UNIVERSITE D'ANTANANARIVO

FACULTE DE DROIT, D'ECONOMIE, DE GESTION, ET DE SOCIOLOGIE

DEPARTEMENT ECONOMIE

ANNEE UNIVERSITAIRE 2006/2007

MEMOIRE DE MAITRISE

LA MODELISATION, L'ECONOMIE, L'ECONOMETRIE, LA MATHEMATIQUE, ET L'INFORMATIQUE

Présenté par : **CHAN ZEN Rollin**

Encadré par : **Monsieur RAZAFINDRAVONONA Jean**

Date de soutenance : 09 Novembre 2007

REMERCIEMENTS

Ce travail, plus précisément ce mémoire, est le fruit d'une entraide pendant cette année universitaire 2007. Sa réalisation n'était possible que grâce à Dieu qui nous a donné la force et le courage pendant son élaboration. On a aussi apporté nos efforts, et nos recherches pour trouver tous les documents et les données nécessaires à sa réalisation.

Je remercie aussi mon honorable encadreur, Monsieur **RAZAFINDRAVONONA Jean** qui m'a encadré et dirigé, il a aussi apporté ses compétences et donné son aide lors de la recherche des documents. Il faut aussi noté ici que la recherche des documents n'était pas facile puisque les documents relatifs à ce thème sont rares, surtout dans le pays comme Madagascar.

Même si on était contraint par le temps qui n'était pas vraiment suffisant pour traiter dans les moindres détails ce thème, on a quand même pu le terminer à temps avec une analyse qu'on peut qualifier de satisfaisante.

Enfin, je remercie tous ceux qui ont contribué à ce mémoire que ce soit d'une manière directe ou d'une manière indirecte. Je souhaite vivement que les analyses qui figurent ici soient, dans les prochaines années, approfondies par d'autres étudiants qui sont intéressés par ce type de thème.

TABLE DES MATIERES

INTRODUCTION

Un des objets de l'économie est d'étudier les réalités et les phénomènes économiques. Des théories économiques se sont inspirées des phénomènes économiques comme le cas de la théorie Keynésienne qui a pu naître grâce à la crise des années 30. Mais l'étude des phénomènes économiques nécessite l'existence des outils capables de déceler les mécanismes des phénomènes économiques. C'est pourquoi dès le début du 20ème siècle, les économistes commencent à adopter et à intégrer la mathématique dans leurs analyses, et ils ont commencé à faire des modélisations des réalités et phénomènes économiques. Mais résoudre un modèle nécessite à son tour l'existence d'une technique et méthodes appropriées, cette technique est l'économétrie.

A vraie dire, la modélisation a déjà existée en économie bien avant le 20ème siècle car Il ne faut pas oublié que Léon WALRAS a déjà fait la modélisation, son modèle est le célèbre modèle d'équilibre général concurrentiel. Ce modèle est basé sur la formalisation mathématique des comportements économiques, des pratiques économiques, et de la logique économique. Léon WALRAS était donc le premier économiste (économiste mathématicien) à faire de l'économie une science dure qui peut se placer au même rang que les autres sciences comme la mathématique et la physique.

C'est à partir des années 70 qu'a commencé l'utilisation de l'outil informatique pour faciliter le travail car les modèles économétriques construits devenaient de plus en plus complexes. Dans la pratique, les modélisateurs émettent toujours des hypothèses simplificatrices, sinon, la complexité est telle qu'on ne pourrait pas construire des modèles.

Pour analyser le thème : « La modélisation, l'économie, l'économétrie, la mathématique, et l'informatique », on va faire une approche progressive, c'est-à-dire que dans un premier temps, on va d'abord analyser les spécificités de la modélisation, de l'économie, et de l'économétrie, et dans un second temps, on va analyser les interactions qui existent entre la modélisation, l'économie, l'économétrie, la mathématique, et l'informatique.

PREMIERE PARTIE

LES SPECIFICITES DE LA MODELISATION, DE L'ECONOMIE, ET DE L'ECONOMETRIE

Cette première partie sert à analyser les différentes spécificités de la modélisation, de l'économie, et de l'économétrie. En d'autre terme, on va donner des définitions, déterminer leurs objets d'études respectifs, connaître ses approches, et déterminer ses évolutions dans le temps et dans l'espace. En un mot, il s'agit de faire une analyse que l'on peut qualifiée de détaillée de ces trois disciplines.

Ces analyses des spécificités de la modélisation, de l'économie, et de la modélisation vont nous permettre en deuxième partie de ce travail, de déterminer les interactions entre la modélisation, l'économie, l'économétrie, la mathématique, et l'informatique.

Chapitre 1 : LA MODELISATION

I- DEFINTION ET STRUCTURE D'UN MODELE

1-1) Définition

D'une manière générale, un modèle est la représentation formalisée et structurée, mais approximative et incomplète, d'un ensemble d'éléments réels, ensemble choisi et délimité par le créateur du modèle. D'après **Christopher Dougherty**, dans son livre intitulé « *Introduction à l'économétrie* » publié en 2002, il définit un modèle comme un ensemble de relations entre différentes variables présentant entre elles un lien de cause à effet. Cet ensemble d'éléments, de relations, et des variables, constitue un « système ». Le système peut être défini comme un ensemble d'éléments et d'événements qui sont en étroite dépendance et qui se soutiennent mutuellement. Les actions du monde extérieur sur le système se traduisent par l'existence des «variables d'entrée ». Par contre, les actions du système sur son environnement se traduisent par des «variables de sortie ». L'étude d'un système consiste, une fois ses variables d'entrée et de sortie définies, à rechercher les liaisons qui existent entre ces variables, c'est-à-dire à établir un modèle du système. Pour se faire, on introduit souvent un certain nombre de variables auxiliaires représentatives d'éléments du système, que l'on appelle « variables d'état ».

1-2) Structure d'un modèle

Un modèle peut se limiter à une seule ou quelques équations. On peut prendre l'exemple de la fonction de consommation keynésienne suivante:

$C = cY + C_O$ c'est-à-dire $C = f(Y)$ avec C : la consommation ; Y : le revenu national ; et c et C_O des constantes.

La consommation est fonction du revenu c'est-à-dire que C dépend de Y. C est donc dans ce cas appelée *variable dépendante* ou *expliquée* et Y la *variable indépendante* ou *explicative*. La ou les variables explicatives sont alors considérées comme faisant partie de l'environnement. Elles sont extérieures aux modèles et on parle à leur propos de

variables *exogènes* par opposition à la variable expliquée dite variable *endogène.* En générale, un modèle va intégrer plusieurs relations et les variables explicatives pour chacune d'entre elle ne sont plus nécessairement des variables exogènes du modèle. Prenons le cas du modèle d'équilibre simple suivant :

$Y = C+I$; $C = f(Y)$; $I = I^*$

Dans la première relation, la variable expliquée qui est le revenu national Y est égale à la somme de la consommation et de l'investissement. Ici, la qualité de l'explication est faible et on parle souvent de *relation comptable ou de définition*.

La seconde relation explique la consommation C en fonction du revenu. Il s'agit là d'une véritable explication ayant à la fois des fondements empiriques et théoriques. Elle reflète la façon dont les agents économiques se comportent pour dépenser leur revenu, on parle donc de *relation de comportement*. La théorie économique et les travaux empiriques vont permettre de préciser certaines caractéristiques de cette fonction, par exemple la consommation augmente quand le revenu augmente. Ceci peut être écrit $dC/dY > 0$ ou encore que la consommation augmente proportionnellement moins que le revenu soit $d^2C/dy^2 < 0$.

La troisième relation n'est là que pour indiquer la limite du modèle qui ne cherche pas à expliquer l'investissement.

Ce modèle possède trois relations (équations), deux variables expliquées ou endogènes; mais on remarquera que dans les deux premières équations, la variable expliquée de l'une est aussi la variable explicative de l'autre soulignant une première forme d'interdépendance. Ecrit de cette manière, le modèle est dit mis sous sa forme structurelle. Les relations économiques fondamentales, ici la fonction de consommation, apparaît explicitement.

On peut dire donc qu'au niveau des équations, on distingue :

- des *équations de définition* des variables à partir d'autres variables, qui se traduisent par des équations comptables. On peut prendre comme exemple la première relation c'est-à-dire $Y=C+I$ avec I supposé constante.
-des *équations d'équilibre* qui reflètent les équilibres entre les variables du système, et se traduisant par des égalités entre deux variables.

Exemple : Offre=Demande

-des équations ou des *relations de comportement* ou *fonctionnelles* aussi, dont la forme algébrique et les coefficients numériques ne sont pas connus à priori mais doivent être déterminés par des méthodes appropriées.
Exemple : C=f(Y)=cY+Co Les coefficients à déterminer sont c et Co.

Il faut connaître maintenant ce que c'est un modèle structurel et un modèle réduit.

Un modèle structurel et un modèle réduit :

Quelle que soit leur diversité, les modèles peuvent être organisés de deux manières distinctes, suivant l'approche choisie dans l'étude d'un système :

-soit le système est constitué par des équations reliant chacune, une variable endogène à un ensemble de variables exogènes :
Exemple : Y1=aX1+bX2

Ce modèle présente un grand intérêt puisqu'il permet de calculer rapidement les valeurs des variables de sortie à partir des valeurs des variables d'entrée. Ce type de modèle s'appelle modèle réduit ;

-soit le système est analysé dans ses structures et ses mécanismes de fonctionnement internes. Dans ce cas le modèle, dit *modèle structurel,* est constitué par des équations pouvant contenir simultanément plusieurs variables endogènes et plusieurs variables exogènes, suivant la nature des liaisons entre les variables du système :

Exemple : $a_1Y_t + a_2 Z_t = b_1X_{1t}+b_2X_{2t}+b3Y_{t-1}$

Cette forme de modèle, plus complexe, permet une vue plus approfondie du système étudié ; elle est plus adaptée à la connaissance scientifique, en particulier à la vérification d'hypothèses théoriques.
En résolvant le système d'équations du modèle structurel, on peut toujours obtenir le modèle réduit correspondant.

II-TYPOLOGIE DES MODELES

On va faire des classifications (des modèles) qui va nous faire apparaître avec précision la nature et l'intérêt particulier d'un type particulier de modèle.

2-1) Modèles théoriques, modèles comptables, modèles économétriques

Cette classification est basée sur le degré de proximité du modèle par rapport à l'analyse théorique et à la description brute des faits concernant le système étudié.
L'analyse théorique de tout phénomène, particulièrement dans le domaine des sciences économiques, permet d'identifier les concepts utiles à la représentation de ce phénomène puis de formuler un corps d'hypothèses cohérentes sur les relations entre les concepts étudiés.

a) *Les modèles théoriques* : ces modèles sont une représentation d'un système sous forme de relation algébrique, qui assure une cohérence et une rigueur mathématique à l'analyse théorique.

Exemple : La théorie keynésienne affirme que la consommation des ménages est fonction de leur revenu. On peut donc formuler le modèle comme suit :
$C=aY+b$

On remarque sur cet exemple qu'il peut exister pour une théorie plusieurs modèles théoriques correspondants à différentes spécifications ou variantes de celle-ci. Néanmoins un modèle théorique est toujours du domaine de l'hypothèse non vérifiée.

b) Les modèles comptables : ces modèles permettent une organisation des informations sur les phénomènes, à partir des concepts fournis par l'analyse théorique. Ces informations sont élaborées à partir des techniques d'observation et de collectes appropriées (enquête,..). Ces modèles assurent une rigueur et une cohérence des informations sur le système étudié. Cependant ces modèles ne formalisent pas vraiment les liaisons qui existent entre les variables endogènes et les variables exogènes.

c) Les modèles économétriques : ces derniers modèles font la synthèse des modèles théoriques et des modèles comptables. Ils ont pour objet de fournir une représentation la plus exacte possible du système étudié. A cette fin, ils confrontent les hypothèses des modèles théoriques avec les chiffres fournis par les modèles comptables, grâce à des méthodes statistiques spécifiées. Ainsi on obtient un modèle dont les coefficients sont connus et qui permet d'obtenir des valeurs chiffrées des variables endogènes à partir des variables exogènes. Par ailleurs des techniques statistiques d'évaluation permettent de

mesurer la précision des résultats, c'est-à-dire les erreurs avec la réalité connue à partir des informations.

On peut aussi classifier les modèles selon le critère d'intégration du temps, d'où les modèles statiques et les modèles dynamiques.

2-2) Les modèles statiques et les modèles dynamiques

A la base de cette classification se trouve la représentation du système par rapport au temps.

a) *Modèles statiques* : ces modèles ne feront pas intervenir, pour la détermination de l'équilibre associé à une période donnée, que les variables de cette période. En d'autre terme, on suppose que l'état du système au temps t est indépendant ou identique à l'état du système en t-1, t+1, etc.… . On fait donc l'hypothèse de l'invariance temporelle des relations du modèle.

On se contente donc de décrire l'état d'un système dont la structure est donnée (les paramètres) et pour un état donné de son environnement (les variables exogènes). Ce type de modèle n'est possible que si, dans cet état, le système n'est soumis à aucune force qui va en modifier la structure et pour cela, on se réfère très souvent à la notion d'équilibre. De plus, cet équilibre doit posséder certaines propriété, et comme le système est à priori soumis à des chocs qui vont l'en écarter, des mécanismes automatiques doivent se mette à jouer pour restaurer un tel équilibre. On parle à ce propos de *stabilité de l'équilibre*.

Supposons un modèle de forme suivante : $Y_t = a + bX_t + \mu_t$ avec μ_t l'erreur.

Ce type de modélisation va servir à évaluer l'impact d'une variation d'une variable exogène sur les variables endogènes du modèle. On peut donc voir dans le modèle que si X varie d'une unité, Y va varier de b unité(s).

b) Modèle dynamique : Dans les modèles dynamiques, le temps joue au contraire un rôle essentiel et la solution du modèle donnera la valeur des variables endogènes à une date donnée comme dans un modèle statique, mais surtout permettra de préciser le comportement dans le temps de ces variables. D'un point de vue formel, on peut travailler en temps discret par référence à une période élémentaire permettant de mesurer la

variation des variables $\Delta X_t = X_t - X_{t-1}$, et qu'on peut se rapprocher d'une période calendaire (jour, mois, trimestre, année selon le cas). On peut également raisonner en temps continu en choisissant une période aussi petite qu'on le désire. Les modèles en temps continu sont des modèles où les variables dans le système d'équation apparaissent comme une fonction continue du temps et les variations deviennent donc des dérivées exactes de celui-ci. Deux cas méritent tout particulièrement l'attention en matière de modélisation dynamique :

- certaines spécifications des relations de comportement d'un modèle vont impliquer, au niveau de sa solution, des conditions de comportement dans le temps des variables endogènes. Partons par exemple d'une condition d'équilibre épargne - investissement S=I en prenant une fonction d'épargne proportionnelle du type S = sY et en spécifiant la fonction de demande d'investissement par un accélérateur I= ßΔY, la condition d'équilibre s'écrit :

$\frac{\Delta Y}{Y} = \frac{s}{\beta}$. Pour que cette condition d'équilibre se maintienne dans le temps, il faut que le revenu national croisse au taux s/ß.

- la spécification de certaines relations de comportement peut conduire à l'introduction de variables endogènes retardées et ceci, afin de tenir compte de délais d'ajustement. Par exemple, on fera dépendre la consommation, non pas du revenu de la période courante, mais du revenu de la période précédente. De même, le niveau de la production ne dépendra pas du prix courant, mais du prix de la période précédente afin de prendre en compte les délais entre les décisions de mise en production et le moment où la production est disponible à la vente sur le marché.

A la période t, la variable endogène retardée est donnée et invariante, on peut donc la considérer comme exogène en raisonnant purement en statique. Les modèles dynamiques utilisent donc pour déterminer l'équilibre d'une période, des variables d'autres périodes. La justification en pourra être :

- théorique : certaines agents seront supposés intégrer dans leur comportement leurs constatations passées ;
- institutionnelle : Par exemple l'impôt sur le revenu payé par le ménage sera basé sur leur revenu de la période précédente ;

- mécanique : par exemple le taux de croissance, le passage du taux de croissance annuel au niveau instantané nécessite la prise en compte du niveau précédant.

