



La politique monétaire dans les modèles économétriques : primat de la théorie sur l'empirie

Abdelali Attioui

► To cite this version:

Abdelali Attioui. La politique monétaire dans les modèles économétriques : primat de la théorie sur l'empirie. Économies et finances. Université de Grenoble, 2014. Français. <NNT : 2014GRENE004>. <tel-01160084>

HAL Id: tel-01160084

<https://tel.archives-ouvertes.fr/tel-01160084>

Submitted on 4 Jun 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Sciences Economiques**

Arrêté ministériel : 7 août 2006

Présentée par

Abdelali ATTIOUI

Thèse dirigée par **Rédouane TAOUIL**

préparée au sein du **Centre de Recherche en Economie de Grenoble**

dans l'**École Doctorale de Sciences Economiques**

La Politique Monétaire dans les Modèles Econométriques : Primat de la Théorie sur l'Empirie

Thèse soutenue publiquement le **4 Décembre 2014**, devant le jury composé de :

M. Jean Pierre ALLEGRET

Professeur des Universités, Université Paris Ouest Nanterre La Défense,
(Rapporteur)

Mme Jézabel COUPPEY-SOUBEYRAN

Maître de Conférences, Habilitée à Diriger les Recherches, Université Paris 1
Panthéon-Sorbonne, (Rapporteur)

M. Guéliffo HOUNTONDI

Professeur des Universités, Université Pierre Mendès-France, Grenoble 2,
(Président de jury)

M. Rédouane TAOUIL

Professeur des Universités, Université Pierre Mendès-France, Grenoble 2,
(Directeur de thèse)



*L'Université n'entend donner aucune approbation
ni improbation aux opinions émises dans la thèse,
celles-ci devront être considérées comme propres
à leur auteur.*

Thèse de doctorat *es* Sciences Economiques

Soutenance le 4 décembre 2014 (salle 110, BATEG)

Laboratoire : Centre de Recherche en Economie de Grenoble

**La Politique Monétaire dans les Modèles Econométriques :
Primat de la Théorie sur l'Empirie**

Doctorant : ATTIOUI Abdelali

Directeur : TAOUIL Rédouane

Jury :

M. Jean Pierre ALLEGRET

Professeur des Universités, Université Paris Ouest Nanterre La Défense, (Rapporteur)

Mme Jézabel COUPPEY-SOUBEYRAN

Maître de Conférences, Habilitée à Diriger les Recherches, Université Paris 1 Panthéon-Sorbonne, (Rapporteur)

M. Guéliffo HOUNTONDI

Professeur des Universités, Université Pierre Mendès-France, Grenoble 2, (Président de jury)

M. Rédouane TAOUIL

Professeur des Universités, Université Pierre Mendès-France, Grenoble 2, (Directeur de thèse)

Résumé

La Politique Monétaire dans les Modèles Econométriques : Primat de la Théorie sur l'Empirie

En s'appuyant sur les limites de l'économétrie mises en évidence dans les débats sur la politique monétaire depuis les années 1960, cette thèse s'attache à montrer le primat de la théorie sur l'empirie et que l'économétrie ne peut pas être décisive dans la remise en cause de la théorie. Nous adoptons une démarche basée sur des arguments épistémologiques pour montrer que ces débats dépassent le clivage théorie/empirie et intègrent une différence de vision quant à l'utilité d'un modèle empirique.

Le programme de recherche de la Commission Cowles s'est constitué autour d'une articulation particulière de trois éléments fondamentaux. Un référentiel théorique issu de la Théorie Générale de Keynes, un modèle formel s'appuyant sur le relatif consensus autour du schéma IS-LM et des techniques économétriques pour estimer les paramètres de ce modèle. C'est la nature et le degré d'interdépendance entre les trois éléments ci-dessus qui sont remis en cause par les monétaristes et les tenants de la modélisation VAR. Alors que les keynésiens établissent une nette distinction entre le modèle théorique et le modèle estimé, pour les monétaristes cette distinction n'est pas claire et ne leur semble pas pertinente.

Sims (1980) reproche aux modèles structurels de la Commission Cowles de comporter trop d'hypothèses théoriques non testées empiriquement. Il propose de soumettre les hypothèses d'exogénéité à des tests économétriques directs et précis. Toutefois, l'indétermination empirique de la causalité dans un modèle VAR, liée au problème de l'équivalence observationnelle (Basman, 1965), impose l'adoption d'un schéma d'identification sur la base d'*a priori* théorique pour identifier les chocs de politique monétaire. Ceci constitue un cas extrême du problème de la sous-détermination de la théorie par les données soulevé par la thèse de Duhem-Quine (Duhem, 1906, Quine, 1951). De plus, Hoover (2009) note que l'analyse des réponses impulsionnelles dans un VAR fournit un bon exemple de ce que Cartwright (2007) qualifie de "contrefactuel imposteur".

Le développement des Modèles à Correction d'Erreurs et des modèles VAR cointégrés a permis de renouveler l'analyse des propositions monétaristes. Toutefois, les liens entre les propositions de cointégration, les notions d'équilibre de long terme et de déséquilibre de court terme sont rarement interprétés dans le cadre d'un modèle théorique rigoureux et complètement spécifié. Pour Faust et Leeper (1994), l'identification d'un modèle par l'imposition de contraintes peut s'avérer non fructueuse lorsque la théorie économique n'établit pas de distinction claire entre les dynamiques de court et de long terme. Faust et Whiteman (1997) relèvent l'absence d'un critère d'arbitrage dans ces démarches en présence de conflit entre le principe théorique et l'ajustement aux données, sinon une subordination de la théorie à l'économétrie.

Parallèlement au problème de l'identification, la critique de Lucas (1976) constitue la seconde critique fondamentale à laquelle se heurtent les modèles économétriques. Lucas (1980, 1986) adopte une nouvelle posture épistémologique en considérant le modèle théorique comme une "fiction" et non plus comme un ensemble de propositions sur le comportement d'une économie réelle. Il défend l'idée d'une explication du cycle en termes de discipline de l'équilibre (Lucas, 1977). Les modèles DSGE, qui constituent les modèles de base de la Nouvelle Synthèse, sont fortement influencés par la méthodologie lucasienne et s'inscrivent dans la continuité des modèles RBC (Taouil, 2011). Benati et Surico (2009) ont établi la supériorité des DSGE par rapport aux VAR structurels (SVAR). Cet échec des SVAR est la conséquence directe des restrictions inter-équations imposées par l'hypothèse des anticipations rationnelles, tel que cela a été initialement soulevé par la critique de Sargent (1979).

Mots-clés : Politique monétaire, Modèles VAR, Modèles DSGE, Causalité, Identification, Critique de Lucas.

Abstract

The Monetary Policy in Econometric Models: Primacy of the Theory over the Empirics

The purpose of this thesis is to show the primacy of the theory over the empirics and prove that econometrics cannot be decisive to question the theory. For this, we rely on the limits of econometrics highlighted in discussions of monetary policy since the 1960s. We adopt an approach based on epistemological arguments to show that these debates go beyond the cleavage theory/empirics and that they integrate a difference of vision as to the usefulness of an empirical model.

The research program of the Cowles Commission was formed around a particular articulation of three fundamental elements: a theoretical repository of Keynes' General Theory, a formal model based on the relative consensus on the IS-LM diagram and econometric techniques to estimate the parameters of this model. It is the nature and the degree of interdependence between these three elements that are contested by the monetarists and supporters of the VAR modeling. While Keynesians make a clear distinction between the theoretical model and the estimated model, this distinction is not clear and does not seem relevant to the monetarists.

Sims (1980) criticizes the structural models of the Cowles Commission for including too many theoretical hypotheses empirically untested. He proposes to review the exogeneity assumptions through direct and specific econometric tests. However, the empirical indeterminacy of causality in a VAR, linked to the problem of observational equivalence (Basman, 1965), requires the adoption of an identification scheme on the basis of a theoretical *a priori* to identify the monetary policy shocks. This is an extreme case of the problem of under-determination of theory by data raised by the Duhem-Quine thesis (Duhem 1906, Quine, 1951). Furthermore, Hoover (2009) notes that the impulse response analysis in a VAR provides a good example of what Cartwright (2007) calls “counterfactual impostor”.

The development of the Error Correction Models and cointegrated VAR models has renewed the analysis of monetarist proposals. However, the links between the proposals for cointegration, the notions of long-term equilibrium and short term disequilibrium are rarely interpreted in the context of a rigorous and fully specified theoretical model. According to Faust and Leeper (1994), the identification of a model by imposing constraints may not be fruitful when economic theory does not clearly distinguish the short-term and long-term dynamics. Faust and Whiteman (1997) note the absence of an arbitration criterion in these approaches apparent in the presence of conflict between the theoretical principle and the adjustment to the data; otherwise subordination of the theory to the econometrics.

Alongside the issue of identification, the Lucas critique (1976) is the second fundamental criticism facing the use of econometric models. Lucas (1980, 1986) adopts a new epistemological posture considering the theoretical model as a 'fiction' and not as a set of proposals on the behavior of a real economy. He supports the idea of explaining the cycle in terms of discipline of equilibrium (Lucas, 1977). The DSGE models, that constitute the fundamental models of the New Synthesis theory, are strongly influenced by Lucas' methodology and are a continuity of the RBC models (Taouil, 2011). Benati and Surico (2009) demonstrated the superiority of a DSGE model with respect to a structural VAR (SVAR). This failure is a direct consequence of inter-equation restrictions imposed by the rational expectations hypothesis, initially raised by Sargent's critics (1979).

Keywords : Monetary policy, VAR models, DSGE models, Causality, Identification, Lucas critique.

REMERCIEMENTS

J'exprime, tout d'abord, ma profonde gratitude et mon entière reconnaissance à M. Rédouane TAOUIL, Professeur à l'Université Pierre Mendès-France, pour sa grande disponibilité, les conseils utiles et les encouragements précieux qu'il m'a prodigués tout au long de son encadrement de cette thèse.

Je présente, ensuite, mes vifs remerciements à Mme Jézabel COUPPEY-SOUBEYRAN, Maître de Conférences et Habilitée à Diriger les Recherches à Université Paris 1 Panthéon-Sorbonne et à M. Jean Pierre ALLEGRET, Professeur à l'Université Paris Ouest Nanterre, pour l'intérêt qu'ils ont porté à cette thèse en acceptant d'en être les rapporteurs. Je remercie également M. Guéliffo HOUNTONDJI, Professeur à l'Université Pierre Mendès-France qui a accepté de participer au jury de cette thèse.

Je tiens, enfin, à remercier tous les membres de ma famille pour leur soutien et leur patience ainsi que tous mes amis qui m'ont de près ou de loin aidé pour la réalisation de ce travail.

Que toutes ces personnes trouvent ici toute mon estime et toute ma considération.

A ma mère,
A mon fils, Abdoullah.

SOMMAIRE

| | |
|---|----|
| <i>Résumé</i> | 6 |
| <i>Abstract</i> | 8 |
| REMERCIEMENTS | 10 |
| SOMMAIRE | 14 |
| INTRODUCTION GENERALE | 18 |
| | |
| PREMIERE PARTIE : LA MODELISATION STRUCTURELLE REMISE EN CAUSE PAR LES MONETARISTES ET LES MODELES VAR | 34 |
| | |
| CHAPITRE 1 : LA CRITIQUE MONETARISTE DE LA POLITIQUE MONETAIRE KEYNESIENNE | 38 |
| | |
| INTRODUCTION | 40 |
| | |
| SECTION 1 : IS-LM COMME CADRE DE LA MODELISATION STRUCTURELLE | 42 |
| Sous-section 1 : Modèles macroéconométriques et souci de validation empirique | 42 |
| Sous-section 2 : L'attaque monétariste des fondements keynésiens de la synthèse | 50 |
| | |
| SECTION 2 : CANAL DU TAUX D'INTERET VERSUS CANAL DES PRIX | 56 |
| Sous-section 1 : La transmission de la politique monétaire | 56 |
| Sous-section 2: La critique monétariste | 62 |
| | |
| CONCLUSION | 70 |

| | |
|---|------------|
| CHAPITRE 2 : LES LIMITES THEORIQUES DE LA POLITIQUE MONETAIRE MONETARISTE..... | 72 |
| INTRODUCTION..... | 74 |
| SECTION 1 : DEMANDE DE MONNAIE ET COMPOSITION DU PORTEFEUILLE | 76 |
| Sous-section 1 : Demande de monnaie monétariste et préférence pour la liquidité | 76 |
| Sous-section 2 : Un modèle dynamique du revenu | 84 |
| SECTION 2 : DECALAGES TEMPORELS ET CAUSALITE MONNAIE-REVENU : L'ARGUMENT DE LA CAUSALITE INVERSEE | 93 |
| Sous-section 1 : Théorie du revenu permanent et stabilité de la demande de monnaie ... | 93 |
| Sous-section 2 : Limites théoriques de la causalité monnaie-revenu..... | 99 |
| SECTION 3 : LA RELECTURE MONETARISTE DE LA COURBE DE PHILLIPS | 108 |
| Sous-section 1 : De la rigidité du salaire à la courbe de Phillips dans IS-LM..... | 108 |
| Sous-section 2 : Objectif de stabilité des prix et règle monétariste | 116 |
| CONCLUSION..... | 122 |
| | |
| CHAPITRE 3 : L'EVALUATION DE LA POLITIQUE MONETAIRE PAR LES MODELES VAR : L'EMPIRIE FACE A L'IMPERATIF THEORIQUE | 124 |
| INTRODUCTION..... | 126 |
| SECTION 1. MODELES VAR ET CAUSALITE MONNAIE-REVENU..... | 128 |
| Sous-section 1. G-causalité et S-causalité..... | 128 |
| Sous-section 2. Critique de Sims et ses implications | 136 |
| SECTION 2. LES VAR STRUCTURELS ET LE RETOUR OBLIGE A LA THEORIE | 147 |
| Sous-section 1. Causalité et exogénéité : questions de méthode..... | 147 |
| Sous-section 2. L'indétermination empirique de la causalité | 155 |

| | |
|--|------------|
| CONCLUSION..... | 165 |
| | |
| DEUXIEME PARTIE : DU RETOUR FORCE DE LA THEORIE A SA PRIMAUTE | 166 |
| | |
| CHAPITRE 4 : TENTATIVES ECONOMETRIQUES DE CONCILIATION ENTRE THEORIE ET EMPIRIE | 170 |
| | |
| INTRODUCTION..... | 172 |
| | |
| SECTION 1 : LE CRITERE PREDICTIF DES MONETARISTES | 175 |
| Sous-section 1 : Le statut des paramètres du modèle..... | 175 |
| Sous-section 2 : Propositions monétaristes et limites des tests empiriques | 181 |
| | |
| SECTION 2. LA COINTEGRATION ET L'ARTICULATION ENTRE COURT ET LONG TERME | 187 |
| Sous-section 1 : La dynamique des variables monétaires | 187 |
| Sous-section 2 : Validation empirique et exigence théorique | 195 |
| | |
| CONCLUSION..... | 201 |
| | |
| CHAPITRE 5 : LA CRITIQUE DE LUCAS ET SES IMPLICATIONS..... | 202 |
| | |
| INTRODUCTION..... | 204 |
| | |
| SECTION 1. LA CRITIQUE DE LUCAS : UNE INNOVATION CARDINALE..... | 206 |
| Sous-section 1. La discipline de l'équilibre | 206 |
| Sous-section 2. Une nouvelle posture épistémologique..... | 211 |

| | |
|---|------------|
| SECTION 2. LA COHERENCE TEMPORELLE DES POLITIQUES ECONOMIQUES : UN RENOUVEAU DE L'ECONOMIE POLITIQUE | 214 |
| Sous-section 1. L'incohérence temporelle de la politique discrétionnaire..... | 214 |
| Sous-section 2. La subordination de la politique budgétaire à la politique monétaire... | 219 |
| CONCLUSION..... | 223 |
| | |
| CHAPITRE 6 : LES MODELES DSGE ET L'EXTENSION DE LA METHODOLOGIE DES CYCLES REELS | 224 |
| | |
| INTRODUCTION..... | 226 |
| | |
| SECTION 1. LES FONDEMENTS DES MODELES DSGE | 228 |
| Sous-section 1. L'influence de la méthodologie Lucasienne..... | 228 |
| Sous-section 2. Demande agrégée et courbe de Phillips..... | 232 |
| | |
| SECTION 2. EVALUATION DE LA POLITIQUE MONETAIRE : SUPREMATIE DES DSGE SUR LES VAR..... | 238 |
| Sous-section 1. Règle de politique monétaire et objectif de stabilité des prix..... | 238 |
| Sous-section 2. La Grande modération : l'échec explicatif des VAR..... | 241 |
| | |
| CONCLUSION..... | 247 |
| | |
| CONCLUSION GENERALE | 248 |
| | |
| REFERENCES BIBLIOGRAPHIQUES..... | 252 |
| | |
| TABLE DES MATIERES | 272 |

Introduction générale

La place de l'économétrie dans la conception et l'évaluation de la politique monétaire n'a cessé de prendre de l'importance depuis le lancement de la controverse entre les monétaristes et les keynésiens au milieu des années 1960. A cette époque, l'économétrie connaît déjà un important essor grâce aux travaux de la Commission Cowles sur le problème de l'identification de la forme structurelle d'un modèle à équations simultanées. Il s'agit de savoir comment retrouver les paramètres de la forme structurelle à partir de l'estimation de ceux de la forme réduite. La solution apportée par les chercheurs de la Commission Cowles à ce problème consiste en l'adjonction d'un nombre suffisant de restrictions, qui doivent être formulées par la théorie économique, sur les paramètres de la forme structurelle du modèle.

L'un des rôles principaux de l'économétrie pour la Commission Cowles est de soumettre les hypothèses formulées *a priori* par la théorie économique à des tests statistiques développés à cet effet. Evaluée à l'aune de ce critère, l'étude empirique menée par Friedman et Meiselman (1963) pour démontrer l'efficacité supérieure de la politique monétaire par rapport à la politique budgétaire est jugée économétriquement hautement déficiente (Ando et Modigliani, 1965; De Prano et Mayer, 1965; Hester, 1964). En effet, les équations estimées par Friedman et Meiselman présentent de mauvaises spécifications et comportent des restrictions sur les paramètres non testées statistiquement.

Dans le même sens, le modèle économétrique dynamique de la Réserve Fédérale de St. Louis, développé par Andersen et Jordan (1968), cherche également à mettre en évidence l'efficacité supérieure de la politique monétaire dans le cadre du débat sur la présence d'un effet d'éviction exercé par les dépenses publiques sur l'investissement privé. L'estimation et l'interprétation des résultats de ce modèle font l'objet d'intenses critiques, car il s'agit d'une forme réduite où l'hypothèse d'exogénéité des dépenses publiques n'est soumise à aucun test statistique de causalité.

L'article de Friedman et Meiselman (1963) s'appuie sur l'existence d'une relation directe entre le revenu nominal et la masse monétaire. Or, avant la publication de sa nouvelle lecture de la courbe de Phillips (Friedman, 1968), Friedman n'a pas réussi à développer une théorie complète pour justifier l'influence causale de la monnaie sur le revenu nominal et la décomposition de ce dernier entre production et prix (Desai, 1981). La mise en évidence de

ces propositions, centrales pour le monétarisme, s'appuie exclusivement sur les études statistiques menées au sein du *National Bureau of Economic Research (NBER)* (Friedman, 1958, Friedman et Schwartz, 1963 a). En particulier, Friedman (1958) s'appuie sur l'observation de corrélations et de décalages temporels entre la masse monétaire et le revenu nominal pour justifier l'influence causale de celle-là sur celui-ci et recommander une règle de politique monétaire qui consiste à adopter un taux de croissance fixe de l'offre de monnaie.

C'est dans le contexte du débat entre les monétaristes et les keynésiens sur l'exogénéité et la causalité de la monnaie sur le revenu que Sims (1972) propose de soumettre les hypothèses d'exogénéité à des tests économétriques directs et précis s'appuyant sur le test de causalité de Granger. L'intérêt du concept de G-causalité ou de S-causalité est, selon Sims, d'offrir la possibilité de tester aisément et directement la présence d'une causalité unidirectionnelle, qu'il identifie avec une hypothèse de stricte exogénéité. L'introduction de la modélisation VAR (Vectoriel AutoRegressif) dans sa célèbre critique a permis à Sims (1980, a, b) d'étendre l'analyse de la causalité au sens de Granger à un vecteur comportant en plus de la monnaie et le revenu, le prix et le taux d'intérêt (Paulré, 2007). Dans cette critique, Sims (1980 a) reproche aux modèles structurels de la Commission Cowles de comporter trop d'hypothèses théoriques non testées empiriquement et présente les modèles VAR comme une alternative à ces modèles.

Le développement des Modèles à Correction d'Erreurs dans la tradition de la *London School of Economics (LES)* (Hendry, 1995) et des modèles VAR cointégrés (Engle et Granger, 1987) a permis de renouveler l'analyse des propositions monétaristes en mettant en évidence les insuffisances des techniques économétriques traditionnelles (Granger et Newbold, 1974; Nelsson et Plosser, 1982). Conçus dans le cadre de l'économétrie des séries non stationnaires, ces modèles ambitionnent d'interpréter les fonctions de comportement comme les fluctuations d'une dynamique de court terme, purement économétrique, autour d'un équilibre de long terme, dicté par la théorie économique. Une large partie de la littérature sur ces modèles s'est intéressée à la proposition centrale des monétaristes relative à la stabilité de la fonction de demande de monnaie, utilisée pour établir une relation directe entre la monnaie et le revenu, et à la question connexe d'exogénéité de la monnaie (Hendry et Mizon, 1978; Hendry et Ericsson, 1991).

Or, la stabilité de la fonction de demande de monnaie exige que ses paramètres soient constants et qu'elle vérifie la même qualité d'ajustement sur les différentes sous-périodes. Hendry et Ericsson (1991) démontrent que l'équation de demande de monnaie construite par Friedman et Schwartz (1982) pour le Royaume-Uni ne vérifie aucune de ces hypothèses. Il s'est, en effet, avéré que la demande de monnaie a connu de larges déviations, au début des années 1970 et 1980, perçues comme le signe d'une instabilité dans cette relation (Laidler, 1981 ; Hafer et Jansen, 1991). Selon Hendry et Ericsson (1990), la constance des paramètres, qui a été remise en cause par cette instabilité de la masse monétaire, est importante pour l'inférence, la prévision et l'évaluation de la politique monétaire.

En dépit de son rôle central pour la politique monétaire, l'économétrie présente plusieurs limites et fait face à de nombreuses critiques, dont on retient les deux plus fondamentales, à savoir le problème de l'identification et la critique de Lucas. Hoover (2009) considère que, depuis l'équivalence établie par Simon (1953) entre la causalité et les conditions d'identification de la forme structurelle, le problème central de l'économétrie a été celui de l'identification. Cette équivalence a été à l'origine du glissement progressif du débat au sein de la Commission Cowles sur la question de la causalité vers le problème de l'identification (Hoover, 2001).

Zellner (1979) relève l'importance du concept de causalité en Economie et en Econométrie et note que ce concept ne peut être associé à une relation temporelle ou réduit à la minimisation de la variance de prévision d'un ensemble d'informations comme le stipule la G-causalité. En effet, Pierce (1977) précise que le test de Granger s'articule autour d'une spécification empirique entre des séries chronologiques et que les données sont utilisées pour révéler l'existence ou l'absence de certaines formes d'interdépendances. Les relations examinées ne sont pas spécifiées conformément à une théorie précise, ce qui risque de conduire à retenir des formes incorrectes des variables ou d'exclure des variables pertinentes (Zellner, 1979). D'ailleurs, Granger (1979) admet la possibilité de l'inférence d'une causalité fallacieuse ou fictive dans le cas de l'omission de variables importantes. A cet effet, Sims (1972) reconnaît que la méthode du test de Granger constitue bien une version sophistiquée du principe du "*post hoc ergo propter hoc*" qui fût rejeté par Tobin (1970) dans sa célèbre critique contre Milton Friedman.

En effet, les vifs débats entre Tobin (1970) et Friedman (1970) ont porté sur le mode d'articulation entre "théorie" et "économétrie" pour la justification de la causalité entre la monnaie et le revenu. Tobin rejette la justification de la causalité par l'antécédence temporelle et développe un modèle théorique keynésien qui génère les décalages observés par Friedman, mais où la monnaie ne joue aucun rôle causal sur l'activité économique. La position de Tobin s'inscrit dans la tradition de modélisation de la Commission Cowles, selon laquelle la relation causale est un concept théorique et non pas empirique.

Le programme de recherche de cette Commission s'est constitué autour d'une articulation particulière de trois éléments fondamentaux. Un référentiel théorique issu de la Théorie Générale de Keynes, un modèle formel s'appuyant sur le relatif consensus autour du schéma IS-LM et des techniques économétriques pour estimer les paramètres de ce modèle. En conséquence, les keynésiens préfèrent l'estimation des paramètres d'un modèle dans le cadre d'équations structurelles simultanées et insistent sur l'importance de l'*a priori* théorique pour interpréter les corrélations observées entre les données (Desai, 1981). Ils conçoivent l'exogénéité comme une forme de restriction imposée sur les paramètres, nécessaire à l'identification de la forme structurelle du modèle.

C'est la nature et le degré d'interdépendance entre les trois éléments ci-dessus qui sont remis en cause par les monétaristes et les tenants de la modélisation VAR. En particulier, la controverse durant les années 1960 et 1970 sur les méthodes d'estimation directes d'une forme réduite ou à travers un modèle structurel illustre parfaitement l'important désaccord entre les keynésiens et les monétaristes sur la méthode adéquate pour l'estimation des effets de la politique monétaire sur l'activité économique. Alors que les keynésiens insistent sur la cohérence théorique et la portée explicative du modèle théorique, pour les monétaristes c'est la capacité prédictive du modèle estimé à partir des données observées qui compte.

A cet effet, les keynésiens établissent une nette distinction entre le modèle théorique et le modèle estimé à l'aide de techniques économétriques à partir des données observées. En revanche, pour les monétaristes cette distinction n'est pas claire et ne leur semble pas pertinente. La principale faiblesse de la démarche des monétaristes porte sur l'impossibilité de définir le modèle structurel à partir duquel leur forme réduite est dérivée. Or, c'est le modèle structurel qui permet de préciser les restrictions sur les coefficients des variables indépendantes à travers les identités, les retards, les variables omises, ... etc. En l'absence de

ce modèle, il n'est pas possible de comprendre la nature des relations causales entre les variables endogènes et exogènes.

Déjà à la fin des années quarante, la démarche de modélisation de la Cowles Commission suscite une vive réaction de la part des chercheurs du *National Bureau of Economic Research* (*NBER*) dans le cadre de la controverse "mesure sans théorie" qui a opposé Koopmans (1947) à Vining (1949). Pour ce dernier, la complexité des modèles de la Commission Cowles cache les faiblesses de la théorie économique sous-jacente qu'il qualifie d'incomplète et de confuse. En revanche, Koopmans pense que l'absence de fondement théorique et de l'inférence statistique des travaux des économistes du *NBER* les empêche de comprendre les phénomènes économiques.

La G-causalité, ainsi que les modèles VAR qui sont des formes réduites, n'intègrent pas la vision de la Commission Cowles sur la simultanéité où il existe une rétroaction instantanée entre deux variables endogènes (Desai, 1981), car la condition d'asymétrie temporelle entre la cause et l'effet qui est à la base de la définition de la G-causalité ne l'autorise pas. Pour Cooley et LeRoy (1985), l'assimilation des corrélations conditionnelles avec un ordre causal dans le cas de l'approche VAR ne peut se justifier que sous une hypothèse de prédétermination qui est non testable en l'absence de restrictions *a priori* dérivées de la théorie.

Pour pouvoir interpréter les innovations canoniques comme des chocs exogènes de politique monétaire, Sims transforme le modèle VAR en un VAR structurel (VARs). Paradoxalement, pour justifier les restrictions d'identification imposées aux innovations contemporaines, Sims se réfère à la notion de chaîne causale de Wold (1954). Ainsi, comme le relève Pagan ((1987), la volonté de la méthode des VAR de s'affranchir de l'*a priori* théorique s'est révélée illusoire. Par ailleurs, la sensibilité des résultats des VAR utilisés par Sims aux choix des spécifications retenus a été mise en évidence par plusieurs auteurs (Eichenbaum et Singleton, 1986; Runkle, 1987; Spencer, 1989) et de vives critiques ont porté sur la fragilité et le manque de robustesse des résultats des tests de G-causalité et de S-causalité (Feige et Pearce, 1979).

