîtra alors que X augmente. On utilisera alors une fonction logarithmique inverse:

$$\text{Log } Y = a - \frac{b}{X}$$

- Et enfin, on peut supposer que l'intercept a et l'élasticité varient sans aucune autre restriction. On peut alors utiliser une fonction quadratique comme:

$$Y = a + b X + c X^2$$

Ainsi tout dépend du constructeur et les raisons qui le poussent à adopter telle ou telle forme. Mais ce qui reste, la difficulté majeure est de savoir pourquoi il faut utiliser cette forme et pas une autre. Comme illustration, les graphiques suivants peuvent nous faire comprendre comment la forme mathématique d'une équation expliquant un même phénomène peut conduire à des conclusions différentes (10).

Ces données sont celles du Kenya pour la période 60-68. Et les équations y relatives satisfont tous les critères de validité d'une équation.

GRAPHIQUE N° 1

Graphique n° 1: Projections et forme mathématique du modèle.

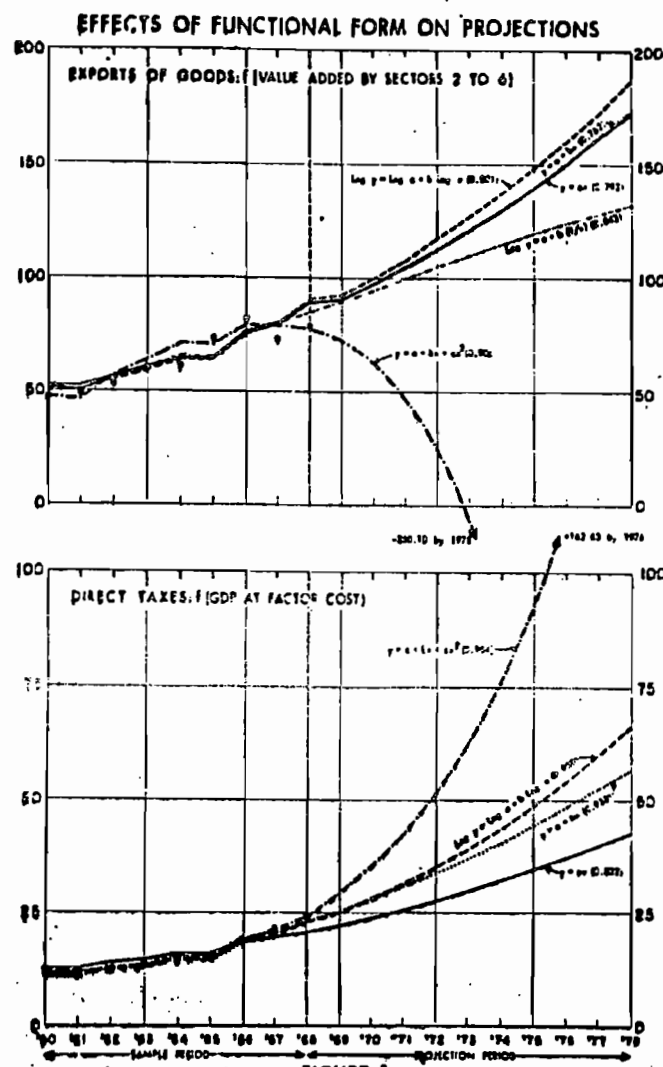


FIGURE 1

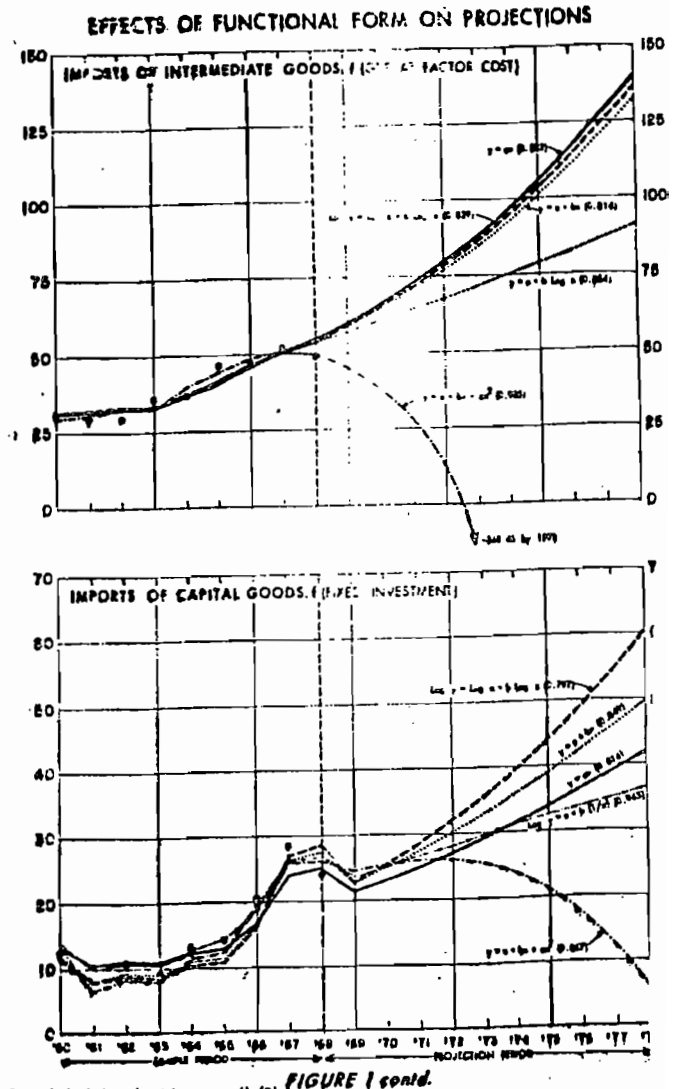


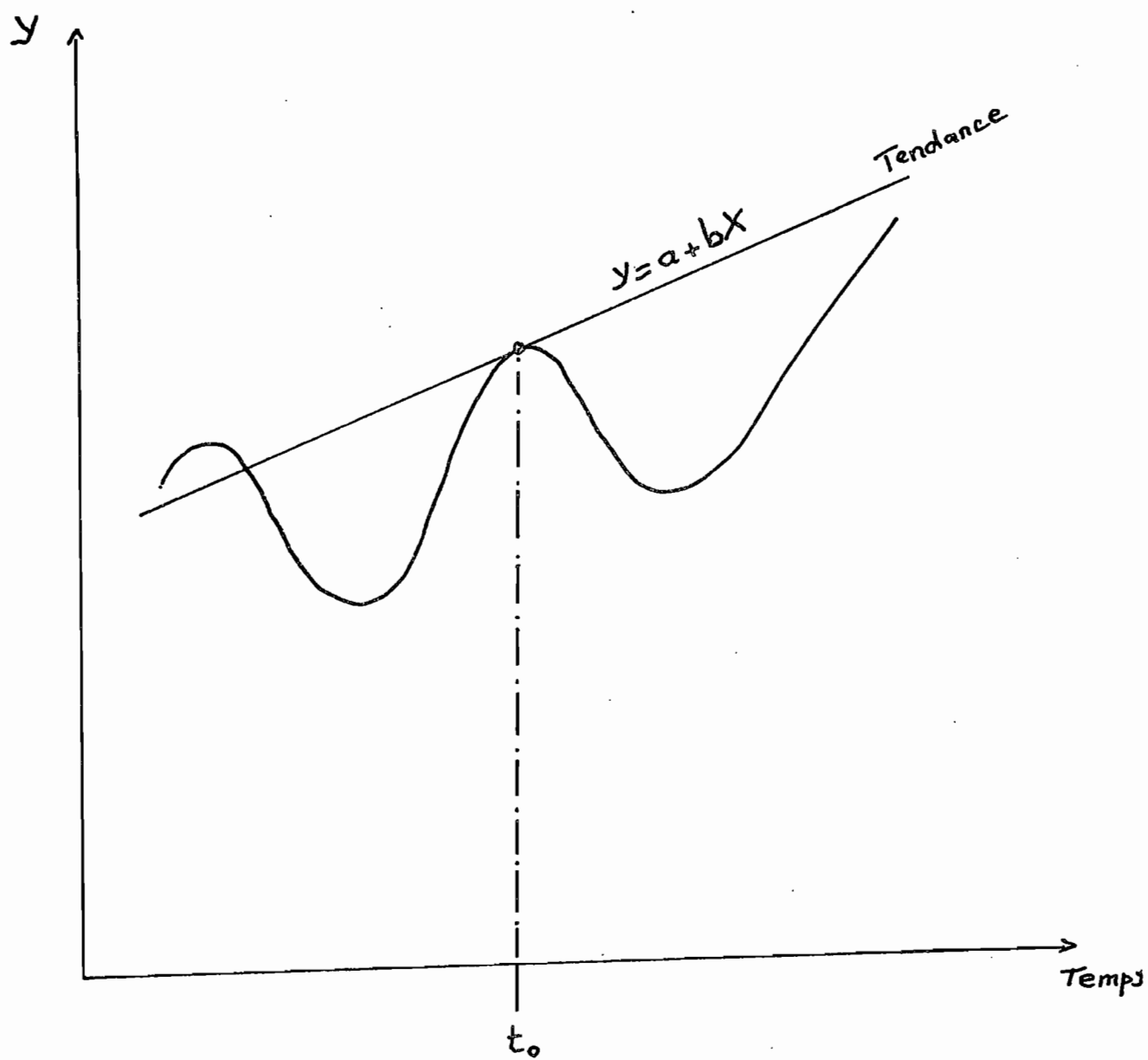
FIGURE 1 contd.