2-3) Les modèles endogènes et modèles explicatifs à variables exogènes

Cette classification repose sur le mode de représentation du système : sa structure interne et ses relations avec son environnement.

a) Un modèle endogène : c'est un modèle constitué uniquement par des variables endogènes du système étudié (variables d'état c'est-à-dire des variables auxiliaires ; et des variables de sorties dues aux actions du système sur son environnement) sans aucune référence aux variables d'entrée qui sont des variables résultant des actions de l'environnement sur le système. De ce fait, la structure interne du système et ses relations avec son environnement ne sont pas explicitées.

Ce modèle se présente sous la forme d'équations indépendantes où chaque variable endogène est fonction du temps et du décalage :

$Y_t = f(y_{t-1}, \ldots, y_{t-i}, \ldots ; .t)$

Ce type de modèle n'est utilisable que pour des prévisions à cour terme, en supposant que l'environnement est stable.

b) Un modèle explicatif à variables exogènes : c'est un modèle constitué de variables endogènes et de variables exogènes liées entre elles par des équations. Ce modèle est orienté vers l'analyse de la structure interne du système et de ses relations avec son environnement. Il permet de vérifier les hypothèses théoriques, de réaliser des prévisions à court et moyen terme, d'établir des simulations.

2-4) Modèles de simulation, de prévision ,et d'optimisation

Un modèle peut faire l'objet de plusieurs types d'utilisations pratiques. En tant que représentation simplifiée, le modèle devra être construit de façon à être adapté le mieux possible aux exigences de ces différentes utilisations. Cette dernière classification permet ainsi d'étudier les caractéristiques spécifiques des modèles suivants leurs utilisations.

a) *Les modèles de simulation* : ils ont pour objet d'étudier le champ des états possibles du système et de ses sorties en fonction :

- de l'impact des modifications de l'environnement,
- de l'impact des décisions concernant les variables de commande des variables exogènes.

Deux procédures de simulation sont couramment utilisées :

- simulations successives du modèle en faisant varier tour à tour chaque variable exogène dans une plage ou bande autour d'une valeur centrale donnée de façon à étudier les déformations marginales du système autour d'un état de référence,
- simulation sur la base de quelque « scénarios » successifs. Chaque scénario correspondant à un ensemble cohérent de valeurs des variables exogènes traduisant un état probable de l'environnement ou un mode de gestion déterminé.

b) *Les modèles de prévision* : ils ont pour objet de prévoir avec la plus grande précision possible l'état le plus probable du système à un instant donné, à partir des prévisions sur l'état de l'environnement et d'un choix déterminé des valeurs des variables de commandes.

Il existe deux types de modèles de prévision :

- un modèle de prévision à court terme qui est basé sur l'étude fouillée des délais de réaction entre les variables et la recherche des effets des variables exogènes les plus instables à court terme (moins d'un an) ;
- un modèle de prévision à moyen et long terme qui est basé sur l'étude détaillée des effets des variables exogènes traduisant les modifications de structure de l'environnement ou la transformation des politiques ou des actions sur le système. Dans ce type de modèle, on devra travailler sur des données de longues périodes.

c) *Les modèles d'optimisation* : ils présentent deux caractéristiques :

-une fonction d'évaluation qui permet de définir les états souhaitables ou optimaux du système parmi les états possibles ;
-un algorithme de résolution du modèle qui permet de repérer les états optimaux et de calculer les valeurs des variables de commande correspondantes.

III- UTILISATIONS DES MODELES

Depuis environ une trentaine d'années, les modèles ont connu un développement considérable en particulier dans le domaine des sciences économiques. Pour que les modélisations puissent être vraiment possibles, trois conditions doivent être réunies :

-l'existence d'informations structurées, de faible coût, et en grande quantité sur le phénomène étudié. Ceci est rendu possible grâce aux développements des systèmes comptables nationaux et d'entreprises, des systèmes informatiques et des réseaux, et des banques de données informatisées ;

-l'existences des moyens de calcul : ordinateurs et programmathèques (logiciel) spécialisées ;

-l'existence d'économistes ou des statisticiens ayant la formation requises.

Nous allons maintenant aborder l'utilisation des modèles. Pour cela, nous allons centrons notre commentaire sur l'exemple des modèles économiques. La première utilisation des modèles économiques est certainement la prévision, et ensuite l'utilisation pour la recherche scientifique.

3-1) **La prévision**

L'utilisation la plus naturelle d'un modèle est de prévoir l'avenir économique. On y injectera des hypothèses concernant le futur, et sa résolution produira le diagnostic recherché.

Exemple : on peut prévoir l'évolution des principaux agrégats de l'économie nationale comme le PIB, les investissements, les importations,… compte tenu d'hypothèses sur l'évolution de l'économie internationale.

La question est maintenant de savoir les différents types de prévisions. Il en existe 2 types: les scénarios et les variantes.

Les scénarios : Dans un scénario, on s'intéresse aux résultats dans l'absolu, c'est-à dire que l'on associera à un ensemble d'hypothèses, une évolution de l'équilibre économique ;

Les variantes : dans une variante, on partira d'une simulation de base dite souvent : « compte de référence » et on mesurera la sensibilité de l'équilibre économique à une modification des hypothèses.

a) les différents types de scénarios :

On distingue 2 types de scénarios :

-les prévisions tendancielles, qui utiliseront les hypothèses les plus possibles lorsqu'on veut baliser les domaines des évolutions envisageables. Ces hypothèses seront souvent élaborées par des experts extérieurs au domaine de la modélisation.

-les prévisions normatives, où l'on s'astreint à respecter un certain nombre de conditions sur les résultats. Ex : un retour progressif à l'équilibre du solde extérieur.

b) les différents types de variantes :

-les variantes analytiques qui auront pour but d'analyser les propriétés du modèle : pour cela, on effectuera sur les principales hypothèses (de manière indépendante, un ensemble de modifications simples permettant de prendre en compte l'ensemble des mécanismes du modèle). L'analyse logique et quantitative des résultats donnera des indicateurs sur la qualité du modèle, par comparaison avec la théorie économique et les propriétés des autres modèles ;

-les variantes complexes : elles chercheront à prévoir les conséquences des modifications des hypothèses par rapport à celles du compte de référence. Dans les variantes complexes, on peut distinguer :

- les variables d'aléas, où l'on cherche à mesurer les conséquences d'événements non maîtrisables ;
- les variantes de politiques économiques, où l'on s'intéresse aux conséquences des choix concernant les instruments institutionnels, variables maîtrisables si l'on considère que l'Etat est le comandataire de l'étude.

3-2) Utilisation pour la recherche scientifique

En macroéconomie, la majorité des domaines sinon l'ensemble ont fait l'objet de travaux de modélisation. Les modèles les plus utilisés sont les modèles structurels, dynamiques, et de type économétrique. Citons parmi les domaines très étudiés : les

fonctions de consommation, d'épargne et d'investissement des ménages ; les fonctions de production, d'investissement des entreprises ; les fonctions d'importation et d'exportation, le domaine monétaire et financier,… .

Chapitre 2 : L'ECONOMIE

L'économie, plus précisément la science économique, a pour objet l'étude de la production, de la répartition, et de la consommation des biens et services. En économie, on fait toujours, dans la plupart du temps, l'hypothèse que nous sommes dans une société où les individus sont des êtres « homo-oeconomicus », c'est à dire que les agents économiques sont rationnels. On peut déjà affirmer que la science économique ne traite pas vraiment d'une manière complète les problèmes économiques car elle émet toujours des hypothèses simplificatrices, mais la question qui nous intéresse ici c'est de savoir si la science économique peut se placer au même rang que les sciences exactes, c'est-à-dire on veut connaître si sa démarche et sa méthodologie sont justifiées scientifiquement. Pour traiter cette question, on a intérêt non seulement à déterminer les limites de la science économique, mais aussi à connaître à quel moment on peut considérer la science économique comme une science dure. Nous verrons dans un premier temps ce que c'est la science économique, après on va déterminer son objet, puis nous analyserons les arguments pour et les arguments contre de l'affirmation que la science économique soit une science dure.

I-RETOUR AUX DEFINITIONS ET OBJETS DE LA SCIENCE ECONOMIQUE

Pour définir l'économie, il vaut mieux d'abord partir de la définition étymologique de ce mot « Economie ». Comme on le sait, ce mot vient des mots grecs « *oïkos* » et «*nomos*» qui signifie respectivement : « maison » et « loi ». Donc au début, l'économie était considérée comme l'art de gérer la maison. Ce n'est qu'en 1615 grâce Antoine de Monchrétien que l'économie est considérée comme loi de l'économie de l'Etat grâce à l'apparition du terme « *économie politique* ». L'économie s'est prétendue comme une science objective pendant la révolution marginaliste et avec leurs raisonnements marginaux.

On peut partir de l'objet fondamental de l'économie qui est la gestion des ressources rares. Cet objet sera précisé au fur et à mesure qu'on fait une analyse sur les différentes écoles de pensées économiques. Pour se faire, on va s'appuyer sur l'idée de **GHISLAIN DELEPLACE** dans son livre intitulé « *Histoire de la pensée économique* » édition Dunod, publié pour la première fois en 1999 et republié en 2002.

Pour les mercantilistes et les physiocrates, l'économie est définie en tant que «science des richesses matérielles ». Pour les mercantilistes, leur principale question était : comment accroître les richesses d'une nation ? Mais pour les physiocrates, la question est plutôt ; quelle est l'origine de cette richesse ? La doctrine économique des physiocrates est fondée sur la terre : c'est la terre qui crée la richesse des nations d'où l'importance accordée à l'agriculture.

A l'arrivée de l'école néoclassique, on a assisté à une autre définition de l'économie : « l'économie est la science de la rareté ». L'économie est définie ici par son objet et par sa méthode. On trouve ici la notion de rareté, c'est-à-dire que les ressources destinées à satisfaire les besoins illimités sont rares. Cette rareté implique le choix c'est-à-dire qu'il y a ici une priorisation des besoins. A cause de cette rareté, il ne faut pas gaspiller les ressources et il faut les utiliser d'une manière efficace et optimale. Cela introduit la notion de « rationalité ». Une personne rationnelle est une personne qui cherche à maximiser sa satisfaction ou ses intérêts en dépensant le moins possible. L'économie est définie alors comme une méthode qui va combiner ensuite différentes techniques comme la mathématique, la statistique, etc….

En conclusion, au fil du temps, la définition et l'objet de la science économique changent mais pas pour autant parce qu'une chose est sûr, c'est que quelque soit la définition donnée à l'économie, son objet reste la résolution des problèmes économiques.

La question est maintenant de savoir si l'économie qui essaie d'utiliser la mathématique, est une science exacte, on va aussi analyser les limites de l'approche scientifique en économie.

II- L'ECONOMIE EST-ELLE UNE SCIENCE EXACTE OU DURE ? QUELLES SONT LES LIMITES DE L'APPROCHE SCIENTIFIQUE EN ECONOMIQUE ?

Avant tout, il faut définir d'abord c'est qu'une science. D'une manière générale, une science au sens pur du terme (science dure) est caractérisée par ces trois éléments suivants : un objet spécifique, un projet cognitif, et enfin une méthode analytique.

Pour analyser cette approche scientifique de l'économie, on va se baser sur l'analyse de **DAUTUME** dans son livre intitulé « *L'économie devient-elle une science dure ?* » publié en 1995.

La combinaison des deux premiers critères d'une science (l'objet spécifique et l'objet cognitif) pourrait nous permettre de savoir distinguer les sciences dures des sciences molles.

Pour les sciences dures, elles peuvent se prévaloir d'une certaine vérité de l'objet. On peut prendre comme exemples des sciences dures la physique et la mathématique.

Mais pour le cas des sciences molles, elles ne peuvent pas offrir cette vérité de l'objet. On peut prendre comme exemples la sociologie et la psychologie. Ces sciences molles doivent dans la majorité des cas composer un objet dont elles analysent par avance la relativité. Pour ces sciences molles, Il n' y a pas de vérité exacte, c'est-à-dire, il n' y a que la vérité relative.

Pour les sciences économiques, elles ont une position qu'on peut qualifier de floue car si on se base sur la formalisation mathématique des méthodes qu'elles utilisent, elles se rapprochent des vraies sciences, mais si on se base sur leur objet, elles se rapprochent des sciences sociales.

Si on fait appel à la définition de **Paul SAMUELSON** dans son livre intitulé « *Les fondements de l'analyse économique* » qui définit les sciences économiques comme des sciences des choix, on peut dire que la science économique a pour objet d'analyser les processus de choix et d'opter pour les choix optimaux. Dans la vie quotidienne, des pratiques économiques sont entretenues par les agents économiques et des phénomènes économiques se produisent, c'est inévitable. Tout le monde constate et se rend compte de cette vérité qu'il soit économiste ou pas. Or comme on a dit tout en haut, une science doit avoir une fonction cognitive, donc si les sciences économiques sont des sciences des choix comme a définit P. Samuelson, cela suppose donc que l'objet de la science économique n'est pas l'acte d'échange mais la finalité de cette acte d'échange.

Compte tenu de toutes ces incohérences dans les sciences économiques, on ne sait pas vraiment où les classer. La thèse selon laquelle les sciences économiques sont des sciences dures peut être contestée car la force des sciences exactes tient dans leur

capacité de prévision issue de la vérité de leur objet. Donc si les sciences économiques veulent être considérées comme des sciences dures, elles doivent satisfaire les caractéristiques des sciences dures .On se pose quelque fois la question da la fiabilité des analyses théoriques et pratiques des phénomènes économiques par les sciences économiques. On peut dire donc que la science économique a tendance à être une science de l'incertain.

Le qualificatif de sciences dures des sciences économiques est menacé par l'existence des diverses incertitudes dans ses analyses. Ce concept d'incertitude est un concept flou. Mais cette notion d'incertitude a une petite nuance avec la notion qui est proche d'elle le risque. A vrai dire, le risque est un cas particulier de l'incertitude. Le risque est un concept qui appartient au monde néoclassique. Un événement futur comporte de risque si cet événement est aléatoire mais les individus connaissent la distribution de probabilité de cet événement. Mais le concept d'incertitude appartient au monde keynésien. Un événement est incertain lorsqu'il est aléatoire dans le futur mais cet aléa peut dépendre des actions stratégiques des agents. Mais on constate que dans les sciences économiques, il y a plus d'incertitude que de probabilité. On va voir maintenant la construction de l'économie comme une science au sens pur du terme à travers les différentes époques.

2-1) L'économie comme une science au sens pur du terme

Selon toujours **DAUTUME** dans son livre intitulé « *L'économie devient-elle une science dure ?* » publié en 1995, il montre que la construction de l'économie comme Science à partir du 19ème siècle s'appuyait sur trois éléments fondamentaux suivants:

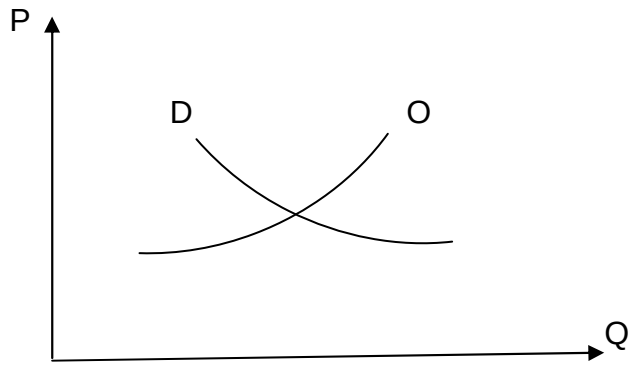
- *les comportements économiques sont rationnels* ;
- *on peut formaliser les comportements économiques ;*
- *et enfin, la dimension monétaire des phénomènes économiques*

a) <u>Les comportements économiques sont rationnels</u>

C'est dans l'analyse néoclassique que cette hypothèse de rationalité des comportements économiques ait une grande place. En effet, les néoclassiques ont pris pour hypothèse que les pratiques et activités économiques sont toujours justifiées par l'intérêt. Pour les agents économiques, les besoins sont nombreux, or les moyens pour satisfaire ces besoins sont limités. Ainsi, chaque individu doit réaliser des choix qui permettent de mieux utiliser au mieux ses moyens pour satisfaire au maximum ses besoins. Comme on a vu précédemment, on peut supposer que le motif décisif du choix est l'intérêt. Les sciences économiques vont faire de cette raison, celle de l'intérêt, leur base d'analyse fondamentale. A la fin du XVIIIème siècle, Adam SMITH suppose, dans son livre "La richesse des nations", que le comportement individuel visant l'intérêt individuel va aboutir à la fin, à la convergence des intérêts de chaque agent économique vers l'intérêt général. On constate donc que les analyses économiques utilisent, dans la plupart du temps, cette hypothèse de la rationalité des comportements économiques.

b) <u>On peut formaliser les comportements économiques</u>

La rationalité économique des agents permet de construire un modèle des comportements économiques. On peut prendre comme exemple de modèle de comportements économiques le modèle de la loi de l'offre et de la demande. Ce sont les néoclassiques qui ont été à la base de ce modèle. Ce modèle peut être schématisé de la manière suivante :



P : prix,
D : Demande des consommateurs,
O : Offre des producteurs,
Q : la quantité des biens.