On retrouve également le problème de l'identification et de l'interprétation causale des relations économétriques, qui exigent le recours à des *a priori* théoriques, dans les modèles

dynamiques qui cherchent à concilier entre la dynamique de court terme et l'équilibre de long terme. Il s'agit, plus précisément, des travaux sur les Modèles à Correction d'Erreurs (MCE) à la Hendry, des modèles VAR cointégrés dans la tradition de Engle et Granger (1987) et des modèles VAR structurels à la Blanchard et Quah (1989). Par exemple, les liens entre les propositions de cointégration, les notions d'équilibre de long terme et de déséquilibre de court terme sont rarement interprétés dans le cadre d'un modèle théorique rigoureux et complètement spécifié.

De fait, l'interprétation des modèles de cointégration en termes de comportement économique devient souvent difficile et elle se focalise le plus souvent sur les relations de long terme et néglige les fluctuations de court terme. Or, Cooley et Hansen (1993,) ont développé un modèle monétaire qui montre que les fluctuations de court terme renseignent mieux sur le comportement de l'économie que les relations de cointégration. En particulier, ce modèle ne permet pas d'interpréter la relation de long terme entre le revenu, la monnaie, le prix et le taux d'intérêt comme une fonction de demande de monnaie.

Pour Faust et Leeper (1994), l'identification d'un modèle par l'imposition de contraintes peut s'avérer non fructueuse lorsque la théorie économique ne fournit pas suffisamment de restrictions sur les variables d'intérêt ou n'établit pas de distinction claire entre les dynamiques de court terme et de long terme. De plus, les contraintes imposées pour l'identification et l'interprétation des relations de long terme détériorent en général les qualités d'ajustement de la dynamique de court terme. Dans ce sens, Faust et Whiteman (1997) relèvent l'absence d'un critère d'arbitrage dans la démarche de la *London School of Economics (LES)* en présence de conflit entre le principe théorique et l'ajustement aux données, sinon une subordination de la théorie à l'économétrie.

Monfort et Rabemanajara (1990) et Hendry et Mizon (1993) ont montré que l'hypothèse d'exogénéité faible, introduite par Engle, Hendry et Richard (1983), ne permet pas de résoudre le problème de l'identification des modèles VAR. Dans le cas d'un modèle VAR cointégré bivarié, l'hypothèse d'exogénéité faible est testable uniquement en présence de contrainte identifiante sur le vecteur de cointégration. En revanche, en dimension supérieure, l'hypothèse d'exogénéité faible ne peut plus être directement vérifiée et devrait être postulée *a priori*.

La critique de Lucas (1976) constitue la seconde critique fondamentale à laquelle fait face l'utilisation des modèles économétriques pour l'étude de la politique monétaire. Initialement, Lucas s'attaque aux modèles macroéconométriques qui s'inscrivent dans la tradition de la Commission Cowles et remet en cause l'usage de ces modèles pour l'évaluation des politiques économiques. Les modèles VAR sont également soumis à la critique de Lucas (1976) relative à l'invariance des paramètres des méthodes économétriques. En particulier, ces modèles sont incapables de rendre compte des ruptures structurelles dans les comportements des agents induites par la modification de la politique monétaire.

La vulnérabilité de la méthode des VAR à la critique de Lucas a été mise en évidence par Benati et Surico (2009) dans le cadre du débat sur les causes de la "Grande Modération" aux Etats-Unis. Ces auteurs estiment que l'utilisation des VAR structurels (SVAR) ne permet pas toujours de bien identifier un changement de règle de politique monétaire. Cet échec des modèles SVAR n'est pas lié à des problèmes d'estimation comme le soutient Canova (2006), mais il est la conséquence directe des restrictions inter-équations imposées par l'hypothèse des anticipations rationnelles tel que cela a été initialement soulevé par la critique de Sargent (1979) à l'encontre de ces modèles.

Le va-et-vient permanent qui existe entre le développement de l'économétrie et les débats théoriques sur la politique monétaire soulève naturellement la question de la nature des rapports qui existent entre les deux disciplines (Pirrotte, 2004). Souvent, l'économétrie prend le pas sur la théorie et le modèle de Klein et Goldberger (1955), qui constitue le premier maillon de la modélisation macroéconométrique, fournit un exemple éclatant où le souci de validation empirique l'emporte sur la rigueur du modèle théorique. Le désaccord entre les monétaristes et les keynésiens incite à penser que la question de l'efficacité relative de la politique monétaire et la politique budgétaire est une question empirique et non théorique. Pourtant, le volume impressionnant des travaux empiriques qui a été produit par chacune des deux parties n'a pas permis de trancher le débat.

En s'appuyant sur les limites de l'économétrie mises en évidence dans les différents débats mentionnés *supra*, cette thèse s'attache à montrer le primat de la théorie sur l'empirie pour l'étude de la politique monétaire et que l'économétrie ne peut pas être décisive dans la remise en cause de la théorie. Nous adoptons une démarche basée sur des arguments

épistémologiques pour montrer que les débats ci-dessus dépassent le clivage théorie/empirie et intègrent une différence de vision quant à l'utilité d'un modèle empirique.

Les monétaristes tentent de déplacer le débat sur le terrain économétrique et les tenants de l'approche des modèles VAR adoptent une position extrême en cherchant à se fier uniquement aux tests économétriques et se passer le plus possible de la théorie économique. Toutefois, ces tentatives n'ont pas été fructueuses et n'ont pas permis de résoudre les divergences en raison de l'impossibilité d'inférer la causalité à partir des données observées.

Pierce (1977) relève que l'analyse statistique est d'une portée limitée pour la détermination de certaines relations économiques, car des propositions incompatibles peuvent toujours être vérifiées par l'utilisation des mêmes données. En l'absence de restrictions sur-identifiantes, les modèles VARS sont tous observationnellement équivalents (Hoover, 2004 ; Hoover et Demiralp, 2003), puisqu'ils ont la même forme réduite qui fait l'objet d'une estimation empirique explicite. L'indétermination empirique de la causalité, liée au problème de l'équivalence observationnelle (Basmann, 1965), impose l'adoption d'un schéma d'identification sur la base d'*a priori* théorique pour identifier les chocs de politique monétaire.

Ceci constitue un cas extrême du problème de la sous-détermination de la théorie par les données soulevé par la thèse de Duhem-Quine (Duhem, 1906, Quine, 1951). Cette thèse est également connue sous le nom de *holisme de la confirmation*, selon lequel l'observation des données ne permet pas d'imposer une seule théorie, mais qu'il existe toujours des théories différentes qui sont toutes compatibles avec ces données. Dans cette perspective, le contrôle empirique soulève une ambiguïté, car une hypothèse ne peut jamais subir l'épreuve empirique isolément, mais le test implique nécessairement des hypothèses auxiliaires.

Zellner (1979) estime que le choix par Sims de retenir des chaînes causales de Wold représente une position *a priori* sur la nature de la causalité qui est non fondée méthodologiquement et qui ne justifie pas l'exclusion des autres formes de causalité. Par exemple, les définitions de la causalité de Basmann et Feigl s'appliquent à l'ensemble des modèles économétriques, qu'ils soient écrits sous une forme récursive, triangulaire ou interdépendante. Comme le rappelle Paulré (1985), Wold a lui-même reconnu l'équivalence

de la forme récursive avec certaines formes non récursives, remettant ainsi en cause la supériorité statistique des chaînes causales.

Plus récemment, Cartwright (2007) note la pluralité des théories de la causalité et rappelle que les théories sont orientées, soit vers la caractérisation des causes, soit vers leur utilisation, mais pas les deux à la fois. Le problème central à la base de cette thèse est l'absence de connexion entre les méthodes de test et de caractérisation des propositions causales et les modes de justification de l'utilisation des structures causales établies. Cette critique s'adresse, particulièrement, à la théorie de la causalité de Woodward (2003) qui défend la causalité en termes de manipulation contrefactuelle et où justement cette connexion n'est pas possible. Dans ces conditions, le décideur de la politique monétaire ne peut pas avoir une confiance dans le contrefactuel que Woodward cherche à évaluer. Pour ces raisons, Cartwright (2007) qualifie ce genre de contrefactuel de "contrefactuel imposteur" et l'analyse des réponses impulsionnelles dans un VAR fournit, selon Hoover (2009), un bon exemple d'un tel type de contrefactuel.

Par ailleurs, malgré l'absence d'une confirmation empirique concluante (Ericsson, Irons, 1995), la critique de Lucas a eu un impact dévastateur sur la recherche en modélisation économétrique s'inscrivant dans la tradition de la Commission Cowles ainsi que sur celle portant sur les modèles VAR. En raison de la préséance de la théorie, cette critique a révolutionné la conception et le rôle d'un modèle et elle a remis en cause les méthodes économétriques traditionnelles (Pirrotte, 2004). Chez les Nouveaux Classiques, il devient essentiel pour les besoins d'identification de distinguer les paramètres de politique économique, ceux des préférences et de la technologie et les autres paramètres du modèle. Sargent (1982) estime que le meilleur moyen d'éviter la critique de Lucas est d'interpréter un modèle comme une fonction-objectif d'un individu représentatif, d'en déduire sa réaction face à un changement de politique économique ou d'autres contraintes et d'utiliser ces résultats pour dériver les comportements agrégés de l'économie.

Lucas s'oppose à la vision traditionnelle de l'équilibre où celui-ci est réalisé uniquement à long terme et défend l'idée d'une explication du comportement du cycle en termes d'équilibre où les deux postulats de comportement optimisateur des agents et de l'apurement des marchés, sont réalisés en chaque point du temps (Lucas et Sargent, 1979). Ces postulats sont qualifiés par Lucas (1977, p. 12) de "*discipline imposée par la théorie de l'équilibre*" et juge qu'un

raisonnement en dehors de l'équilibre ouvre la voie à "l'arbitraire" en l'absence de règle précise et laisse une grande marge pour le modélisateur pour fixer les grandeurs du modèle (De Vroey, 2009).

Le nouveau programme de recherche initié par Lucas et adopté par les nouveaux Classiques s'appuie sur une vision épistémologique constituée de deux éléments essentiels portant sur la nature d'un modèle théorique (De Vroey, 2009). D'un côté, Lucas abandonne la distinction entre la notion de théorie et de modèle, dans le sens où les deux sont fusionnées. Cette position tranche radicalement avec la conception antérieure qui considérait la théorie comme un ensemble de propositions à propos de la "réalité", et que le modèle constitue une tentative pour exposer certaines implications de la théorie. D'un autre côté, Lucas conçoit le modèle théorique comme un "système parallèle" ou une "imitation de l'économie" et non pas un ensemble de propositions sur le comportement d'une économie réelle (Lucas, 1980, 1986).

En dépit de son caractère abstrait et du faible contenu empirique de ses hypothèses, le modèle doit être testé, estime Lucas, par rapport aux données d'observation (Lucas, 1977). Ainsi, le test empirique de la théorie constitue une étape cruciale puisque le modèle s'appuie sur des axiomes abstraits qui sont nécessairement faux (Lucas, 1986). Ce sont Kydland et Prescott (1982) qui ont procédé à la mise en œuvre du programme de Lucas pour la confrontation de la théorie avec les données à travers l'application de la technique du calibrage aux modèles des Cycles Réels (RBC). Cette technique est présentée comme une stratégie alternative aux méthodes économétriques traditionnelles pour le calcul des paramètres structurels de ces modèles.

A cet effet, Faust et Whiteman (1997) relèvent deux questions fondamentales qui opposent profondément les approches des modèles RBC, des modèles VAR et des méthodes du général au particulier dans la tradition de la *London School of Economics (LES)*. La première porte sur les caractéristiques des données à reproduire par le modèle et la seconde concerne l'ampleur des modifications théoriques à introduire dans la spécification empirique pour améliorer la qualité de l'ajustement.

Les modèles DSGE (*Dynamic Stochastic General Equilibrium*), qui constituent les modèles de base de la théorie de la Nouvelle Synthèse, sont fortement influencés par la méthodologie lucasienne et s'inscrivent dans la continuité des modèles RBC (Taouil, 2011). Ils sont une

référence de base pour l'étude des conditions d'exercice de la politique monétaire par les Banques centrales (Woodford, 2003) dans le cadre du Nouveau Consensus Monétaire (Mishkin, 2008). Dans la perspective lucasienne, les principales hypothèses de ce modèle, comme celle des anticipations rationnelles ou la fiction de l'agent représentatif, constituent des énoncés analytiques et des principes de modélisation plutôt que des assertions à propos de la réalité.

Même si elle a bénéficié de tests empiriques positifs, l'hypothèse des anticipations adaptatives retenue par Friedman (1968) a été supplantée par celle des anticipations rationnelles, même si celle-ci a été critiquée en raison de son écart par rapport au comportement observé des agents économiques (Lovell, 1986 ; Keane et Runkle, 1990). De même, en raison de la présence de la théorie, la règle de Taylor (1993) et la courbe de Phillips, initialement conçues comme des catégories descriptives, sont devenues des catégories théoriques centrales des modèles DSGE, en dépit des conclusions contrastées quant à leur portée empirique (Taouil, 2014 b).

A ce titre, la vision de Lucas rejoint celle de Friedman (1953) concernant la récusation du réalisme des hypothèses d'un modèle macroéconomique (Taouil, 2014 a). Les hypothèses les plus fécondes pour la construction théorique sont celles qui ont un faible contenu descriptif. En conséquence, une théorie ne peut être remise en cause en raison de l'absence de pertinence empirique de ses hypothèses.

Les fortes restrictions simplificatrices sur la structure d'un DSGE altèrent naturellement la qualité de l'ajustement du modèle aux données (Vo Phuong et *ali*, 2012). Les paramètres de ce modèle sont déterminés à partir d'une conciliation du calibrage avec les méthodes d'estimation bayésiennes, voire les méthodes d'inférence simulée, telles que le bootstrap et l'inférence statistique indirecte. En effet, se situant dans la continuité de la tradition des Cycles Réels, les tenants des modèles DSGE s'abstiennent de l'utilisation d'une démarche descriptive et formelle de l'économétrie et privilégient le recours aux techniques du calibrage (Pirrotte, 2004).

Dans cette optique, on ne peut pas opposer les modèles VAR aux modèles DSGE car ils ne se situent pas sur le même plan et n'ont pas la même visée. Un VAR ne peut être considéré comme étant supérieur à un DSGE tout simplement parce qu'il offre un meilleur ajustement aux données. En revanche, la critique de Benati et Surico (2009) établit la supériorité d'un

modèle DSGE par rapport à un modèle VAR structurel (SVAR) pour l'évaluation de l'effet d'un changement de règle de politique monétaire. Cet échec des modèles SVAR est la conséquence directe des restrictions inter-équations imposées par l'hypothèse des anticipations rationnelles tel que cela a été initialement soulevé par la critique de Sargent (1979) à l'encontre des modèles VAR.

Le plan de cette thèse s'articule autour de deux parties, chacune est déclinée en trois chapitres. La première partie étudie la remise en cause de l'analyse de la politique monétaire dans les modèles de la Commission Cowles par les monétaristes et les tenants de l'approche VAR et relève les limites du recours exclusif à l'économétrie pour l'évaluation de la politique monétaire. Après avoir mis en évidence les difficultés des tentatives économétriques de conciliation entre "théorie" et "empirie", la seconde partie s'attache à montrer le primat de la première sur la seconde. On montre à cet effet que, s'inscrivant dans la continuité du programme de recherche des Nouveaux Classiques, les modèles DSGE de la Nouvelle Synthèse permettent de mieux identifier les changements de règles de politique monétaire que les modèles VAR structurels.

Nous présentons dans le premier chapitre la critique monétariste de la politique monétaire keynésienne. La première section de ce chapitre expose les écarts entre les cas keynésien et classique des versions de Hicks (1937) et de Modigliani (1944) du modèle IS-LM. L'écart important qui existe entre ces cadres théoriques et les modèles estimés, dans la continuité de celui de Klein et Goldberger (1955), favorise la position de Friedman qui cherche à minimiser l'importance de l'écart théorique entre les keynésiens et les monétaristes et de déplacer le débat sur le terrain empirique. De plus, on précise que les particularités qui caractérisent les cas classique et keynésien sont au cœur du débat entre les monétaristes et les keynésiens. En particulier le principe de la demande effective, l'autonomie de l'épargne par rapport à l'investissement et l'origine monétaire de la formation du taux d'intérêt font l'objet d'une critique virulente de la part des monétaristes.

Dans la seconde section, on relève avec Mayer (1975) que la présence d'un désaccord à propos de la gamme d'actifs n'est pas décisive pour expliquer la divergence sur le canal de transmission de la politique monétaire entre les keynésiens et les monétaristes. Comme le note Tobin (1972), c'est plutôt l'interprétation différente de la formation du taux d'intérêt qui est à la base de ce désaccord. On examine, ensuite, la critique de l'hypothèse de la trappe à

liquidité qui, selon Friedman (1971), est à l'origine de la divergence entre les adeptes et les opposants à la théorie quantitative de la monnaie. On analyse également la proposition monétariste stipulant que le rôle crucial du taux d'intérêt dans le système keynésien est responsable de l'apparition de l'effet d'éviction.

Le second chapitre identifie les propositions centrales du monétarisme et étudie les critiques keynésiennes qu'elles ont suscitées. La section 1 présente la fonction de demande de monnaie friedmanienne qui est interprétée comme un cas particulier de la théorie du capital. La volonté de Friedman de rattacher cette demande de monnaie à la tradition quantitativiste est fortement remise en cause par les auteurs de la synthèse qui la conçoivent plutôt comme une reformulation de la théorie keynésienne de la préférence pour la liquidité (Patinkin, 1969). A partir de cette fonction, Friedman (1970) élabore une relation reliant directement le revenu nominal à la masse monétaire, moyennant l'utilisation de la relation de Fisher pour expliquer la formation du taux d'intérêt nominal.

La section 2 est consacrée à la présentation du débat entre Friedman et Tobin quant à la pertinence de l'utilisation des décalages temporels entre les cycles du revenu et de la monnaie pour justifier l'influence causale de celle-ci sur celui-là. Cette causalité empirique est la base, jusqu'en 1968, de la justification de la politique monétaire monétariste d'une règle de croissance constante de la masse monétaire. De même, on montre l'échec des monétaristes à formuler un modèle théorique complet capable d'expliquer les dynamiques observées de toutes les variables nominales.

La section 3 revient sur la critique monétariste de l'introduction de la courbe de Phillips empirique dans le modèle IS-LM et sur le rôle de la fixité du niveau général des prix dans ce modèle. Friedman (1968) défend une nouvelle lecture de la courbe de Phillips qui lui permet d'expliquer théoriquement l'influence à court terme de la monnaie sur le revenu réel et de restaurer la dichotomie classique à long terme.

Le chapitre 3 vise à montrer les limites économétriques et les problèmes méthodologiques auxquels fait face l'approche des modèles VAR pour l'analyse de la politique monétaire. La section 1 rappelle que, suite à la critique de Sims (1980 a), l'objectif de cette démarche est de s'appuyer exclusivement sur l'économétrie et de se passer de la théorie pour l'analyse de la politique monétaire. Ensuite, on discute les limites économétriques des concepts de G-

causalité et de S-causalité utilisés par Sims (1972) pour confirmer la proposition monétariste d'une influence causale de la monnaie sur le revenu et pour vérifier empiriquement l'hypothèse d'exogénéité de la monnaie. Puis, on expose les critiques économétriques concernant la robustesse et la sensibilité des résultats des modèles VAR aux choix de spécifications retenus.

La seconde section identifie les critiques d'ordre méthodologique qui portent sur l'incapacité de l'approche des VAR à fournir des résultats qui puissent être interprétés comme étant structurels. On relève, d'abord, les limites du caractère a-théorique du concept de G-causalité et on traite également les liens entre les concepts de causalité, exogénéité et identification. Ensuite, on note la situation paradoxale dans laquelle se retrouvent les modèles VAR. Ceux-ci sont contraints de spécifier une hypothèse causale, en faisant référence aux chaînes causales de Wold, qui *in fine* conditionne la nature des effets attendus d'une politique monétaire. Puis, on remarque avec Hoover (2009) que les modèles VAR sont soumis à la critique générale de Cartwright et qu'ils peuvent être qualifiés de "contrefactuels imposteurs". Enfin, on aborde le problème de l'équivalence observationnelle des modèles VARS lié à la sous-détermination de la théorie par les observations et à la thèse de Duheim-Quine.

Après avoir montré dans le chapitre 3 les limites du recours exclusif à l'économétrie pour l'évaluation de la politique monétaire, le chapitre 4 vise à mettre en évidence les difficultés des tentatives économétriques de conciliation entre théorie et empirie. La première section de ce chapitre revient sur le débat économétrique des années 1960 entre les monétaristes et les keynésiens. Elle montre que la plupart des propositions centrales du monétarisme, identifiées dans les chapitres 1 et 2, s'appuie surtout sur des considérations empiriques en partant de la fonction théorique de la demande de monnaie de Friedman (1956). Toutefois, l'utilisation de formes réduites, du coefficient de corrélation et de restrictions non testées sur les paramètres pour établir ces propositions est fortement critiquée par les keynésiens. De plus, les équations des monétaristes souffrent d'importants problèmes de spécification et sont sensibles au choix des variables monétaires.

La seconde section discute les résultats de la théorie de la cointégration et des modèles à Correction d'Erreur qui cherchent à concilier entre la dynamique de court terme, purement économétrique, et l'équilibre de long terme, dicté par la théorie économique. Toutefois, ces nouvelles techniques souffrent de limites méthodologiques liées à l'absence de cadre

théorique rigoureux et complètement spécifié décrivant l'articulation du court terme avec le long terme de la politique monétaire. Ceci favorise la subordination de la théorie à l'économétrie en l'absence d'un critère d'arbitrage entre le principe théorique et l'ajustement aux données. De plus, ces méthodes font face, à l'instar des modèles VAR, à des problèmes d'identification et à des difficultés pour tester directement l'hypothèse d'exogénéité sans *a priori* théorique (Faust et Leeper, 1994).

Le chapitre 5 examine les apports de la révolution des Nouveaux Classique pour l'analyse de la politique monétaire et pour l'articulation de la "théorie" avec "l'empirie". La première section explique la portée dévastatrice de la critique de Lucas (1976) pour la modélisation économétrique, qu'il s'agit de celle de la Commission Cowles ou celle dans la tradition des modèles VAR et de la cointégration. Suite à cette critique, Lucas (1977) explique le comportement du cycle économique dans le cadre de la discipline de l'équilibre qui s'articule autour des postulats de l'optimisation intertemporelle et de l'apurement des marchés. La nouvelle posture épistémologique adoptée par Lucas (Lucas, 1980, 1986) lui permet de rompre définitivement avec les visions précédentes de modélisation et l'amène à concevoir le modèle comme une fiction et non une description de l'économie. Dans cette perspective, le test du modèle se fait par la méthode du calibrage (Kydland et Prescott, 1982).

On montre dans la seconde section que ce programme de recherche conduit à un renouveau de la conception de la politique monétaire. Inscrivant l'équilibre économique dans le cadre d'un jeu stratégique entre décideur public et agents privés, Kydland et Prescott (1977) déplacent le débat sur l'efficacité de la politique monétaire vers celui des institutions en charge de celle-ci. L'argument de la crédibilité est avancé pour conférer la responsabilité de la politique monétaire à une Banque centrale indépendante avec un Gouverneur conservateur (Barro et Gordon, 1983 ; Rogoff, 1985). Sargent et Wallace (1981) remettent en cause la pertinence de la proposition monétariste pour la conduite du *policy-mix* et recommandent l'adoption d'un régime de politique monétaire dominante où la Banque centrale se comporte en meneur du jeu et impose une discipline à la politique budgétaire.

Le dernier chapitre vise à montrer la supériorité des modèles DSGE sur les modèles VAR pour l'évaluation du changement de la règle de taux d'intérêt. La section 1 s'intéresse, d'abord, aux fondements des modèles DSGE et note leur filiation à la méthodologie lucasienne, en particulier, leur adhésion aux principes de modélisation des Cycles Réels

(RBC). Toutefois, on assiste à une évolution de l'évaluation des DSGE vers une réconciliation du calibrage avec les méthodes d'estimation bayésiennes et les techniques d'inférence simulée. Puis, on rappelle le rôle central des hypothèses de rigidité des prix et des imperfections de marché pour la justification de l'utilisation de la politique monétaire dans un DSGE. On présente, enfin, la maquette de base de ce type de modèles, qui s'appuie sur une relation IS et une courbe de Phillips néokeynésienne.

La deuxième section précise que la place et la conception de la politique monétaire dans les modèles DSGE sont fortement influencées par les travaux des Nouveaux Classiques (Barro et Gordon, 1983 ; Kydland et Prescott, 1979 ; Sargent et Wallace, 1981). Dans le cadre d'un régime à politique monétaire dominante, s'articulant autour de la crédibilité monétaire et de la discipline budgétaire afin d'éviter l'incohérence temporelle de la politique économique, la stabilité des prix est érigée au rang d'objectif ultime. Une règle de Taylor, déduite de la fonction-objectif de la Banque centrale, décrit le comportement de celle-ci qui cherche à influencer l'évolution à court terme de la demande agrégée dans un contexte de rigidité des prix. En raison des restrictions inter-équations (Sargent, 1979) imposées par l'hypothèse des anticipations rationnelles, Benati et Surico (2009) montrent que les modèles DSGE permettent mieux que les VAR structurels d'identifier un changement dans cette règle de politique monétaire.

Première Partie

**La Modélisation Structurelle Remise en Cause par les
Monétaristes et les Modèles VAR**

Introduction de la Première Partie

La première partie de cette thèse a pour objet d'analyser les termes des débats théoriques entre les monétaristes et les keynésiens et d'étudier les limites de la remise en cause par les tenants de l'approche des modèles VAR de la représentation de la politique monétaire dans les modèles structurels de la Commission Cowles. La justification théorique du mode de détermination de l'offre de monnaie, du taux d'intérêt nominal, du produit réel, du niveau général des prix et l'évaluation empirique des effets des variations de ces variables les unes sur les autres sont au cœur de ces débats. De fait, le concept de la causalité, qui fût au centre des travaux de la Commission Cowles (Hoover, 2001), et les notions connexes d'exogénéité et d'identification, sont à la base des divergences à propos de la modélisation des interdépendances entre ces variables.

Le traitement de ces questions est effectué par les chercheurs de la Commission Cowles à partir de l'articulation d'un référentiel théorique issu de la théorie Générale de Keynes, d'un modèle formel s'appuyant sur le relatif consensus autour du schéma IS-LM et de techniques économétriques pour estimer les paramètres de ce modèle. En insistant sur l'importance de l'*a priori* théorique, les keynésiens de la synthèse conçoivent l'exogénéité comme une forme de restriction imposée sur les paramètres, nécessaire à l'identification de la forme structurelle du modèle.

Les particularités qui distinguent la version keynésienne du modèle IS-LM de sa version classique, les conséquences qui en découlent pour l'étude des canaux de transmission de la politique monétaire vers l'activité économique et l'utilité de l'intervention de l'Etat pour la stabilisation conjoncturelle constituent des éléments clés du débat théorique entre les monétaristes et les keynésiens. A cet effet, la première critique fondamentale des monétaristes à l'encontre du modèle IS-LM concerne l'hypothèse de la trappe à liquidité qu'ils considèrent comme constituant un élément central du système keynésien (Friedman, 1971). L'alternative monétariste pour la description des mécanismes de transmission de la politique monétaire s'appuie sur la demande de monnaie friedmanienne (Friedman, 1956) qui cherche à spécifier les déterminants du choix de la composition optimale d'un portefeuille formé d'une large gamme d'actifs. Toutefois, Mayer (1975) note que l'existence d'un désaccord à propos de la gamme d'actifs et des dépenses associées n'est pas décisive pour expliquer la divergence

entre les monétaristes et les keynésiens. Pour Tobin (1972), c'est plutôt l'interprétation différente de la formation du taux d'intérêt qui est à la base de cette divergence.