Un autre exemple peut être tiré des mouvements cycliques. Le graphique suivant peut l'illustrer:

Voir graphique page suivante.

On risque de projeter de fausses valeurs. Tout est de savoir déterminer la forme du modèle en question.

Graphique n° 2: Mouvements Cycliques et Projections.



Le choix d'une forme particulière a des conséquences sur l'efficacité de la prévision et de l'instauration de réforme économique basée sur ces prévisions.

Nous abordons maintenant le deuxième point de cette partie à savoir les problèmes techniques que rencontre l'économètre dans la construction du modèle de planification, basée sur ses prévisions.

2.2. Les problèmes techniques.

Nous avons abordé, au premier point les techniques à utiliser pour résoudre certaines équations suivant que le modèle est sous sa forme structurelle ou réduite. Parmi elles, il y a une méthode, la plus répandue et la plus facile manuellement - du moins pour des équations qui ne dépassent pas quatre variables - qu'on appelle la méthode des moindres carrés ordinaires (MCO).

Pour résoudre des équations, elle se base sur toute une série d'hypothèses qui facilitent leur résolution. Celles-ci concernent essentiellement le terme - erreur u_t dont on a parlé dans le premier point et les autres variables du modèle. Ainsi, elle se fonde sur les hypothèses suivantes (11):

a) Hypothèse concernant le terme-erreur u_t :

- La variable u_t est une variable aléatoire, dont les valeurs sont déterminées par le hasard.
- Son espérance mathématique est égale à zéro: $E(u_t) = 0$
- Sa variance est constante:

$$\text{Var}(u_t) = \sigma_u^2$$
- Elle a une distribution normale de moyenne 0 et de variance σ_u^2 : $u_t \sim N(0, \sigma_u^2)$.

- Les μ_t de différentes périodes sont indépendantes.
- Les valeurs sont indépendantes de celles des variables explicatives.

b) Hypothèses concernant les variables explicatives.

- les variables explicatives sont mesurées sans erreur.
- elles ne sont ni totalement ni hautement corrélées.
- elles sont complètement agrégées.

c) Hypothèses concernant le modèle.

- le modèle est identifié.
- le modèle à estimer est correctement spécifié.

Face à ces restrictions, la question naturelle à se poser ici est celle de savoir si les résultats que nous obtenons concernent ou non leurs propriétés lorsqu'une (ou plusieurs) hypothèse (s) est (sont) violée (s).

On ne retiendra que les trois principaux cas à savoir:

- 1) L'autocorrélation qui provient de la violation de l'hypothèse qui voulait que le terme-erreur soit temporairement indépendant des autres.
- 2) L'hétéroscédasticité qui a pour origine la violation de l'hypothèse stipulant que la variance de μ_t est constante et finie pour tout t .
- 3) La multicollinéarité, problème qui résulte de la non-satisfaction de l'hypothèse qui supposait que les variables explicatives ne sont ni hautement ni totalement corrélées.

Reprenons en détail ces trois points.