Du coté des consommateurs, la demande est une fonction décroissante du prix, c'est-à-dire que lorsque les prix augmentent, la demande diminue. En effet, lorsque les prix augmentent, les dispositions à payer des consommateurs diminuent ainsi que leur surplus, donc ils ont intérêt à réduire leur demande pour ne pas être trop lésés.

Par contre du coté des producteurs, l'offre est une fonction croissante du prix, c'est-à-dire que si les prix se trouvent en hausse, la quantité offerte par les producteurs augmente aussi car l'augmentation des prix qui, en provoquant une augmentation des recettes et les bénéfices, incite les producteurs à augmenter leur quantité offerte.

On peut aussi interpréter cette courbe de la manière suivante : lorsque la quantité demandée augmente, les producteurs ont tendance à fixer un prix élevé car ils savent que les consommateurs veulent à tout prix acquérir les biens, soit parce que ces biens leurs procurent plus d'utilité, soit parce qu'ils ont une disposition à payer très élevée.

Cette loi de l'offre et de la demande nous montre donc que les comportements économiques peuvent être formalisés, et ça renforce le statut scientifique des sciences économiques.

c) <u>Les dimensions monétaires des phénomènes économiques</u>

Si l'économie veut être considérée comme une science dure, il faut qu'elle ait une unité de mesure comme les cas dans la mathématique et dans la physique. Pour l'économie, la monnaie constitue cette unité de mesure. Pour l'économiste **Milton FRIEDMAN** dans son livre intitulé « *Inflation: Causes et conséquences* », publié en 1963, considère que toute activité économique dépend essentiellement du volume de monnaie en circulation. Donc pour Friedman, le système économique dépend de la monnaie.

Cependant, il y a des cas ou les activités économiques ne font pas intervenir la monnaie, on peut prendre l'exemple du troc qui ne nécessite pas d'échange monétaire mais qui est considéré comme une activité économique.

Les sciences économiques ont toujours cherché à s'intégrer au statut des sciences exactes. Elles ont utilisé les trois éléments fondamentaux analysés précédemment pour justifier leur statut scientifique.

La question maintenant est de savoir : que se passe t-il lorsque l'un des axiomes fondamentaux cités précédemment est remis en question ?

2-2) L'existence de l'incertitude et le statut scientifique des sciences économiques

Comme on a déjà vu en haut, on peut définir l'incertitude économique comme une situation où des événements économiques futurs sont aléatoires. Mais cet aléa peut dépendre des actions stratégiques des agents économiques.

L'existence de cette incertitude économique peut remettre en cause la rationalité des agents économiques qui est considérée comme un des éléments qui font les sciences économiques des sciences dures. Donc le statut scientifique des sciences économiques est menacé.

On a remarqué qu'à partir du début du 20ème siècle, les économistes mathématiciens ont intégré le concept d'incertitude dans leurs analyses pour consolider le statut scientifique de l'économie.

On peut prendre l'exemple du modèle de **Cox-Ross-Rubinstein** qui est un modèle de valorisation qui se base sur l'équilibre concurrentiel de **WALRAS** et d'**Arrow-Debreux.** Ce modèle a intégré l'incertitude économique sous forme des états contingents de la nature et il a déterminé les différents équilibres concurrentiels pour chacun de ces états contingents.

Par contre, si on prend l'exemple de la théorie des jeux qui est une théorie initiée par les travaux de Von Neumann et Morgenstern dans leur livre intitulé « *Theory of games and economic behavior »,* elle pourrait remettre en cause la rationalité des agents. La théorie des jeux étudie les interactions stratégiques des agents économiques. Elle essaie d'analyser toutes les issues possibles d'un jeu déterminé et met en évidence les processus de décision des agents économiques. Si nous prenons le cas du dilemme des prisonniers qui est un cas célèbre dans la théorie des jeux, il nous montre que, quelque fois, les comportements des individus ne sont pas toujours rationnels. Ce dilemme des prisonniers peut être énoncé de la manière suivante : deux coupables sont interrogés séparément au commissariat de police. Chacun d'eux a deux stratégies : soit il dit que c'est l'autre qui a commis le délit, soit il se tait et ne dit rien. Si les deux coupables se taisent, ils pourront être libérés ou du moins, seront mis en prison pour quelques mois seulement. Mais s'ils se trahissent, ils vont purger des lourdes peines. La théorie des jeux a montré que la stratégie dominante de chacun des deux coupables est de trahir l'autre. Donc à la fin, ils vont se trahir et vont purger des lourdes peines, alors que si les deux n'avaient parlé, ils auraient purgé le minimum de peine. Ce dilemme des prisonniers a montré donc qu'il y a un problème de coordination dans le processus de décision des agents économiques. Cela veut dire donc que dans la réalité, la rationalité des agents économiques n'est pas vérifiée.

Ainsi, dans la réalité, la rationalité des agents économiques peut ne pourra être vérifiée par l'existence de l'incertitude. Pourtant, la prise en compte de la situation d'incertitude a renforcé le statut scientifique des sciences économiques car elles utilisent de plus en plus l'outil mathématique pour traiter le problème.

La prise en compte de l'incertitude va constituer pour les sciences économiques un moyen de se différencier des autres sciences. Les sciences économiques sont considérées comme des sciences capables de prévoir l'avenir même si celui-ci comporte des aléas et des incertitudes.

Ainsi, la scientificité des sciences économiques devance l'incertitude du temps en participant à la construction de l'avenir. On peut considérer donc que l'économie est plus prédictive que descriptive.

La science économique a voulu imiter les méthodes des sciences exactes, et cela a fragilisé son objet. Or dans ce que nous avons vu en haut, ce qui distingue les sciences exactes des sciences molles se trouve dans la vérité de l'objet. En économie, l'objet est flou et n'est pas très explicite. A cause de l'incertitude, l'utilisation de la mathématique en économie n'est qu'un habillage. Les capacités prédictives en économie sont donc faibles. Pour remédier à ce problème, l'économie commence à adopter un outil, cet outil est l'économétrie. On est donc amené à étudier cet outil indispensable de l'économie.

Chapitre 3 : L'ECONOMETRIE

Tout d'abord il faut définir en premier lieu ce que c'est l'économétrie. Il y a plusieurs définitions de l'économétrie qui évolue avec le temps :

-littéralement, on peut définir l'économétrie comme la mesure de l'économie ;
-c'est une application de la statistique mathématique aux données économiques en vue de tester empiriquement les modèles d'économie mathématique et d'en obtenir les résultats numériques ;
-c'est une analyse quantitative des phénomènes économiques actuels basée sur le développement de la théorie et de l'observation lié au méthode d'inférence ;
-l'économétrie est aussi concernée par la détermination empirique des lois économiques ;
- l'art d'un économètre consiste à émettre des hypothèses qui sont à la fois suffisamment spécifiques et réalistes lui permettant de tirer le maximum avantage des données qui lui sont disponibles.

Mais d'une manière générale, l'économétrie est le principal outil d'analyse quantitative utilisé par les économistes dans divers domaines d'application, comme la macroéconomie ou la finance. Les méthodes économétriques permettent de vérifier l'existence de certaines relations entre des phénomènes économiques et de mesurer concrètement ces relations sur des observations réelles.

Pour **FOURGEAUD,** dans son livre intitulé « *Econométrie* » publié en 1978, il définit l'économétrie comme un ensemble de techniques utilisant la statistique mathématique qui vérifient la validité empirique des relations supposées entre les phénomènes économiques et mesurent les paramètres de ces relations. Au sens large, l'économétrie est l'art de construire et d'estimer des modèles empiriques adéquats par rapport aux caractéristiques de la réalité, et cohérents au regard de la théorie économique.

La question qui se pose est de savoir comment les économètres procèdent-ils pour analyser un phénomène économique ?

D'après **Christopher DOUGHERTY**, dans son livre intitulé « *Introduction à l'économétrie* » publié en 2002, la méthodologie classique en économétrie est la suivante :

1. écriture de la théorie ou de l'hypothèse ;
2. spécification en modèle mathématique de la théorie ;
3. recherche des données et analyse des données collectées ;
4. estimation des paramètres du modèle;
5. tests d'hypothèse ;
6. prédiction ou prévision ;
7. et enfin, utilisation du modèle à des fins de contrôle ou de politique économique.

On va analyser une à une ces différentes étapes de la méthodologie économétrique.

I- ECRITURE DE LA THEORIE OU DE L'HYPOTHESE (première étape)

La première étape consiste en premier lieu, à énoncer la théorie ou l'hypothèse. Il faut que cette théorie énoncée ou hypothèse émise soit cohérente. Prenons l'exemple du modèle de consommation Keynésien. L'hypothèse émise est la suivante : « Lorsque le revenu augmente, la consommation augmente aussi mais faiblement proportionnelle.

Il faut donc faire une formalisation pour que les hypothèses émises puissent établir des séries d'équations c'est-à-dire un ensemble de liaisons entre des variables endogènes et exogènes.

II- SPECIFICATION EN MODELE MATHEMATIQUE DE LA THEORIE (deuxième étape)

Une fois que les théories énoncées, les hypothèses émises, il faut maintenant établir les liaisons entre les variables du système qui va constituer le modèle. Il s'agit donc ici de la formalisation mathématique des hypothèses. Il ne faut pas se tromper dans cette formalisation mathématique car si on fait une erreur de spécification (erreur due à la mauvaise liaison systémique utilisée ou aux choix des variables utilisées), le modèle ne sera pas utilisable.

Prenons l'exemple du modèle de consommation keynésien que nous avons vu précédemment. On va donc faire la formalisation mathématique de l'hypothèse émise. On a supposé que la consommation des ménages augmente avec leur revenu mais faiblement proportionnelle.

Dans ce cas, on peut faire la formalisation suivante :

C = f (Y) avec $\frac{\delta C}{\delta Y} > 0$ où C correspond à la consommation, Y au revenu, et la fonction f désigne une fonction quelconque, linéaire ou non linéaire.

$\frac{\delta C}{\delta Y} > 0$ signifie donc qu'une augmentation du revenu va provoquer une augmentation de la consommation.

Il faut remarquer ici que la supposition de départ est cohérente avec la réalité car l'hypothèse émise est vérifiée dans la vie pratique. Cela veut dire donc que si on veut que le modèle soit utilisable, il faut que les hypothèses émises correspondent bien à la réalité.

Maintenant, il faut donner une forme fonctionnelle du modèle, c'est-à-dire une forme quelconque de la fonction *f* qui constitue la liaison entre les variables. Il faut savoir donc qu'on a deux choix : soit choisir un modèle linéaire, soit choisir un modèle non linéaire.

2-1) Choix d'une relation linéaire

Il faut savoir qu'on peut utiliser le modèle linéaire lorsqu'on peut envisager que la variation de la variable expliquée, provoquée par une variation de une unité de l'une des variables indépendantes, est toujours la même.

Si nous supposons que la fonction f est linéaire, voici la formalisation mathématique de l'hypothèse émise :

$C = f(Y) = \beta_o + \beta_1 Y \quad avec \quad \beta_1 > 0$

On remarque facilement ici que : $\beta_1 = \frac{\delta C}{\delta Y} > 0$.

On peut faire l'interprétation suivante: β_1 est la variation de la consommation C quand Y augmente de une unité.

On peut dire donc que si la relation liant les variables endogènes et les variables exogènes est supposée linéaire, chaque paramètre mesure donc la variation de la variable expliquée provoquée par une augmentation de une unité de la variable indépendante concernée, toutes choses égales par ailleurs.

2-2) <u>Choix d'une relation non linéaire</u>

Il est vrai que la linéarité du modèle facilite le travail et le raisonnement, mais dans la réalité, il y a des cas où les liaisons entre les variables du modèle ne sont pas linéaires, c'est-à-dire que la supposition de l'invariance de la variation de la variable expliquée, provoquée par une variation de une unité de l'une des variables indépendantes n'est pas vérifiée. Ici donc, les dérivées partielles par rapport aux variables indépendantes ne sont plus fixes, mais fonction des variables indépendantes.

On peut prendre l'exemple du modèle de croissance de la population. Il serait insensé de penser que la croissance de la population chaque année est constante. En réalité, la population s'accroît de façon exponentielle. Ici donc, la relation du modèle n'est plus linéaire mais non linéaire.

On formalise ce modèle de croissance de la population de la manière suivante :

P = $N_o e^{\alpha t}$ avec P le nombre de la population, t le temps, et N_o et α sont des paramètres avec N_o>0 et α>0

On a ici donc $\frac{\delta P}{\delta t} = \alpha\ N_o\ e^{\alpha t} > 0$.

Cela signifie que la croissance de la population se fait d'une façon exponentielle d'intensité α par rapport à *t.* On voit bien donc ici que la dérivée partielle de la variable

dépendante par rapport à la variable indépendante n'est pas constante mais fonction de la variable indépendante *t.*

Il faut noter qu'il y a des relations non linéaires qui peuvent être transformées en relations linéaires après quelques transformations, par exemple en passant les variables en logarithmes. On peut prendre l'exemple du modèle de production de type Cobb-Douglas suivante:

$Y = \alpha K^{\beta} L^{\gamma}$ où Y, L , K sont respectivement la production, le travail, le capital, et α , β , et γ les paramètres. Après une linéarisation par l'intermédiaire de la fonction log, le modèle donne une relation linéaire suivante:

$\ln (Y) = \ln (\alpha K^{\beta} L^{\gamma}) = \ln(\alpha) + \beta \ln(K) + \gamma \ln(L)$

L'intérêt de cette équation logarithmique réside dans le fait qu'elle permet de voir la variation en pourcentage de la variable expliquée Y, suite à une variation de 1% de l'une des variables explicatives, en supposant que les autres variables restent inchangées.

En d'autre terme, cette équation logarithmique permet de voir les élasticités de la variable expliquée Y par rapport aux variables explicatives K et L. Les paramètres constituent ces élasticités, en effet :

$\beta = \frac{\delta \ln Y}{\delta \ln K} = \frac{\delta Y}{\delta K} \times \frac{K}{Y} = \varepsilon_{Y/K}$ et $\gamma = \frac{\delta \ln Y}{\delta \ln L} = \frac{\delta Y}{\delta L} \times \frac{L}{Y} = \varepsilon_{Y/L}$. On remarque que ces élasticités qui sont β et γ sont constantes.

Il y a une nuance entre la valeur de la dérivée partielle et l'élasticité, pour la dérivée partielle, il s'agit d'une variation en unités, tandis que pour l'élasticité, il s'agit d'une variation en pourcentage.

Il ne faut pas oublier qu'il y a des relations non linéaires qui ne peuvent être transformées en relations linéaires.

III- RECHERCHE DES DONNEES ET ANALYSE DES DONNEES COLLECTEES (troisième étape)

Après avoir fait la spécification en modèle mathématique de la théorie ou de l'hypothèse, c'est-à-dire après avoir fait la représentation en formes fonctionnelles du modèle, il faut faire la confrontation du modèle aux données observées. Pour faire cela, il

faut donc rechercher des données, et après faire l'analyse des données collectées. Il s'agit donc ici d'intégrer la réalité au modèle théorique qu'on ne connaît pas à priori si c'est un modèle fiable ou non.