La seconde critique fondamentale des monétaristes porte sur le manque de fondement théorique pour la courbe de Phillips utilisée dans les modèles structurels de la Commission Cowles. La reformulation de cette courbe en 1968 permet à Milton Friedman d'établir un lien entre sa théorie de court terme du revenu nominal et la dichotomie classique de long terme entre les variables monétaires et les variables réelles (Desai, 1981). Avant cette date, les monétaristes s'appuient sur l'observation de corrélation et de décalages temporels entre la masse monétaire et le revenu nominal pour justifier l'influence causale de celle-là sur celui-ci et formuler leur règle de politique monétaire. S'inscrivant dans la tradition de la Commission Cowles, Tobin (1970) rejette la justification de la causalité par l'antécédence temporelle et développe un modèle théorique keynésien qui génère les décalages observés par Friedman (1958), mais où la monnaie ne joue aucun rôle causal sur l'activité économique.

L'introduction des modèles VAR s'inscrit dans le prolongement du débat entre Friedman et Tobin à propos de la causalité monnaie-revenu en cherchant à étendre l'analyse de la causalité au sens de Granger à un vecteur de plus de deux variables (Sims, 1980, a, b). Dans sa célèbre critique, Sims reproche aux modèles structurels de la Commission Cowles de comporter trop d'hypothèses théoriques non testées empiriquement et propose de soumettre les hypothèses d'exogénéité et de la présence d'une causalité à des tests économétriques directs et précis. Toutefois, l'indétermination empirique de la causalité dans un VAR, liée au problème de l'équivalence observationnelle (Basmann, 1965), impose l'adoption d'un schéma d'identification sur la base d'*a priori* théorique pour identifier les chocs de politique monétaire. Ainsi, l'évolution de la modélisation VAR conduit à une situation paradoxale où, pour justifier les restrictions d'identification imposées aux innovations contemporaines, Sims se réfère à la notion de chaîne causale introduite par Wold (1954).

Chapitre 1

**La Critique Monétariste
de la Politique Monétaire Keynésienne**

Introduction

La modélisation macroéconométrique structurelle connaît son âge d'or de la fin des années 1950 jusqu'au début des années 1970. Cette appellation est généralement utilisée pour désigner le programme de recherche développé, dès le début des années 1940, au sein de la Commission Cowles autour des modèles à équation simultanées. Les modèles de Klein (1950) et de Klein et Goldberger (1955) constituent le premier maillon de cette nouvelle discipline dont l'initiateur est Tinbergen (1937).

Le cadre théorique de référence utilisé par les keynésiens dans cette modélisation structurelle est constitué par le modèle IS-LM de Ando Modigliani (1944), considéré comme l'incarnation de la macroéconomie de la synthèse (De Vroey, 2000). Depuis les premiers modèles de Klein, le souci de la validation empirique du modèle théorique et de la désagrégation de ses équations a été très présent dans la tradition de la Commission Cowles, ce qui a conduit à d'importants amendements des relations théoriques initiales.

La plupart des questions soulevées dans le débat des années 1960 et 1970 entre les monétaristes et les keynésiens concernent des relations déjà incorporées dans le modèle IS-LM. La question centrale de ce débat porte sur l'effet d'une variation de l'offre de monnaie sur le revenu et les prix. L'existence de larges écarts entre les relations estimées et le modèle théorique de la synthèse favorise la position de Friedman qui cherche à minimiser l'importance de l'écart théorique entre les keynésiens et les monétaristes et de déplacer le débat sur le terrain empirique.

Une large partie de ce débat concerne l'étude des canaux de transmission de la politique monétaire vers l'activité économique et l'utilité de l'intervention de l'Etat pour la stabilisation conjoncturelle. De fait, les particularités qui distinguent la version keynésienne du modèle IS-LM de sa version classique et les conséquences qui en découlent en termes de politique économique deviennent au cœur des débats entre les monétaristes et les keynésiens. En particulier, les monétaristes cherchent à remettre en cause l'autonomie de l'investissement par rapport à l'épargne, l'indépendance de celle-ci vis-à-vis du taux d'intérêt et la nature monétaire de la formation de ce dernier.

Chaque courant adopte une version particulière du processus d'ajustement du portefeuille comme cadre de référence pour décrire les effets de la politique monétaire sur l'économie réelle (Park, 1972). Toutefois, l'existence d'un désaccord à propos de la gamme d'actifs et de dépenses associées n'est pas décisive pour expliquer la divergence entre ces deux courants de pensée (Mayer, 1975). Pour Tobin (1972), c'est plutôt l'interprétation différente de la formation du taux d'intérêt qui est à la base de cette divergence.

La critique fondamentale de Friedman (1971) à l'encontre du modèle IS-LM concerne l'hypothèse de la trappe à liquidité qu'il considère comme constituant un élément central du système keynésien. Plus généralement, Friedman présente cette hypothèse dans son article de 1956 comme l'une des principales différences entre les adeptes et les opposants à la théorie quantitative de la monnaie. L'effet faible d'une variation de l'offre de monnaie sur le revenu réel est considéré par les monétaristes comme la conséquence de l'hypothèse keynésienne de fixité des prix et de la forte sensibilité de la demande de monnaie au taux d'intérêt. De plus, le rôle crucial du taux d'intérêt dans le modèle de la synthèse est jugé par les monétaristes à l'origine de l'effet d'éviction exercé par les dépenses publiques sur l'investissement privé. Pour remédier à ce problème, les monétaristes recommandent de financer le déficit public par création monétaire au lieu de l'endettement par émissions d'obligations.

Section 1 : IS-LM comme cadre de la modélisation structurelle

En partant du cadre théorique de la synthèse, Klein et Goldberger (1955) construisent le premier modèle macroéconométrique dans la tradition de la Commission Cowles. L'objectif de Klein est de prouver empiriquement la supériorité de la version keynésienne du modèle IS-LM par rapport à sa variante classique et d'utiliser son modèle pour l'élaboration des prévisions de l'activité économique et la simulation des mesures de politique économique (De Vroey et Malgrange, 2012). Toutefois, les relations estimées dans ce modèle présentent un important écart par rapport aux équations théorique des deux versions de Hicks et de Modigliani du schéma IS-LM.

Les monétaristes s'appuient sur l'ambiguïté introduite entre les versions théoriques et empiriques dans la modélisation de la Commission Cowles pour remettre en cause les principes keynésiens du modèle de la synthèse. Il s'agit, en particulier, de l'autonomie de l'investissement par rapport à l'épargne, de l'indépendance de celle-ci du taux d'intérêt et de la nature monétaire de la formation de ce dernier.

Sous-section 1 : Modèles macroéconométriques et souci de validation empirique

1. Le modèle économétrique de Klein- Goldberger

C'est à Klein et Goldberger (1955) que revient la construction du premier modèle macroéconométrique et son application à l'analyse des fluctuations économiques et des politiques économiques. Ce modèle est à la base de la modélisation macroéconométrique structurelle keynésienne des années 1960 et 1970. Selon Klein, la construction de ce modèle macroéconométrique est motivée par la volonté de vérifier par les données observées laquelle des deux versions classique et keynésienne du modèle IS-LM est vraie. En effet, Klein s'efforce de montrer, en s'appuyant sur le cadre théorique fourni par le modèle IS-LM développé par Modigliani, la supériorité empirique de la version keynésienne de ce modèle (De Vroey, 2000, De Vroey, Malgrange, 2012).

Toutefois, Klein exprime deux réserves contre le modèle IS-LM, toutes les deux justifiées par des arguments empiriques. La première concerne la rigidité à la baisse du salaire nominale qui ne constitue pas, selon lui, la cause principale du chômage. Celui-ci résulterait surtout de la difficulté de l'ajustement de l'investissement au niveau de l'épargne déterminée par le revenu de plein-emploi. L'absence de cette égalisation est expliquée par l'inélasticité empirique de la fonction d'investissement au taux d'intérêt. La seconde réserve porte sur la nature statique du modèle qui empêcherait son utilisation empirique. La volonté de rendre le modèle IS-LM dynamique a conduit Klein à y introduire une équation dynamique de formation du salaire.

Les relations théoriques de base estimées par Klein et Goldberger s'appuient sur les principales fonctions macroéconométriques introduites par Keynes, à savoir les fonctions de consommation C , d'investissement I et une demande d'encaisses réelles M/p , avec M la masse monétaire et p le niveau général des prix :

$$C = C(i, Y)$$

$$I = I(i, Y)$$

$$Y + T + D = C + I + G$$

$$M / p = M(i, Y)$$

où T désigne les impôts, G les dépenses du Gouvernement en biens et services, D l'amortissement physique (Artus, Deleau, Malgrange, 1986).

La description du comportement des marchés des biens et services et de la monnaie est enrichie par une formalisation théorique du marché du travail à travers une offre de travail N^S , une demande de travail N^D et une variation du taux de salaire nominal w :

$$N^D = N^D(w / p, K)$$

$$N^S = N^S(w / p)$$

$$dw / dt = F(N^S - N^D)$$

avec K le stock de capital.

L'existence d'un écart important entre la forme finale des équations estimées par rapport aux relations théoriques initiales prouve à quel point le souci empirique et la qualité de l'ajustement du modèle aux données étaient dominants dans les travaux de Klein. Cette tendance s'est poursuivie dans la modélisation macroéconométrique structurelle de la Commission Cowles au cours des années 1960 et 1970 (modèle de la Brookings, modèle de Wharton, modèle MPS, ...) avec une attention plus importante accordée au développement de la partie réelle du modèle et à la régulation de la demande. Les versions empiriques correspondant aux relations théoriques ci-dessus, telles que estimées par Klein et Goldberger, sont données par :

$$C = -34,5 + 0,62W + 0,46P + 0,39A + 0,23C_{-1} + 0,024(L_1)_{-1} + 0,36N_p$$

$$I = -16,8 + 0,76P_{-1} - 0,14K_{-1} + 0,14(L_2)_{-1}$$

$$L_1 = 0,14(W + P + A) + 75.0(i_L - 2)^{-0.84}$$

$$L_2 = -0,77 + 0,24W_1 - 0,69i_s - 0,27(p - p_{-1}) + 0,64(L_2)_{-1}$$

$$N^s = N$$

$$W_1 = -2,70 + 0,36(Y + T + D - W_2) + 0,14(Y + T + D - W_2)_{-1} + 0,16t$$

$$w - w_{-1} = 4,11 - 0,75(N - N_w - N_E) + 0,56(p_{-1} - p_{-2}) + 0,56t$$

où W est la masse salariale totale, W_1 est la masse salariale privée, W_2 est la masse salariale publique, P le revenu des entreprises non agricoles, A le revenu des entreprises agricoles, L_1 les liquidités des ménages, L_2 les liquidités des entreprises, N_p la population, N_w l'emploi salarié, N_E l'emploi non salarié, i_L le taux d'intérêt de long terme, i_s le taux d'intérêt de court terme.

Ce niveau de désagrégation des variables du modèle comparativement à la forme théorique, qui est de plus en plus important dans les versions ultérieures des modèles macroéconométriques, témoigne à son tour de la volonté des auteurs de reproduire les données observées.

2. Les versions monétaire et à prix fixé de IS-LM

Le modèle de la synthèse néoclassique s'est constitué autour de deux lectures particulières de la Théorie Générale de Keynes, celle de Hicks (1937) et celle de Modigliani (1944). Le modèle IS-LM qui a dominé la macroéconomie durant les années 1960 et auquel se réfèrent les monétaristes est celui que présente Modigliani dans son article de 1944. Le modèle IS-LL proposé par Hicks en 1937 constitue le point de départ du travail de Modigliani et la première étape dans le développement du modèle de la synthèse. Les relations de base du modèle IS-LM sont la demande agrégée et la préférence pour la liquidité.

Dans son article séminal de 1937, Hicks se fixe comme objectif de comparer les macroéconomies monétaires de Keynes et des classiques. Le cadre d'analyse général adopté par le modèle de la synthèse prend en compte la monnaie, supposée exogène, et retient une économie à un seul bien homogène utilisé à la fois pour la consommation et l'investissement. Le raisonnement est conduit à court terme, c'est-à-dire, à capital installé donné. Le taux de salaire nominal est supposé exogène. Le modèle fait référence à onze variables dont les notations sont les suivantes :

Y , le revenu nominal global

I , l'investissement nominal global

C , la consommation nominale globale

S , l'épargne nominale globale

i , le taux d'intérêt nominal

M , la demande de monnaie

\bar{M} , l'offre de monnaie

La version monétaire du modèle de base de la synthèse, telle qu'initialement développée par Hicks (1937), est composée de six équations qui déterminent les 6 variables endogènes M , Y , C , I , S , i .

- (1) $M = \bar{M}$ équilibre entre offre exogène et demande de monnaie,
- (2) $M = L(i, Y)$ demande de monnaie,
- (3) $I = I(i)$ fonction d'investissement décroissante du taux d'intérêt,
- (4) $S = S(i, Y)$ fonction d'épargne croissante du taux d'intérêt, pour Y donné,
- (5) $I = S$ égalité entre épargne et investissement,

(6) $C = Y - I$ relation de définition comptable (la consommation est un solde)

Généralement, les équations (1) et (6) ne sont pas écrites de façon aussi explicite dans ce modèle. De même, on rencontre des versions dans lesquelles les relations (4) et (5) sont remplacées, respectivement, par les équations (4') et (5') suivantes :

(4') $C = C(i, Y)$ fonction de consommation décroissante du taux d'intérêt,

(5') $Y = C + I$ égalité entre le revenu et la dépense totale

La relation (6) est superflue dans cette écriture du modèle puisqu'elle est nécessairement réalisée à l'équilibre lorsque Y et S sont déterminés. La consommation est ainsi une variable de sortie dans le sens où elle n'est pas une variable explicative d'une autre variable endogène.

La nature de l'ajustement de l'équilibre (1) sur le marché de la monnaie dépend du rôle joué par le taux d'intérêt et du revenu dans la demande de monnaie donnée par la relation (2) et du mode de leur détermination. Les keynésiens de la synthèse adoptent une interprétation de l'équilibre (1) du marché monétaire comme résultant de la confrontation entre une offre exogène de monnaie \bar{M} et la demande de monnaie totale $L(i, Y)$ présentée par Keynes (1936) dans le Chapitre 15 de sa Théorie générale sous la forme :

(2') $M = L(i, Y) = L_1(i) + L_2(Y)$

où $L_1(i)$ exprime la demande de monnaie de spéculation et $L_2(Y)$ représente la demande de monnaie pour motif de transaction et de précaution. Toutefois, avec cette lecture de la théorie keynésienne, l'équilibre sur le marché monétaire ne permet pas à lui seul de déterminer le taux d'intérêt puisqu'il fait intervenir le niveau du revenu, et la connaissance de ce dernier exige à son tour, d'après les équations (3), (4) et (5), la détermination du taux d'intérêt.

Sous les conditions (1) et (2'), ce modèle monétaire de base de la synthèse est indécomposable dans le sens où les équations (2), (3) et (4) sont interdépendantes et les variables Y , i , I et S sont déterminées de façon simultanée. La détermination du taux d'intérêt et de l'investissement n'est pas indépendante de celle de l'épargne et du revenu et réintroduit ainsi la dépendance "classique", qui a été fortement critiquée par Keynes, de l'investissement vis-à-vis de l'épargne.

Les variables utilisées ici pour présenter le modèle de Modigliani ont la même signification que dans le modèle de Hicks, soient les 6 variables endogènes monétaires M , Y , C , I , S , i , et l'offre de monnaie \bar{M} toujours supposée exogène. Cette liste est complétée par les cinq variables suivantes, ce qui fait en tout 11 variables endogènes :

p , le niveau général des prix, c'est-à-dire le prix nominal du bien unique

w , le taux de salaire nominal

Q , la quantité produite et offerte du bien

N^- , la demande de travail des producteurs

Modigliani complète le modèle monétaire de base de la synthèse constitué des équations (1) à (6) par trois équations, toutes présentes dans l'article de Hicks; à savoir, une fonction de production qui relie le niveau de l'emploi à la production, une relation qui détermine le prix à partir du coût marginal du travail, et une identité qui permet de calculer la production en fonction du revenu réel :

(7) $N^- = f^{-1}(Q)$ fonction de production inverse de court terme,

(8) $p = w / f'(N^-)$ prix d'offre du bien homogène égal au coût marginal,

(9) $Q = Y / p$ relation comptable entre la production et le revenu réel.

La modification fondamentale apportée par Modigliani consiste à transformer le modèle monétaire de Hicks constitué des équations (1) à (6) en un modèle réel exprimant les variables en termes du niveau général des prix :

$$(1.2) \quad M = \bar{M}$$

$$(2.2) \quad M = p.L(i, Q)$$

$$(3.2) \quad \frac{I}{p} = \frac{I}{p}(i)$$

$$(4.2) \quad \frac{S}{p} = \frac{S}{p}(i, Q)$$

$$(5.2) \quad \frac{I}{p} = \frac{S}{p}$$

$$(6.2) \quad \frac{C}{p} = Q - \frac{S}{p}$$

A cette étape, on ne décrit pas la modélisation de Modigliani du marché du travail qui sera présentée au niveau du débat sur la courbe de Phillips dans le chapitre 2. De fait, on suppose ici que le niveau général des prix p est une donnée fixée.

Contrairement à la relation (2') ci-dessus, la demande de monnaie (2.2) est écrite ici sous la forme d'encaisses réelles :

$$M = L_1(i) + p.L_2(Q)$$

Dans cette écriture, la demande de monnaie pour motif de spéculation $L_1(i)$ est toujours monétaire mais la fonction $L_2(Q)$ est maintenant une grandeur réelle et non plus monétaire. Pour une offre exogène de la masse monétaire \bar{M} et un niveau donné du niveau général des prix, on en déduit une première relation entre le taux d'intérêt et la production :

$$(2.2') \quad L_1(i) = \bar{M} - p.L_2(Q) \\ = S_p(Q)$$

où $S_p(Q)$ est une fonction décroissante de la production. La figure 1 présente les solutions d'équilibre sur le marché monétaire, pour des niveaux de revenus réels donnés $Q_1 < Q_2 < Q_3$, obtenues par l'intersection de la courbe de demande de monnaie pour motif de spéculation $L_1(i)$ et les courbes de l'offre de monnaie $S_p(Q)$ correspondantes.

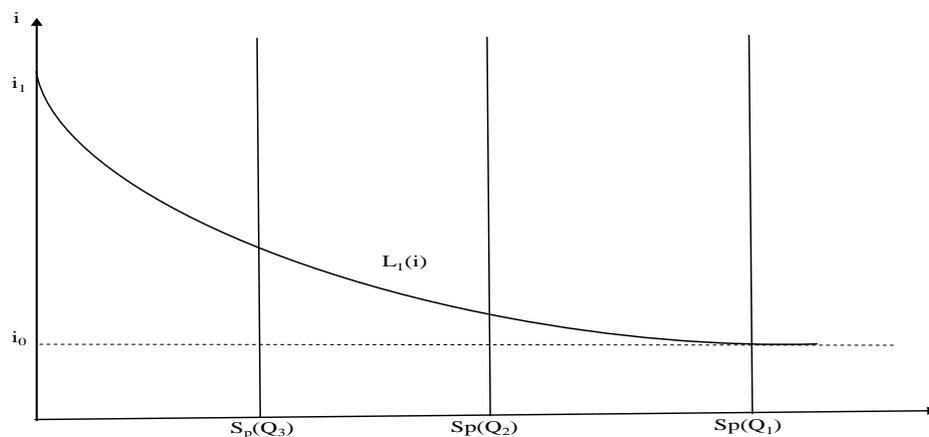


Figure 1 : courbe de la préférence pour la liquidité

Modigliani reprend l'hypothèse faite par Hicks concernant l'existence d'un taux d'intérêt minimum i_0 , qui sera désigné plus tard de trappe à liquidité par Denis Robertson (1940), pour lequel les individus n'accepteraient de détenir leurs actifs que sous forme monétaire. Une fois ce minimum atteint, la demande de monnaie pour motif de spéculation devient infiniment élastique et le taux d'intérêt ne peut plus descendre au-dessous de i_0 . Dans ce cas, $L_1(i)$ tend vers l'infini pour tout $i \leq i_0$ et la courbe de la fonction $L_1(i)$ devient parallèle à l'axe des abscisses. De même, Modigliani suppose l'existence d'un taux d'intérêt i_1 pour lequel les individus sont prédisposés à détenir toute leur épargne sous forme de titres et se défaire de toute leur demande de monnaie pour motif de spéculation. Ainsi, pour tout $i \geq i_1$, on a $L_1(i) = 0$.

Etant donné que l'offre de monnaie $S_p(Q)$ est une fonction décroissante du revenu réel et que la préférence pour la liquidité $L_1(i)$ est une fonction décroissante du taux d'intérêt, on obtient ainsi une courbe LM (voir Figure 2) croissante entre le revenu réel et le taux d'intérêt nominal:

Courbe LM : $i = LM(Q)$

Une deuxième relation IS est construite entre le taux d'intérêt et le revenu réel sous l'hypothèse d'une fonction d'épargne (4.2) indépendante du taux d'intérêt, c'est-à-dire conforme à l'analyse keynésienne d'une propension à épargner le revenu donnée :

$$(4.2') \quad \frac{S}{p} = \frac{S}{p}(Q)$$

La combinaison des équations (3.2), (5.2) et (4.2') permet de déterminer une relation décroissante entre le revenu réel et le taux d'intérêt nominal :

Courbe IS : $i = IS(Q)$

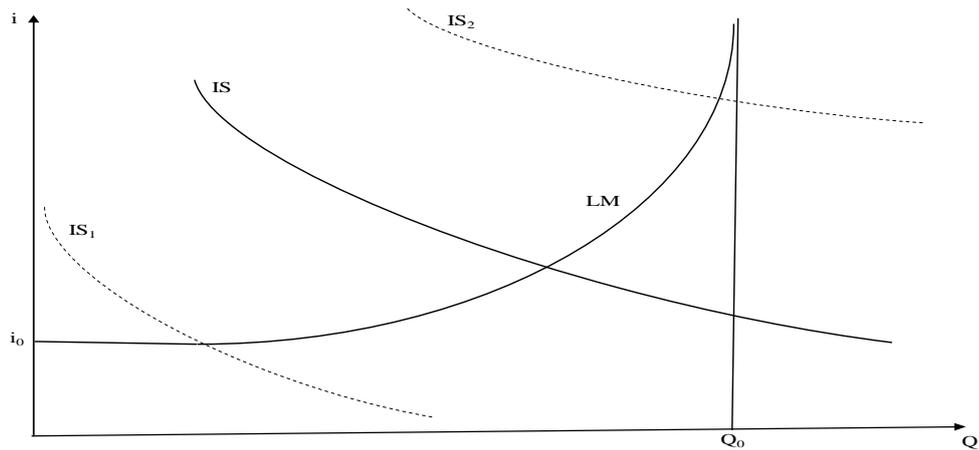


Figure 2 : Schéma IS-LM à prix fixé

On obtient au point d'intersection des deux courbes IS et LM le revenu réel d'équilibre et le taux d'intérêt d'équilibre. Les deux formes asymptotiques de la courbe LM trouvent leur origine dans la valeur de l'élasticité de la demande de monnaie pour motif de spéculation par rapport au taux d'intérêt. Dans cette lecture hicksienne de la Théorie générale de Keynes, la version keynésienne du modèle est associée aux points d'intersection de la courbe IS avec la partie horizontale de la courbe LM (cas de la courbe IS_1), alors que le modèle classique correspond à l'intersection de la courbe IS avec la partie verticale de la courbe LM (cas de la courbe IS_2).

Sous-section 2 : L'attaque monétariste des fondements keynésiens de la synthèse

1. Demande effective et autonomie de l'investissement et l'épargne

Une partie des débats qui ont opposé les monétaristes aux keynésiens autour du modèle de la synthèse a porté sur les conditions de réalisation des équilibres donnés par les équations (1.2) et (5.2) du modèle de Modigliani présenté dans la sous-section 1 ci-dessus. S'inscrivant dans la tradition classique, les monétaristes considèrent la relation (5.2) comme une condition d'équilibre sur le marché financier obtenue par la confrontation d'une épargne des ménages avec une demande d'investissement des entreprises moyennant l'ajustement du taux d'intérêt. Sur cette question, les keynésiens s'opposent radicalement aux monétaristes, soulignant que Keynes a rejeté catégoriquement cette lecture en termes de théorie dite des fonds prêtables

défendue par les économistes classiques qui établit une dépendance de la fonction d'investissement par rapport aux comportements de l'épargne (Deleplace, 1999).

D'une part, Keynes (1936) brise, à travers son concept de la préférence pour la liquidité, l'identité automatique entre l'épargne et l'investissement en montrant qu'une partie de cette épargne peut être thésaurisée et détenue sous forme monétaire plutôt que d'actifs financiers. En effet, le motif de spéculation reflète ce que Keynes appelle la préférence pour la liquidité qui détermine la forme de la détention de l'épargne en distinguant deux décisions. La première concerne le niveau de l'épargne et porte sur le partage du revenu monétaire entre consommation et épargne. Cette décision ne dépend pas du taux d'intérêt mais elle est fonction du revenu, des prix, de la préférence de l'individu pour le présent et de facteurs psychologiques. En revanche, la seconde décision fait intervenir le taux d'intérêt car elle porte sur la forme de détention de cette épargne. L'individu effectue un arbitrage, en fonction du niveau du taux d'intérêt, entre la détention de son épargne sous formes d'actifs financiers ou sous forme d'encaisses monétaires.

D'autre part, la détermination de l'investissement se fait pour un niveau donné du taux d'intérêt qui se forme sur le marché monétaire et non pas sur le marché financier. Dans ce cas, la présence d'une épargne n'est pas une condition préalable à la réalisation de l'investissement qui est financé par les crédits fournis aux entreprises par le système bancaire. Ainsi, pour les keynésiens, la relation (5.2) ne constitue plus une condition d'équilibre sur le marché financier, mais une condition d'équilibre global qu'il formule en termes de théorie de la demande effective.

Pour analyser la nature de la relation qui existe entre le taux d'intérêt et le revenu à travers la condition d'équilibre (5.2), il faut rappeler que dans le système keynésien le revenu est déterminé par la demande agrégée qui est anticipée par les entrepreneurs. Autrement dit, dans une perspective keynésienne, l'équilibre entre l'épargne et l'investissement serait mieux décrit par la relation :

$$(6.3) \quad Q = \frac{C}{p} + \frac{S}{p}$$

avec des spécifications particulières pour la fonction de consommation et d'investissement.

D'un côté, la fonction d'épargne donnée par la relation (4.2), ou de manière équivalente celle de la consommation, ne dépend pas du taux d'intérêt. Elle est uniquement fonction du revenu selon la formule suivante :

$$(6.4) \quad \frac{C}{p} = c.Q \quad \text{avec } c \text{ la propension constante à consommer le revenu}$$

D'un autre côté, l'investissement contribue bien dans la détermination du revenu, mais il est indépendant par rapport à ce revenu en raison de la nature des anticipations faites par les entrepreneurs. Les déterminants de l'investissement sont constitués du taux d'intérêt et de l'efficacité marginale du capital. Cette dernière est définie comme étant le taux de rendement escompté d'un capital nouveau et, de fait, il s'agit d'une grandeur monétaire qui est anticipée par les entrepreneurs. Puisque la concurrence exercée par la multiplication des projets d'investissement réduit les recettes attendues, l'efficacité marginale du capital est une fonction décroissante du niveau de l'investissement. Pour un état donné des anticipations, l'entrepreneur pousse le niveau de son investissement jusqu'au niveau où l'efficacité marginale du capital soit égale à au taux d'intérêt.