2.2.1. L'autocorrélation.

Appelée aussi corrélation sériale, elle est tout simplement un cas spécifique de la corrélation, cas où les divers éléments d'une même série sont corrélés. Or, les variables économiques ont plutôt tendance à évoluer de pair. En économétrie, le terme d'autocorrélation est lié à la variable stochastique μ_t . La question que l'on peut se poser est celle de savoir si

l'hypothèse de la présence de l'autocorrélation dans les modèles est plausible et/ou fréquente.

Ce phénomène se rencontre surtout dans les cas où nous travaillons sur des séries temporelles car - comme nous venons de le dire - ont tendance à s'influencer d'une période à l'autre.

Or, comme le terme-erreur est sensé prendre en considération toutes les variables omises, le phénomène d'autocorrélation est fréquent. Toutefois, cela n'est pas toujours le cas étant donné que rien n'empêche, a priori, que les variables omises aient des effets contradictoires parmi les autres sources d'autocorrélation, on peut citer la mauvaise spécification de la forme mathématique du modèle et certaines transformations mathématiques comme celles de Koyk (12).

On détecte normalement l'autocorrélation par le test de Durbin Watson.

Y a-t-il des solutions à ce problème?

Cela est fonction des causes, notamment, là où il y a eu mauvaise spécification - variables omises, forme mathématique - il suffit d'inclure l'une ou l'autre variable ou de modifier la forme mathématique.

2.2.2. L'hétéroscédasticité.

Techniquement, avec le problème, la variance du terme-erreur devient une variable comme toutes les autres. elle peut s'écrire:

$$\text{Var} (\mu_t) = \sigma_{ut}^2.$$

Le problème se situe au niveau de la propriété de la variance minimum qui est une des conditions exigées pour la variance du terme-erreur. En effet, il faut que celle-ci demeure constante tout au long de la période d'observation et aussi celle de prévision. Ce qui n'est pas toujours le cas. On obtient alors des estimateurs qui ne sont plus les meilleurs.

Pour résoudre le problème, on transforme l'équation de façon que la variance du terme-erreur devienne une constante. On la détecte par le test de corrélation des rangs ou celui de Goldfeld-Quandt (13).

2.2.3. La multicollinéarité.

Lorsque deux ou plusieurs variables explicatives sont hautement corrélées, les techniques statistiques ne réussissent plus à séparer les effets de chacune des variables concernées. Il y a deux raisons à la présence de multicollinéarité dans les modèles économiques:

- 1) La tendance des variables économiques à évoluer ensemble dans le temps,
- 2) L'utilisation en tant que variables explicatives des variables retardées conjointement avec les variables actuelles.

Il n'existe malheureusement pas de test précis pour détecter ce problème. Ceux proposés sont plus ou moins basés sur le tâtonnement comme les deux suivants:

- le test de L. Klein (14),
- le test de R. Frish (15).

Elle a donc pour conséquence de rendre les paramètres incertains (grande variance) et peu stables; ce qui peut entraîner des décisions erronées; comme l'exclusion de variables économiquement intéressantes.

Quant aux solutions, elles dépendent des moyens dont nous disposons. Nous pouvons suggérer toute une série de possibilités comme:

- l'incorporation d'informations externes,
- l'accroissement de la grandeur de l'échantillon,
- l'introduction d'équations additionnelles dans le modèle,
- la transformation du modèle.

CONCLUSION

Il existe d'autres problèmes mais ces derniers sont les plus importants. Ainsi, même si les modèles économétriques consistent en la représentation formelle d'idées ou de relations économiques, et qu'ils sont indissociables de leurs objectifs, on ne peut les juger que par référence au but qu'ils entendent servir.

Or nous savons que dans tout modèle, il y a "mutilation de la réalité", car certaines variables sont privilégiées par rapport à d'autres.

Cette insuffisance nous a permis de cerner plus ou moins exhaustivement les principales difficultés que rencontrent les économistes qui planifient; ainsi que les limites de ces modèles qui peuvent nous conduire dans nos investigations à des résultats erronés et partant à une planification défectueuse. Il est par conséquent conseillé, même si ce sont des supports techniques pour le planificateur, de les manier avec prudence et discernement.