En général, on peut assister à 3 types de données observées: les séries temporelles ou chronologiques, les données en coupe instantanée, et les données de panel :

– une série temporelle ou chronologique représente une variable observée à intervalles réguliers ;
– pour les données en coupe instantanée, les données sont observées au même instant et concernent les valeurs prises par la variable pour un groupe d'individus spécifiques ;
– et enfin pour les données de panel, la variable représente les valeurs prises par un échantillon d'individus à intervalles réguliers.

IV-ESTIMATION DES PARAMETRES DU MODELE (quatrième étape)

Lorsque l'ensemble des relations du modèle est formulé mathématiquement, et lorsqu'on dispose des données observées de taille assez grande, on peut passer à l'estimation du modèle. Il s'agit de donner des valeurs estimées aux paramètres pour que l'on puisse utiliser le modèle. Mais pour se faire, on a besoin de méthodes.

Divers méthodes sont proposées par les économètres et les statisticiens si on ne cite que la méthode du moindre carré ordinaire et la méthode du maximum de vraisemblance. Mais il faut tout simplement déterminer la méthode la plus adaptée au modèle théorique spécifié.

Aujourd'hui, l'estimation des paramètres d'un modèle peut se réaliser, soit en recourant aux différents programmes informatiques qui proposent diverses méthodes d'estimation, soit en faisant le calcul manuellement.

V- TESTS DES HYPOTHESES (cinquième étape)

Les tests des hypothèses ont pour objet de savoir si les variables explicatives choisies pourront vraiment expliquer la ou les variables expliquées. Ces tests

correspondent aux tests de significativité des paramètres. Il faut également aussi tester les équations du modèle après les estimations des paramètres pour savoir si les équations du modèle peuvent refléter la réalité.

VI- PREDICTIONS OU PREVISIONS (sixième étape)

Avant de faire des vraies prévisions, il faut tout d'abord faire des simulations ex-post et des prévisions ex-post.

6-1) Les simulations ex-post

Les simulations ex-post sont effectuées sur l'ensemble de la période utilisée pour l'estimation des paramètres ; on peut penser que meilleur est le modèle si plus cette simulation donnera des résultats proches de la réalité observée.

6-2) Les prévisions ex-post :

L'utilisation des prévisions ex-post est par contre plus valable ; celles-ci consistent à simuler spontanément le modèle sur un espace de temps (en générale les derniers périodes connues) n'ayant pas été utilisé pour l'estimation. Le défaut de cette méthode est qu'elle exclut de l'estimation un certain nombre d'observations, et réduit donc la significativité des paramètres.

Lorsqu' après ces différentes simulations, le modèle ne s'écarte pas trop de la réalité, on peut donc l'utiliser pour faire des prévisions pour le futur.
Dans le cas contraire, c'est-à-dire lorsque le modèle s'écarte trop de la réalité, cela veut dire qu'on ne peut pas faire la prévision car le modèle n'est pas fiable. Un travail de mise au point s'avère alors nécessaire qui a pour but de corriger les équations non satisfaisantes, ou en cas d'erreur globale, de modifier la structure du modèle c'est-à-dire l'enchaînement des équations. Pour réaliser ces corrections, on recommence les étapes depuis le début.

VII- UTILISER LE MODELE A DES FINS DE CONTROLE ET DE POLITIQUE ECONOMIQUE(Septième étape)

Une fois le modèle est fiable, c'est-à-dire lorsqu'il reflète vraiment la réalité, on peut l'utiliser aux différentes prévisions. Les gouvernements peuvent profiter de cette capacité de prévision du modèle pour assurer l'efficacité des politiques économiques et pour exercer des contrôles. Aujourd'hui, beaucoup d'organismes gouvernementaux tentent de construire des modèles pour permettre de faire des prévisions de la conjoncture économique, et après d'entretenir des actions pour diriger la conjoncture à une direction déterminée.

Après avoir vu les différentes spécificités de la modélisation, de l'économie, et de l'économétrie dans cette la première partie, on va maintenant analyser en deuxième partie, les différentes interactions entre la modélisation, l'économie, l'économétrie, la mathématique, et l'informatique.

DEUXIEME PARTIE

LES INTERACTIONS ENTRE LA MODELISATION, L'ECONOMIE, L'ECONOMETRIE, LA MATHEMATIQUE, ET L'INFORMATIQUE

Dans cette deuxième partie, on va maintenant analyser les différentes interactions entre ces 5 disciplines car certaines de ces disciplines utilisent les autres disciplines comme des outils dans leurs analyses. On peut prendre comme exemple l'économie qui utilise la mathématique et l'économétrie comme outils, et à son tour, l'économétrie va utiliser la mathématique et l'informatique. L'intérêt de cette deuxième partie est donc de savoir l'efficacité et l'aptitude d'une discipline lorsqu'on l'a combinée avec d'autres.

Chapitre 1 LA MODELISATION ET L'ECONOMETRIE

L'économétrie est inséparable avec la modélisation. La combinaison de la modélisation et l'économétrie forme ce qu'on appelle « la modélisation économétrique ». Lorsqu'on fait de la modélisation économétrique, il faut ainsi faire l'estimation du modèle et après, il faut tester le modèle. Le test du modèle sert à vérifier l'hypothèse fixé préalablement c'est à dire vérifier si les variables explicatives pourraient bien expliquer la variable expliquée. Quant à l'estimation du modèle, il existe plusieurs méthodes d'estimation comme la méthode de moindre carré ordinaire, la méthode de maximum de vraisemblance, etc..... Mais la méthode la plus utilisée est la méthode du moindre carré ordinaire (MCO). Dans cette section, on va d'abord étudier la méthode du MCO, ensuite le test du modèle, et enfin les violations des hypothèses classiques du MCO car les hypothèses classiques du MCO ne sont pas toujours respectées.

I - LA METHODE DU MOINDRE CARRE ORDINAIRE :

Le MCO est une méthode fréquemment utilisée pour estimer un modèle. Il consiste à minimiser les erreurs quadratiques c'est-à-dire la somme des carrés des erreurs.

1-1) Modèle linéaire simple

Soit un modèle linéaire suivant : $Y_i = \beta_0 + \beta_1 Xi + \varepsilon_i$ avec ε_i le terme aléatoire c'est-à-dire l'erreur. Ce terme ε_i indique donc qu'il existe des aléas ou qu'on considère qu'on est incapable de comprendre l'infinité des causes qui déterminent le phénomène étudié. Que ce soit pour un modèle linéaire simple ou pour un modèle linéaire multiple, le raisonnement de l'estimation reste le même c'est-à-dire on minimise toujours la somme des carrés des erreurs.

Mais pour pouvoir appliquer le MCO, il faut que les hypothèses suivantes soient respectées.

Voici ces hypothèses :

1- ε_i est une variable aléatoire dont l'espérance est nulle et que sa variance est égale à σ^2 c'est à dire :

$E(\varepsilon_i) = 0$ et $V(\varepsilon_i) = \sigma^2$

2- il faut que les erreurs soient non corrélées c'est-à-dire :

Pour i # j , $Cov(\varepsilon_i, \varepsilon j) = 0$

Cette deuxième hypothèse signifie donc que certaines ou tous les variables Xi ne sont pas liées par une relation affine ;

3- ε_i est une variable aléatoire identiquement et indépendamment distribuée normalement, c'est-à-dire : $\varepsilon_i \rightarrow$ iid N (0 , σ^2) ;

4- Toutes les erreurs ε_i ont la même variance σ^2 et dans ce cas on parle de « homoscédasticité des erreurs » ;

5- La moyenne de Y est sur la droite de régression. En d'autre terme, on fait l'hypothèse qu'il s'agit d'un modèle linéaire c'est-à-dire :
E (Yi/Xi) = β_0 + β_1Xi ;

6- Les Xi sont non stochastiques pour des échantillons répétés. En d'autre terme, on peut dire que la nature de X ne change pas quel que soit le mode de tirage.

Pour que le MCO soit applicable, il faut impérativement que ces hypothèses soient respectées sinon les estimateurs du modèle ne seront pas sans biais ou leurs variances ne seront pas minimales, c'est-à-dire le théorème de Gauss-Markov ne sera pas vérifié. Voici ce théorème : « Les estimateurs sont les meilleurs estimateurs sans biais et la variance est le minimum des variances parmi les classes d'estimateurs ».
C'est vraiment important d'insister sur ces hypothèses parce que même s'il y a une seule hypothèse non respectée, il ne faut pas utiliser le MCO mais il faut faire d'abord des retraitement(c'est ce qu'on va voir dans les cas des violations des hypothèses).
Comme on a dit, le principe du MCO est de minimiser la somme des erreurs. L'erreur est l'écart entre la valeur réelle de la variable à expliquée c'est-à-dire Y, et la valeur estimée de Y que nous notons $\widehat{Y}$.

Donc on a : Min $\sum_{i=1}^{n} \varepsilon_i^2$ = min $\sum_{i=1}^{n}$ (Yi - $\widehat{Y}$)^2 avec n le nombre d'observations.

$$= \min \sum_{i=1}^{n} (Yi - \beta_0 - \beta_1 Xi)^2$$

Comme il s'agit d'une minimisation en vue de chercher les valeurs estimées de β_0 et β_1, il faut faire la dérivation par rapport aux paramètres β_0 et β_1.

On a : $d(\sum_{i=1}^{n} \varepsilon_i^2) / d\beta_0 = -\sum_{i=1}^{n} 2 (Yi - \beta_0 - \beta_1 Xi)$

et $d\left(\sum_{i=1}^{n} \varepsilon_i^2\right)/d\beta_1 = -\sum_{i=1}^{n} 2Xi(Yi - \beta_0 - \beta_1 Xi)$

Donc on a donc le système d'équations suivantes :

$$\begin{cases} -2\sum_{i=1}^{n} (Yi - \beta_0 - \beta_1 Xi) = 0 \quad (1) \\ -2\sum_{i=1}^{n} Xi(Yi - \beta_0 - \beta_1 Xi) = 0 \quad (2) \end{cases}$$

On va utiliser la méthode de substitution pour résoudre ce système.

L'équation (1) donne : $\sum_{i=1}^{n} Yi - n\beta_0 - \sum_{i=1}^{n} \beta_1 Xi = 0$ ce qui donne $\beta_0 = \overline{Y} - \beta_1 \overline{X}$

Avec $\overline{Y}$ et $\overline{X}$ sont les moyennes respectives de X et Y, c'est-à-dire : $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Yi$ et $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} Xi$

L'équation (2) donne : $\sum_{i=1}^{n} XiYi - \beta_0 \sum_{i=1}^{n} Xi - \beta_1 \sum_{i=1}^{n} Xi^2 = 0$

En reportant la valeur de β_0 de l'équation (1) dans cette équation, on a :

$$\sum_{i=1}^{n} XiYi - (\overline{Y} - \beta_1 \overline{X})\sum_{i=1}^{n} Xi - \beta_1 \sum_{i=1}^{n} Xi^2 = 0$$

A partir de cette équation, on peut facilement trouver la valeur de β_1. Voici donc la valeur estimée de β_1 :

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(Xi - \overline{X})(Yi - \overline{Y})}{\sum_{i=1}^{n}(Xi - \overline{X})^2} \quad \text{c'est-à-dire } \hat{\beta}_1 = \frac{Cov(X,Y)}{Var(X)}$$

Donc les valeurs estimées par MCO des paramètres du modèle linéaire simple $Y_i = \beta_0 + \beta_1 Xi + \varepsilon_i$ sont :

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(Xi - \overline{X})(Yi - \overline{Y})}{\sum_{i=1}^{n}(Xi - \overline{X})^2} \quad \text{et} \quad \hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$$

Ces estimateurs de β_0 et β_1 sont des estimateurs sans biais et de variance minimale comme a dit le théorème de Gauss-Markov.

En effet :

$$E(\hat{\beta}_1) = E\left[\frac{\sum_{i=1}^{n}(Xi-\overline{X})(Yi-\overline{Y})}{\sum_{i=1}^{n}(Xi-\overline{X})^2}\right]$$

$$= E\left[\frac{\sum_{i=1}^{n}(Xi-\overline{X})Yi}{\sum_{i=1}^{n}(Xi-\overline{X})^2} + \overline{Y}\,\frac{\sum_{i=1}^{n}(Xi-\overline{X})}{\sum_{i=1}^{n}(Xi-\overline{X})^2}\right] \quad \text{avec } \sum_{i=1}^{n}(Xi-\overline{X}) = 0$$

$$= E\left[\frac{\sum_{i=1}^{n}(Xi-\overline{X})(\beta_0 + \beta_1 Xi + \varepsilon_i)}{\sum_{i=1}^{n}(Xi-\overline{X})^2}\right]$$

$$= E\left[\beta_1\,\frac{\sum_{i=1}^{n}(Xi-\overline{X})Xi}{\sum_{i=1}^{n}(Xi-\overline{X})^2}\right] \quad \text{car } \sum_{i=1}^{n}(Xi-\overline{X}) = 0,\ \sum_{i=1}^{n} Xi\,\varepsilon_i = 0, \text{ et que } \sum_{i=1}^{n}\varepsilon_i = 0$$

Or $\sum_{i=1}^{n}(Xi-\overline{X})^2 = \sum_{i=1}^{n} Xi^2 - n\,\overline{X}^2 = \sum_{i=1}^{n}(Xi-\overline{X})Xi$

Donc $E(\hat{\beta}_1) = E(\beta_1) = \beta_1$ ce qui prouve que $\hat{\beta}_1$ est un estimateur sans biais de β_1.

Pour β_0, on a :

$E(\hat{\beta}_0) = E(\overline{Y} - \hat{\beta}_1\overline{X})$

$= E(\beta_0 + \beta_1\overline{X} + \overline{\varepsilon} - \hat{\beta}_1\overline{X})$

$= E(\beta_0) + E(\beta_1\overline{X}) + E(\overline{\varepsilon}) - E(\hat{\beta}_1\overline{X})$ avec $E(\overline{\varepsilon}) = 0$

$= \beta_0 + \beta_1\overline{X} - \beta_1\overline{X}$ car $E(\hat{\beta}_1) = \beta_1$

$E(\hat{\beta}_0) = \beta_0$ Donc $\hat{\beta}_0$ est un estimateur sans biais de β_0. Donc les deux estimateurs sont tous deux, des estimateurs sans biais.

Mais dans la réalité, il est rare qu'un modèle soit un modèle linéaire simple, mais dans la plupart du temps, les modèles sont des modèles linéaires multiples voire même non linéaires. Ainsi, il est nécessaire d'analyser le cas des modèles linéaires multiples.

1- 2) Modèle linéaire multiple

Pour le modèle linéaire multiple, le principe reste le même mais il faut quand même arranger les écritures pour pouvoir mieux utiliser la méthode, c'est-à-dire il faut donc faire une approche matricielle.

En présence d'un modèle linéaire multiple, en plus des 6 hypothèses classiques dans le modèle linéaire simple, il faut ajouter 2 autres hypothèses supplémentaires suivantes :

- le nombre d'observations doit être supérieur au nombre des variables indépendantes ;
- les variables indépendantes du modèle ne sont pas liées, autrement dit, c'est l'absence de la multicollinéarité.

Voici la forme d'un modèle linéaire multiple :

$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots\ldots + \beta_k X_{ki} + \varepsilon_i$

Sous forme matricielle, on a : $Y = X\beta + \varepsilon$

avec $Y = \begin{bmatrix} Y_1 \\ . \\ . \\ Y_n \end{bmatrix}$ $X = \begin{bmatrix} 1 & X_{11} & . & . & X_{k1} \\ 1 & X_{12} & . & . & X_{k2} \\ . & . & & & . \\ . & . & & & . \\ 1 & X_{1n} & . & . & X_{kn} \end{bmatrix}$; $\beta = \begin{bmatrix} \beta_0 \\ . \\ . \\ \beta_k \end{bmatrix}$; et $\varepsilon = \begin{bmatrix} \varepsilon_1 \\ . \\ . \\ \varepsilon_n \end{bmatrix}$

On va maintenant appliquer le MCO à ce modèle matricielle. Donc on a :

Min $\sum_{i=1}^{n} \varepsilon_i^{\,2} = \min \varepsilon' \varepsilon$ avec ε' est la transposée de ε

$= \min(Y - X\hat{\beta})'(Y - X\hat{\beta})$

$= \min(Y' - \hat{\beta}'X')(Y - X\hat{\beta})$

$= \min (Y'Y - Y'X\hat{\beta} - \hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta})$ avec $Y'X\hat{\beta} = \hat{\beta}'X'Y$

$= \min(Y'Y - 2\,\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta})$

En dérivant par rapport à $\hat{\beta}$, on trouve :

$d(\varepsilon'\varepsilon)/d\hat{\beta} = -2X'Y + X'X\hat{\beta} + \hat{\beta}'X'X$

$= -2X'Y + 2X'X\hat{\beta}$

ce qui nous donne : $\hat{\beta} = (X'X)^{-1}X'Y$

Comme dans le modèle linéaire simple, l'estimateur $\hat{\beta}$ est encore sans biais.