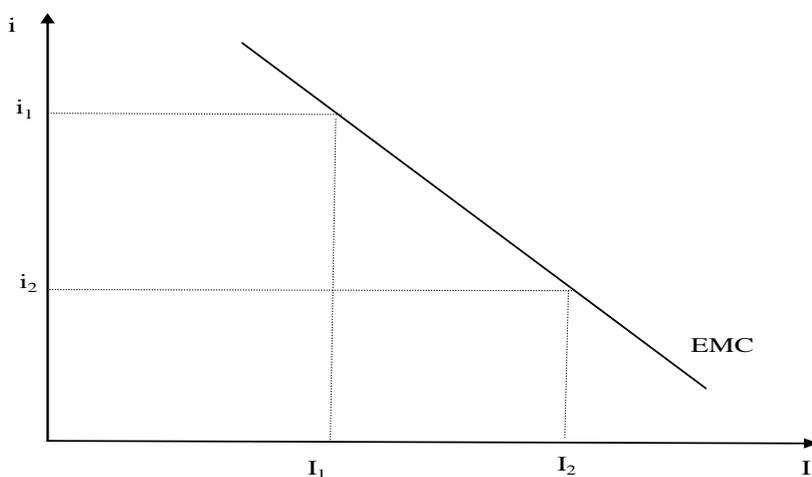


Figure 3 : Courbe de l'efficacité Marginale du Capital(EMC)

A partir de ces éléments, la relation (6.3) peut être transformée pour faire apparaître le concept du multiplicateur qui relie directement le revenu à l'investissement et ainsi au taux d'intérêt :

$$(6.5) \quad Q = \frac{1}{(1-c)} \cdot \frac{I(i)}{p}$$

Etant donné que $(1-c)Y = S$, l'équation (6.5) est exactement l'équivalent de la relation (5.2), mais avec une épargne qui ne dépend que du revenu et qui est indépendante du taux d'intérêt. Il s'agit d'une relation qui représente l'équilibre global de l'économie, mais où l'ajustement entre l'épargne et l'investissement ne se fait pas par le taux d'intérêt comme chez les monétaristes ou les classiques. C'est plutôt l'épargne qui s'ajuste au niveau donné de l'investissement par l'efficacité marginale du capital et du taux d'intérêt de façon à ce que le revenu se fixe à son niveau d'équilibre donné par (6.5). Keynes défend ainsi l'idée d'une autonomie de l'investissement par rapport à l'épargne car la présence de celle-ci n'est pas une condition préalable à la réalisation de celui-là (Deleplace, 1999).

2. Nature monétaire du taux d'intérêt et récursivité du modèle

Les monétaristes retiennent au niveau théorique la même formalisation que le modèle IS-LM pour l'équilibre monétaire (1.2) avec une offre exogène de monnaie \bar{M} qui satisfait la demande de monnaie (2.2) (Friedman, 1970 a). Mais, ils introduisent en plus l'identité de Fisher pour expliquer le comportement du taux d'intérêt nominal à travers celui de l'inflation et du taux d'intérêt réel qui égalise l'épargne et l'investissement (voir chapitre 2). A ce niveau, les monétaristes se démarquent significativement de la tradition classique en abandonnant la formulation de l'équation quantitative de la monnaie.

Or, Keynes (1936) déplace la définition du taux d'intérêt de celle, retenue par les classiques, puis par les monétaristes, qui le considère comme une récompense des décisions de ne pas consommer vers celle qui le voit comme la récompense des décisions de ne pas thésauriser ou la récompense pour la renonciation à la liquidité. En effet, Keynes assimile le concept de préférence pour la liquidité à la tendance des individus à thésauriser, autrement dit, à conserver des encaisses monétaires oisives. Il explique cette tendance des agents à détenir de l'argent liquide comme réserve de valeur ou comme moyen de conserver leur richesse alors qu'il ne fournit aucune rémunération par l'existence d'une incertitude sur le niveau futur du taux d'intérêt.

Dans cette perspective, le désir de détenir la monnaie pour motif de spéculation naît de l'incertitude qu'ont les individus sur la valeur future des actifs financiers. En effet, contrairement à ceux-ci, la monnaie a l'avantage de conserver une valeur nominale future connue avec certitude. De fait, le taux d'intérêt sera d'autant plus élevé que l'incertitude concernant son niveau futur est grande. Ces éléments renvoient au débat entre Friedman (1972) et Davidson (1972) sur l'incertitude et sur l'opportunité de l'intégration de la monnaie dans un modèle walrasien.

Il existe une seconde spécificité keynésienne de l'analyse de l'équilibre du marché monétaire, défendue par le courant post-keynésien et basée sur le motif de financement, qui renforce la dimension monétaire du taux d'intérêt. Dans le même temps, cette interprétation permet de casser la simultanéité de résolution du modèle IS-LM, présentée dans le point 2 de la sous-section 1 ci-dessus, et d'introduire une récursivité dans cette résolution. Ces deux interprétations opposées de l'équilibre du marché monétaire résultent de l'ambiguïté de l'analyse keynésienne de l'ajustement de l'offre exogène de monnaie aux deux composantes de la demande de monnaie $L_1(i)$ et $L_2(Y)$. D'autant plus que Keynes présente dans le chapitre 15 de la Théorie générale des arguments qui renforcent chacune des deux interprétations (Deleplace, 1999). On retrouve cette ambiguïté à la base de la controverse entre les monétaristes et les keynésiens sur la détermination et le rôle du taux d'intérêt.

Keynes développe le motif de financement comme un motif distinct de la demande de monnaie dans deux articles publiés après la Théorie générale (Keynes, 1937a, 1937b) pour insister sur la différence de sa théorie monétaire du taux d'intérêt avec la théorie des fonds prêtables défendue par Hicks (1936), Robertson (1940) et Ohlin (1937) et adoptée par les monétaristes. Il s'appuie sur ce motif pour introduire le crédit bancaire dans la satisfaction de la demande de monnaie et dans l'analyse du processus de la création monétaire et pour renforcer son hypothèse d'autonomie de l'investissement par rapport à l'épargne (Deleplace, 1999).

Dans cette perspective, les entreprises peuvent recourir au crédit bancaire mis à leur disposition par le système bancaire pour financer leurs projets d'investissement. Ces investissements n'exigent pas que leur financement soit déjà disponible sur le marché financier à travers une épargne préalable. Le crédit bancaire permet ainsi de couvrir l'intervalle de temps qui sépare le lancement de la production et l'apparition de l'épargne qui

suit l'augmentation du revenu. Le financement bancaire dont il s'agit dans cette analyse est un financement de court terme que les entreprises doivent consolider par l'émission d'obligations de long terme.

La demande de monnaie qui émerge de cette analyse keynésienne du financement de court terme de l'investissement est satisfaite de façon endogène, à un taux d'intérêt de court terme donné, par le système bancaire puisque les liquidités s'adaptent à cette demande. Elle se distingue de la demande de monnaie pour motif de spéculation présentée ci-dessus et qui est satisfaite par une offre de monnaie centrale contrainte par le taux d'intérêt conventionnel de long terme.

L'analyse du motif de financement se couple à la rupture introduite par Keynes par rapport aux classiques dans l'analyse du rôle du taux d'intérêt dans la détermination de l'investissement par le multiplicateur. D'une part, elle porte sur le rôle de la monnaie à travers la prise en compte de l'incertitude sur l'évolution future du taux d'intérêt, contrainte par les conventions du marché, et l'arbitrage des agents entre la détention de monnaie et celle des actifs financiers. D'autre part, en établissant une distinction entre les activités d'entreprise et les activités de spéculation, la théorie monétaire de Keynes dégage deux façons distinctes de satisfaire les demandes de monnaie correspondantes, celle de spéculation par l'offre de monnaie centrale, et celle de financement de l'investissement par le crédit bancaire.

Section 2 : Canal du taux d'intérêt versus canal des prix

Les keynésiens et les monétaristes adoptent des versions différentes pour la description des mécanismes de transmission de la politique monétaire vers la production. Les monétaristes retiennent une classe d'actifs et de prix plus large que les keynésiens pour la description de ces mécanismes. Selon Mayer (1975), cet écart n'est, toutefois, pas décisif pour expliquer la divergence entre les deux courants de pensée. C'est plutôt la conception différente de la formation du taux d'intérêt qui est à la base de cette différence (Tobin, 1972).

Friedman (1971) explique la portée limitée de l'effet de la politique monétaire dans le modèle de la synthèse par l'hypothèse keynésienne de fixité des prix et par la forte sensibilité de la demande de monnaie au taux d'intérêt. Les monétaristes s'appuient également sur le rôle central que joue le taux d'intérêt dans le modèle IS-LM pour expliquer l'effet d'éviction qu'engendre une hausse de ce taux sur le revenu. Pour éviter les conséquences perverses de cet effet, les monétaristes proposent de financer le déficit budgétaire par création monétaire au lieu du recours à l'endettement par émission d'obligations.

Sous-section 1 : La transmission de la politique monétaire

1. Courbe LM et multiplicateur monétaire

Le canal par lequel la politique monétaire exerce un effet sur le revenu dans la version standard du modèle de la synthèse à prix fixé est le taux d'intérêt. Dans ce cadre, les autorités monétaires peuvent décider d'adopter une politique monétaire expansionniste par une augmentation de l'offre nominale de monnaie \bar{M} pour stimuler l'activité à travers l'effet du taux d'intérêt sur l'investissement.

Rappelons, en effet, que la courbe LM est définie sous l'hypothèse d'une offre réelle de monnaie donnée \bar{M} / p . Une augmentation de cette offre, qu'elle soit le fait d'une hausse de la masse monétaire nominale \bar{M} ou de la baisse du niveau général des prix exogène p , aura pour conséquence une hausse de l'offre de monnaie $S_p(Q)$, pour un revenu réel Q donné. D'après l'équation (2.2') il en résulte une augmentation de la demande de monnaie pour motif

de spéculation $L_1(i)$ et donc une baisse du taux d'intérêt et un déplacement de la courbe LM vers le bas.

L'augmentation de l'investissement qui résulte de cette baisse du taux d'intérêt entraîne, par le jeu du multiplicateur, une hausse du revenu. Mais, cet accroissement du revenu engendre à son tour une élévation de la demande de monnaie pour motif de transaction qui exige une hausse du taux d'intérêt pour rétablir l'équilibre sur le marché monétaire. Ce second effet réduit alors l'impact initial sur le revenu. Au total, l'effet d'un accroissement de la masse monétaire sur le revenu réel, pour un niveau donné des prix p , est donné par le multiplicateur monétaire qu'on peut calculer à partir des équations de l'équilibre sur le marché monétaire et du multiplicateur d'investissement suivantes :

$$\begin{aligned}\bar{M} &= L_1(i) + p.L_2(Q) \\ p.Q &= \frac{I(i)}{(1-c)}\end{aligned}$$

La différentiation de ces équations donne :

$$\begin{aligned}d\bar{M} &= L_1'(i).di + p.L_2'(Q).dQ \\ (1-c).p.dQ &= I'(i).di\end{aligned}$$

où $L_1'(i)$ désigne la dérivée de $L_1(i)$ par rapport au taux d'intérêt, $L_2'(Q)$ la dérivée de $L_2(Q)$ par rapport au revenu réel et $I'(i)$ la dérivée de l'investissement par rapport au taux d'intérêt. La seconde équation fournit une première relation entre les différentielles du taux d'intérêt et de la production :

$$di = \frac{(1-c).p}{I'(i)}.dQ$$

En remplaçant cette expression de di dans l'équation donnant $d\bar{M}$, on obtient :

$$d\bar{M} = L_1'(i). \frac{(1-c).p}{I'(i)}.dQ + p.L_2'(Q).dQ$$

En réaménageant les paramètres, cette équation peut se mettre sous la forme suivante :

$$\frac{dQ}{d\bar{M}} = \frac{1}{p} \frac{I'(i)}{(1-c).L_1'(i) + I'(i).L_2'(Q)}$$

Cette formule diffère par le paramètre $1/p$ de celle présentée dans l'analyse de Lavalie (2003).

Ce multiplicateur monétaire, qui est un nombre toujours positif, admet deux valeurs extrêmes qui ont été à l'origine d'intenses débats entre les monétaristes et les keynésiens. La première valeur est nulle et correspond, soit au cas où $L_1'(i)$ tend vers l'infini, soit au cas où l'investissement est insensible au taux d'intérêt. Le premier cas a été qualifié de "cas spécial de Keynes" par Hicks et correspond à la partie du schéma IS-LM où la courbe IS croise la courbe LM au niveau de sa partie horizontale de gauche (voir figure 2 ci-dessus).

A ce niveau de la courbe, tout accroissement de l'offre de monnaie est conservé pour des motifs de spéculation, de sorte que le taux d'intérêt ne varie pas. La politique monétaire ne permet pas de faire baisser le taux d'intérêt en dessous de son taux planché et elle est donc inefficace pour stimuler l'investissement et porter le revenu à son niveau de plein-emploi. La rigidité du taux d'intérêt à la baisse est présentée par Hicks, et plus tard par les monétaristes, comme une explication de l'émergence de la configuration keynésienne dans ce modèle où, uniquement un déplacement à droite de la courbe IS permet une augmentation du revenu. Dans cette situation de trappe à liquidité, le seul moyen pour augmenter le revenu est de provoquer un déplacement vers la droite de la courbe IS à travers une action sur les dépenses publiques.

Le multiplicateur admet une seconde valeur maximale qui correspond au cas où $L_1'(i)$ est nulle, c'est-à-dire le "cas classique" de Hicks, là où la courbe IS coupe la courbe LM au niveau de sa partie verticale de droite. Dans ce cas, la valeur du multiplicateur monétaire est égale à :

$$\frac{dQ}{d\bar{M}} = \frac{1}{p} \frac{1}{L_2(Q)}$$

La demande de monnaie est insensible au taux d'intérêt et l'équilibre sur le marché de la monnaie est donné uniquement par la demande de transaction :

$$\bar{M} = p.L_2(Q)$$

Au niveau de cette portion de la courbe, une hausse de l'offre de monnaie se transmet entièrement au revenu réel, pour un niveau donné des prix. Un déplacement de la courbe IS permet de déterminer le taux d'intérêt, mais n'exerce aucun effet sur le revenu. Cette situation

est considérée par Tobin (1972) comme représentant le modèle monétariste du revenu nominal.

2. Choix de portefeuille et canal monétariste

Pour rendre compte des effets de la variation de la masse monétaire sur le revenu nominal, les monétaristes privilégient un raisonnement en termes d'ajustement d'un portefeuille composé d'une large gamme d'actifs financiers et réels (Friedman et Schwartz, 1963 ; Friedman et Meiselman ; 1963, Friedman, 1961 ; Friedman, 1971).

Le mécanisme de transmission retenu par les monétaristes opère à travers la déformation des prix relatifs des actifs du bilan. L'augmentation de la masse monétaire *par* une opération d'open-market crée un excès de monnaie comparativement aux autres actifs dans le portefeuille des agents non bancaires. Dans la perspective de rétablir l'équilibre de leur bilan, ces agents affectent cet excès de monnaie pour l'achat des actifs financiers, poussant par là même leurs prix à la hausse. Cette augmentation des prix des actifs financiers incite, tôt ou tard, les individus à diversifier leur portefeuille par l'acquisition d'autres types d'actifs, y compris les actions, les biens immobiliers, les biens de productions et les biens durables de consommation. De la même façon, ce comportement provoque une élévation des prix des actifs non financiers déjà existants, ce qui augmente la richesse et donc la consommation et favorise la production des nouveaux actifs. A travers la transformation de la structure du bilan, la stimulation de la masse monétaire se transmet d'abord au marché financier, puis elle se propage au marché des biens et services.

Dans cette perspective, les monétaristes cherchent à établir un lien direct entre la masse monétaire et le revenu nominal conformément à leur principale thèse quantitativiste qui privilégie la prédominance des variations de l'offre de monnaie dans l'explication des variations du revenu nominal. Toutefois, comme cela est relevé par Mayer (1975), il n'est pas sûr que le mécanisme de transmission décrit par les monétaristes puisse se traduire par un effet de la monnaie sur le revenu qui soit plus important que dans le cas du modèle IS-LM. En effet, malgré la multiplication des variables affectées par les variations de l'offre de monnaie dans le cas monétariste, l'effet global sur le revenu peut être faible. A l'opposé, cet effet est d'autant plus élevé dans le modèle keynésien que l'élasticité au taux d'intérêt est forte pour

l'investissement et faible pour la demande de monnaie pour motif de spéculation. Ainsi, l'équivalence entre la thèse quantitativiste et le processus de transmission monétaristes, dans le sens où on peut accepter l'un sans retenir nécessairement l'autre, n'est pas établie.

Friedman considère que le raisonnement monétariste peut être conduit en termes de taux d'intérêt, étant donné la relation inverse qui relie les prix et les rendements des actifs financiers, mais à condition d'envisager une plus large classe de taux d'intérêt implicites et explicites que la conception keynésienne étroite de taux d'intérêt et de titres financiers. Il en déduit que la différence entre les monétaristes et les keynésiens concernant le mécanisme de transmission d'une variation de la masse monétaire au revenu se situe davantage au niveau de la gamme d'actifs prise en compte dans le raisonnement que dans la nature du processus.

Cette position de Friedman, qui cherche à minimiser l'importance de l'écart théorique entre les keynésiens et les monétaristes, s'inscrit dans la stratégie de cet auteur qui tente de déplacer le débat entre les monétaristes et les keynésiens sur le terrain empirique. D'ailleurs, de l'aveu de Friedman lui-même, il existe une différence théorique fondamentale entre les deux courants concernant le comportement des prix (Friedman, 1971). En effet, Friedman prétend que c'est l'hypothèse de fixité des prix qui contraint les keynésiens à adopter un canal étroit de transmission de la politique monétaire. En revanche, l'hypothèse de flexibilité des prix permet aux monétaristes d'interpréter le mécanisme de transmission des variations de l'offre de monnaie à travers l'ajustement des prix relatifs d'une large classe d'actifs financiers et réels.

Le mécanisme de transmission monétariste s'appuie de façon cruciale sur l'hypothèse d'une réaction du prix d'un actif à la hausse de sa demande. Ce raisonnement est valable uniquement parce que les monétaristes supposent que l'économie se situe au voisinage de l'équilibre de plein emploi. Dans leur article de 1963, Friedman et Schwartz supposent explicitement qu'à l'état initial du raisonnement, l'économie est caractérisée par une croissance régulière du revenu réel, du stock de monnaie et du niveau général des prix, et que le taux d'intérêt réel et le ratio de la richesse par rapport au revenu sont constants (Friedman et Schwartz, 1963). Dans un tel contexte, tout économiste, y compris les keynésiens, admet que la hausse de la demande de n'importe quel actif se traduit par une augmentation de son prix. Par contre, ce que les keynésiens n'acceptent pas, et que Keynes s'est forcé de démontrer, c'est que la flexibilité des prix puisse conduire à un équilibre de plein emploi. De

fait, le mécanisme de transmission décrit dans le modèle keynésien ne suppose pas que l'économie se situe au voisinage d'un tel équilibre.

L'interprétation différente du mécanisme de transmission de la politique monétaire par les keynésiens et les monétaristes est la conséquence de deux visions fondamentalement opposées de la nature et des déterminants du taux d'intérêt. Nous avons montré dans la section 1 ci-dessus que le taux d'intérêt constitue dans l'analyse keynésienne le prix du renoncement à la liquidité et il est déterminé sur le marché monétaire en fonction de l'arbitrage entre la détention de monnaie et la détention d'obligations. En revanche, chez les monétaristes, le taux d'intérêt nominal est égal à la somme du taux d'intérêt réel et du taux d'inflation anticipé, où le taux d'intérêt réel est déterminé par la productivité marginale du capital à l'équilibre de plein emploi. Comme l'a relevée Tobin, cette approche reflète un point de vue Fisherien où l'arbitrage s'effectue entre les biens et les obligations et non plus entre la monnaie et les obligations (Tobin, 1972).

Mayer (1975) illustre cet écart d'analyse du canal de transmission par la définition différente adoptée par les uns et les autres pour le prix de la monnaie. Les keynésiens considèrent le taux d'intérêt comme le prix de la monnaie, car ils supposent que la monnaie peut être soit détenue comme un fonds soit prêtée. En revanche, le prix de la monnaie est égal, pour les monétaristes, à l'inverse du niveau général des prix puisqu'ils considèrent que la monnaie est utilisée pour acheter des biens. Friedman reconnaît l'importance de la différence des points de vue à propos de la distinction entre la monnaie et le crédit et de la définition du prix de la monnaie. Il confirme que les quantitativistes interprètent le taux d'intérêt comme étant le prix du crédit et l'inverse du niveau général des prix le prix de la monnaie. Cette conception différente sur le rôle du taux d'intérêt est à la base de l'utilisation différente de la courbe de demande de monnaie par les deux courants.

Parmi les conséquences de ces différences d'interprétation, le taux d'intérêt nominal ne réagit pas de la même façon, dans un cas ou dans l'autre, à une augmentation de la masse monétaire (Mayer, 1975). Dans le modèle de la synthèse, le taux d'intérêt nominal baisse suite à une hausse de l'offre de monnaie. Dans le cas monétariste, cette baisse ne peut être que temporaire puisque, pour un taux d'intérêt réel donné, le taux d'intérêt nominal finit par augmenter en raison de l'effet Fisher par les prix. De plus, alors que l'analyse keynésienne tient compte à la fois du comportement de l'offre et de la demande de monnaie dans la détermination du taux

d'intérêt, l'analyse monétariste néglige la réaction de la demande de monnaie et se focalise uniquement sur l'offre de monnaie.

Sous-section 2: La critique monétariste

1. Fixité des prix et trappe à liquidité

Dans son article de 1971, Friedman adopte la version réelle du modèle de la synthèse présentée dans la section précédente pour comparer son cadre théorique avec celui des keynésiens. Il reprend la classification de Hicks (1937) pour caractériser l'économie keynésienne en l'assimilant au cas de la préférence absolue pour la liquidité, c'est-à-dire, celui de la trappe à liquidité correspondant à la partie horizontale de la courbe LM.

Les monétaristes considèrent que l'effet de la politique monétaire sur le revenu dans l'analyse keynésienne est faible et d'une portée limitée car s'appuyant sur un canal étroit du taux d'intérêt. Ils estiment que le mode de transmission de la politique monétaire dans le modèle keynésien ne permet pas d'expliquer les relations observées entre la masse monétaire, le niveau général des prix et le revenu monétaire. Pour Friedman, cette situation est expliquée par la tendance des keynésiens à se concentrer sur une classe réduite d'actifs financiers, de considérer que l'influence de la monnaie sur le revenu se fait uniquement à travers l'investissement, et qu'elle est, de plus, sensible à l'élasticité de celui-ci au taux d'intérêt.

Keynes est, selon Friedman (1971), enclin à considérer qu'une variation de l'offre de monnaie est davantage reflétée dans la préférence pour la liquidité que dans le revenu réel. Dans le cas extrême de son analyse, Keynes suppose que cet effet est complètement absorbée dans la préférence pour la liquidité et n'exerce aucun effet sur le revenu réel. Pour Friedman, le rejet par Keynes de la théorie quantitative de la monnaie s'explique par l'existence en situation de sous-emploi d'une trappe à liquidité qui rend la demande de monnaie infiniment élastique (Bordo et Schwartz, 2004). Dans ces conditions, lorsque le taux d'intérêt est suffisamment bas, une variation de l'offre réelle de monnaie, qu'elle soit le résultat d'une modification des prix ou d'un changement de la quantité nominale de monnaie, n'exerce aucun effet sur le revenu.

Friedman justifie le rôle de la préférence absolue pour la liquidité dans le raisonnement des keynésiens par la tendance de ceux-ci à accorder une grande importance à la forte sensibilité et à l'ajustement passif de la demande de monnaie au taux d'intérêt (Friedman, 1971). En revanche, le chef de file des monétaristes n'accorde pas d'attention, dans son analyse, aux effets du déplacement de la demande de monnaie sur le taux d'intérêt et minimise l'importance de celui-ci comme canal de transmission des variations de l'offre de monnaie vers la dépense (Friedman, 1972). Bien entendu, les auteurs de la synthèse, en particulier Tobin, Patinkin et Davidson, rejettent catégoriquement cette lecture monétariste de Keynes et rappellent à quel point l'analyse monétaire de Keynes est riche et plus générale que ce que prétend Friedman en l'associant à l'économie de la dépression (voir la section 1 ci-dessus).

Etant donné que le taux d'intérêt constitue le seul canal reliant la masse monétaire au revenu réel dans le modèle de la synthèse, et que le niveau général des prix est une donnée institutionnelle, une variation de l'offre de monnaie affecte soit le motif de spéculation soit le revenu réel, mais pas les prix (Friedman, 1971). C'est cette hypothèse de fixité des prix qui pousse les keynésiens à insister sur le rôle du taux d'intérêt dans la détention de la monnaie. Dans ce cas, la politique monétaire agit uniquement sur le taux d'intérêt, alors que chez les monétaristes elle agit sur l'inflation.

Friedman place l'hypothèse de fixité des prix au centre de la divergence entre les monétaristes et les keynésiens, et estime que cette hypothèse conduit ces derniers à confondre une variation nominale avec une variation réelle de l'offre de monnaie. A cet effet, il prétend que la préférence absolue pour la liquidité joue un rôle central dans le système keynésien car elle empêche la réalisation du plein-emploi même lorsque les prix sont flexibles (Friedman, 1972).

Dans le modèle de Hicks, la présence d'une trappe à liquidité dans la version keynésienne rend inefficace le recours à la politique monétaire pour remédier au chômage involontaire alors qu'une augmentation de l'offre de monnaie permet de stimuler l'emploi dans la version classique, retenue par les monétaristes. Ce dernier cas est, selon Friedman, compatible avec les résultats des études empiriques qui rendent compte de situations où une expansion monétaire exerce des effets réels.

Pour évaluer les effets de la politique monétaire, Friedman privilégie de s'appuyer sur un modèle qui relie directement la masse monétaire au revenu nominal qui a la même forme que

l'équation (2.2) mais avec un taux d'intérêt qui est donné par l'identité de Fischer. Ainsi, le taux d'intérêt nominal ne se modifie pas sous l'influence de la rencontre d'une offre de monnaie exogène avec la demande de monnaie, mais il réagit à la hausse de l'inflation générée par une augmentation de l'offre de monnaie (voir le chapitre 2 pour de plus amples détails).

2. Effet d'éviction et prééminence de la politique monétaire

Le concept de "effet d'éviction" est apparu dans le débat entre les monétaristes et les keynésiens suite aux résultats du modèle de la Réserve Fédérale de Saint Louis développé par Andersen et Jordan (Andersen et Jordan, 1968). A partir de l'estimation d'une équation reliant le revenu nominal aux valeurs présentes et passées de la masse monétaire et des dépenses gouvernementales, les auteurs montrent qu'une augmentation des dépenses publiques exerce un effet négligeable sur le revenu au bout d'un an. Depuis, la terminologie "effet d'éviction" est utilisée pour désigner la situation dans laquelle l'effet d'une hausse des dépenses publiques sur le revenu, nominal ou réel, est nul.

La question de l'efficacité supérieure de la politique monétaire sur la politique budgétaire constitue déjà un argument des monétaristes contre le modèle IS-LM depuis le travail économétrique initié par Friedman et Meiselman (1963). La nouveauté de l'étude de Andersen et Jordan c'est qu'elle introduit une dimension dynamique dans le débat à travers la distinction entre les effets de court terme et de long terme. Les auteurs reconnaissent l'effet positif de la politique budgétaire sur le revenu à court terme, mais ils lui dénie tout rôle à long terme. Dans le même temps, ils déclarent que l'action sur la masse monétaire influence significativement et durablement le revenu monétaire. Ces résultats de Andersen et Jordan ont fait l'objet de vives critiques économétriques et ont été à l'origine d'un intense débat sur les fondements théoriques de cet effet d'éviction empirique.

Pour justifier l'apparition de l'effet d'éviction, les monétaristes insistent sur la nécessité d'introduire la dimension dynamique dans le raisonnement théorique et de distinguer entre l'effet initial et l'effet final d'une expansion budgétaire. Dans cette perspective, l'effet initial positif d'une augmentation des dépenses publiques est compensé et complètement neutralisé par les effets négatifs qu'exerce sur le revenu la hausse du taux d'intérêt induite par le

financement du déficit budgétaire par émissions d'obligations. De fait, l'endettement du Gouvernement a tendance à réduire la capacité d'endettement du secteur privé et de fait à baisser les capacités productives de long terme de l'économie (Friedman, 1972). Les monétaristes pensent que seul un financement du déficit public par création monétaire est susceptible de préserver l'effet positif initial.