NOTES ET REFERENCES

- (1) Cet article est un corrigé, revu et augmenté d'une Conférence que l'auteur a donné lors d'un Séminaire sur les systèmes de planification économique organisé par la Fondation Allemande pour le Développement International à Bujumbura en septembre 1986.
- (2) Schéma tiré et traduit de INTRILIGATOR, M. (1978). Ecometric Models, Techniques and Applications. Amsterdam and Oxford: North - Holland Publishing Company, pp. 2 et ss.
- (3) Schéma tiré de DAYER, M.G. (1980). Econométrie I: 1979-1980. Cours n°4.07, Volume n° 1, Bujumbura, p. 99.
- (4) Exemple tiré de Dayer, M.G., (1980), Op. cit., p. 132.
- (5) Les lettres grecques représentent les coefficients des variables. Elles sont notées ainsi parce qu'elles n'ont pas encore été estimées.
- (6) Un modèle est dit se trouver dans une forme structurelle si les variables endogènes sont fonction des variables exogènes, des variables endogènes et du terme-erreur alors qu'il est dans une forme réduite s'il n'est fonction que des variables exogènes et du terme-erreur.
- (7) L'identification est la procédure par laquelle on essaie d'établir si les équations et partant les coefficients estimés par les techniques économétriques sont effectivement ceux que nous voulions estimer. Ainsi, un modèle peut être sous identifié ou non identifié; il peut être juste identifié et suridentifié.
- (8) Exemple tiré de SHOURIE, A. (1970). "The relevance of Econometric models for medium - and longer-term Forecasts and Policy Description". Washington D.C.: International Bank for Reconstruction and Development. Economic Department Working Paper N° 75, p. 13.

- (9) Une simple corrélation entre les deux révèle un coefficient de corrélation de 0.9982. SHOURIE, A. (1970). Op. cit., p.13.
- (10) Ibidem, pp. 30-31.
- (11) DAYER, M.G., Op. cit., pp. 135-138.
- (12) Voir à ce sujet: DUTTA, M. (1975). Econometric Methods. Brighton: South-Western Publishing.
- (13) Voir à sujet GOLDFELD, S.M. and QUANDT, R.E. (1965). "Some tests of Homoscedasticity" in Journal of the American Statistical Association, Vol. 64, September, pp. 539-541.
- (14) KLEIN, L.R. (1962). An Introduction to Econometrics. Englewood Cliffs, N.J., Printice-Holl, pp. 64-101.
- (15) FRISH, R. (1934). Statistical Confluence Analysis by Means of Complete Regression systems. Oslo: University Press.

BIBLIOGRAPHIE

- (1) DAYER, M.G. (1980). Econométrie I: 1979-1980. Cours n° 4.07. Volume n°1, Bujumbura: FSEA, pp. 361.
- (2) DUTTA, M. (1975). Econometric Methods. Cincinatti: South-Western Publishing Company.
- (3) FRISH, R. (1934), Statistical Confluence Analysis by Means of Complete Regression Systems. Oslo: University Press.
- (4) GOLDFELD, S.M. and QUANDT, R.E. (1965). "Some tests of Homoscedasticity", in Journal of american Statistical Association. Vol. 64, September, pp. 539-541.
- (5) GOLDFELD, S.M. and QUANDT, R.E. (1976). Nonlinear Methods in Econometrics. Amsterdam: North-Holland Publishing Company, pp. 273.
- (6) INTRILIGATOR, M. (1978). Econometric Models, Techniques and Applications. Amsterdam and Oxford: North-Holland Publishing Company, pp. XV, 638.
- (7) KLEIN, L.R. (1956). A testbook of Econometrics. Evanston, Ill.: Row, Peterson and Company, pp. 355.
- (8) KLEIN, L.R. (1962). An Introduction to Econometrics, Englewood Cliffs, N.J.: Prentice-Hall, pp. 280.
- (9) SHOURIE, A. (1970). "The relevance of Econometric Models for Medium - and longer-term Forecasts and Policy Description". Washington, D.C.: International Bank for Reconstruction and Development. Economic Department Working Paper n° 75.



Centre Universitaire de Recherche pour le Développement Economique et Social

Référence bibliographique des Cahiers du CURDES

Pour citer cet article / How to cite this article

NIBITEGEKA Emmanuel, Les limites des modèles économétriques dans la prévision, pp. 1-29, Cahiers du CURDES n° 7, Juin 1989.

Contact CURDES : curdes.fsea@yahoo.fr