En effet : $\hat{\beta} = (X'X)^{-1}X'Y$

$= (X'X)^{-1}X'(X\beta + \varepsilon)$

$= \beta + (X'X)^{-1} \varepsilon$

Donc on a : $E(\hat{\beta}) = E(\beta + (X'X)^{-1} \varepsilon) = \beta$ car $E(\varepsilon) = 0$

Après l'estimation du modèle, il faut tester le modèle pour savoir si tous les paramètres sont significatifs et il faut aussi connaître la qualité du modèle.

II-LE TEST DU MODELE

Il s'agit ici de tester la significativité des paramètres c'est-à-dire on veut donc ici savoir si les variables explicatives de départ peuvent vraiment expliquer la variable expliquée. Pour se faire, on procède en deux étapes. La première étape consiste à faire les tests de significativité individuelle, et lorsque après avoir fait ces tests de significativité individuelle de chaque paramètre, on a constaté qu'il existe des paramètres non significatifs, il faut dans ce cas là recourir au test de significativité globale.

2-1) Le test de significativité individuelle des paramètres

Pour se faire, il faut d'abord connaître les valeurs estimées des paramètres et les variances ou plus précisément l'écart-type de ces paramètres.
Pour le modèle linéaire simple de la forme : $Y_i = \beta_0 + \beta_1 Xi + \varepsilon_i$, voici les variances des paramètres :

$$V(\beta_0) = \frac{\sigma^2 \sum_{i=1}^{n} Xi^2}{\sum_{i=1}^{n} (Xi - \overline{X})^2} \quad \text{et} \quad V(\beta_1) = \frac{\sigma^2}{n \sum_{i=1}^{n} (Xi - \overline{X})^2}$$

Si on ne connaît pas la valeur de σ^2, on doit l'estimer. L'estimateur de σ^2 est s^2 qui est donné par la formule suivante :

$$s^2 = \frac{\sum_{i=1}^{n} (Yi - \hat{Y})^2}{n-k-1} = \frac{\sum_{i=1}^{n} \varepsilon_i^2}{n-k-1}$$

k est le nombre de paramètres à estimer. Il peut aussi être le nombre de variables indépendantes.

Pour le modèle linéaire multiple, il faut trouver la matrice variance-covariance des paramètres. La matrice variance-covariance du modèle matricielle associé au modèle linéaire multiple que nous avons vu précédemment est la suivante :

$$\text{Varcov}(\beta) = \begin{bmatrix} Var\beta_0 & Cov(\beta_1;\beta_0) & . & . & Cov(\beta_k;\beta_0) \\ Cov(\beta_1;\beta_0) & Var\beta_1 & . & . & Cov(\beta_k;\beta_1) \\ . & . & & & . \\ . & . & & & . \\ Cov(\beta_k;\beta_0) & Cov(\beta_1;\beta_k) & . & . & Var\beta_k \end{bmatrix} = \sigma^2 (X'X)^{-1}$$

avec $\hat{\sigma}^2 = s^2 = \dfrac{\varepsilon'\varepsilon}{n-k-1}$

Les valeurs dans la diagonale représentent les variances des estimateurs.

Voici donc les hypothèses à tester pour le premier paramètre :

$$\begin{cases} \text{Ho : } \beta o = \beta_{10} = 0 \\ \text{H1 : } \beta o \neq 0 \end{cases}$$

Pour faire le test, il faut calculer la valeur du t de student observé et ensuite la comparer avec la valeur du t de student théorique.

Voici la formule de la valeur observée du t de student : $t_{observé} = \dfrac{\beta_0 - \beta_{10}}{\hat{\sigma}_{\beta 0}}$ avec $\beta_{10} = 0$

Il faut regarder dans la table de student la valeur de t de student théorique qui est $t_{(n-k-1;1-\alpha/2)}$ avec α le risque d'erreur.

Voici donc la règle de décision :

- Si la valeur absolue du t de student observé est supérieure à la valeur du t de student théorique, il faut dans ce cas là rejeter l'hypothèse Ho qui est $\beta o = 0$, c'est-à-dire donc que le paramètre βo est significative.
- Par contre si la valeur absolue du t de student observé est inférieure à la valeur du t de student théorique, on accepte donc l'hypothèse Ho.

Pour les autres paramètres, c'est la même chose sauf que la valeur du t de student observé change.

Si le paramètre c'est-à-dire le coefficient d'une variable explicative est significatif, cela veut dire que la variable explicative en question peut très bien expliquer la variable expliquée.

Lorsqu' après avoir fait les tests de significativité individuelle des paramètres, il y a des paramètres non significatifs, il faut dans ce cas recourir au test de significativité globale des paramètres.

2- 2) Le test de significativité globale des paramètres

Il s'agit ici de tester la significativité de l'ensemble des paramètres. On utilise ici le F-test ou le test de Fisher. Le principe reste le même c'est-à-dire il faut comparer la valeur observée et la valeur théorique sauf que dans ce test de significativité globale, on utilise le F-Fisher observé mais pas le t de student.

Voici les hypothèses à tester :

$$\begin{cases} \text{Ho : } \beta_0 = \beta_1 = \beta_2 = \ldots = \beta_k = 0 \\ \text{H1 : } \beta_1 \neq 0 \text{ ; } \beta_2 \neq 0 \text{ ; } \ldots \text{ ; } \beta_k \neq 0 \end{cases}$$

Mais avant de donner la formule de F-Fisher observé, il faut connaître les notions suivantes :

Somme carré résidu (SCRés) = Σi (Yi - $\hat{Y}_i$)^2 = $\sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon'\varepsilon$ = Y'Y - $\hat{\beta}$' X'Y

Somme carré régression (SCRég) = Σi ($\hat{Y}_i - \overline{Y}$)^2 = $\hat{\beta}$' X'Y - n$\overline{Y}^2$

Somme carré totale (SCTotale) = Σi (Yi - $\overline{Y}$)^2 = Y'Y - n$\overline{Y}^2$

La formule de F-Fisher observé est la suivante : F observé = $\frac{SCRég}{SCRés} \times \frac{n-k-1}{k}$

avec k le dégré de liberté de la SCRég et n-k-1 pour la SCRésidu.

Pour avoir le F théorique, il faut regarder dans la table de Fisher la valeur de F(k,n-k-1)

Règle de décision :

- Si la valeur absolue du F observé est supérieure à la valeur de F théorique, il faut dans ce cas rejeter l'hypothèse Ho qui est $\beta_0 = \beta_1 = \beta_2 = \ldots = \beta_k = 0$, c'est-à-dire donc que tous les paramètres sont significatifs.
- Par contre si la valeur absolue du F observé est inférieure à la valeur du F-théorique, on accepte l'hypothèse Ho.

Les tests de significativité ne sont pas suffisants pour savoir la pertinence des variables explicatives, il faut aussi connaître la qualité du modèle à partir du coefficient de détermination R^2.

2-3) Le coefficient de détermination

Ce coefficient est le reflet de l'analyse de la corrélation et il indique la qualité de l'adéquation. Il mesure la part de la déviation de la variable expliquée Y par le modèle, ou en d'autre terme, c'est le pouvoir explicatif du modèle de régression.

Voici la formule de ce coefficient de détermination R^2 : $R^2 = \frac{SCRég}{SCTotal}$

Plus ce coefficient est proche de 1, plus la qualité du modèle est bonne, et inversement c'est-à-dire plus il est proche de 0, plus la qualité du modèle est mauvaise.

Mais comme ce coefficient est un bon indicateur de la qualité d'ajustement, certains économistes ont tendance à trouver un moyen d'augmenter la valeur de ce coefficient par exemple en ajoutant d'autres variables explicatives. Pour contrôler donc la valeur de ce coefficient, il faut calculer le coefficient de détermination ajusté qu'on note $\overline{R}^2$, sa formule est la suivante :

$$\overline{R}^2 = 1 - (1- R^2)\frac{n-1}{n-k-1} = 1- \frac{SCRés}{SCTotale}\times\frac{n-1}{n-k-1}$$

Si la valeur de ce R^2 ajusté est stable c'est-à-dire proche de R^2, cela veut dire que l'ajout d'autres variables explicatives n'influe pas vraiment la qualité du modèle.

La question qui se pose maintenant est donc de savoir, qu'est ce qui se passe lorsque les hypothèses classiques du MCO ne sont pas respectées, quelles sont les influences sur l'estimation et quels sont les retraitements à faire pour qu'on puisse à nouveau utiliser le MCO.

III- LA VIOLATION DES HYPOTHESES CLASSIQUES DU MCO

Lorsqu'on construit un modèle, il est possible que les principales hypothèses du MCO ne puissent pas être vérifiées ou respectées. Par exemple, il se peut la moyenne des erreurs n'est pas nulle, il se peut aussi que les variables explicatives ne sont pas fixes. Il y a aussi d'autres cas de violation des principales hypothèses comme la multicollinéarité, le non respect de l'hypothèse de normalité des erreurs, l'homoscédasticité, et l'autocorrélation des erreurs. Il est intéressant d'analyser ces violations des hypothèses classiques du MCO, de connaître leurs conséquences sur les

estimations et la qualité du modèle, et aussi de pouvoir traiter le problème en vue d'éviter des éventuelles répercussions sur la qualité des estimations et du modèle.

3-1) Cas où l'espérance des erreurs n'est pas nulle

Lorsque l'espérance des erreurs n'est pas nulle, on a deux cas :
-soit l'espérance des erreurs est égale à une constante mais qui est différent de 0 ;
-soit cette espérance est variable c'est-à-dire varie en fonction de l'erreur de chaque observation.

Dans le premier cas, on a : $E(\varepsilon_i) = \varepsilon$. Donc il faut trouver un moyen pour ramener cette espérance à 0.

On peut poser $\varepsilon_i^* = \varepsilon_i - \varepsilon$, On a : $E(\varepsilon_i^*) = E(\varepsilon_i - \varepsilon) = E(\varepsilon_i) - E(\varepsilon) = \varepsilon - \varepsilon = 0$.

Dans ce cas, le modèle devient

$Y_i = \beta_0 + \beta_1 Xi + \varepsilon_i^* + \varepsilon$ car on a posé $\varepsilon_i^* = \varepsilon_i - \varepsilon$

$= (\beta_0 + \varepsilon) + \beta_1 Xi + \varepsilon_i^*$ on peut poser $\beta_0^* = \beta_0 + \varepsilon$

Donc $Y_i = \beta_0^* + \beta_1 Xi + \varepsilon_i^*$

Dans le deuxième cas, c'est à dire dans le cas où l'espérance de l'erreur varie en fonction des erreurs des observations, on a $E(\varepsilon_i) = \varepsilon_i$, on peut quand même faire un changement de variable $\alpha_i = \beta_0 + \varepsilon_i$ dans le modèle.

3-2) Violation de l'hypothèse de normalité de ε_i

La loi suivie par le terme aléatoire n'a pas vraiment d'importance pour le besoin d'estimation dans la pratique, c'est-à-dire elle n'influe pas vraiment l'estimation. Mais l'intérêt de l'hypothèse de normalité se trouve dans l'analyse de la qualité des coefficients estimés de régression.

Si le terme aléatoire suit la loi normale, on peut définir deux mesures :
la variation de Skewness qui mesure la symétrie de la distribution du modèle ; et
la variation de Kurtosis qui est une mesure d'aplatissement.
Voici leurs formules respectives :

S (Skewness) = $\frac{[E(X-m)^3]^2}{[E(X-m)^2]^3}$ avec m la moyenne de X

Et K (Kurtosis) = $\frac{E(X-m)^4}{[E(X-m)^2]^2}$

Il faut cependant noter que l'hypothèse de normalité ne sera pas valable lorsqu'il s'agit d'un échantillon de taille limité car dans la plupart du temps, l'hypothèse de normalité vient du théorème centrale limite qui suppose des observations de taille grande.

3- 3) La multicollinéarité

Ce problème de multicollinéarité est l'un des problèmes le plus fréquent dans la modélisation économétrique. Ici, certains ou toutes des variables explicatives ou indépendantes sont corrélés entre eux. On peut donc ici établir une combinaison linéaire entre certaine ou toutes les variables explicatives.

Donc $\exists \ \lambda_i \in$ IR tel que pour i = 1 à k ; $\sum_{i=1}^{k} \lambda_i Xi = 0$

Si tous les λ_i sont égaux à 0, on parle dans ce cas de parfaite collinéarité, par contre si certains λ_i seulement sont différents de 0, on parle de moin parfaite colinéarité ou tout simplement une multicollinéarité.

Dans le cas d'une parfaite colinéarité, les estimateurs sont indéterminés car la matrice X'X n'est pas inversible, et que les variances des estimateurs sont infinis.

Mais dans le cas d'une simple multicollinéarité, quelles sont ses conséquences, comment la détecter, et enfin quelles sont les solutions pour remédier à ce problème ?

a)Conséquence de la multicollinéarité

En présence de multicollinéarité :

- la matrice variances-covariances des estimateurs sera très élevée ;
- les intervalles de confiances des estimateurs deviennent en conséquences très larges ;
- la plupart des *t* de Student seront non significatifs ;
- le coefficient de détermination va être très élevé avec quelques t de Student significatifs ;
- et enfin les estimations et les variances des estimateurs sont très sensibles à un petit changement des données utilisées.

Voilà les conséquences les plus significatives de la multicollinéarité, mais qu'en est-il de sa détection ?

b) Détection de la multicollinéarité

Les présomptions de l'existence de multicollinéarité peuvent être évoquées dès lorsque les éléments suivants apparaît dans le modèle plus précisément dans le processus d'estimation :

i) le déterminant de la matrice des variables explicatives est égal à 0 et par conséquent cette matrice n'est pas inversible, ce qui empêche la détermination des estimateurs ;
ii) le coefficient de détermination est très élevé alors que la plupart des t de Student sont non significatifs ;
iii) et enfin les coefficients de corrélations partielles des variables explicatives sont très élevés par rapport au coefficient de détermination. Cette constatation est due au test de KLEIN qui a proposé la comparaison des coefficients des corrélations partielles et le coefficient de détermination.

c) Les solutions proposées pour résoudre le problème de multicollinéarité

Divers solutions sont proposées par les économètres lorsqu'on est en présence de multicollinéarité, voici quelques unes :

i) on peut utiliser les transformations à priori sur les paramètres afin de minimiser les effets de la multicollinéarité ;
ii) on peut aussi combiner les variables qui sont corréléés en utilisant la ou les équations qui les lient ;
iii) on peut aussi enlever certaines variables des variables corrélées si c'est possible. Par exemple lorsqu'on a utilisé comme variables explicatives les deux variables PIB par tête et le nombre de la population, effectivement les deux variables sont corrélées, et il faut enlever la variable population pour résoudre le problème ;
iv) on peut très bien aussi transformer les variables, par exemple en utilisant la différence de premier ordre s'il s'agit du temps ;
v) il y aussi une solution qui consiste à ajouter des nouveaux vecteurs de données.

Concernant la combinaison des variables, voici un exemple :

Soit un modèle suivant : $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$

Supposons que X_{1i} et X_{2i} sont corrélées et que par ailleurs, $X_{2i} = \alpha + \delta X_{1i}$

Donc il faut remplacer X_{2i} par l'équation $X_{2i} = \alpha + \delta X_{1i}$, et on a donc :

$Y_i = (\beta_0 + \alpha \beta_2) + (\beta_1 + \delta \beta_2) X_{1i} + \varepsilon_i$

et on peut poser $\beta_0^* = \beta_0 + \alpha \beta_2$ et $\beta_1^* = \beta_1 + \delta \beta_2$

Le seul inconvénient de cette solution est qu'on peut estimer β_0^* et β_1^* mais pas les autres paramètres.