Blinder et Solow (1972) notent qu'il y a une absence de désaccord entre les keynésiens et les monétaristes quant aux conséquences expansionnistes d'un financement du déficit public par création monétaire. De même, les keynésiens reconnaissent l'existence d'une concurrence sur le marché financier exercée par l'Etat vis-à-vis du secteur privé pour financer sa dette, mais ils n'admettent pas que cette concurrence puisse conduire automatiquement à un "effet d'éviction" complet. Afin de restituer les arguments keynésiens quant aux effets d'une augmentation exogène des dépenses publiques nominales G sur le revenu réel Q , considérons la réécriture suivante de l'équation (5') présentée dans la première section ci-dessus :

$$Q = \frac{C}{p}(Q) + \frac{I}{p}(i) + \frac{G}{p}$$

En notant par c la propension à consommer, on pose :

$$\frac{C}{p}(Q) = c.Q$$

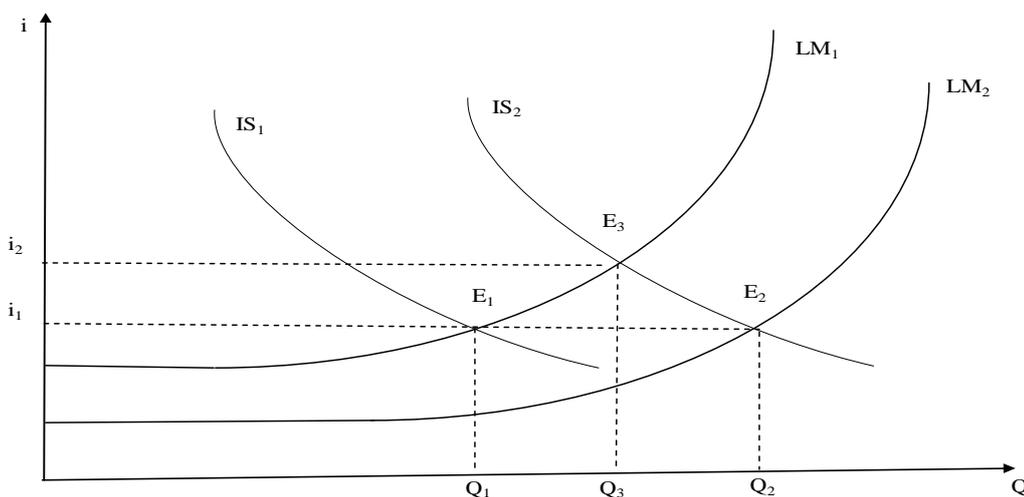
Pour un niveau donné du taux d'intérêt et du niveau général des prix, ces deux formules nous permettent de calculer le multiplicateur simple des dépenses budgétaires auquel font référence les monétaristes lorsqu'ils parlent d'effet initial :

$$\frac{dQ}{dG} = \frac{1}{1-c}$$

Etant donné que la propension à consommer est inférieure à 1, ce multiplicateur est supérieur à l'unité. Autrement dit, l'effet sur le revenu de l'augmentation des dépenses publiques est amplifié par la réaction de la consommation des ménages. Sur le graphique 4, on se déplace du point initial E_1 au point E_2 .

Toutefois, dans le cadre du modèle IS-LM, cette hausse du revenu induit un accroissement de la demande de monnaie pour motif de transaction et par conséquent une augmentation de la demande de monnaie globale. Le déséquilibre sur le marché de la monnaie qui en résulte peut être restauré de deux façons qui ont des conséquences différentes sur la valeur du multiplicateur initial. La première solution privilégie d'accompagner la hausse des dépenses publiques par une augmentation de la masse monétaire de façon à rétablir l'équilibre sur le marché de la monnaie et éviter une hausse du taux d'intérêt. Dans ce cas, la courbe LM_1 se déplace vers LM_2 et l'effet initial de l'expansion budgétaire est maintenu au point E_2 .

La seconde possibilité pour rétablir l'équilibre sur le marché monétaire consiste à maintenir l'offre de monnaie constante et autoriser une hausse du taux d'intérêt pour compenser la baisse de la demande de monnaie pour motif de spéculation. Ce deuxième mécanisme est interprété comme reflétant le recours de l'Etat au marché financier pour financer son déficit résultant de l'augmentation de ses dépenses. L'achat des obligations publiques émises par le Gouvernement entraîne une baisse de la demande de monnaie pour motif de spéculation et une hausse du taux d'intérêt, pour une offre de monnaie inchangée. L'augmentation du taux d'intérêt de i_1 à i_2 réduit le niveau de l'investissement et induit un fléchissement du revenu de Q_2 à Q_3 .



Graphique 4 : Politique budgétaire financée par emprunt ou création monétaire

L'objet du débat entre les monétaristes et les keynésiens sur l'occurrence de l'effet d'éviction total est de savoir dans quelles conditions le revenu Q_3 reste ou non supérieur au revenu initial Q_1 . Blinder et Solow (1972) considèrent que le débat sur l'effet d'éviction se ramène

finalement à un débat empirique sur les valeurs des élasticités de la demande de monnaie et de l'investissement au taux d'intérêt.

Pour mettre en évidence cette dernière proposition, on procède au calcul du multiplicateur global associé à l'augmentation des dépenses publiques à travers la différentiation simultanée des deux équations définissant l'équilibre global et l'équilibre sur le marché de la monnaie du modèle IS-LM. En reprenant les notations de la sous-section précédente, on aura ainsi pour une offre de monnaie constante :

$$dQ = c.dQ + I'(i).di + dG$$

$$0 = L_1'(i).di + p.L_2'(Q).dQ$$

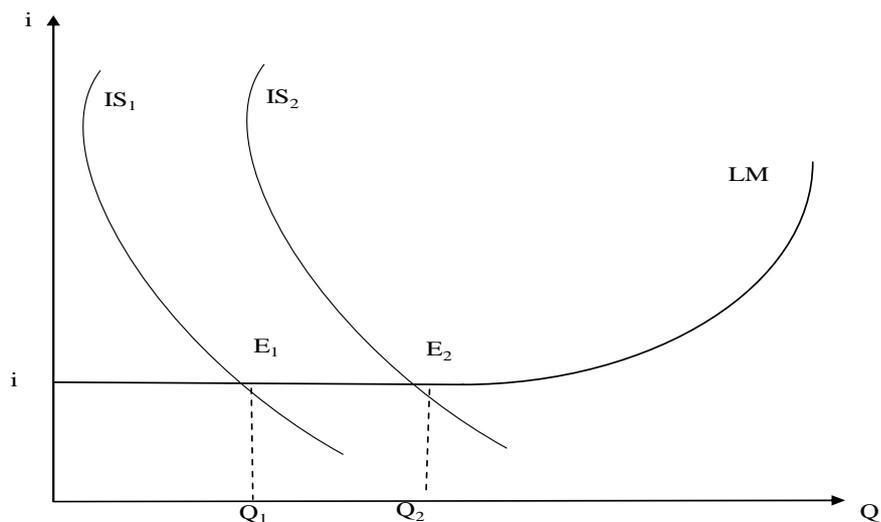
On tire à partir de la seconde équation une relation entre les différentielles du taux d'intérêt et de la production :

$$di = -p \frac{L_2'(Q)}{L_1'(i)} .dQ$$

On intègre cette expression de di dans la première équation et on réaménage les paramètres pour calculer la valeur du multiplicateur budgétaire :

$$\frac{dQ}{dG} = \frac{1}{(1-c) + p.I'(i) \frac{L_2'(Q)}{L_1'(i)}}$$

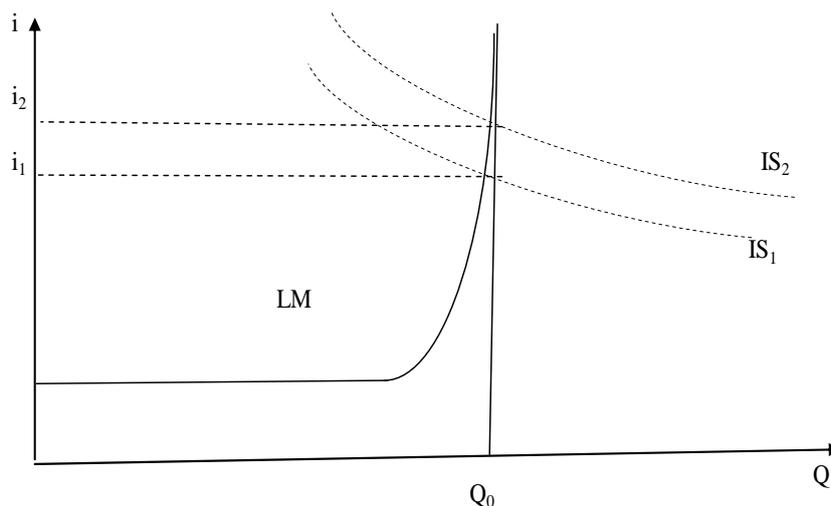
La valeur de ce multiplicateur, qui est toujours positive, est inférieure à celle du multiplicateur initial calculé ci-dessus. On peut retrouver la valeur de ce dernier dans deux situations particulières. La première correspond au cas où l'investissement est insensible au taux d'intérêt et la courbe IS est verticale. Dans la suite du raisonnement on exclut cette situation et on suppose que l'investissement varie en fonction du taux d'intérêt. La deuxième situation est celle où la courbe LM est horizontale, c'est-à-dire lorsque le taux d'intérêt ne bouge pas et seule le déplacement de la courbe IS qui détermine le niveau du revenu qui passe de Q_1 à Q_2 (voir graphique 5). Ce cas est assimilé à une efficacité maximale de la politique budgétaire et il est identifié par Friedman au cas du modèle keynésien.



Graphique 5 : Politique budgétaire avec trappe à liquidité

En revanche, au niveau de la partie verticale de la courbe LM, la demande de monnaie devient inélastique par rapport au taux d'intérêt et la valeur du multiplicateur budgétaire tend vers zéro. C'est uniquement dans ce cas que "l'effet d'éviction" est total et où la politique budgétaire n'exerce aucun effet sur le revenu.

C'est cette position que défend Tobin (1972) en soutenant que l'apparition de l'effet d'éviction s'explique par l'inélasticité de la demande de monnaie au taux d'intérêt, autrement dit par l'hypothèse d'une courbe LM verticale. Dans cette situation particulière du schéma IS-LM, un déplacement de la courbe IS, sous l'influence de la politique budgétaire ou d'un comportement exogène des fonctions de consommation et d'investissement, n'exerce aucun effet sur le revenu. Ainsi, une hausse des dépenses publiques conduit à un déplacement vers le haut de la courbe IS avec une augmentation du taux d'intérêt de i_1 à i_2 (voir graphique 6). La demande de monnaie étant insensible au taux d'intérêt, cette hausse du taux d'intérêt provoque une baisse de la demande privée d'un montant équivalent à l'augmentation des dépenses publiques de façon à ce que le niveau du revenu ne change pas, mais uniquement sa structure qui change.



Graphique 6 : Politique budgétaire avec demande de monnaie inélastique

L'effet d'une expansion budgétaire financée par emprunt est généralement présenté comme pouvant être décomposé sous la forme des deux phases qu'on vient de présenter, ce qui sous-entend la succession de deux équilibres. De même, en distinguant l'effet initial de l'effet final dans ce processus, Friedman (1972) tente d'introduire une séquentialité dans la transmission des effets de la politique budgétaire sur le revenu. Or, le modèle IS-LM offre un cadre d'analyse en statique comparative et n'intègre aucune dimension dynamique. Dans le graphique 4, le passage du point E_1 vers le point E_2 puis le retour au point E_3 ne décrit pas la transition entre des équilibres successifs mais uniquement une décomposition analytique pour faciliter la compréhension des mécanismes en œuvre. Rappelons que dans le modèle IS-LM le taux d'intérêt et le revenu sont déterminés de façon simultanée.

Dans la perspective de rendre compte des effets de long terme de la politique budgétaire et de la politique monétaire, deux modèles dynamiques s'appuyant sur le cadre IS-LM, ont été développés, d'une part, par Blinder et Solow (1972), et, d'autre part, par Buiter et Tobin (1976). Les deux articles montrent que la proposition monétariste d'effet d'éviction de long terme n'est pas compatible avec l'hypothèse de stabilité de l'économie telle que reflétée par le modèle IS-LM.

Conclusion

A partir d'un exposé du modèle de Klein et Goldberger (1955) et des versions monétaire et réelle du modèle de la synthèse, ce chapitre a mis en évidence la présence d'un important souci de validation empirique dans la modélisation économétrique keynésienne. S'appuyant sur ces divergences, Friedman tente de déplacer le débat avec les auteurs de la synthèse sur le rôle de la politique monétaire vers le terrain empirique. De même, les principes de base du cas keynésien du modèle IS-LM sont fortement remis en cause par les monétaristes.

Les monétaristes placent la diversité de la gamme d'actifs qu'ils retiennent dans leur analyse à l'origine de leur divergence avec les keynésiens à propos du canal de transmission de la politique monétaire. Pour ces derniers, c'est plutôt la nature monétaire du taux d'intérêt contre le mode fishérien de sa détermination qui est à la base de cet écart. De même, les keynésiens rejettent la réduction par les monétaristes de l'apport de Keynes au cas de la trappe à liquidité. Pour Tobin (1972), c'est l'adoption par les monétaristes d'une demande de monnaie inélastique par rapport au taux d'intérêt qui explique leur argument d'effet d'éviction. On montre dans le chapitre 2, que cette inélasticité est la base de la proposition centrale du monétarisme reliant directement le revenu nominal à la masse monétaire.

Chapitre 2

Les Limites Théoriques de la Politique Monétaire Monétariste

Introduction

Jusqu'en 1968, Friedman n'a pas réussi à développer une théorie complète qui décrit l'influence de la monnaie sur la production réelle et les prix. Jusqu'alors, ses contributions étaient surtout empiriques et avaient porté sur la nature et l'ampleur de la relation entre la masse monétaire et le revenu nominal (Desai, 1981). Les mécanismes de transmission de la politique monétaire vers l'activité réelle n'étaient pas encore rigoureusement identifiés.

Les arguments sur lesquels s'appuient les monétaristes pour recommander l'adoption d'une règle de politique monétaire qui consiste à fixer la croissance de la masse monétaire entre 3% et 5% concernent la stabilité de la demande de monnaie friedmanienne (Friedman, 1956) et l'inférence d'une influence causale de la masse monétaire vers le revenu nominal à partir de l'observation de décalages temporels entre ces deux agrégats (Friedman, 1958 ; Friedman et Schwartz, 1963 a). C'est seulement à l'occasion de son discours de 1967 devant '*l'American Economic Association*' que Friedman présente sa nouvelle lecture de la courbe de Phillips, ce qui lui permet de fournir une théorie pour la décomposition de la variation du revenu nominal entre une variation de la production et une variation du niveau général des prix (Desai, 1981).

La demande de monnaie friedmanienne porte sur les déterminants du choix de la composition optimale d'un portefeuille formé d'une large gamme d'actifs. Friedman (1956) interprète sa fonction de demande de monnaie comme une reformulation de la théorie quantitative (Friedman, 1968, 1970), mais, cette filiation quantitativiste est fortement remise en cause par les auteurs de la synthèse. La question centrale de ce débat porte sur l'influence du taux d'intérêt sur la demande de monnaie, considérée par Patinkin comme '*l'une des questions centrales de l'économie monétaire*' (1972, p. 886). En particulier, Patinkin (1969) conçoit la demande de monnaie de Friedman comme une exposition '*élégante et sophistiquée*' de la théorie keynésienne de la préférence pour la liquidité, avec l'hypothèse empirique supplémentaire d'une relative inélasticité par rapport au taux d'intérêt.

Friedman (1970) utilise cette fonction de la demande de monnaie, moyennant plusieurs hypothèses empiriques, pour élaborer une équation reliant le revenu nominal aux valeurs présentes et passées de la masse monétaire. De plus, il présente un modèle, vivement critiqué

par Tobin (1972), pour expliquer la dynamique de la décomposition du revenu nominal entre la production et les prix le long d'un équilibre walrasien.

La différence entre le comportement procyclique de la vitesse-revenu à l'intérieur des cycles et son évolution contracyclique à très long terme est expliquée par l'écart entre les valeurs permanentes et mesurées du revenu et du prix (Friedman, 1959). Toutefois, Friedman reconnaît l'échec de ce modèle à rendre compte de ce qu'il qualifie comme l'un des résultats les plus importants de ses investigations statistiques, à savoir que la masse monétaire tend à devancer et à diriger le revenu nominal à la fois aux sommets et aux creux du cycle. Les keynésiens rejettent l'utilisation de corrélation statistique comme preuve de la causalité et insistent sur l'importance de l'*a priori* théorique (Kareken et Solow, 1963 ; Tobin, 1970). Tobin (1970) développe, à cet effet, deux modèles théoriques, l'un keynésien et le second monétariste, qui tous les deux génèrent une structure de décalages temporels. Dans le modèle keynésien, la monnaie est créée de façon endogène et n'exerce aucune influence causale sur le revenu.

La seconde critique théorique fondamentale adressée par les monétaristes à l'encontre de la modélisation structurelle¹ porte sur le manque de fondement théorique pour la détermination du prix et du salaire. De plus, Friedman (1971) estime que la courbe de Phillips est construite de façon *ad hoc* et souffre de plusieurs limites théoriques. La reformulation monétariste de cette courbe la conçoit comme l'expression d'un processus de négociation salariale reflétant le comportement de l'offre de travail (Friedman, 1968). La nouvelle courbe montre qu'il existe bien un arbitrage à court terme entre l'inflation et le chômage, mais qu'à long terme ce dernier retourne à son niveau naturel avec la persistance d'une accélération de l'inflation. Ainsi, cette lecture permet à Friedman d'établir un lien entre sa théorie de court terme du revenu monétaire et la dichotomie classique de long terme entre les variables monétaires et les variables réelles (Desai, 1981). De même, on retrouve les mêmes recommandations de politique monétaire que celles formulées par Friedman à la fin des années 1950, à savoir une règle de croissance constante de la masse monétaire pour assurer l'objectif ultime de la stabilité des prix.

¹ La première critique concerne l'assimilation du modèle keynésien au cas de la trappe à liquidité (voir chapitre 1).

Section 1 : Demande de monnaie et composition du portefeuille

La demande de monnaie friedmanienne est dérivée à partir de la théorie du choix de consommateur. Friedman (1956) rattache cette fonction aux auteurs quantitativistes qui supposent que les variations de la quantité de monnaie affectent d'abord les prix. Cette filiation quantitativiste est, toutefois, fortement remise en cause par les auteurs de la synthèse. Friedman (1971, 1972) utilise cette fonction de la demande de monnaie pour construire un modèle de court terme du revenu nominal. De plus, il développe un modèle, vivement critiqué par Tobin (1972), pour la décomposition de la dynamique du revenu nominal entre la production et les prix.

Sous-section 1 : Demande de monnaie monétariste et préférence pour la liquidité

1. Théorie quantitative et demande de monnaie

La position de Friedman vis-à-vis de l'influence de la théorie keynésienne de la préférence pour la liquidité sur ses travaux a beaucoup évolué entre 1956 et la fin des années 1960 (Patinkin, 1972). En 1956, Friedman inscrit, sans aucune référence aux travaux de Keynes, son cadre analytique de la fonction de demande de monnaie et sa vision de la théorie quantitative de la monnaie dans la lignée des travaux des auteurs de l'école de Chicago durant les années 1930 et 1940, tels que Simons (1948), Mints (1945), Knight (1941) et Viner (1937). Dans ses travaux ultérieurs, Friedman reconnaît que sa présentation de la fonction de la demande de monnaie fût influencée par l'analyse keynésienne de la préférence pour la liquidité, mais tout en continuant à considérer cette fonction comme une '*reformulation de la théorie quantitative*' (Friedman, 1956).

Dans son article fondateur de 1956, il présente la théorie quantitative comme une théorie de demande de la monnaie et non pas une théorie de la production, du revenu monétaire ou du niveau des prix (Friedman, 1956). Au niveau macroéconomique, la demande de monnaie est envisagée comme une agrégation des demandes de monnaie des ménages et des entreprises. La monnaie constitue pour un ménage un actif comme les autres, un moyen de détenir de la richesse. Pour les entreprises, la monnaie représente un bien capital, un actif facteur de

production. La théorie de la demande de monnaie est ainsi interprétée comme un cas particulier de la théorie du capital.

L'analyse de la demande de monnaie des ménages s'effectue suivant le modèle de la théorie du choix du consommateur, et elle admet trois principaux déterminants ; a/ la richesse totale, b/ le prix et le rendement de la monnaie et des autres types d'actifs constituant cette richesse, c/ les goûts et les préférences des ménages. En plus de la monnaie, les autres formes de détention de la richesse considérées par Friedman concernent les obligations, les actions, les biens physiques et le capital humain. Toutes les formes de richesse sont exprimées en termes monétaires aux prix en vigueur.

La richesse totale W joue le rôle de la contrainte budgétaire dans la théorie du choix du consommateur. Elle est appréhendée par le flux total de revenus Y qu'elle procure moyennant un taux d'intérêt r , soit $Y = rW$. Le revenu dont il s'agit ici ne doit pas être identifié au revenu mesuré qui est sujet à de fortes fluctuations, mais plutôt du revenu permanent qu'il a introduit dans l'étude sur la consommation des ménages (Friedman, 1957 ; Friedman, 1970).

Chaque ménage cherche à répartir sa richesse totale sous ses différentes formes de façon à en maximiser l'utilité. Le choix de détenir une forme particulière de richesse plutôt qu'une autre entraîne une différence dans le flux de revenus reçus, et cette différence est, pour Friedman, fondamentale pour l'appréciation de "l'utilité" d'une structure particulière de la richesse. Ainsi, Friedman retient, pour la description des formes alternatives d'actifs qui composent la richesse d'un individu, non seulement les prix de marché de ces actifs, mais également les flux de revenus qu'ils engendrent.

Friedman assimile l'analyse sur les taux de rendement attendus de la monnaie et des autres actifs à celle, qui est effectuée dans la théorie du consommateur, sur les prix d'un produit et ses substituts. Le taux de rendement nominal de la monnaie sous forme fiduciaire est nul alors qu'il est positif pour les dépôts générant des intérêts (Friedman, 1970). Il suppose que la monnaie peut produire un rendement en nature, sous forme de commodité ou de sécurité. En termes "réels", le niveau de ce rendement pour une unité de monnaie dépend du volume de marchandises qu'elle peut acheter, ou encore du niveau général des prix P (Friedman, 1956).

Le taux de rendement nominal sur les autres actifs est constitué de deux éléments ; le revenu généré et la variation du prix nominal. En notant par r_o et r_a , respectivement, le taux d'intérêt nominal attendu sur une obligation et le taux d'intérêt nominal attendu sur une action, Friedman montre que le flux des revenus nominaux généré par la détention de chacun de ces actifs financiers peut être approximé par :

$$r_o - \frac{1}{r_o} \frac{dr_o}{dt}$$

$$r_a + \frac{1}{P} \frac{dP}{dt} - \frac{1}{r_a} \frac{dr_a}{dt}$$

L'apparition de l'inflation attendue dans la seconde formule s'explique par le fait que le flux de revenus fournis par les actions est censé être fixé en unités réelles constantes, à l'opposé de celui produit par les obligations qui s'exprime en unités nominales constantes. La conversion en unités nominales du revenu réel d'une action s'obtient par la multiplication de ce dernier par le différentiel, par rapport à la période de référence, du niveau général des prix P . Le terme en taux de croissance du taux d'intérêt dans les deux formules mesure l'effet de la variation du prix du titre considéré sur le revenu qu'il génère.

La possession de biens physiques fournit pour les ménages des rendements en nature dont la valeur dépend du comportement des prix (Friedman, 1956). Ces biens produisent également un rendement monétaire correspondant à l'appréciation ou la dépréciation attendue de leur valeur et qui est ainsi fonction de la variation anticipée du niveau général des prix :

$$\frac{1}{P} \frac{dP}{dt}$$

En raison de l'absence de prix de marché pour le capital humain, il est difficile de définir une valeur pour ce dernier ou sa substitution avec les autres formes de capital. Pour tenir compte de l'influence du capital humain sur la demande de monnaie d'un individu, Friedman (1956) retient la variable w mesurant la part du capital non humain par rapport au capital humain et qu'il considère comme constante dans le temps.

Les goûts et les préférences des ménages pour les flux de services provenant de leurs différentes formes de richesse influencent la demande de monnaie. Afin de donner à la théorie un contenu empirique, Friedman (1956) suppose que ces préférences sont constantes.

Toutefois, dans certaines circonstances, lorsque par exemple un individu se déplace géographiquement ou fait face à une incertitude inhabituelle, sa préférence pour la détention d'une plus grande partie de sa richesse sous forme monétaire s'accroît. Les variables qui affectent les goûts et les préférences des individus sont regroupées dans le terme u .

La combinaison de toutes les variables présentées ci-dessus permet de formuler la fonction théorique friedmanienne de la demande de monnaie M :

$$M = f\left(\frac{Y}{r}, P, r_o - \frac{1}{r_o} \frac{dr_o}{dt}, r_a + \frac{1}{P} \frac{dP}{dt} - \frac{1}{r_a} \frac{dr_a}{dt}, \frac{1}{P} \frac{dP}{dt}, w, u\right)$$

Cette forme de la demande de monnaie est simplifiée moyennant quelques hypothèses. En particulier, les taux d'intérêt r_o et r_a sont supposés constants dans le temps et l'effet du taux d'intérêt r est supposé pris en compte dans ces derniers. La fonction de demande de monnaie se réécrit alors sous la forme suivante :

$$M = f\left(Y, P, r_o, r_a, \frac{1}{P} \frac{dP}{dt}, w, u\right) \quad (1)$$

Cette forme de la fonction de demande de monnaie est établie dans le cas d'un ménage. Or, comme mentionné *supra*, les entreprises détiennent également la monnaie comme une ressource productive. La quantité de monnaie détenue par les entreprises dépend du coût des services productifs rendus par la monnaie, du coût des services productifs substituables à la monnaie et la valeur du produit fourni par le service productif. Les entreprises ne subissent, toutefois, pas de contrainte comparable à celle des ménages pour la composition de leur richesse totale (Friedman, 1970). La possibilité qu'ont les entreprises d'acquérir plus de capital sur le marché des capitaux leur permet de moduler le volume total de capital intégré dans leurs actifs productifs, y compris la monnaie, de façon à maximiser leur profit. Friedman considère que les variables retenues au niveau de la demande de monnaie pour un ménage, à l'exception de celle relative à la richesse, constituent de bonnes approximations pour expliquer également la demande de monnaie pour une entreprise.

La fonction de demande de monnaie pour l'ensemble de l'économie s'obtient par agrégation de celles de la forme (1) qui ont été élaborées dans le cas d'un ménage et d'une entreprise (Friedman, 1956). Friedman reconnaît la difficulté de passer, par agrégation, de ces fonctions

de demande de monnaie microéconomique à une fonction de demande de monnaie macroéconomique. L'agrégation ne pose pas de problèmes pour le niveau général des prix ni pour les taux d'intérêt qui peuvent être considérés comme semblables pour l'ensemble des ménages et toutes les entreprises. La difficulté d'agrégation concerne principalement l'inflation anticipée, le revenu et la richesse, variables qui diffèrent d'un agent à un autre. Toutefois, ces problèmes sont négligés et l'équation (1) est considérée comme étant valable au niveau macroéconomique.

En supposant cette fonction de demande de monnaie comme étant dérivée du principe de maximisation d'une fonction d'utilité exprimée en termes réels, Friedman considère qu'elle doit être indépendante de tout changement de l'unité monétaire utilisée pour mesurer les variables monétaires. Pour cela, il introduit une hypothèse fondamentale, à savoir l'homogénéité de degré un de la fonction M en Y et P :

$$f(\lambda \times Y, \lambda \times P, r_o, r_a, \frac{1}{P} \frac{dP}{dt}, w, u) = \lambda \times f(Y, P, r_o, r_a, \frac{1}{P} \frac{dP}{dt}, w, u)$$

Le choix de valeurs particulières pour le scalaire λ permet à Friedman de réécrire sa fonction de demande de monnaie sous deux formes particulières déjà connues. D'une part, si $\lambda = \frac{1}{P}$ alors l'équation (1) exprime une demande d'encaisses réelles :

$$\frac{M}{P} = f\left(\frac{Y}{P}, r_o, r_a, \frac{1}{P} \frac{dP}{dt}, w, u\right) \quad (2)$$

D'autre part, en posant $\lambda = \frac{1}{Y}$, Friedman retrouve l'équation quantitative usuelle :

$$Y = V\left(\frac{Y}{P}, r_o, r_a, \frac{1}{P} \frac{dP}{dt}, w, u\right) \times M \quad (3)$$

Cette équation est interprétée par Friedman comme une *''reformulation de la théorie quantitative de la monnaie''* (Friedman, 1956). L'argument avancé par Friedman pour justifier cette filiation quantitativiste est que *''Fisher et les autres auteurs de la théorie quantitative ont explicitement reconnu que la vitesse de circulation de la monnaie serait affectée, entre autres facteurs, par le taux d'intérêt ainsi que le taux de variation des prix''*

(in Patinkin, 1969, p. 14). Cette filiation est fortement remise en cause par les économistes keynésiens.

2. *L'influence keynésienne*

Patinkin (1969) remet en cause la validité de l'interprétation que fait Friedman de la théorie quantitative de la monnaie et de la filiation de sa fonction de demande de monnaie à cette théorie. Il conclut que l'analyse de la demande de monnaie effectuée par Friedman n'est qu'une exposition "*élégante et sophistiquée*" de la théorie monétaire keynésienne moderne. En effet, la fonction de demande de monnaie développée par Friedman (1956, 1970) s'appuie sur l'analyse des déterminants du choix de la composition optimale d'un portefeuille constitué d'une classe élargie d'actifs. Cette fonction de demande de monnaie est explicitement décrite par Friedman (1970) comme "*faisant partie d'une théorie du capital ou de la richesse, concernée par la composition du bilan ou du portefeuille d'actifs*".

Le principal débat concernant cette filiation oppose Friedman à Patinkin et porte sur la justification théorique de l'influence du taux d'intérêt sur la demande de monnaie, influence que Patinkin (1972, p. 886) considère comme "*l'une des questions centrales de l'économie monétaire*". A ce titre, ce dernier estime que la fonction de demande de monnaie constitue la caractéristique centrale du cadre théorique développé par Friedman dans son article de 1970, mais que cette fonction doit être reliée à la théorie de la préférence pour la liquidité plutôt qu'à la "tradition de l'école de Chicago". Plus généralement, Patinkin critique "*l'interprétation erronée*" que fournit Friedman de la théorie quantitative pour défendre sa thèse fondamentale qui consiste à relier son cadre théorique à cette théorie.

En effet, pour Patinkin, l'inclusion du taux d'intérêt dans la fonction de demande de monnaie est bien une caractéristique de la théorie keynésienne. Certes, les théoriciens quantitativistes reconnaissent l'influence du taux d'intérêt sur la vitesse de circulation de la monnaie, mais ils ne l'intègrent pas dans leur cadre d'analyse et ne tiennent pas compte dans leurs travaux des implications de cette influence (Patinkin, 1969, 1972). De même, les quantitativistes accordent peu d'importance aux effets sur le taux d'intérêt résultant des changements dans les préférences des individus quant à la forme sous laquelle ils souhaitent détenir leurs actifs.

Par ailleurs, l'absence du taux d'intérêt des études empiriques des quantitativistes est encore plus révélatrice du manque d'intérêt qu'ils accordent à cette variable dans leur système de pensée. En effet, de par leur nature, les études empiriques exigent de dresser la liste de toutes les variables théoriques potentiellement susceptibles d'expliquer les données disponibles, même si certaines de ces variables pourraient être par la suite rejetées comme étant statistiquement non significatives. Il est également intéressant de noter à cet égard, que la première étude empirique qui traite explicitement de l'influence du taux d'intérêt sur la demande de monnaie est d'inspiration keynésienne et a été réalisée par Brown (1940).

En plus de la question du taux d'intérêt, Patinkin relève une seconde différence fondamentale qui sépare le cadre théorique de Friedman de celui de l'école de Chicago des années 1930 et 1940, représentée notamment par Henry Simons. En effet, cette école n'est pas particulièrement intéressée par les propriétés de la fonction de demande de monnaie et ses études sont menées dans les termes de l'équation traditionnelle de Fisher, $MV = PT$, pour expliquer le niveau observé des prix. D'ailleurs, H. Simons n'a jamais utilisé le concept de fonction de demande de monnaie. Par ailleurs, l'hypothèse de base de l'école de Chicago consiste à considérer que c'est l'instabilité de la vitesse de circulation de la monnaie qui est à l'origine de l'instabilité de l'économie, ce qui contraste fortement avec l'approche de Friedman.

Dans le même sens, la théorie quantitative et l'école de Cambridge sont principalement concernées par la relation directe entre le stock de monnaie et le flux des dépenses en biens et services. En revanche, la théorie de la demande de monnaie présentée par Friedman s'intéresse à la relation entre le stock de monnaie et le stock des autres actifs, exactement comme dans le cas de la théorie keynésienne de la préférence pour la liquidité.

Pour justifier la pertinence de son interprétation de la théorie quantitative en termes de demande de monnaie, Friedman présente la version revenu de cette théorie, $MV = Py$ où y est le revenu réel, comme une évolution dans ce sens de l'équation de Fisher. Pour cela, Friedman (1970) considère que ces deux versions de la théorie quantitative renferment des conceptions différentes sur le rôle de la monnaie. Celle de Fisher devrait insister sur le transfert de la monnaie alors que pour la seconde version, le plus important serait la détention de la monnaie. Cette différence se serait accentuée pour la version cambridgienne des encaisses.

Toutefois, la littérature de l'entre-deux guerres sur la théorie quantitative révèle que la version revenu a été utilisée uniquement comme une variante de l'équation des transactions (Patinkin, 1972). De plus, la conception du rôle de la monnaie n'a pas beaucoup évolué entre les deux approches qui insistent sur la propriété du transfert de la monnaie plutôt que sur le motif de sa détention. Le recours durant cette époque à la version revenu se justifie uniquement par des considérations de disponibilité de l'information ; l'utilisation du volume de la production et de son prix est jugée plus significative que celle du nombre de transaction et de leur prix.

Lors de sa réplique aux critiques de Patinkin, Friedman (1972) persiste et réaffirme que son article de 1956 constitue bel et bien une reformulation de la théorie quantitative de la monnaie et qu'il n'est nullement keynésien. De même, il soutient (Friedman, 1970) que le cadre théorique qui sous-tend ses travaux empiriques est celui de la théorie quantitative de la monnaie.

Friedman (1972) affirme qu'il a tendance à minimiser l'importance des variations des taux d'intérêt comme principal canal par lequel les variations de la quantité de monnaie affecte la dépense, la production et les prix. Il refuse d'admettre que l'inclusion du taux d'intérêt dans la fonction de demande de monnaie puisse constituer une caractéristique de la théorie keynésienne, et il pense que la prise en compte du taux d'intérêt est une caractéristique mineure de son cadre théorique. Friedman (1972,) considère aussi que le fait d'inclure ou d'exclure le taux d'intérêt de la fonction de demande de monnaie n'est pas la raison qui relie son travail à la "tradition de Chicago". Cette tradition a surtout influencé son travail pour l'interprétation des fluctuations de court terme et le rôle des politiques monétaire et budgétaire.

La position de Friedman vis-à-vis du rôle que joue le taux d'intérêt dans ses études pousse Patinkin à penser que le cadre conceptuel utilisé par Friedman pour analyser la demande de monnaie est celui de la théorie keynésienne de la préférence pour la liquidité, avec l'hypothèse empirique supplémentaire d'une relative inélasticité par rapport au taux d'intérêt (Patinkin, 1972). A l'opposé, Friedman (1970) cherche à présenter la forte sensibilité de la demande de monnaie au taux d'intérêt (élasticité infinie) comme étant la différence fondamentale qui distingue son cadre théorique de celui des keynésiens.

Mais, Friedman (1972) relie l'importance qu'accordent les keynésiens aux effets du taux d'intérêt à leur hypothèse sur la fixité des prix, considérés comme une donnée institutionnelle. En raison de cette hypothèse, les keynésiens interprètent le taux d'intérêt comme le déterminant de la quantité plus ou moins grande de monnaie que les individus souhaitent détenir. Par contre, les quantitativistes supposent que les variations de la quantité de monnaie affectent d'abord les prix, et ils envisagent la formation du taux d'intérêt comme étant influencée par le comportement de l'épargne et de l'investissement ou encore par les opérations de prêt et d'emprunt. C'est en produisant de l'inflation (ou de la déflation) que les variations de la quantité de monnaie influencent le taux d'intérêt. Le processus inflation/déflation déplace les fonctions d'investissement et d'épargne en poussant les prêteurs à réclamer des taux d'intérêt nominaux plus élevés ou plus bas et les emprunteurs à payer des taux d'intérêt nominaux plus élevés ou plus bas.

Sous-section 2 : Un modèle dynamique du revenu

1. Demande de monnaie et relation de Fisher

Tout au long de ses débats avec les keynésiens, Friedman rappelle et insiste toujours sur le fait qu'il utilise la théorie quantitative de la monnaie pour dériver une théorie de court terme du revenu nominal et non pas une théorie des prix ou du revenu réel. Ainsi, le modèle du revenu nominal développé par Friedman (1970, 1971) cherche à établir une relation directe entre le revenu nominal et le stock de monnaie. Il adopte, à cet effet, une fonction de demande de monnaie, exprimée sous la forme d'encaisses réelles, dépendant du revenu réel et du taux d'intérêt nominal et qu'il considère comme une simplification de l'équation (2) :

$$\frac{M^d}{P} = L\left(\frac{Y}{P}, r\right) \quad (4)$$

avec Y le revenu nominal, M^d la demande de monnaie nominale, P le niveau général des prix et r le taux d'intérêt nominal.

L'élasticité de la demande réelle de monnaie par rapport au revenu réel est supposée égale à l'unité. Cette hypothèse est introduite ici, selon Friedman, pour éliminer le niveau général des

prix et le revenu réel des équations du bloc monétaire de son modèle. Il s'agit là d'une hypothèse qui est basée sur des considérations empiriques et non théoriques et dont la pertinence est justifiée par la capacité du modèle à prévoir le comportement du revenu nominal. Dans ces conditions, l'équation (4) se réécrit sous la forme suivante :

$$M^d = Y.I(r) \quad (5)$$

En considérant l'offre de monnaie M comme étant exogène, déterminée par les autorités monétaires, la demande de monnaie s'ajuste à tout instant à l'offre de sorte à ce que l'équilibre sur le marché de la monnaie soit toujours réalisé. En introduisant le temps, l'équation (5) se transforme sous la forme de l'équation quantitative en termes de revenu nominal :

$$Y_t = V(r).M_t \quad (6)$$

où $V(r)$ désigne la vitesse de circulation de la monnaie. Toutefois, la connaissance du stock de monnaie M_t dans l'équation (6) ne suffit pas à elle seule pour déterminer le revenu nominal Y . Il faut de plus spécifier le comportement du taux d'intérêt nominal r dans la vitesse de circulation de la monnaie V .

Pour transformer l'équation (6) en une relation directe entre la masse monétaire et le revenu nominal, Friedman introduit une nouvelle relation, qu'il qualifie d'équation manquante, pour la détermination du taux d'intérêt nominal r . Il s'appuie pour cela sur ce qu'il considère comme étant les apports de Keynes et de Fisher relatifs à la détermination du taux d'intérêt nominal. L'idée keynésienne reprise par Friedman est celle qui consiste à supposer que le taux d'intérêt r est déterminé par le taux d'intérêt anticipé de long terme r^* . Friedman considère le cas limite de cette hypothèse, à savoir :

$$r = r^* \quad (7)$$

La contribution de Fisher exploitée par Friedman concerne la relation qui relie le taux d'intérêt nominal au taux d'intérêt réel, écrite pour des valeurs anticipées ou permanentes :

$$r^* = \rho^* + \left(\frac{1}{P} \frac{dP}{dt}\right)^* \quad (8)$$

où r^* et ρ^* désignent, respectivement, le taux d'intérêt nominal anticipé et le taux d'intérêt réel anticipé, et $(\frac{1}{P} \frac{dP}{dt})^*$ représente la variation anticipée du niveau général des prix. La combinaison des relations (7) et (8) donne l'équation du taux d'intérêt suivante :

$$r = \rho^* + (\frac{1}{P} \frac{dP}{dt})^* \quad (9)$$

Pour éviter d'introduire là encore le raisonnement sur le niveau général des prix, Friedman remplace le taux de croissance de celui-ci par l'écart entre la croissance du revenu nominal et celle du revenu réel. Cette décomposition permet de transformer la relation (9) sous la forme suivante :

$$r = \rho^* - g^* + (\frac{1}{Y} \frac{dY}{dt})^* \quad (10)$$

où $g^* = (\frac{1}{y} \frac{dy}{dt})^*$ représente le taux de croissance du revenu réel anticipé ou permanent.

Friedman introduit ici une seconde hypothèse empirique qui porte cette fois-ci sur la constance de l'écart entre le taux d'intérêt réel anticipé et le taux de croissance anticipé du revenu réel, noté k_0 :

$$k_0 = \rho^* - g^* \quad (11)$$

Cette hypothèse permet à Friedman d'éliminer l'effet d'une variable réelle sur le bloc monétaire de son modèle et d'expliquer les mouvements du taux d'intérêt nominal par ceux de l'inflation anticipée. La justification de la constance de k_0 , ou de ses faibles variations devant celles de $(\frac{1}{P} \frac{dP}{dt})^*$, s'appuie essentiellement sur des arguments empiriques. D'abord l'étude de Gupta (1964) qui établit la similarité des taux d'intérêt nominaux sur les périodes caractérisées par des comportements semblables des prix. Ensuite, le taux d'intérêt réel estimé par la Réserve Fédérale de Saint-Louis est très stable malgré les changements importants du taux d'intérêt nominal. Enfin, la croissance réelle moyenne calculée sur différentes période et pour plusieurs pays est plutôt constante, même si elle diffère d'un pays à un autre.

Le modèle monétaire proposé par Friedman pour l'analyse des fluctuations de court terme est composé du système des deux équations (12) et (13) suivantes :

$$Y_t = V(r).M_t \quad (12)$$

$$r = k_0 + \left(\frac{1}{Y} \frac{dY}{dt}\right)^* \quad (13)$$

Dans son article de 1970, Friedman explique le taux de croissance anticipé du revenu nominal par les valeurs passées du niveau du revenu nominal. Or, il semble, en principe, plus naturel de relier le taux de croissance anticipé du revenu nominal aux valeurs passées du taux de croissance du revenu nominal plutôt qu'à celles de son niveau. C'est, en effet, cette seconde approche qu'adopte Friedman dans son livre co-écrit avec A. Schwartz et publié en 1982 '*Monetary Trends in the United States and the United Kingdom*'. Les auteurs retiennent une hypothèse d'anticipations adaptatives qui exprime le taux de croissance anticipé du revenu nominal \dot{Y}_t^* par une moyenne pondérée du taux de croissance du revenu nominal courant \dot{Y}_t et passé \dot{Y}_{t-1} :

$$\dot{Y}_t^* = w.\dot{Y}_t + (1-w).\dot{Y}_{t-1} \quad (14)$$

Toutefois, les auteurs retiennent une écriture du taux de croissance du revenu nominal qui prouve la forte influence des considérations empiriques sur le cadre théorique de Friedman :

$$\dot{Y}_t = (\log(Y_t) - \log(Y_{t-1})) / n_t \quad (15)$$

où n_t est la longueur moyenne de la période séparant les milieux de deux phases successives du cycle. De ce fait, le taux de croissance anticipé du revenu nominal se retrouve déterminer par les valeurs passées du niveau du revenu nominal. Dans ces conditions, la valeur du taux d'intérêt fournit à chaque instant par l'équation (13) est une grandeur prédéfinie en fonction des valeurs présentes et passées du revenu nominal et l'équation (12) se réécrit :

$$Y_t = V_{k_0}(Y_{t-j}).M_t, \quad j = 1, \dots, t \quad (16)$$

Selon cette dernière équation, les valeurs passées du revenu nominal sont une fonction des valeurs passées du stock de monnaie. On peut alors exprimer le revenu nominal uniquement en fonction des valeurs présentes et passées du stock de monnaie supposé exogène :

$$Y_t = f_{k_0}(M_{t-j}).M_t, \quad j = 1, \dots, t \quad (17)$$

En raison de la dépendance de la vitesse V et de la fonction f du paramètre k_0 , Friedman interprète les équations (16) et (17) comme représentant le mouvement du revenu nominal autour d'une tendance de long terme au cours de laquelle k_0 et ses composantes ρ^* et g^* s'ajustent à des forces de long terme représentées par les changements dans les ressources et la technologie disponibles.

En termes du schéma IS-LM, le modèle du revenu monétaire de Friedman ne fait que refléter le "cas classique" du modèle IS-LM (Tobin, 1972; Patinkin, 1972). Rappelons que cette classification de Hicks correspond au cas où la courbe LM est verticale, autrement dit, au cas où la demande de monnaie est indépendante du taux d'intérêt. Dans ces conditions, seule une augmentation de la masse monétaire permet une hausse du revenu nominal ou encore, une augmentation de l'offre réelle de monnaie entraîne une hausse du revenu réel. Le déplacement de la courbe IS, quel qu'en soit l'origine, n'exerce aucun effet sur le revenu (Tobin, 1972). Mais, Friedman considère l'hypothèse de constance du taux d'intérêt réel dans la demande de monnaie comme étant complètement satisfaisante. Celui-ci constitue dans ce cas au mieux une variable de support et c'est l'inflation qui compte.

Tobin (1972) critique l'hypothèse de constance du taux d'intérêt réel retenue par Friedman pour construire son modèle du revenu monétaire. Le taux d'intérêt réel constitue, en effet, une variable fondamentale pour la détermination de l'investissement à court terme. Tobin souligne l'incohérence de l'analyse de Friedman qui suppose que le taux d'intérêt nominal varie sous l'influence de l'inflation, mais en même temps néglige la réaction de la demande de monnaie aux effets du taux d'intérêt.

2. Le problème de la dynamique du revenu

Friedman estime que l'interprétation de la différence des vitesses d'ajustement des prix et de la production se trouve au cœur des divergences entre les keynésiens et les non-keynésiens. Il considère que la théorie quantitative ne spécifie pas le sentier de la dynamique de cet ajustement, autrement dit, cette théorie ne précise pas si une augmentation de la masse monétaire se transmet aux prix ou à la production ou encore à la vitesse de circulation. Par

contre, les keynésiens supposent que l'ajustement se fait par la production et la vitesse de circulation, mais pas par les prix qui sont supposés constants (Friedman, 1971).

La décomposition de la dynamique d'ajustement du revenu nominal entre celle de la production et celle des prix constitue pour Friedman l'apport fondamental de son approche. Il présente un modèle qui peut être considéré comme une formalisation alternative du processus de transition du court au long terme décrit par sa courbe de Phillips (voir section 3 ci-dessous).

Ainsi, le mécanisme d'ajustement qu'il propose consiste à décomposer les variations du revenu nominal, résultant de l'écart entre l'offre et la demande de monnaie, entre des variations des prix et de la production. Plus précisément, il explique la déviation du taux de croissance du revenu nominal par rapport à son taux de croissance anticipé par l'écart entre l'offre et la demande de monnaie selon la formule :

$$\frac{d\text{Log}(Y)}{dt} = \left(\frac{d\text{Log}(Y)}{dt}\right)^* + \psi \cdot \left[\frac{d\text{Log}(M^s)}{dt} - \left(\frac{d\text{Log}(M^d)}{dt}\right) \right] + \varphi \cdot [\text{Log}(M^s) - \text{Log}(M^d)] \quad (18)$$

où M^s et M^d désignent, respectivement, l'offre et la demande de monnaie. La demande de monnaie est supposée une fonction du revenu permanent anticipé, du niveau et du taux de croissance anticipés des prix et des taux d'intérêt. En raison des fortes fluctuations ayant caractérisées l'offre de monnaie au cours de l'histoire, celle-ci est supposée exogène et non influencé par les effets de rétroaction des autres variables du modèle (Friedman, 1971).

La décomposition du taux de croissance du revenu nominal entre prix et production obéit à la même logique. L'écart des taux de croissance des prix et de la production par rapport à leurs valeurs anticipées est expliqué par la dynamique d'ajustement du revenu nominal. Cette décomposition, exprimée par les équations (19) et (20), s'explique par deux facteurs principaux à savoir les anticipations sur le comportement des prix et l'écart entre le niveau courant et le niveau permanent ou de plein-emploi de la production et de l'emploi :

$$\frac{d\text{Log}(P)}{dt} = \left(\frac{d\text{Log}(P)}{dt}\right)^* + \alpha \cdot \left[\frac{d\text{Log}(Y)}{dt} - \left(\frac{d\text{Log}(Y)}{dt}\right)^* \right] + \gamma \cdot [\text{Log}(y) - \text{Log}(y)^*] \quad (19)$$

$$\frac{d\text{Log}(y)}{dt} = \left(\frac{d\text{Log}(y)}{dt}\right)^* + (1-\alpha) \left[\frac{d\text{Log}(Y)}{dt} - \left(\frac{d\text{Log}(Y)}{dt}\right)^* \right] - \gamma [\text{Log}(y) - \text{Log}(y)^*] \quad (20)$$

où les contraintes sur les paramètres α et γ reflètent l'identité $Y = Py$, avec Y le revenu nominal, y la production et P le niveau général des prix.

Les perturbations qui affectent l'équilibre de long terme du modèle sont progressivement intégrées dans les anticipations des agents sur les variables du modèle de manière à éliminer tout écart entre les valeurs observées et anticipées de ces variables. Ce mode de formation des anticipations, présenté dans les équations ci-dessous, assure la transition entre l'équilibre de court terme et l'équilibre de long terme du modèle. A long terme, les anticipations sont validées, c'est-à-dire, les valeurs anticipées et les valeurs observées coïncident.

$$\left(\frac{d\text{Log}(P_t)}{dt}\right)^* = f\left(\frac{d\text{Log}(P_{t-j})}{dt}\right) \quad (21)$$

$$\left(\frac{d\text{Log}(Y_t)}{dt}\right)^* = g\left(\frac{d\text{Log}(Y_{t-j})}{dt}\right) \quad (22)$$

$$y_t^* = h(y_{t-j}) \quad (23)$$

$$P_t^* = k(P_{t-j}) \quad (24)$$

L'équilibre de long terme est donné par le modèle classique du macro-équilibre qui est composé de la combinaison, d'une part, des équations de l'équilibre général walrasien, et, d'autre part, de l'équation quantitative de la monnaie :

$$\frac{C}{P} = f\left(\frac{Y}{P}, \rho\right) \quad (25)$$

$$\frac{I}{P} = g(\rho) \quad (26)$$

$$\frac{Y}{P} = \frac{C}{P} + \frac{I}{P} \quad (27)$$

$$\rho = \rho^* = \rho_0 \quad (28)$$

Et le niveau général des prix par :

$$P = \frac{Y}{y} \quad (29)$$

Friedman considère que cette façon de combiner les secteurs monétaire et réel comme insatisfaisante pour deux raisons : la constance du taux d'intérêt réel et l'absence d'importants arguments relatifs à la fonction de consommation.

Le mécanisme de transmission peut être décrit de la façon suivante : une perturbation de l'équilibre de long terme introduit un écart dans les deux derniers termes de l'équation (18) et provoque une déviation du taux de croissance du revenu nominal par rapport à sa valeur anticipée. Ce dernier écart engendre à son tour, d'après les équations (19) et (20) un écart du taux de croissance du niveau général des prix et de la production par rapport à leur taux de croissance anticipé. Comme le taux de croissance anticipé du revenu nominal est, d'après l'équation (22), complètement déterminé par ses valeurs passées, il sera progressivement modifié, ce qui influence à nouveau le taux de croissance observé donné par la formule (18) qui rétroagit sur le prix et la production, et ainsi de suite.

Tobin (1972) fait remarquer que la décomposition de la déviation du taux de croissance du revenu nominal par rapport à sa valeur anticipée entre les déviations des taux de croissance de la production et du prix par rapport à leurs valeurs anticipées est présentée par Friedman sans aucune justification et n'a aucun fondement. En combinant les équations (19) et (20), Tobin dérive une nouvelle formulation de la courbe de Phillips qui relie positivement la déviation du niveau observé de l'inflation par rapport à sa valeur anticipée aux écarts, en niveau et en taux de croissance, de la production par rapport à sa valeur anticipée :

$$\frac{d\text{Log}(P)}{dt} = \left(\frac{d\text{Log}(P)}{dt}\right)^* + \frac{\alpha}{1-\alpha} \left[\frac{d\text{Log}(y)}{dt} - \left(\frac{d\text{Log}(y)}{dt}\right)^* \right] + \frac{\gamma}{1-\alpha} [\text{Log}(y) - \text{Log}(y)^*]$$

où les deux derniers termes de cette équation peuvent être liés au taux de chômage, en variation et en niveau.

Tobin considère que le problème du lien de court terme entre le salaire, le prix, le chômage et la production est bien trop complexe pour qu'il soit expliqué par les écarts entre les valeurs observées des variables et leurs valeurs anticipées. En faisant sans doute implicitement référence au lien de l'équation ci-dessus avec le discours présidentiel de Friedman en 1968 sur sa relecture de la courbe de Phillips, Tobin remet en cause la pertinence de généraliser au

niveau agrégé de l'économie les résultats trouvés dans le cas d'un produit homogène ou du marché du travail.

Tobin relève également une contradiction entre les deux cadres théoriques présentés par Friedman en 1959 et 1971. Le modèle monétaire de 1971 s'appuie sur deux piliers. D'une part, une relation de dépendance dynamique du revenu nominal par rapport à l'offre de monnaie. D'une autre part, un processus de décomposition de la variation du revenu nominal entre variation des prix et variation de la production. Dans ce modèle, la relation entre les variations de l'offre de monnaie et les variations du revenu nominal semble indépendante de la décomposition de ce dernier entre prix et production. En revanche, dans le modèle de la demande de monnaie de 1959 utilisant la théorie du revenu permanent, la vitesse de circulation de la monnaie dépend à la fois du revenu et du prix.

Section 2 : Décalages temporels et causalité monnaie-revenu : l'argument de la causalité inversée

Friedman (1959) explique la différence de comportement de la vitesse-revenu par l'écart entre les valeurs permanentes et mesurées du revenu et du prix. Il s'appuie sur l'existence de décalages temporels entre les cycles de la monnaie et du revenu pour inférer une influence causale de celle-là sur celui-ci (Friedman, 1958). Toutefois, Tobin (1970) développe un modèle keynésien où la monnaie est créée de façon endogène et n'exerce aucune influence causale sur le revenu.

Sous-section 1 : Théorie du revenu permanent et stabilité de la demande de monnaie

1. Revenu permanent et décalages temporels

Friedman (1970) déclare que les tenants de la théorie quantitative adoptent l'hypothèse empirique de la forte stabilité de la demande de monnaie, plus stable que la fonction de consommation keynésienne. Il s'agit là de la stabilité empirique, mais Friedman parle aussi de la stabilité théorique, dans le sens où la demande de monnaie dépend d'un petit nombre de variables. Etant donné cette stabilité de la vitesse-revenu de long terme, toute variation de court terme dans le stock de monnaie se transmet au revenu mesuré plutôt qu'au revenu permanent (Desai, 1981, Mayer, 1975).