3-4) L'hétéroscédasticité

Ici, l'hypothèse de l'homoscédasticité des erreurs a été violée, ce qui veut dire que les variances des estimateurs ne sont plus constantes mais varient en fonction des observations.

L'hétéroscédasticité peut se présenté dans les modèles d'apprenti à erreurs, ou lorsque la qualité des données collectées sont mauvaises, ou lorsqu'il y a une erreur de spécification du modèle.

En présence d'hétéroscédasticité, on a Var (ε_i) = σ_i^2

Lorsqu'on applique le MCO dans ce cas, les estimateurs des paramètres sont toujours sans biais mais ils ne sont plus efficaces c'est-à-dire que leurs variances ne sont plus minimales.

Donc pour résoudre ce problème, il faut d'abord faire des retraitements.

a) Traitement de l'hétéroscédasticité

Il faut faire en sorte que les erreurs soient homoscédastiques. Pour se faire, il faut diviser l'équation du modèle par l'écart-type des erreurs.

On a : $Y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i$

En divisant par σ_i, on a : $\frac{Yi}{\sigma_i} = \frac{\beta o}{\sigma_i} + \frac{\beta_1}{\sigma_i} X_{1i} + \frac{\varepsilon_i}{\sigma_i}$

et on peut poser $Yi^* = \frac{Yi}{\sigma_i}$; $\beta_0^* = \frac{\beta o}{\sigma_i}$; $\beta_1^* = \frac{\beta_1}{\sigma_i}$; et $\varepsilon_i^* = \frac{\varepsilon_i}{\sigma_i}$

On remarque ici que ε_i^* est homoscédastique, en effet $E(\varepsilon_i^*) = E(\frac{\varepsilon_i}{\sigma_i}) = \frac{1}{\sigma_i^2} E(\varepsilon_i) = 1$

Cette méthode est appelée le moindre carré généralisé (MCG) ou moindre carré pondéré.

En raisonnant matriciellement, on peut utilisé l'estimateur de AITKEN qui est défini de la manière suivante :

$Y = X\beta + \mathcal{E}$

Dans le cas de l'homoscédasticité des erreurs, $Var(\mathcal{E}) = \sigma^2 I$ avec I la matrice identité.

Mais dans le cas de l'hétéroscédasticité, $Var(\mathcal{E}) = \sigma^2 C$ avec C est une matrice symétrique de même dimension que I mais qui est différente de I.

Il existe une matrice B telle que $BC({}^tB) = I$, et que par la suite ${}^tBB = C^{-1}$.

En posant $Y_1 = BY$; $X_1 = BX$, et $\mathcal{E}_1 = B\mathcal{E}$, on a $Y_1 = X_1\beta + \mathcal{E}_1$ avec $\mathcal{E}_1$ homoscédastique car $Var(\mathcal{E}_1) = {}^tBB\ Var(\mathcal{E}_1) = {}^tBB\sigma^2 C = \sigma^2\ C^{-1}C = \sigma^2 I$.

On appelle donc estimateur d'AITKEN, l'estimateur du modèle $Y_1 = X_1\beta + \mathcal{E}_1$ qui est :

$\hat{A} = ({}^tX_1\ X_1)^{-1}\ \ {}^t X_1\ Y_1 == ({}^tX\ C^{-1}X)^{-1}\ \ {}^t X\ C^{-1}\ Y$

avec $E(\hat{A}) = A$ et $Var(\hat{A}) = \sigma^2\ ({}^tX\ C^{-1}X)^{-1}$

b) <u>Détection de l'hétéroscédasticité</u>

Il faut noter tout d'abord que l'hétéroscédasticité apparaît souvent dans les données de panel. Pour détecter l' hétéroscédasticité, on peut dans un premier temps raisonner graphiquement. Lorsqu'on fait la représentation graphique de $\hat{\varepsilon}_i^{\,2}$ en fonction de $\hat{Y}_i$, il y a hétéroscédasticité des erreurs lorsque les nuages des points sur le graphiques ne sont pas dans une bande de fluctuation, dans le cas contraire, il n'y a pas d'hétéroscédasticité.

Pour détecter l' hétéroscédasticité, on peut aussi recourir au *test de Park*. Lorsque que les σ_i ne sont pas connus, on suppose qu' ils sont en fonction de Xi de la manière suivante :

$\sigma_i^{\,2} = \sigma^2\ X_i^{\beta} e^{Vi}$ et en linéarisant, on a : $\ln\sigma_i^{\,2} = \ln\sigma^2 + \beta \ln X_i + Vi$ avec Vi suit N (0 ;1).

On suppose par ailleurs aussi que les $\sigma_i^{\,2}$ peuvent être estimés par $\hat{\varepsilon}_i^{\,2}$.

Donc on a : $\ln\hat{\varepsilon}_i^{\,2} = \ln\sigma^2 + \beta \ln X_i + Vi$,on peut poser $\alpha = \ln\sigma^2$.

Le test de PARK propose donc de calculer d'abord l'erreur $\hat{\varepsilon}_i^{\,2}$, ensuite d'utiliser les valeurs de $\hat{\varepsilon}_i^{\,2}$ pour pouvoir estimer α et β.

Pour connaître s'il y a hétéroscédasticité ou pas, il faut tester β par le test de Student. Si β est significatif, cela veut dire qu'il existe une phénomène d'hétéroscédasticité.

A titre d'information, on peut aussi utiliser d'autres tests pour détecter l'hétéroscédasticité, comme le Test de Goldfeld-Quandt, le test d'égalité des variances, le test de corrélation du rho de Spearman, le test de White, et le Test ARCH

3- 5) L'autocorrélation des résidus

L'autocorrélation signifie que les erreurs sont corrélées entres eux. On assiste à une autocorrélation des résidus lorsqu'il existe une corrélation entre les éléments d'une série temporelle ou des données de panel.
Considérons un modèle linéaire des données en série temporelle suivant :

$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \ldots\ldots + \beta_k X_{kt} + \varepsilon_t$

et on suppose que $\varepsilon_t = \rho\, \varepsilon_{t-1} + v_t$ et que v_t suit la loi $N(0, \sigma_v^2)$, $Cov(\varepsilon_t, \varepsilon_{t-1}) = 0$, $/\rho/<1$
Si on calcule la variance de l'erreur, on obtient :

$Var(\varepsilon_t) = Var[(\rho\, \varepsilon_{t-1} + v_t)^2]$, ce qui nous donne $Var(\varepsilon_t) = \frac{\sigma_v^2}{1-\rho^2}$ car $Var(\varepsilon_t) = Var\, \varepsilon_{t-1}$

a)Conséquences de l'existence de l'autocorrélation des résidus

Il faut savoir tout d'abord que lorsqu'on estime directement le modèle sans tenir compte de l'autocorrélation, les estimateurs seront sans biais mais pas efficaces. Lorsqu'on estime s^2, il est sous évalué. Par contre le coefficient de détermination va être surestimé. Et enfin, les test de Student et de Fisher ne sont plus utilisables ou plus précisément, n'ont pas de sens.

b)Processus d'estimation en présence d'autocorrélation des résidus :

Pour estimer le modèle, il faut partir de l'équation de différence généralisée qui est obtenue par la différence entre Y_t et $\rho\, Y_{t-1}$:

$Y_t - \rho\, Y_{t-1} = (1-\rho)\, \beta_0 + \beta_1 (X_{1t} - \rho\, X_{1t-1}) + \ldots. + \beta_k (X_{kt} - \rho\, X_{kt-1}) + \varepsilon_t - \rho\, \varepsilon_{t-1}$

Posons $Y_t^* = Y_t - \rho\, Y_{t-1}$; $\beta_0^* = (1-\rho)\, \beta_0$; $X_{1t}^* = X_{1t} - \rho\, X_{1t-1}$,…….,

et $\varepsilon_t^* = \varepsilon_t - \rho\, \varepsilon_{t-1}$

On a : $Y_t^* = \beta_0^* + \beta_1 X_{1t}^* + \ldots.. + \beta_k X_{1t}^* + \varepsilon_t^*$

Cette équation de différence généralisée peut être estimé par MCO mais sans le couple (X1,Y1). Cependant on peut remplacer ce couple par ($X1\sqrt{1-\rho^2}$;$Y1\sqrt{1-\rho^2}$).

En effet, si on multiplie l'équation par $\sqrt{1-\rho^2}$, on a une un nouveau terme aléatoire $\mu_t = \varepsilon_t \sqrt{1-\rho^2}$ dont l'espérance est : $E(\mu_t) = E(\varepsilon_t \sqrt{1-\rho^2}) = (1-\rho^2) * \frac{\sigma_v^2}{1-\rho^2} = \sigma_v^2$

On peut utiliser la méthode de Cochrane Orcutt en utilisant l'équation de différence généralisée pour estimer un modèle en présence d'autocorrélation des résidus dont le processus est autorégressif d'ordre 1. En outre, on peut aussi utiliser la méthode de Durbin en 2 étapes pour estimer ρ .

c) Détection de l'autocorrélation :

Le test le plus utilisé pour détecter l'autocorrélation des résidus est le Test de Durbin-Watson. Eventuellement, il y a aussi d'autres test de détection de l'autocorrélation comme le test de Geary .

Pour effectuer le test de Durbin-Watson, il faut calculer la valeur DW qui est égale à :

$$DW = \frac{\sum_{i=2}^{n} (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{i=1}^{n} \hat{\varepsilon}_t^{\,2}}$$

En développant cette formule de DW, et en supposant que $\sum_{t=2}^{n} \hat{\varepsilon}_t^{\,2} = \sum_{t=2}^{n} \hat{\varepsilon}_{t-1}^{\,2}$, on trouve :

$DW = 2(1-\rho)$ avec $\rho = \frac{\sum_{t=2}^{n} \hat{\varepsilon}_t \hat{\varepsilon}_{t-1}}{\sum_{t=1}^{n} \hat{\varepsilon}_t^{\,2}}$

Il faut aussi regarder dans la table de Durbin-Watson les valeurs de DL et DU.

Voici les règles de décisions :

- $\rho > 0$ (autocorrélation positive) si $0 \leq DW \leq DL$
- $\rho = 0$ (Absence d'autocorrélation) si $DU \leq DW \leq 4 - DU$
- Il y a incertitude dans les 2 cas suivants : $DL \leq DW \leq DU$ et $4-DU \leq DW \leq 4-DL$
- $\rho < 0$ (autocorrélation négative) si $4-DL \leq DW \leq 4$

En définitive donc, l'existence ou non de l'autocorrélation peut être analysée par la valeur de DW.

La modélisation économétrique qui est la combinaison de la modélisation et l'économie est devenue un outil très puissant pour la formalisation, l'estimation, et l'interprétation d'un modèle.

On a bien vu que la modélisation économétrique utilise beaucoup la mathématique, et ce n'est pas seulement la modélisation économétrique qui l'utilise mais les autres disciplines aussi comme l'économie, la géographie, etc.…. Il serait intéressant donc d'analyser ces utilisations de la mathématique et économie et en économétrie aussi.

Chapitre 2 : L'UTILISATION DE LA MATHEMATIQUE EN ECONOMETRIE ET EN ECONOMIE

Plusieurs disciplines utilisent la mathématique dans leurs raisonnements et dans leurs études. Même la sociologie commence aujourd'hui à utiliser la mathématique pour étudier le phénomène d'interaction stratégique entres différentes sociétés ou population. Pour l'économétrie, elle ne peut pas se passer de la mathématique car l'économétrie se base sur une formalisation mathématique que nous avons vue dans la section précédente concernant la modélisation et l'économétrie. Concernant l'économie, contrairement à l'économie du 18ème et du 19ème siècle, on assiste aujourd'hui à une tendance vers l'économie quantitative c'est-à-dire l'économie qui utilise la mathématique pour le raisonnement et l'analyse. Les mathématiques utilisées en économétrie et en économie sont souvent l'algèbre linéaire et l'analyse, on utilise aussi la géométrie mais rarement. On va donc étudier l'utilisation de l'algèbre linéaire en économétrie et en économie, et l'utilisation de l'analyse en économétrie et en économie aussi.

I-L'UTILISATION DE L'ALGEBRE LINEAIRE

On va ici analyser séparément l'utilisation de l'algèbre linéaire en économétrie et l'utilisation de l'algèbre linéaire en économie.

1- 1) L'utilisation de l'algèbre linéaire en économétrie

Comme on a vu lorsqu'on a étudié la modélisation et l'économétrie, on a remarqué l'utilisation de l'algèbre linéaire lorsqu'on a évoqué le problème de multicollinéarité. En effet, lorsqu'il existe une multicollinéarité dans le modèle, toutes ou certaines des variables explicatives sont liées par une relation linéaire.

Donc $\exists \lambda_i \in$ IR tel que pour i = 1 à k avec k le nombre de variables indépendantes, $\sum_{i=1}^{k} \lambda_i Xi = 0$.

Si certaines λ_i sont différents de 0, il y donc une multicollinéarité entre un certain nombre de variables indépendantes.

Pour connaître s'il y a multicollinéarité, on peut aussi utiliser la méthode de rang d'un système. Le système qu'on va déterminer le rang c'est le système S{X_{1i} , X_{2i} ,.., X_{ki}}. On va d'abord définir ce que c'est un rang d'un système. Nous appelons rang du système

S{ X_{1i} , X_{2i} ,.., X_{ki} }, la dimension du sous espace engendré par { X_{1i} , X_{2i} ,.., X_{ki} }, c'est-à-dire le nombre maximum de vecteurs linéairement indépendants qu'on peut extraire de { X_{1i} , X_{2i} ,.., X_{ki} }.

Pour mieux appréhender cette notion de rang d'un système, prenons un exemple. Supposons qu'on veuille déterminer le rang du système S suivant : S= $\{U_1, U_2, U_3\}$

avec $U_1 = \begin{bmatrix}1\\0\\2\end{bmatrix}$ $U_2 = \begin{bmatrix}2\\1\\0\end{bmatrix}$ $U_3 = \begin{bmatrix}1\\3\\1\end{bmatrix}$

Pour connaître le rang, il faut faire en sorte que le système ressemble à une matrice triangulaire inférieure ou à une matrice triangulaire supérieure.

On a : S U_1 U_2 U_3

$\begin{bmatrix}1\\0\\2\end{bmatrix}$ $\begin{bmatrix}2\\1\\0\end{bmatrix}$ $\begin{bmatrix}1\\3\\1\end{bmatrix}$

S' : U_1 $U_2' = U_2 - 2U_1$ $U_3' = U_3 - U_1$

$\begin{bmatrix}1\\0\\2\end{bmatrix}$ $\begin{bmatrix}0\\1\\-4\end{bmatrix}$ $\begin{bmatrix}0\\3\\-1\end{bmatrix}$

S'' U_1 $U_2'' = U_2'$ $U_3'' = U_3' - 3U_2'$

$\begin{bmatrix}1\\0\\2\end{bmatrix}$ $\begin{bmatrix}0\\1\\-4\end{bmatrix}$ $\begin{bmatrix}0\\0\\11\end{bmatrix}$

On a transformé le système S à un système S'' dont le triangle supérieur est nulle. Comme les éléments de la diagonale sont tous différents de 0, on peut affirmer que le rang du système S est 3. S'il y avait des éléments nuls dans la diagonale, le rang du système sera le nombre de vecteurs constituant le système diminué du nombre des éléments de la diagonale qui n'était pas nuls. Ce n'est pas seulement dans le problème de multicollinéarité que l'algèbre linéaire est utilisée en économétrie, mais c'est à travers la multicollinéarité qu'il est le plus explicitement utilisé. On se ramène donc à se demander le cas de l'économie, c'est-à-dire quelles sont les utilisations de l'algèbre linéaire en économie ?

1-2- Utilisation de l'algèbre linéaire en économie

L'algèbre linéaire est très utilisée dans les diverses branches de l'économie. Mais on va prendre ici un exemple de l'utilisation de l'algèbre linéaire dans l'une des branches de l'économie qui est la finance. En finance, on peut voir l'utilisation de l'algèbre linéaire dans la recherche des conditions pour qu'un marché financier déterminé puisse être complet.