Tobin (1972) considère que les propositions des monétaristes s'appuient sur l'hypothèse de l'inélasticité de la vitesse de circulation de la monnaie au taux d'intérêt. Toutefois, dans son article de 1970, Friedman utilise la sensibilité de la vitesse au taux d'intérêt pour montrer son comportement procyclique. Ce résultat est conforme à celui obtenu dans l'article de 1959 qui utilise le concept du revenu permanent. Ces deux cadres d'analyse de Friedman ne permettent, toutefois, pas d'expliquer les constatations empiriques où la monnaie devance le revenu dans les pics et les creux du cycle (Friedman, 1959 ; Friedman et Schwartz, 1963 a). En particulier,

la théorie du revenu permanent ne permet pas d'expliquer les décalages temporels entre la monnaie et le revenu

Les travaux menés par Friedman et Schwartz au sein du NBER les ont conduits à dégager d'importantes conclusions sur le comportement de la vitesse-revenu à court et à long terme sur une longue période (1870-1954). A très long terme, le stock de monnaie augmente à un rythme plus rapide que le revenu nominal. Dans ce cas, la vitesse-revenu, mesurée par le ratio du revenu nominal par rapport au stock de monnaie, baisse pendant les phases d'expansion du revenu réel. De plus, les auteurs attribuent les longues oscillations constatées au niveau du revenu nominal aux mouvements du stock nominal de la monnaie et à ceux de la vitesse-revenu. Or, le stock nominal de la monnaie évolue en sens opposé à celui de la vitesse-revenu, mais avec des amplitudes plus grandes de sorte que c'est le stock de monnaie qui domine et commande le mouvement du revenu nominal (Friedman, 1959).

En revanche, à l'intérieur des cycles, l'accroissement du stock de monnaie est inférieur à celui du revenu nominal durant une phase d'expansion, de sorte que la vitesse-revenu augmente dans le sens du revenu réel. Là encore, les mouvements cycliques du revenu monétaire sont attribués à ceux du stock de la monnaie et de la vitesse-revenu, qui évoluent dans le même sens et avec des amplitudes comparables. De plus, les sommets et les creux de la vitesse-revenu sont à peu près conformes à ceux du cycle.

Il ressort, ainsi, de ces observations une opposition entre le comportement procyclique de la vitesse-revenu à l'intérieur des cycles et son évolution contracyclique à très long terme. Friedman interprète ces comportements de la vitesse-revenu comme étant des déviations par rapport à un comportement relativement stable de la demande de monnaie. Ces déplacements sont eux-mêmes provoqués par des modifications du comportement des variables explicatives de cette demande de monnaie (Friedman, 1959).

A très long terme, le revenu semble être la variable dominante dans l'explication de la demande d'encaisses réelles. Les variables du taux d'intérêt et de l'inflation sont écartées car leur effet est supposé trop faible à cet horizon. La variable du taux d'intérêt est également écartée à court terme, car son comportement cyclique est considéré non compatible avec celui de la vitesse-revenu.

S'inspirant de ses travaux sur la consommation des ménages (Friedman, 1957), Friedman introduit la distinction entre revenu mesuré et revenu permanent pour construire un modèle qui puisse rendre compte du comportement de la vitesse-revenu à la fois à court et à long termes. Pour ce faire, Friedman interprète la monnaie comme un bien de consommation durable détenue pour les services qu'elle rend et rapportant un flux de services proportionnels au stock (Friedman, 1959). Le modèle proposé relie la demande d'encaisses réelles permanentes par tête au revenu réel permanent par tête :

$$\frac{M}{NP_p} = \gamma \left(\frac{Y_p}{NP_p} \right)^\delta \quad (30)$$

où N désigne la population, P_p le niveau permanent des prix, Y_p le niveau permanent du revenu nominal. L'estimation des paramètres γ et δ permet de calculer les valeurs estimées pour les encaisses nominales :

$$M = 0,00323 \cdot \left(\frac{Y_p}{NP_p} \right)^{1,810} \cdot NP_p \quad (31)$$

et la vitesse-revenu :

$$V = \frac{1}{0,00323} \cdot \left(\frac{Y_p}{NP_p} \right)^{-0,810} \cdot \frac{Y}{Y_p} \quad (32)$$

A partir de ces résultats, Friedman interprète la différence entre le comportement cyclique de la vitesse-revenu et son mouvement de long terme comme étant le résultat de l'écart entre les valeurs mesurées du revenu et des prix et les valeurs permanentes qui déterminent le comportement de détention des encaisses. A court terme, les encaisses monétaires s'ajustent au revenu permanent plutôt qu'au revenu mesuré (Friedman, 1958).

Friedman reconnaît l'échec de son modèle à rendre compte de ce qu'il qualifie comme étant l'un de ses résultats les plus importants concernant le comportement des encaisses monétaires. En effet, d'après l'équation (31), les encaisses monétaires s'ajustent positivement au cycle et lui sont synchrones ou légèrement décalées aux points de retournement. L'auteur rappelle que le résultat majeur qui se dégage des vastes investigations statistiques qu'il a menées avec sa collègue Anna Schwartz sur le comportement de la vitesse-revenu aux Etats-Unis révèle que

le stock nominal de la monnaie tend à devancer et à diriger le revenu nominal à la fois aux sommets et aux creux du cycle (Friedman, 1959).

Ainsi, il y a un élément résiduel dans le comportement cyclique de la vitesse-revenu dont le modèle ne rend pas compte. De fait, les mouvements cycliques des encaisses monétaires ne peuvent pas être entièrement expliqués par la déviation relative à une demande de monnaie stable répondant uniquement au mouvement cyclique du revenu permanent. Ce constat remet fortement en cause l'hypothèse de la stabilité de la demande de monnaie, dans le sens où le modèle du revenu permanent ne suffit pas à décrire l'ensemble du comportement de cette demande.

L'explication de cet élément résiduel exige d'introduire dans la demande de monnaie des variables qui rendent compte de la détention des actifs substituables à la monnaie. Les candidats proposés par l'auteur concernent le taux d'intérêt, comme mesure du rendement sur la détention des titres, et le taux de croissance des prix qui est associé à la détention des biens physiques.

2. Incohérence théorique de l'explication de la vitesse-revenu

Dans son article de 1970, Friedman expose un modèle alternatif à celui du revenu permanent pour expliquer également le comportement procyclique de la vitesse-revenu. Pour cela, il procède à une dynamisation du modèle (12) et abandonne ainsi l'hypothèse d'exogénéité du taux d'intérêt qui lui a permis de dériver la relation directe (17) entre le revenu nominal et la masse monétaire. La différentiation de la version logarithmique de l'équation (12) et de la relation (13) fournit ainsi la relation suivante :

$$\frac{1}{Y} \frac{dY}{dt} = s \cdot \frac{d}{dt} \left(\frac{1}{Y} \frac{dY}{dt} \right)^* + \frac{1}{M} \frac{dM}{dt} \quad (33)$$

où $s = \frac{1}{V} \frac{dV}{dr}$ est supposé non nul, autrement dit, la vitesse de circulation de la monnaie évolue en fonction du taux d'intérêt. Friedman suppose que le taux de croissance anticipé du revenu nominal est déterminé par un modèle d'anticipations adaptatives :

$$\frac{d}{dt} \left(\frac{1}{Y} \frac{dY}{dt} \right)^* = \beta \cdot \left[\frac{1}{Y} \frac{dY}{dt} - \left(\frac{1}{Y} \frac{dY}{dt} \right)^* \right] \quad (34)$$

L'équation (34) devient dans ces conditions :

$$\frac{1}{Y} \frac{dY}{dt} = \left(\frac{1}{Y} \frac{dY}{dt} \right)^* + \frac{1}{1 - \beta \cdot s} \left[\frac{1}{M} \frac{dM}{dt} - \left(\frac{1}{Y} \frac{dY}{dt} \right)^* \right] \quad (35)$$

Pour Tobin (1972), la relation (35) reflète l'idée monétariste de base de la dynamique du lien entre le revenu nominal et l'offre de monnaie, à savoir qu'à long terme le taux de croissance de l'offre de monnaie et le taux de croissance anticipé du revenu nominal sont égaux. Un accroissement de la masse monétaire plus rapide que son taux d'équilibre se traduit par un taux de croissance observé du revenu nominal supérieur à son taux de variation anticipé.

De même, l'écart entre le taux de croissance du revenu nominal donné par l'équation (35) et celui de la masse monétaire permet de calculer le taux de croissance de la vitesse de circulation de la monnaie :

$$\frac{1}{V} \frac{dV}{dt} = \frac{\beta \cdot s}{1 - \beta \cdot s} \left[\frac{1}{M} \frac{dM}{dt} - \left(\frac{1}{Y} \frac{dY}{dt} \right)^* \right] \quad (36)$$

avec $0 < \beta \cdot s < 1$.

Si la masse monétaire croît plus vite que le revenu nominal anticipé, alors, d'après l'équation (35), le taux de croissance du revenu nominal est supérieur son taux de croissance anticipé, et, d'après la relation (36), la vitesse de circulation augmente. Friedman (1971) considère cette interprétation du mouvement de la vitesse de circulation de la monnaie, en termes de déviation du taux de croissance du revenu nominal par rapport à son taux de croissance anticipé, comme étant complémentaire à ses résultats antécédents. En effet, dans son article de 1959, Friedman explique le mouvement procyclique de la vitesse-revenu par l'écart entre le revenu permanent et le revenu mesuré. Toutefois, les équations (35) et (36) n'intègrent pas explicitement le raisonnement autour des phases d'évolution du revenu réel.

Tobin note que cette explication par Friedman du mouvement procyclique de la vitesse de circulation par rapport au revenu nominal renvoie, implicitement, à l'analyse keynésienne du

comportement procyclique du taux d'intérêt (Tobin, 1972). En effet, l'équation (34) n'est rien d'autre que l'expression de la dérivée du taux d'intérêt par rapport au temps :

$$\frac{dr}{dt} = \frac{d}{dt} \left(\frac{1}{Y} \frac{dY}{dt} \right)^* = \beta \cdot \left[\frac{1}{Y} \frac{dY}{dt} - \left(\frac{1}{Y} \frac{dY}{dt} \right)^* \right] \quad (37)$$

et, l'équation (36) peut se réécrire sous la forme :

$$\frac{1}{V} \frac{dV}{dt} = s \cdot \frac{dr}{dt} \quad (38)$$

Friedman reconnaît, comme ce fût le cas de sa théorie du revenu permanent, l'incapacité de sa théorie du revenu monétaire à expliquer le résultat empirique portant sur la réaction retardée de la vitesse de circulation de la monnaie au niveau des points de retournement du taux de croissance de la masse monétaire. En revanche, l'équation (17) reliant le revenu monétaire aux valeurs présentes et passées de la masse monétaire, permet d'expliquer les décalages temporels entre les mouvements de la masse monétaire et du revenu nominal (Tobin, 1970). Considérons une écriture de l'équation (17) sous la forme suivante :

$$\dot{Y}_t = \sum_{j=0}^{j=s} b_j \dot{M}_{t-j} \quad (39)$$

où \dot{Y}_t et \dot{M}_t expriment le taux de croissance du revenu et de la masse monétaire, et les b_j sont des paramètres positifs de somme égale à l'unité. Ce modèle permet d'expliquer les décalages constatés par Friedman puisque les fluctuations cycliques du taux de croissance de la masse monétaire précèdent les taux de croissance du revenu nominal. Mais, ce modèle ne permet pas de rendre compte du comportement procyclique ou contracyclique de la vitesse-revenu, dont le taux de croissance peut être déduit de la relation (39) :

$$\dot{V}_t = \dot{Y}_t - \dot{M}_t = \sum_{j=0}^{j=s} b_j (\dot{M}_{t-j} - \dot{M}_t) \quad (40)$$

D'après la relation (39), lorsque les \dot{M}_{t-j} augmentent, \dot{Y}_t s'accroît également, mais moins rapidement que \dot{M}_t . Dans le même temps, la vitesse-revenu décroît, c'est-à-dire affiche un comportement contracyclique qui est contredit par les données observées.

L'explication du comportement d'une seule variable, en l'occurrence la vitesse-revenu, exige le recours à deux modèles différents. Tobin critique vivement la démarche de Friedman qui consiste à changer à chaque fois de modèle et de ne pas disposer d'un seul et même modèle cohérent qui puisse rendre compte de tous les comportements observés de la vitesse de circulation.

Sous-section 2 : Limites théoriques de la causalité monnaie-revenu

1. De la causalité empirique monnaie-revenu ...

Dans son article de 1958, Friedman recommande l'adoption d'une politique monétaire passive, s'appuyant sur une règle de croissance fixe de la masse monétaire, et il met en garde contre l'utilisation de la politique monétaire pour la stabilisation conjoncturelle. Cette recommandation s'appuie sur deux arguments empiriques qui cherchent à prouver le rôle actif de la masse monétaire dans l'activité économique. Le premier concerne l'hypothèse, constatée sur l'histoire, de la stabilité de la vitesse de circulation de la monnaie à court terme et son évolution lente à long terme. La demande de monnaie, fonction du revenu permanent, permet de fournir une réconciliation théorique entre le comportement de court terme (ou cyclique) et de très long terme de la vitesse-revenu.

Le second argument porte sur les décalages temporels observés au niveau des creux et des sommets du cycle de l'activité économique et du taux de croissance de la masse monétaire. Ce dernier argument est utilisé par Friedman pour justifier l'influence causale de la monnaie sur le revenu monétaire. Mais, comme mentionné dans la sous-section précédente, le cadre théorique de Friedman ne permet pas d'expliquer les décalages temporels.

Friedman s'appuie fortement sur les travaux qu'il a réalisés en partenariat avec A. Schwartz pour la description et l'analyse des mouvements cycliques et de très long terme du stock de la monnaie. Ces travaux ont révélé l'existence, à long terme, d'une forte corrélation positive entre les mouvements des prix et de la masse monétaire. Friedman admet que cette liaison statistique n'apprend rien sur le sens de l'influence ou de la causalité entre ces deux variables.

Toutefois, en raison de la variété des régimes monétaires sous lesquels cette liaison a été observée dans l'histoire, Friedman privilégie l'existence d'une influence de la monnaie sur les prix, de sorte à ce qu'un contrôle de la masse monétaire entraîne un contrôle des prix.

Il existe également à court terme une corrélation positive entre la masse monétaire et les prix, mais moins forte que celle de long terme. De plus, le sens de la liaison entre ces deux variables est moins net et plus complexe que dans le cas du long terme. Les variations du stock de monnaie constituent une conséquence et une cause indépendante pour les mouvements des prix et du revenu, mais lorsque ces variations se produisent, elles exercent des effets supplémentaires sur les prix et le revenu. Ce phénomène de rétroaction brouille la liaison entre la monnaie et les prix, mais n'inverse pas son sens, car les décalages temporels entre ces deux variables prouvent qu'au cours du cycle le stock de monnaie joue un rôle en grande partie indépendant.

Le stock de monnaie possède une tendance haussière très forte qui domine son comportement cyclique. De plus, la temporalité cyclique aux points de retournement n'apparaît pas clairement lorsque c'est le niveau du stock de monnaie qui est utilisé, car ce dernier affiche un petit nombre de points de retournement (Friedman, Schwartz, 1963 a). Pour faciliter la visualisation et l'analyse des décalages temporels entre les cycles du revenu et de la masse monétaire, Friedman et Schwartz comparent le taux de croissance du stock de monnaie aux creux et aux sommets du cycle d'activité.

Le taux de croissance de la masse monétaire affiche un cycle étroitement semblable à celui de l'activité économique mais il devance ce dernier avec un long délai. En moyenne, l'écart entre les sommets du taux de croissance de la masse monétaire et du cycle est de 16 mois, et l'écart entre les creux est de 12 mois. Pour Friedman, cette observation constitue une preuve solide de l'indépendance de l'influence de la masse monétaire, mais elle montre aussi que cette influence prend beaucoup de temps avant de se faire sentir. De plus, le délai des décalages varie considérablement d'un cycle à un autre ; l'intervalle de temps le plus court est de 13 mois pour les sommets et 5 mois pour les creux, l'intervalle de temps le plus long est de 24 mois pour les sommets et 21 mois pour les creux.

Du point de vue de la politique économique, la variation de ces délais est problématique, car une décision prise aujourd'hui pour agir sur un aspect particulier de l'activité économique

n'exercera ses effets que dans 6 mois ou peut être au bout d'un an et demi. L'ampleur des délais entraîne une mauvaise interprétation et des erreurs de conception dans la politique monétaire. Les effets de la politique monétaire ne se produisant pas instantanément, celle-ci est perçue comme étant inefficace.

La variation de l'ampleur des délais signifie qu'il existe une dérive importante dans la relation liant à court terme les prix et la masse monétaire, reflétant l'implication d'autres facteurs. Lorsque les fluctuations de la masse monétaires sont importantes, elles dominent ces autres facteurs et elles imposent leur sens d'évolution aux prix et au revenu, mais, si la masse monétaire évolue régulièrement, l'effet des autres facteurs réapparaît. Si le niveau de connaissance sur ces facteurs le permettait, la politique monétaire pourrait être mobilisée pour lutter contre leurs effets. Il y a, ainsi, de véritables limites quant à la possibilité d'un contrôle fin du niveau général des prix par des ajustements de la masse monétaire.

Friedman s'appuie sur son analyse statistique de la relation entre la masse monétaire, les prix et la production pour formuler plusieurs implications pour la politique monétaire en termes de croissance et de stabilité. Les fortes fluctuations des prix représentent une perturbation pour la croissance économique et sa stabilité. Il estime que pour assurer la stabilité à long terme des prix, ceux-ci doivent croître à un taux qui, d'un côté, soit semblable à celui de la production et de la population, et, d'un autre côté, puisse permettre aux agents économiques d'augmenter le rapport de leurs encaisses par rapport à leur revenu. Partant de ces normes et des observations historiques, Friedman recommande d'adopter la règle d'un taux de croissance de la masse monétaire compris entre 3% et 5% et d'éviter de mener une politique monétaire discrétionnaire qui, selon lui, risque d'aggraver la situation plutôt que de l'améliorer.

2. ... à l'a priori théorique

L'analyse des séries historiques sur le revenu nominal et la masse monétaire, menées en collaboration avec Anna Schwartz, ont permis à Milton Friedman de mettre en évidence une corrélation entre ces deux agrégats avec une avance de quelques trimestres de la monnaie. L'auteur interprète ce constat empirique comme constituant la preuve de la causalité exercée par l'offre exogène de la monnaie sur le revenu nominal. Autrement dit, les fluctuations de ce

dernier sont la conséquence de la manipulation de la masse monétaire par les autorités monétaires.

Les keynésiens insistent sur le rôle de l'*a priori* théorique pour interpréter les corrélations observées entre les séries chronologiques (Desai, 1981). Solow et Kareken (1963) refusent de retenir l'observation des décalages temporels entre la masse monétaire et le revenu, mis en évidence par Friedman, comme la preuve de l'existence d'une relation causale entre ces deux variables qui puisse être mobilisée contre l'utilisation de la politique monétaire pour la stabilisation conjoncturelle. Ils estiment que plusieurs hypothèses peuvent être formulées pour expliquer la succession des sommets et des creux du cycle et de la masse monétaire. A cet effet, ils considèrent qu'il est nécessaire de préciser d'abord la structure du modèle qui est utilisé pour estimer les effets de la politique monétaire.

Tobin remet en cause cette conclusion en soutenant que les décalages en question peuvent très bien être interprétés comme la conséquence d'une causalité inversée qui va du revenu vers la masse monétaire. Tobin présente en 1970 deux modèles théoriques, l'un "ultra-keynésien" et le second "friedmanien", intégrant des mécanismes de transmission qui génèrent une structure de retards entre les variables (Tobin, 1970).

Le modèle keynésien est constitué de trois équations, la première détermine le revenu agrégé, la seconde décrit le comportement de la demande de monnaie et la troisième est l'identité du déficit du budget du gouvernement. La première équation relie le revenu monétaire Y aux dépenses du gouvernement G et aux dépenses autonomes nominales d'investissement² K' à l'aide du multiplicateur m :

$$Y = m.(G + K') \quad (41)$$

Le multiplicateur m est déduit à partir de l'identité suivante :

$$\text{Epargne} + \text{Taxes} = \text{Dépenses du gouvernement} + \text{Investissement net}$$

qui s'écrit encore :

$$s(1 - \mu).Y + \mu.Y = G + K' \quad (42)$$

² Pour une variable X , la notation X' désigne la dérivée par rapport au temps de X .

avec s la propension à épargner et μ le taux de taxation. Le multiplicateur m est alors donné par :

$$m = \frac{1}{s(1-\mu) + \mu} \quad (43)$$

Les dépenses du gouvernement et celles de l'investissement sont supposées exogènes.

La demande de monnaie s'écrit comme la somme d'une demande pour transaction, proportionnelle au revenu, et d'une demande d'actifs qui est liée au taux d'intérêt. La formulation de cette dernière s'appuie sur une description particulière de l'ajustement du portefeuille. Quatre types d'actifs sont considérés, le stock de capital K , la dette privée vis-à-vis du système bancaire L , les engagements monétaires du privé à l'égard du gouvernement et du système bancaire, incluant les dépôts à terme, M et le stock des obligations sur la dette du gouvernement B . La richesse privée W est constituée de la somme du stock de capital K et du stock de la dette du gouvernement D :

$$W = K + D \quad (44)$$

Le bilan du public s'écrit alors :

$$K + D = K - L + M + B \quad (45)$$

Etant donné que les fluctuations autonomes de l'investissement réel constituent l'origine des fluctuations du cycle de l'activité (Tobin, 1970), la répartition de la richesse entre les quatre actifs est fortement influencée par le comportement exogène du stock du capital. Ainsi, durant les périodes d'expansion/recul de l'investissement, le capital devient plus/moins attractif relativement à la monnaie et aux obligations. De la même façon, les emprunts auprès du système bancaire sont supposés augmenter pendant les expansions de l'investissement et baisser durant les chutes de celui-ci. Dans ces conditions, la dette privée à l'égard du système bancaire est considérée comme une proportion fixe du stock du capital :

$$L = \tau.K \quad (46)$$

avec $0 < \tau < 1$.

L'unique décision relative au choix de la composition du portefeuille qui reste à faire concerne le partage du solde de la richesse restante, à savoir :

$$W - K + L \quad (47)$$

entre les deux actifs restants M et B en fonction du taux d'intérêt r sur les obligations. La relation (47) peut s'écrire aussi sous la forme :

$$M + B = D + \tau.K \quad (48)$$

La demande de monnaie peut finalement se mettre, conformément à la théorie keynésienne de la préférence pour la liquidité, sous la forme :

$$M = a_0(r).(D + \tau.K) + a_1.Y \quad (49)$$

où a_1 est un paramètre positif et $a_0(r)$ est négative. Tobin suppose que l'autorité monétaire fournit les réserves nécessaires au système bancaire de façon à maintenir constant le taux d'intérêt r . Dans ces conditions, la fonction $a_0(r)$ peut être considérée comme constante. Ainsi, dans ce modèle, la monnaie n'exerce aucune influence causale sur le revenu. Les autorités monétaires adaptent l'offre de monnaie aux fluctuations de l'investissement et de la demande de monnaie. La hausse anticipée du revenu incite les agents à demander du crédit au système bancaire ce qui va accroître la masse monétaire de façon endogène.

Le modèle keynésien est alors formé du système des trois équations suivantes :

$$Y = m.(G + K') \quad (50)$$

$$M = a_0.(D + \tau.K) + a_1.Y \quad (51)$$

$$D' = G - \mu.Y \quad (52)$$

Lorsqu'elle est exprimée en termes de flux, la relation (51) devient :

$$M' = a_0.(D' + \tau.K') + a_1.Y' \quad (53)$$

En utilisant les relations (50) et (52), l'équation (53) devient :

$$M' = a_0.(1 - \tau).G + a_0\left(\frac{\tau}{m} - \mu\right).Y + a_1.Y' \quad (54)$$

Pour des dépenses du gouvernement G données, l'équation (54) exprime la variation de la masse monétaire M' comme une combinaison linéaire de Y et de Y' , et l'évolution de ces deux dernières variables s'explique uniquement par le comportement de K' et de K'' . En

supposant de plus que $(\frac{\tau}{m} - \mu) < 0$, l'équation (54) permet de générer une relation négative entre M' et Y . Pour introduire des mouvements cycliques, Tobin suppose que K' est généré par une onde sinusoïdale et peut s'écrire sous la forme :

$$K' = K_0 + \alpha.Sin(t) \quad (55)$$

où K_0 et α sont des constantes positives, et t désigne le temps. Dans ce cas, Y et Y' s'écrivent :

$$Y = Y_0 + m\alpha.Sin(t) \quad (56)$$

$$Y' = m\alpha.Cos(t) \quad (57)$$

$$M' = A + B.Sin(t) + C.Cos(t) \quad (58)$$

où Y_0 , A , B et C sont des constantes, avec B négative. Les creux et les sommets du cycle correspondent aux cas où $Y' = 0$, c'est-à-dire, lorsque $t = (2k+1)\frac{\pi}{2}$, où k est un nombre entier. Ainsi, lorsque Y est à son sommet, $M' = A + B$, et quand Y est à son creux, $M' = A - B$. Puisque B est négative, le sommet de Y correspond au creux de M' , et inversement. Autrement dit, il existe un décalage d'une demi phase (c'est-à-dire $\frac{\pi}{2}$) entre les sommets et les creux des deux variables. Ainsi, en exploitant les propriétés cycliques et la succession des sommets et des creux des fonctions cosinus et sinus, Tobin réussit à mettre en évidence à l'aide d'un modèle théorique, où la monnaie ne joue aucun rôle causale sur l'activité économique, les décalages observés par Friedman entre les sommets et les creux de la variation de la masse monétaire M' et du cycle de l'activité économique.

Pour le cas du modèle "friedmanien", Tobin utilise la théorie du revenu permanent de la demande de monnaie développée par Friedman dans son article de 1959 et qu'il écrit sous la forme :

$$Log(M) = A + \delta.Log(Y_p^*) \quad (59)$$

Cette formulation retient uniquement des grandeurs nominales et n'intègre pas le niveau général des prix. On suppose que le revenu permanent Y_p^* assure l'ajustement entre la demande de monnaie et l'offre de monnaie M supposé exogène. A est une constante et δ est l'élasticité de la demande de monnaie au revenu permanent dont la valeur estimée par

Friedman est de 1,8. Le revenu et le revenu permanent sont supposés augmenter à très long terme au taux exponentiel de paramètre β . Ainsi, corrigé de sa tendance de long terme, le revenu permanent Y_p s'écrit :

$$\text{Log}(Y_p) = \text{Log}(Y_p^*) - \beta.t - C \quad (60)$$

L'équation (59) peut alors se mettre sous la forme :

$$\text{Log}(M) = B + \delta.\text{Log}(Y_p) + \delta.\beta.t \quad (61)$$

où B et C sont des constantes et t désigne le temps. La différentiation par rapport au temps de l'équation (61) fournit la relation suivante entre le taux de croissance de la masse monétaire et le taux de croissance du revenu permanent corrigé de sa tendance :

$$\frac{M'}{M} = \delta.\frac{Y_p'}{Y_p} + \delta.\beta \quad (62)$$

Le revenu permanent corrigé de sa tendance est une moyenne pondérée, avec des poids décroissants exponentiellement, des revenus actuels et passés également corrigés de leurs tendances. Lorsque le revenu mesuré s'écarte du revenu permanent, le public révisé son estimation de ce dernier par une proportion ω de cet écart :

$$\frac{Y_p'}{Y_p} = \omega.(\text{Log}(Y) - \text{Log}(Y_p)) \quad (63)$$

D'après les calculs faits par Friedman dans son travail sur la fonction de consommation, la valeur du paramètre ω est égale à 0,4.