Considérons un marché financier comportant s états contingents de la nature et un nombre A de titres financiers ou actifs. On veut donc savoir quand est ce que ce marché est complet ? On note α_1 le vecteur de rendement du premier actif et α_A pour le dernier actif. Il faut noter que les α_i comportent s coordonnées car il y a s états contingents sur le marché. Voici la matrice de l'ensemble des revenus de tous les actifs présents sur le marché, on le note R :

$$R=\begin{bmatrix} 1+\alpha_1(1) & . & . & 1+\alpha_A(1) \\ . & & & . \\ . & & & . \\ 1+\alpha_1(s) & . & . & 1+\alpha_A(s) \end{bmatrix}$$

Supposons qu'on compose une portefeuille d'actif composé de x_i actif i. Donc la portefeuille est constitué par $x = (x_1, \ldots, x_A)$. A la fin d'une période, ce portefeuille rapporte :

$$x_1\begin{bmatrix} 1+\alpha_1(1) \\ . \\ . \\ 1+\alpha_1(s) \end{bmatrix} +\ldots..+ x_A\begin{bmatrix} 1+\alpha_A(1) \\ . \\ . \\ 1+\alpha_A(s) \end{bmatrix} = Rx$$

$$Rx = \begin{bmatrix} 1+\alpha_1(1) & . & . & 1+\alpha_A(1) \\ . & & & . \\ . & & & . \\ 1+\alpha_1(s) & . & . & 1+\alpha_A(s) \end{bmatrix}\begin{bmatrix} x_1 \\ . \\ . \\ x_A \end{bmatrix}$$

L'ensemble de revenu qu'on peut atteindre avec les actifs existant sur le marché est égal à Img(R) c'est-à-dire à l'image de la matrice de revenu. Si nous considérons *f* l'application dont la matrice est R, donc Img R $= \{ y \in IR^s / \ni x \in IR^A, f(x) = y\}$.
Le marché est complet lorsque toute configurations de revenu peut être obtenue, c'est-à-dire lorsque Img R = IR^s ou ker *f = 0.* En d'autre terme, le marché est complet lorsque les actifs existants constituent une famille génératrice de IR^s ($A \geq s$) et que par ailleurs , parmi les A actifs financiers existants, on peut tirer une base de IR^s.

Voilà quelques exemples de l'utilisation de l'algèbre linéaire en économétrie et en économie, mais qu'en est–il de l'utilisation de l'analyse ?

II- ***L'UTILISATION DE L'ANALYSE***

Comme dans le cas de l'utilisation de l'algèbre linéaire, on va voir d'abord l'utilisation de l'analyse en économétrie et après l'utilisation de l'analyse en économie.

2-1- L'utilisation de l'analyse en économétrie

L'utilisation de l'analyse est prépondérante en économétrie si l'on ne cite que les différentes minimisations et les traitements des processus autorégressifs. Concernant les processus autorégressifs, on voit bien l'utilisation de l'analyse dans le traitement de l'autocorrélation. En présence d'autocorrélation, les erreurs sont corrélées, en effet :

$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \ldots\ldots + \beta_k X_{kt} + \varepsilon_t$

$\varepsilon_t = \rho\, \varepsilon_{t-1} + v_t$ avec que v_t suit la loi $N(0, \sigma_v^2)$

Calculons la covariance de ε_t et de ε_{t-k} :

$\varepsilon_t = \rho\, \varepsilon_{t-1} + v_t$ avec $\varepsilon_{t-1} = \rho\, \varepsilon_{t-2} + v_t$

$\varepsilon_t = \rho(\rho\, \varepsilon_{t-2} + v_t) + v_t = \rho^2 \varepsilon_{t-2} + v_t + \rho v_t$ avec $\varepsilon_{t-2} = \rho\, \varepsilon_{t-3} + v_t$

……

A la fin, on trouve : $\varepsilon_t = \rho^k \varepsilon_{t-k} + \sum_{i=0}^{k-1} \rho^i v_t$ avec $k \leq t$

$\sum_{i=0}^{k-1} \rho^i$ est une suite géométrique de raison $q = \rho$.

Donc $\sum_{i=0}^{k-1} \rho^i = \dfrac{1-\rho^k}{1-\rho}$

On a donc : $\varepsilon_t = \rho^k \varepsilon_{t-k} + \dfrac{1-\rho^k}{1-\rho}$

$Cov(\varepsilon_t ; \varepsilon_{t-k}) = E(\varepsilon_t\, \varepsilon_{t-k}) - E(\varepsilon_t)E(\varepsilon_{t-k})$ avec $E(\varepsilon_t) = E(\varepsilon_{t-k}) = 0$

$= E\left[\left(\rho^k \varepsilon_{t-k} + \dfrac{1-\rho^k}{1-\rho}\right)\varepsilon_{t-k}\right]$

$= \rho^k E(\varepsilon_{t-k}^2) + \dfrac{1-\rho^k}{1-\rho} E(\varepsilon_{t-k})$ avec $E(\varepsilon_{t-k}^2) = Var(\varepsilon_{t-k}) = \dfrac{\sigma_v^2}{1-\rho^2}$

Donc : $Cov(\varepsilon_t ; \varepsilon_{t-k}) = \rho^k \times \dfrac{\sigma_v^2}{1-\rho^2}$

Si $|\rho| < 1$, la $Cov(\varepsilon_t ; \varepsilon_{t-k})$ converge vers 0 lorsque k tend vers l'infini.

On voit bien ici donc le rôle de l'utilisation de l'analyse en économétrie dans la résolution d'un problème donnée. Ce n'est pas seulement dans le traitement du problème de l'autocorrélation que l'analyse est utilisée mais elle est un outil de l'économétrie. L'analyse est donc devenue inséparable de l'économétrie. On a vu l'utilisation de l'analyse en économétrie, on va voir maintenant, l'utilisation de l'analyse en économie.

2-2- L'utilisation de l'analyse en économie

Comme il y a aujourd'hui la tendance vers l'économie quantitative, tous les économistes essaient d'intégrer dans leurs raisonnements la mathématique car la mathématique permet la rigueur du raisonnement. L'utilisation de l'analyse est nombreuse en économie sans ne citer que les procédées d'optimisation des différentes fonctions (fonction de production, la fonction du coût, ..), et la résolution des équations différentielles stochastiques liant des variables économiques comme le PIB, la consommation, le nombre de la population,etc.…. L'utilisation de la mathématique va sûrement s'intensifier dans les prochaines années à venir car les économistes chercheront toujours à rendre leurs raisonnements incontestables et ça par la formalisation mathématique. Ce n'est pas seulement la mathématique qui pourrait influencer l'économétrie mais l'informatique aussi qui a beaucoup révolutionné le monde depuis ces 20 dernières années. On va donc analyser les apports de l'informatique dans l'économétrie et savoir pourquoi les économètres ont tendance à recourir aux outils informatiques.

Chapitre 3 : L'INFORMATIQUE DANS LA MODELISATION ET DANS L'ECONOMETRIE

Aujourd'hui, l'informatique est présente dans presque tous les domaines que ce soit dans la médecine, dans l'aéronautique, dans le transport aérien, et dans les domaines qui, auparavant, n'ont pas utilisé l'informatique. En un mot, l'informatique devient un outil incontournable dans le monde. La modélisation et l'économétrie ou plus précisément la modélisation économétrique n'échappent à ce pouvoir de l'informatique. En effet, la modélisation économétrique a recours à l'outil informatique pour traiter les données et pour faire l'estimation des paramètres. L'utilisation de l'informatique peut être justifiée par le fait qu'il permet de traiter un nombré élevé d'équations des modèles. Supposons qu'un modèle comporte une cinquantaine d'équations voire par centaine, un être humain prendra beaucoup de temps à traiter tous ces équations et le risque de commettre des erreurs est élevé. C'est l'informatisation qui permet d'éviter ce problème, et en plus, l'informatique mettra quelques secondes pour traiter l'ensemble des données, donc il permet la rapidité du travail.

L'informatique intervient dans la modélisation économétrique à travers les différents logiciels de traitement des données, qui sont proposés sur le marché. Les plus connus de ces logiciels sont : SAS, STATA, SPSS, et E-views. La plupart des concepteurs des logiciels économétriques sont en majorité des sociétés américaines, mais il y a aussi aujourd'hui des sociétés européennes qui se lancent dans la conception des logiciels économétriques et de traitement des données.
On est amené à se demander quels sont les critères possibles du choix de logiciel pour qu'on puisse déterminer le logiciel le plus adapté aux besoins.

I- LE CRITERE DU CHOIX DU LOGICIEL :

On peut avancer les différents critères suivants pour pouvoir choisir au mieux le logiciel d'estimation et de modélisation qu'on va utiliser :

- l'étendue des méthodes d'estimation proposées : il serait mieux que le logiciel choisi comporte plus de méthodes d'estimation que possible ;
- la convivialité présentée par l'interactivité entres les tâches de spécification des équations et d'estimation proprement dites ; présentée aussi par l'accès implicite dans un ordre choisi et par des statistiques et des graphiques claires et faciles à produire ; et enfin présentée par un langage souple d'écriture des équations, comportant tous les opérateurs logiques et mathématiques nécessaires ;

- il faut qu'il y ait une normalisation automatique des équations ;
- la rapidité des algorithmes d'estimations, qui concernera surtout l'emploi des méthodes les plus complexes ;
- la facilité de la gestion des données.

Pour mieux connaître les apports de l'informatique dans la modélisation économétrique, on va prendre un logiciel et après on va déterminer tous ce que ce logiciel est capable de faire, ensuite on va prendre un exemple de modèle et de l'estimer à partir du logiciel. Ici, on va prendre le logiciel SPSS version 13.0.1 qui est un logiciel reconnu par sa capacité à traiter les données et par la simplicité de l'utilisation.

II- CAS DU LOGICIEL SPSS 13.0.1 :

Cette version de SPSS 13.0.1 a été conçue par « The Apache Software Foundation » en Novembre 2004. Il a été le mis à jour du version 13.0. Ce logiciel SPSS a beaucoup été utilisé dans les traitement des données et dans la modélisation économétrique pour plusieurs raisons : sa puissance de traiter les données, son utilisation facile et très simple, sa rapidité, et pour beaucoup d'autres raisons.

1- Qu'est-ce qu'on peut faire avec SPSS 13.0.1 ?

On peut faire beaucoup de chose avec SPSS 13.0.1. En effet, on peut faire :

- les régressions linéaires, les régressions logarithmiques, et les régressions des fonctions de puissances ;
- l'ajustement des fonctions ;
- le T-Test pour les échantillons indépendants, pour les échantillons appariés, et pour un échantillon unique ;
- l'ANOVA à 1 facteur : la procédure de l'analyse de variance ANOVA à 1 facteur permet d'effectuer une analyse de variance univariée sur une variable quantitative dépendante par une variable critère simple (indépendant). L'analyse de variance sert à tester l'hypothèse d'égalité des moyennes. Cette technique est une extension du test t pour deux échantillons ;
- l'analyse GLM univairié : le GLM – Univarié fournit un modèle de régression et une analyse de la variance pour plusieurs variables dépendantes par un ou plusieurs facteurs ou variables. Les variables actives divisent la population en groupes. Cette procédure de régression linéaire généralisée permet de tester les hypothèses nulles à propos des effets des autres variables sur la moyenne des différents regroupements de la variable dépendante. On peut rechercher les interactions entre les facteurs ainsi que les effets des différents facteurs, certains d'entre eux

étant aléatoires. En outre, les effets et les interactions des covariables avec les facteurs peuvent être inclus. Pour l'analyse de la régression, les variables indépendantes (explicatives) sont spécifiées comme covariables ;

- les corrélations bivariées : la procédure de corrélations bivariées calcule le coefficient de corrélation de Pearson, le rho de Spearman et le tau-b de Kendall avec leurs seuils de signification. Les corrélations mesurent comment les variables ou les ordres de rang sont liés. Avant de calculer un coefficient de corrélation, on doit parcourir les données pour rechercher les valeurs éloignées (qui peuvent provoquer des résultats erronés) et les traces d'une relation linéaire. Le coefficient de corrélation de Pearson est une mesure d'association linéaire. Deux variables peuvent être parfaitement liées, mais si la relation n'est pas linéaire, le coefficient de corrélation de Pearson n'est pas une statistique appropriée pour mesurer leur association ;
- la corrélation partielle : la procédure des corrélations partielles calcule les coefficients de corrélation partielle qui décrivent le rapport linéaire entre deux variables tout en contrôlant les effets d'une ou plusieurs autres variables. Les corrélations sont des mesures d'association linéaire. Deux variables peuvent être parfaitement liées mais, si leur rapport n'est pas linéaire, un coefficient de corrélation n'est pas une statistique adaptée pour mesurer leur association ;
- l'analyse discriminante : l'analyse discriminante est utile pour les cas où on veut construire un modèle de prévision de groupe d'affectation basé sur les caractéristiques observés de chaque observation. La procédure génère une fonction discriminante (ou, pour plus de deux groupes, un ensemble de fonctions discriminantes) basée sur les combinaisons linéaires des variables explicatives qui donnent la meilleure discrimination entre groupes. Les fonctions sont générées à partir d'un échantillon d'observations pour lesquelles le groupe d'affectation est connu. Les fonctions peuvent alors être appliquées aux nouvelles observations avec des mesures de variables explicatives, mais de groupe d'affectation inconnu ;
- l'analyse factorielle : elle essaie d'identifier des variables sous-jacentes, ou facteurs, qui permettent d'expliquer l'ensemble des corrélations à l'intérieur d'un ensemble de variables observées. L'analyse factorielle est souvent utilisée pour réduire un ensemble de données. L'analyse factorielle est souvent utilisée dans la factorisation, en identifiant un petit nombre de facteurs qui expliquent la plupart des variances observées ;
- les tests non paramétriques qui comprenent le test de Khi-deux, le test binomial, le test de Kolmogorov-Smirnov pour un échantillon ; et les autres tests non paramétriques ;
- l'analyse des réponses multiples.

Il y aussi d'autres chose que le logiciel SPSS 13.0.1 peut faire autres que ceux déjà cités. Donc on peut remarquer que ce logiciel est capable de faire l'ensemble des analyses des données et l'estimation des modèles. On se demande maintenant quelle est la procédure d'estimation et de test pour ce logiciel, pour se faire, on va faire un cas pratique de l'utilisation du logiciel.

2- Cas pratique de l'utilisation du logiciel SPSS 13.0.1 :

Supposons qu'on veut faire l'estimation du modèle d'investissement en vue de faire une prévision à court terme des investissements. Voici la forme du modèle :
$I_t = a + bPIB_t + cR_t + d\,T_t + \varepsilon_t$ avec I_t les investissements, PIB_t les Produits intérieurs bruts, R_t le taux directeur de la banque centrale, T_t le temps, et ε_t le terme aléatoire.
Par ailleurs, on dispose des 14 observations suivantes pour le pays de Madagascar :

Années	PIB Courant	Inflations	Investissements nominaux	Taux d'interêt
1990	4604,1	11,5	781,4	12
1991	4913,8	13,9	401,4	12
1992	5593,1	12,5	631,9	12
1993	6450,9	13	738,5	12
1994	9131,4	41,6	995,6	20,5
1995	13478,7	45,1	1474,9	33
1996	16224,3	17,8	1888	17
1997	18050,9	7,3	2139,6	11,9
1998	20349,5	8,5	2678	10
1999	23384	9,7	3374	15
2000	26242	7,1	4250	12
2001	29843	7,3	5340	9
2002	30042	40,5	4019	9
2003	33893	2,8	5489	7
2004	40778	14,3	10193	16

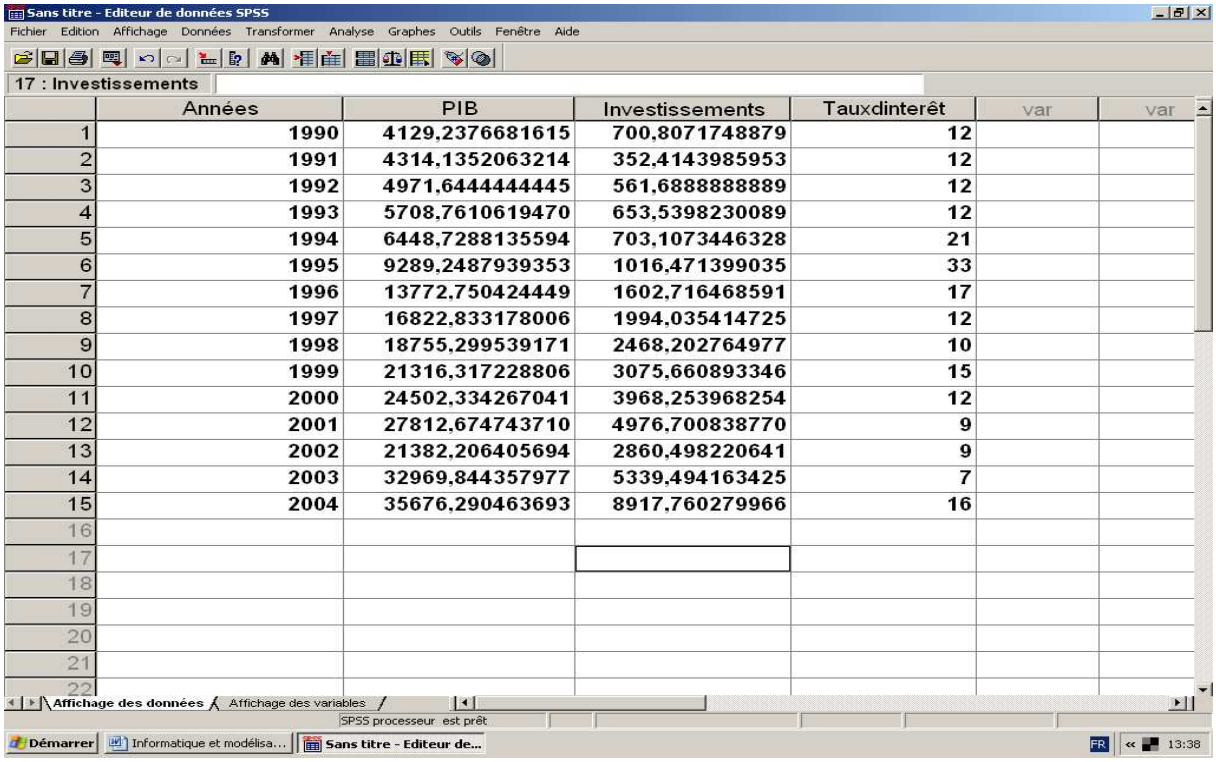
Source : **Rapport annuel de la Banque Centrale de Madagascar, Année 2006.**
NB : Les PIB et les Investissements sont en milliards de Fmg

Pour éviter l'existence d'une éventuelle multicollinéarité entre les variables, il faut mettre les PIB et les investissements en termes réels, pour se faire, il suffit de diviser le PIB courant et les investissements nominaux par les indices des prix et après les multiplier par 100.

Après le retraitement, on a le tableau suivant :

Années	PIB	Investissement	Taux d'interêt
1990	4129,23767	700,8071749	12
1991	4314,13521	352,4143986	12
1992	4971,64444	561,6888889	12
1993	5708,76106	653,539823	12
1994	6448,72881	703,1073446	20,5
1995	9289,24879	1016,471399	33
1996	13772,7504	1602,716469	17
1997	16822,8332	1994,035415	11,9
1998	18755,2995	2468,202765	10
1999	21316,3172	3075,660893	15
2000	24502,3343	3968,253968	12
2001	27812,6747	4976,700839	9
2002	21382,2064	2860,498221	9
2003	32969,8444	5339,494163	7
2004	35676,2905	8917,76028	16

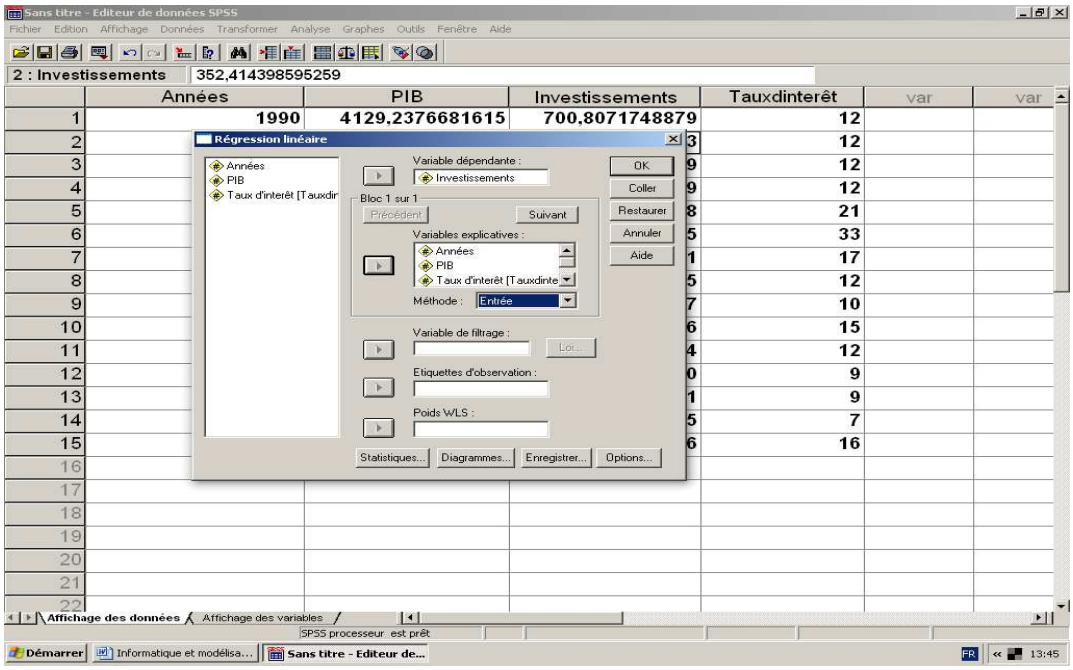
On va maintenant introduire les données dans le logiciel SPSS. On a deux choix : soit saisir les données dans une feuille microsoft excel et après on l'importe, soit on saisit directement les données dans l'éditeur de données de SPSS.
Voici ce qu'on voit lorsqu'on a introduit les données dans SPSS :



	Années	PIB	Investissements	Tauxdinterêt	var	var
1	1990	4129,2376681615	700,8071748879	12		
2	1991	4314,1352063214	352,4143985953	12		
3	1992	4971,6444444445	561,6888888889	12		
4	1993	5708,7610619470	653,5398230089	12		
5	1994	6448,7288135594	703,1073446328	21		
6	1995	9289,2487939353	1016,471399035	33		
7	1996	13772,750424449	1602,716468591	17		
8	1997	16822,833178006	1994,035414725	12		
9	1998	18755,299539171	2468,202764977	10		
10	1999	21316,317228806	3075,660893346	15		
11	2000	24502,334267041	3968,253968254	12		
12	2001	27812,674743710	4976,700838770	9		
13	2002	21382,206405694	2860,498220641	9		
14	2003	32969,844357977	5339,494163425	7		
15	2004	35676,290463693	8917,760279966	16		
16						
17						
18						
19						
20						
21						
22						

Il ne faut pas oublier de bien définir les types de chaque variable dans l'affichage des données qui se trouve en bas de l'écran.

On va maintenant estimer notre modèle $I_t = a + bPIB_t + cR_t + d\ T_t + \varepsilon t$. Pour se faire, on entre dans Menu Analyse, ensuite Régression, et après Linéaire, et on voit quelque chose comme celle-ci :



Il faut maintenant choisir la variable indépendante et les variables indépendantes. Pour notre cas ici, l'Investissement est la variable dépendante et les autres variables sont des variables indépendantes, et après on valide. Après avoir validé, le logiciel donne directement le résultat de l'estimation (MCO) comme on peut le voir dans le schéma suivant :

Variables introduites/éliminées (b)

Modèle	Variables introduites	Variables éliminées	Méthode
1	Taux d'interêt, Années, PIB(a)	.	Introduire

a Toutes variables requises introduites
b Variable dépendante : Investissements

Récapitulatif du modèle

Modèle	R	R-deux	R-deux ajusté	Erreur standard de l'estimation
1	,959(a)	,920	,898	761,519240740498000

a Valeurs prédites : (constantes), Taux d'interêt, Années, PIB

ANOVA (b)

Modèle		Somme des carrés	ddl	Carré moyen	F	Signification
1	Régression	72939289,936	3	24313096,645	41,926	,000(a)
	Résidu	6379027,094	11	579911,554		
	Total	79318317,030	14			

a Valeurs prédites : (constantes), Taux d'interêt, Années, PIB
b Variable dépendante : Investissements

Coefficients(a)

Modèle		Coefficients non standardisés		Coefficients standardisés	t	Signification
				B	Erreur standard	Bêta
1	(constante)	665951,790	368026,252		1,810	,098
	Années	-335,380	184,980	-,630	-1,813	,097
	PIB	,354	,079	1,588	4,476	,001
	Taux d'interêt	40,651	34,978	,108	1,162	,270

a Variable dépendante : Investissements

Ici donc, le logiciel donne tout concernant le modèle comme les coefficients de détermination, le tableau ANOVA, les valeurs des estimateurs et les t de student observé.
Grâce à l'informatique, une fraction de seconde permet d'estimer un modèle quelque soit les nombres des données. Mais si les nombres des variables sont beaucoup par exemple par centaine ou par millier ; il faut déterminer les variables les plus pertinentes qui caractérisent la population statistique étudiée, et pour se faire, on peut utiliser l'analyse factorielle des correspondances, on peut aussi utiliser l'analyse en composantes principales.
L'informatique a beaucoup permis de développer le monde de l'analyse des données et de la modélisation économétrique, et sa place dans ces domaines ne cesse de s'accroître d'en années en années

CONCLUSION

On peut dire que la modélisation et l'économétrie ont beaucoup permis aux sciences économiques de mieux étudier les phénomènes économiques. Les études de ces phénomènes économiques ont pour objet de connaître les mécanismes et les causes, et de pouvoir mieux contrôler ces phénomènes économiques. Aujourd'hui, la modélisation économétrique s'est beaucoup développée, et elle est maintenant utilisée dans presque tous les domaines que ce soit dans le domaine du transport aérien, de la démographie, et dans d'autres domaines aussi.

Cependant, les modèles peuvent présenter des inconvénients. Le premier inconvénient est le problème d'agrégation, c'est-à-dire que pour obtenir une forme fonctionnelle opérationnelle dans les comportements des divers agents économiques, le modélisateur utilise de manière courante un certain nombre d'hypothèses simplificatrices telles que celles de l'agent représentatif pour en dériver le comportement normal ou désiré pour le court, le moyen, et le long terme. Il y a aussi le problème de raisonnement car le modélisateur raisonne bien souvent, non en terme d'agent, mais en terme de grandes fonctions macroéconomiques. Il y a ici donc un antagonisme car on veut déterminer les comportements des agents économiques qui sont du domaine de la microéconomie, en raisonnant en terme macroéconomique. Enfin, il y a le critique de Sims qui dit que quelque fois, on traduit un modèle qui n'est ni justifié par la théorie économique, ni par l'économétrie.

Quoiqu'il en soit, une chose est sûre, c'est que les modélisateurs et les économètres essaient toujours de se rapprocher le plus près possible, par tous les moyens, de la réalité en réduisant les possibilités d'erreurs. Pour se faire, ils ont intégré les outils informatiques pour minimiser ces possibilités d'erreurs et pour permettre la rapidité et l'efficacité du travail. Mais la question qui se pose est de savoir quand est-ce que la modélisation économétrique va atteindre son but ultime qui est d'être parfait, c'est-à-dire de pouvoir faire des prévisions exactes. Est-ce dans 10 ans ? Dans 20 ans ? Ou jamais ?

ANNEXE

Les PIB Courants, les IN (Investissements nominaux), les C. PUB (Consommations publiques), et les C. Privées (Consommations privées) sont en milliards de Fmg.

Années	PIB Courant	Inflations	IN	TI	C. PUB	C. Privées
1990	4604,1	11,5	781,4	12	383,6	3965,1
1991	4913,8	13,9	401,4	12	436,6	4493,7
1992	5593,1	12,5	631,9	12	498,3	4929,9
1993	6450,9	13	738,5	12	561,2	5751
1994	9131,4	41,6	995,6	20,5	723,9	8102,8
1995	13478,7	45,1	1474,9	33	904,2	12120,8
1996	16224,3	17,8	1888	17	731,9	14462,9
1997	18050,9	7,3	2139,6	11,9	1097,7	16294,7
1998	20349,5	8,5	2678	10	1621	17726
1999	23384	9,7	3374	15	1741,1	5867
2000	26242	7,1	4250	12	2064	7984
2001	29843	7,3	5340	9	2558	8735
2002	30042	40,5	4019	9	2510	26221
2003	33893	2,8	5489	7	3551	29368
2004	40778	14,3	10193	16	3726	34152

TI : Taux d'intérêt de la Banque Centrale

Source: **INSTAT**

LISTE DES ABREVIATIONS

ARCH : AutoRegressive Conditional Heteroscedasticity

ARMA : AutoRegressive with Moving Average

EDS : Equation Différentielle Stochastique

GLM : General Linear Model

INSTAT : Institut National de la Statistique

MCO : Moindre Carré Ordinaire

MCG : Moindre Carré Généralisé

PIB : Produits Intérieurs Bruts

SAS : Statistical Analysis System

SPSS : Statistical Package for the Social Science

STATA : Statistical Analysis

VAR : Vectorial AutoRegressive

BIBLIOGRAPHIE

- DELEPLACE Ghislain*, « Histoire de la pensée économique »,* Dunod, 1999, france
- DAUTUME, « *L'économie devient- elle une science dure ?* », Routledge, 1995, France
- Orley Ashenfelter**,** "*Statistics and Econometrics: Methods and Applications*" Wiley, 2001, USA
- Johnston, J. & J. DiNardo, "*Econometric Methods*", 4th Edition.MacGraw Hill, 1997, USA
- Lardic, S. & V.Mignon, « *Econométrie des séries temporelles macroéconomiques et financieres"*, Economica, 2002, Paris
- Kevin M. Currier , "*Comparative Statics Analysis in Economics*", World Scientific, 2000, USA
- Christopher Dougherty , "*Introduction to Econometrics, Pearson Education"*, 2000, USA
- George G. Judge, R. Carter Hill, William E. Griffiths , "*Introduction to the Theory and Practice of Econometrics*" , Wiley, 1988, USA
- Cheng Hsiao , « *Analyses des données de Panel* », WIE,1997, France

RESUME ANALYTIQUE

Nom: CHAN ZEN

Prénom : Rollin

Encadreur: Monsieur RAZAFINDRAVONONA Jean

Intitulé du mémoire: « La modélisation, l'économie, l'économétrie, la mathématique, et l'informatique

Résumé:

Des théories économiques se sont inspirées des phénomènes économiques comme le cas de la théorie Keynésienne qui a pu naître grâce à la crise des années 30. Mais l'étude des phénomènes économiques nécessite l'existence des outils capables de déceler les mécanismes des phénomènes économiques. C'est pourquoi dans le début du 20ème siècle, les économistes commencent à adopter et à intégrer la mathématique dans leurs analyses, et ils ont commencé à faire des modélisations des réalités et phénomènes économiques. Mais résoudre un modèle nécessite à son tour l'existence d'une technique et méthodes appropriées, cette technique est l'économétrie.

La modélisation et l'économétrie ont beaucoup permis aux sciences économiques de mieux étudier les phénomènes économiques. Les études de ces phénomènes économiques ont pour objet de connaître les mécanismes et les causes, et de pouvoir mieux contrôler ces phénomènes économiques.

L'intérêt de ce mémoire est de donc de comprendre pourquoi les économistes ont du recourir à la modélisation, à l'économétrie, à la mathématique, et à l'informatique pour faire les analyses économiques.

Nombre de pages : 60

Nombre de figures :3

Nombre de tableau :8

Adresse : N°85 Villa ARLES Nord, Cité provence, Fort-Voyron , Ouest ambohijanahary

Contact : 033 11 828 75

E-mail : rollin.zen@hotmail.fr