L'équation (63) permet de calculer le revenu mesuré en fonction du revenu permanent :

$$\text{Log}(Y) = \frac{1}{\omega} \frac{Y_p'}{Y_p} + \text{Log}(Y_p) \quad (64)$$

La combinaison des équations (61), (62) et (64) permet d'obtenir une relation entre le revenu, la masse monétaire et sa variation :

$$\text{Log}(Y) = \frac{1}{\delta.\omega} \frac{M'}{M} + \frac{1}{\delta} \text{Log}(M) - \beta.t - \frac{\beta}{\omega} - \frac{A}{\delta} \quad (65)$$

La relation entre le niveau du revenu et celui de la masse monétaire est positive, ainsi que celle entre le niveau du revenu et le taux de croissance de la masse monétaire. La différentiation de l'équation (65) par rapport au temps fournit la relation suivante :

$$g_Y = \frac{Y'}{Y} = \frac{g'_M}{\delta \cdot \omega} + \frac{g_M}{\delta} - \beta \quad (66)$$

où g_M désigne le taux de croissance de la masse monétaire et g'_M est sa dérivée par rapport au temps. La relation entre le taux de croissance du revenu et le taux de croissance de la masse monétaire est également positive.

Si l'on suppose que le taux de croissance de la masse monétaire g_M est généré par une fonction sinusoïdale, alors le taux de croissance du revenu est égal, d'après l'équation (66), à une moyenne pondérée d'une fonction cosinus et d'une fonction sinus avec des coefficients positifs. Dans un repère cartésien, le taux de croissance du revenu est lié au taux de croissance de la masse monétaire par une ellipse dont l'axe a une pente égale à $\frac{1}{\delta} = \frac{1}{1,8} = 0,6$ par rapport à l'axe des abscisses.

A l'aide de ce modèle, Tobin démontre que g_Y devance g_M dans le sens où g_Y atteint son sommet et son creux avant g_M . De même, le niveau du stock de monnaie lui-même atteint son sommet et son creux après le niveau du revenu. Certes, le niveau du revenu Y atteint son sommet et son creux avec un retard par rapport à g_M , mais ce retard est faible alors que Friedman déclare que cet écart est important et atteint en moyenne 16 mois (Friedman, 1958).

En définitive, il apparaît que le modèle du revenu permanent de Friedman génère une antécédence de la monnaie par rapport au revenu moins importante que celle qui est mise en évidence dans le modèle keynésien. En effet, ce dernier reproduit les résultats empiriques constatés par Friedman aussi bien dans l'ordre de leur succession que dans la longueur de leurs décalages, ce qui n'est pas le cas du modèle du revenu permanent.

Section 3 : La relecture monétariste de la courbe de Phillips

L'hypothèse de rigidité à la baisse du salaire nominal dans le modèle IS-LM est jugée par les monétaristes sans fondement théorique et à l'origine de l'efficacité de la politique monétaire. De plus, Friedman (1970, 1971) estime que la construction de la courbe de Phillips keynésienne est faite de façon *ad hoc*. Friedman (1968) propose une reformulation de cette courbe, qui entraîne un arbitrage à court terme entre l'inflation et le chômage, mais qui est verticale à long terme. On retrouve les mêmes recommandations de politique monétaire que celles formulées par Friedman à la fin des années 1950, à savoir une règle de croissance constante de la masse monétaire pour assurer l'objectif ultime de la stabilité des prix.

Sous-section 1 : De la rigidité du salaire à la courbe de Phillips dans IS-LM

1. Chômage involontaire et efficacité de la politique monétaire

Le cadre théorique qui sous-tend la détermination des prix et ses implications pour la politique monétaire est une question centrale dans le débat entre les keynésiens et les monétaristes. Ces derniers placent l'hypothèse de fixité des prix, retenu dans le modèle IS-LM, au cœur de la divergence qui les oppose aux keynésiens (Friedman, 1971). Comme cela a été démontré dans le chapitre 1, c'est cette hypothèse qui, aux yeux des monétaristes, pousse les keynésiens à trop insister sur le rôle du taux d'intérêt dans la détermination de la monnaie et justifie la transmission de la politique monétaire par ce canal (Friedman, 1972).

Le débat sur la rigidité du salaire nominal s'appuie sur le modèle IS-LM développé par Modigliani (1944) et non pas celui de Hicks (1937). L'hypothèse de rigidité à la baisse du salaire nominal, considérée par Friedman comme étant arbitraire, ou *deus ex machina*, et sans sous-bassement théorique (Friedman, 1971), est responsable de l'apparition d'un équilibre de sous-emploi dans ce modèle. La remise en cause de cette hypothèse conduit automatiquement à nier l'efficacité, supposée par Modigliani, de la politique monétaire pour lutter contre le chômage involontaire.

La principale distinction entre le modèle IS-LM de Modigliani par rapport à son prédécesseur IS-LL de Hicks est l'introduction explicite du comportement du marché du travail ce qui rend le salaire une variable endogène. Modigliani retient l'hypothèse de la rigidité à la baisse du salaire nominal, au lieu de celle de la trappe à liquidité retenue par Hicks, pour caractériser le cas keynésien par rapport au cas classique. Chez Hicks l'exogénéité du salaire constitue une hypothèse commune aux deux variantes classique et keynésienne qui présentent de ce fait la même situation d'équilibre de sous-emploi, alors que chez Modigliani le chômage involontaire devient une propriété de la version keynésienne du modèle. Dans ces conditions, une augmentation de l'offre de monnaie devient dans le modèle de Modigliani, contrairement à celui de Hicks, efficace dans le cas keynésien et inefficace dans le cas classique pour stimuler l'emploi³.

L'analyse des conséquences de l'endogénéisation du salaire sur l'efficacité de la politique monétaire dans le modèle de la synthèse est conduite à partir du couplage du modèle IS-LM de Modigliani présenté dans le chapitre 1 avec un sous-système qui décrit le comportement du marché du travail. Dans la version classique de son modèle, Modigliani retient le sous-système suivant :

(67) $Q = f(N^-)$, fonction de production de court terme

(68) $f'(N^-) = w/p$, demande de travail, fonction décroissante du taux de salaire réel

(69) $N^+ = g(w/p)$, offre de travail, fonction croissante du taux de salaire réel

(70) $N^+ = N^-$, équilibre sur le marché du travail

La fonction de production donnée par (67) ne fait intervenir que le niveau de l'emploi puisqu'il s'agit d'un raisonnement à court terme et que le stock de capital installé est supposé donné.

Ce second sous-système fournit indépendamment du premier modèle IS-LM de base les valeurs d'équilibre de grandeurs réelles, à savoir la demande et l'offre de travail, la quantité produite et le taux de salaire réel. Selon Modigliani (1944), l'équation (69), absente de l'analyse de Hicks, reflète le comportement rationnel des offreurs de travail. Une fois la

³ Voir à ce titre l'article de De Vroey (2000) sur les différences entre les modèles IS-LL de Hicks et IS-LM de Modigliani.

valeur de Q est déterminée, elle est portée dans les équations (3.2) à (6.2) du modèle IS-LM de la section 1 du chapitre 1 pour déterminer les valeurs réelles d'équilibres de I/p , S/p , C/p et de taux d'intérêt i . Dans ces conditions, les équations (1.2) et (2.2), exprimant l'équilibre sur le marché monétaire, permettent uniquement de déterminer le niveau général des prix p et la résolution du modèle se fait de manière récursive. La forme de la demande de monnaie n'exerce aucune influence sur la détermination de la solution d'équilibre.

L'offre de monnaie n'exerce ainsi aucun effet sur le niveau d'équilibre de l'emploi, de la production et du taux d'intérêt. Ce dernier est assimilé à un taux d'intérêt réel et son rôle se limite à décomposer le revenu réel entre investissement et consommation sans exercer aucune influence sur la détermination du niveau de ce revenu. Nous retrouvons ainsi les résultats traditionnels du modèle classique standard du macro-équilibre dans lequel une expansion de l'offre de monnaie se traduit uniquement par une hausse du niveau général des prix. Cette expression du modèle en termes réels signifie, selon Patinkin (197) l'absence d'illusion monétaire qui constitue une condition nécessaire pour la validité de la théorie quantitative.

Dans le cas keynésien, Modigliani abandonne l'équation (70) de l'équilibre sur le marché du travail et considère au lieu de l'équation (69) la forme particulière de l'offre de travail suivante :

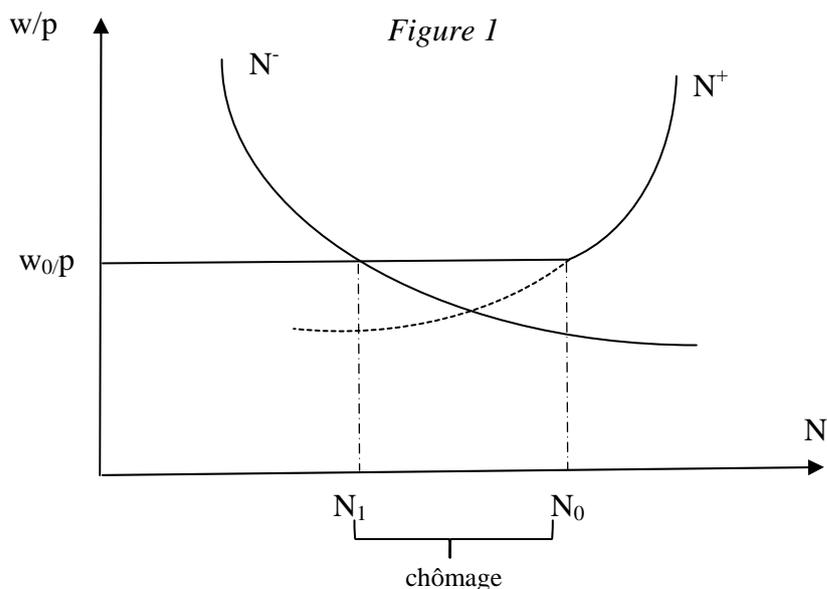
$$w = \alpha.w_0 + \beta.p.g^{-1}(N^+) \quad (71)$$

où w_0 est un taux de salaire nominal donné, α et β sont des constantes vérifiant les conditions du système suivant :

$$\left[\begin{array}{l} \alpha = 1 \text{ et } \beta = 0 \text{ pour } N^+ \leq N_0, \\ \alpha = 0 \text{ et } \beta = 1 \text{ pour } N^+ > N_0, \end{array} \right.$$

avec N_0 le niveau du plein-emploi au sens de Modigliani, c'est-à-dire, l'intersection de la partie horizontale de la courbe d'offre de travail avec sa partie ascendante. Au niveau de la partie horizontale de la courbe de l'offre de travail, le taux de salaire w_0 , fixé par des considérations historiques ou de politique économique, n'est pas une variable du modèle mais

une donnée. C'est l'absence de fondement théorique à cette rigidité à la baisse du salaire que critique Friedman et qui explique à la fois l'apparition du chômage et l'efficacité de la politique monétaire pour stimuler l'emploi. Il prétend que l'approche keynésienne de l'inflation insiste sur les facteurs institutionnels et non pas sur les facteurs monétaires (Friedman, 1971).



Dans la figure 1, aussi longtemps que la demande de travail reste inférieure à N_0 , le taux de salaire demeure à son niveau historique w_0 . Dans cette configuration apparaît un niveau de chômage égal à la différence entre N_0 et la demande de travail N_1 correspondant au taux de salaire w_0 et défini par :

$$N_1 = f^{-1}(w_0 / p)$$

Le taux de salaire devient flexible à la hausse et constitue une variable du modèle lorsque tous ceux qui acceptent d'être employés au taux de salaire réel w_0 / p sont embauchés. Dès que le taux de salaire nominal s'accroît par rapport au niveau du prix, l'offre de travail augmente.

Le niveau général des prix donné par l'équation (2) exprime le prix d'offre du produit fixé par les entreprises de façon à ce que la productivité marginale du travail en nominal soit égale au taux de salaire nominal donné. A chaque niveau de la productivité marginale du travail correspond, pour ce taux de salaire donné, un niveau des prix p .

Lorsque le taux de salaire nominal est égale à w_0 , la solution du modèle est déterminée de façon simultanée par la forme réduite du système d'équations suivant :

$$\bar{M} = p.L(i, Q)$$

$$\frac{I}{p}(i) = \frac{S}{p}(Q)$$

$$p = w_0 / f'(f^{-1}(Q))$$

En tenant compte de la relation (67) qui lie la production et l'emploi, ce sous-système permet d'analyser l'effet d'une augmentation de l'offre de monnaie sur l'emploi. En effet, la combinaison de la première et de la troisième relation ci-dessus donne :

$$\frac{\bar{M}}{w_0} = \frac{L(i, Q)}{f'(N)} = \frac{L(i, f(N))}{f'(N)}$$

On pose $g(i, N) = \frac{L(i, f(N))}{f'(N)}$. La fonction g est une fonction croissante en sa deuxième variable puisque la fonction de production f est croissante et que la productivité marginale du travail f' est décroissante. Dans ces conditions, une hausse de la masse monétaire \bar{M} induit une augmentation du niveau de l'emploi N .

Dans ce modèle de Modigliani, la compatibilité entre l'équilibre économique et la situation de chômage involontaire s'appuie sur l'hypothèse de rigidité à la baisse du salaire nominale. Pourtant, dès le début du chapitre 19 de la Théorie Générale, Keynes critique la théorie classique car elle rend cette rigidité responsable du non ajustement du système économique. Dans le même chapitre, Keynes montre que la flexibilité du salaire n'assure pas le plein-emploi, car une baisse du salaire réduit la demande globale qui est le principal déterminant de l'emploi d'équilibre. Ainsi, comme le relève Patinkin (1971), la rigidité du salaire ne constitue pas une hypothèse dans la Théorie Générale, mais plutôt une conclusion que Keynes dégage après un examen détaillé des effets d'une baisse du salaire.

2. La base empirique de l'arbitrage inflation-chômage

Friedman reconnaît que les keynésiens intègrent dans leurs modèles empiriques des relations qui déterminent le prix et le salaire, mais il estime que ces équations sont le plus souvent construites de façon *ad hoc* (Friedman, 1971, 1972). Il considère que c'est à la suite de la publication de l'article de Phillips (Phillips, 1958) que les keynésiens ont entrepris des tentatives sérieuses pour fonder théoriquement une relation entre les variables réelles du modèle IS-LM et l'évolution des prix.

Pour Patinkin(1972), les prémices de la courbe de Phillips étaient déjà présentes dans les chapitres 19 et 21 de la Théorie Générale. Il rappelle également l'intérêt porté par les keynésiens à la question de l'endogénéisation du prix et du salaire dans leurs modèles bien avant la publication de l'article de Phillips en 1958. Il cite à cet égard le travail pionnier réalisé par Klein (Klein et Goldberger, 1955, voir la section 1 du chapitre 1) qui comporte une formulation explicite pour le comportement du salaire nominal :

$$\frac{dw}{dt} = f(\bar{N} - N)$$

Dans cette équation le salaire nominal w évolue en fonction de l'écart entre l'offre exogène de travail \bar{N} et la demande de travail N . De plus, Klein suppose que les travailleurs prennent en compte l'évolution du niveau général des prix dans leurs négociations salariales.

C'est seulement dans la version révisée de son modèle (Klein et Goldberger, 1966) que Klein remplace la formulation ci-dessus du salaire par la courbe de Phillips. Des développements supplémentaires concernant le comportement des prix sont introduits dans le modèle de Wharton, en particulier l'introduction d'une équation du *mark-up* (Evans, Klein, Schink, 1967).

Dans sa version initiale, l'équation de Phillips (1958) établit, pour les données du Royaume-Uni sur la période 1861-1957, un lien entre la variation du taux de salaire monétaire et le taux du chômage. Cette relation qui est non linéaire à la fois dans les variables et dans les paramètres, prend la forme simple suivante :

$$w' = \alpha + \beta.U^{-\lambda}$$

A la lumière des débats présentés ci-dessus, la relation décroissante qu'exprime cette équation entre la variation du taux de salaire nominal et le niveau du taux de chômage n'est pas nouvelle. Ce qui est nouveau c'est l'affirmation de la stabilité d'une telle relation sur une longue période.

D'après cette équation, il existe un niveau de chômage U_0 qui correspond à une stabilisation de l'évolution du taux de salaire. De même, cette relation montre que, lorsque le taux de chômage croît indéfiniment, il existe un niveau planché égal à α pour la variation du taux de salaire. La courbe de Phillips semble fournir une preuve empirique à tous ceux qui supposent que la rigidité à la baisse du salaire monétaire est une caractéristique de l'économie keynésienne et qu'elle doit, de ce fait, être intégrée dans le modèle IS-LM.

L'équation initiale de Phillips a subi deux principales modifications avant de servir, durant la seconde moitié des années soixante et le début des années soixante-dix, comme un outil de politique économique, offrant la possibilité d'un arbitrage entre un niveau réduit de chômage contre une inflation modérée (Desai, 1981). La première contribution importante est fournie par Lipsey (1960) qui a procédé à l'estimation de l'équation suivante pour les données du Royaume Uni sur la période 1882-1959 :

$$w'_t = \alpha_0 + \alpha_1.U_t^{-1} + \alpha_2.U_t^{-2} + \alpha_3.U_t + \alpha_4.p'_t + \varepsilon_t$$

Cette équation constitue une approximation linéaire en les paramètres de l'équation initiale de Phillips dans le sens où le terme $U^{-\lambda}$ est remplacé par U_t^{-1} et U_t^{-2} . Dans ces conditions, l'équation peut être estimée par la méthode des moindres carrés ordinaires, d'où la présence du terme d'erreur ε_t . Lipsey retient une explication théorique de la courbe de Phillips à partir de la combinaison de deux comportements. D'un côté, une relation positive entre la variation du salaire et un excès de demande sur le marché du travail, et, d'un autre côté, une relation négative entre l'excès de demande de travail et le taux de chômage.

La seconde principale contribution est issue des travaux Solow et Samuelson (1960) qui ont formalisé ces travaux empiriques pour fournir au modèle keynésien, dans un contexte de

hausse de l'inflation, l'équation manquante pour expliquer le comportement de l'évolution des prix. La solution consiste en un système constitué d'une forme simplifiée de la courbe de Phillips et d'une équation qui relie la croissance du prix à celle du coût unitaire du travail :

$$w_t' = \alpha_0 + \alpha_1 U_t^{-1} + \alpha_2 \cdot p_t' \quad (72)$$

$$p_t' = w_t' - \pi_t' \quad (73)$$

où π_t' représente le taux de croissance de la productivité du travail. La substitution dans la deuxième équation de w_t' par sa valeur tirée de la première équation permet de relier directement le niveau de l'inflation mesuré par p_t' au niveau du taux de chômage par la formule suivante :

$$p_t' = \frac{\alpha_0 + \alpha_1 U_t^{-1} - \pi_t'}{(1 - \alpha_2)}$$

En intégrant cette relation dans le modèle IS-LM, la politique monétaire, et plus généralement la politique économique, ne peut plus avoir pour seul objectif la réalisation du plein-emploi. La réduction du chômage devient contrainte par l'inflation et pour préserver la stabilité des prix, il est nécessaire d'accepter un certain niveau de chômage. Dans ces conditions, l'objectif de la politique économique se déplace et devient contraint par le choix social de la combinaison optimale entre inflation et chômage. Ainsi, la courbe de Phillips offre au décideur politique un cadre théorique de référence qui lui permet d'effectuer un arbitrage discrétionnaire qui permet de maintenir l'économie à des niveaux désirés de chômage et d'inflation.

Cette courbe de Phillips keynésienne est considérée par Friedman (1975) comme le reflet d'une demande de travail qui relie le salaire nominal au taux de chômage. A cet effet, il estime que cette courbe souffre d'un défaut fondamental, à savoir qu'elle ne fait pas de distinction entre le taux de salaire nominal et le taux de salaire réel (Friedman, 1968). Phillips est dans ce cas accusé par Friedman (1975) de confondre taux de salaire nominal et taux de salaire réel sous l'influence d'un environnement keynésien qui considère les prix comme une donnée institutionnelle. Friedman qualifie l'analyse de Phillips de fallacieuse, car en reliant le taux de chômage au taux de salaire nominal, la courbe de Phillips serait incompatible avec la

théorie néoclassique pour laquelle c'est le taux de salaire réel qui doit être intégré dans cette relation.

Ainsi, la courbe de Phillips a été considérée, à tort, comme l'équation manquante qui permet de relier le secteur monétaire et le secteur réel dans le modèle de la synthèse. En fait, le modèle IS-LM a besoin d'une équation qui détermine un niveau d'équilibre du prix et non pas la variation des prix (Friedman, 1975). Dans ces conditions, la courbe de Phillips serait vérifiée lorsque les prix ou leurs anticipations sont stables, car dans ces conditions les salaires nominaux et les salaires réels évoluent de la même façon. De fait, il y aurait pour chaque niveau moyen de croissance des prix une position donnée de la courbe de Phillips, et cette position de la courbe de Phillips serait d'autant plus élevée que la croissance moyenne ou anticipée des prix est élevée (Friedman, 1968).

D'un point de vue économétrique, les critiques de Friedman portent sur la valeur du paramètre α_2 dans l'équation (72). Autrement dit, Phillips est accusé de supposer dans son travail empirique, soit que les prix sont constants, c'est-à-dire p_t' est nul, soit que le paramètre α_2 est nul alors qu'il doit être égal à 1. Autrement dit, l'équation originelle de Phillips serait mal spécifiée. Pourtant, la plupart des études économétriques montre que ce paramètre α_2 est inférieur à 1 et Friedman accepte ces résultats.

Sous-section 2 : Objectif de stabilité des prix et règle monétariste

1. Courbe de Phillips verticale à long terme

A partir d'une critique de la version keynésienne de la courbe de Phillips, Friedman propose une reformulation de cette courbe en la considérant comme l'expression d'un processus de négociation salariale reflétant le comportement de l'offre de travail en distinguant le court terme et le long terme (Friedman, 1968, 1975). En effet, étant donné qu'un contrat de travail couvre généralement une longue période, les employeurs et les employés sont conduits à évaluer à l'avance quel serait le niveau du salaire réel correspondant à un taux de salaire nominal donné et ils doivent par conséquent faire des anticipations à propos du niveau futur des prix. Dans ces conditions, le salaire qui permet l'ajustement entre l'offre et la demande de

travail est le salaire réel anticipé et non plus le salaire nominal ou le salaire réel observé (Friedman, 1975). A court terme, la courbe de Phillips doit alors se réécrire :

$$\dot{\omega} - \dot{p}^* = f(U) \quad (74)$$

où $\dot{\omega}$ désigne la variation du salaire nominal, \dot{p}^* est la variation du niveau anticipé des prix et U le niveau du taux de chômage.

Friedman introduit également le concept de "taux naturel de chômage" qui constitue une norme par rapport à laquelle l'évaluation du niveau élevé ou faible du taux de chômage observé est effectuée, et qu'il qualifie par ailleurs de chômage de "friction" ou de "transition" (Friedman, 1975). Le "taux naturel de chômage" est le taux de chômage qui correspond au taux de salaire réel donné par un modèle d'équilibre général walrasien qui tient compte des imperfections sur les marchés des biens et du travail, des changements aléatoires de la demande et de l'offre, le coût de la recherche d'information sur les emplois vacants et les disponibilités en main d'œuvre et les coûts de mobilité (Friedman, 1968). Dans ce cadre, un niveau de chômage plus élevé, correspondant à une offre de travail excédentaire, se traduit par une croissance moins rapide des salaires nominaux que celle des prix anticipés et exerce ainsi une pression à la baisse sur le taux de salaire réel et non sur le taux de salaire nominal.

En paramétrant chaque courbe de Phillips correspondant à l'équation (74) par le niveau anticipé de l'inflation, on peut réécrire cette courbe sous une forme qui permet de tracer cette courbe dans un repère qui relie le salaire nominal au taux de chômage :

$$\dot{\omega} = a + b.U^{-1} + \dot{p}^* \quad (75)$$

où a et b sont des paramètres. Dans ce cas, le taux naturel du chômage U_0 , atteint lorsque $\dot{p}^* = \dot{p}$, est donné par :

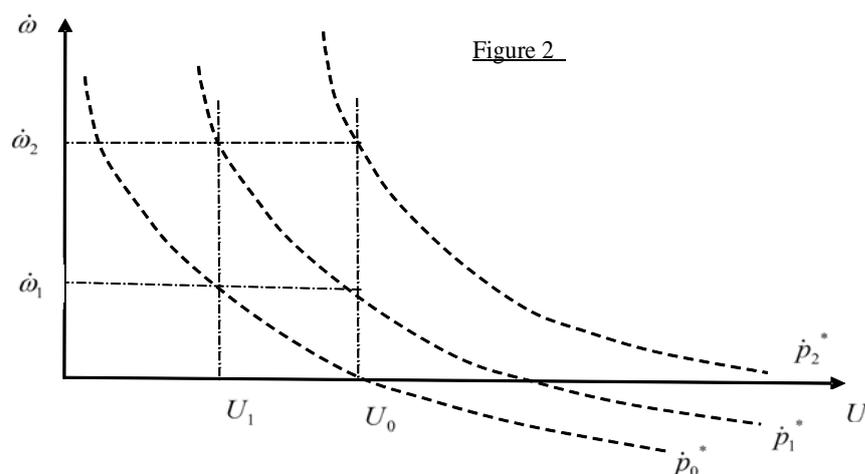
$$U_0 = \frac{b}{\dot{\pi} - a} \quad (76)$$

si l'on suppose que le taux de croissance de la productivité du travail $\dot{\pi}$ est non nul. Graphiquement, l'équation (75) est représentée par une branche d'hyperbole dont la pente dépend de la valeur du paramètre b (voir figure 2 ci-dessous).

Cette nouvelle version de la courbe de Phillips est utilisée par Friedman (Friedman, 1968) pour analyser les effets d'une politique monétaire expansionniste sur l'emploi et la production. Le raisonnement se fait à partir d'une situation de stabilité des prix et où le taux de chômage se trouve à son niveau naturel (le point U_0 sur la figure 1). La hausse de la masse monétaire à un niveau supérieur à celui qui est désiré par les agents se traduit par une augmentation du revenu nominal, de la dépense agrégée et par une progression des prix.

A cette étape de l'analyse, Friedman introduit une distinction entre le comportement des producteurs et celui des salariés à l'égard de la perception de l'augmentation de l'inflation. D'un côté, les salariés ne perçoivent pas cette augmentation des prix et maintiennent leur anticipation de stabilité des prix. Ils assimilent la hausse de leur revenu nominal à une augmentation de leur salaire réel anticipé et augmentent par conséquent leur offre de travail. D'un autre côté, étant plus informés sur les prix de leur production, les producteurs interprètent la hausse de la demande et de son prix comme une augmentation du prix relatif de leurs produits et donc comme une baisse du salaire réel, ce qui les conduit à augmenter leur demande de travail. La combinaison de ces deux situations se traduit par une hausse de l'emploi et une baisse du chômage de son taux naturel U_0 à U_1 .

Progressivement, se rendant compte de l'augmentation des prix des biens qu'ils achètent et de la baisse de leurs salaires réels, les travailleurs vont réajuster à la hausse leur anticipation de l'inflation, la portant de \dot{p}_0^* à \dot{p}_1^* et revendiquer une revalorisation de leurs salaires à $\dot{\omega}_1$ (Friedman, 1968, 1975). Le salaire réel va ainsi augmenter et obliger les producteurs à réduire leur excès de demande de travail. La courbe de Phillips se déplace de la position où le taux d'inflation anticipé est \dot{p}_0^* à celle où ce taux augmente à \dot{p}_1^* et le taux de chômage retourne à son niveau initial U_0 (voir figure 2).



Afin de maintenir le taux de chômage à son niveau U_1 , les autorités monétaires doivent augmenter à nouveau la masse monétaire, ce qui *in fine* conduit à un nouveau déplacement de la courbe de Phillips, correspondant à une inflation plus élevée \dot{p}_2^* . Ainsi, on aura une suite de courbes de Phillips qui se déplacent vers le haut au fur et à mesure que l'inflation anticipée augmente. Le taux de chômage ne peut être temporairement réduit qu'au prix d'une inflation de plus en plus élevée puisqu'il aura toujours tendance à retourner à son niveau "naturel" U_0 . Il existe alors bel et bien un arbitrage entre une inflation non anticipée et le chômage, mais cet arbitrage n'est pas permanent, il est seulement temporaire. De plus, l'absence d'anticipation de l'inflation conduit à un processus d'accélération dans les hausses successives des taux d'inflation. Graphiquement, ce processus accélérationniste découle de la forme particulière de la pente de la courbe de Phillips.

A long terme, une expansion monétaire se traduit entièrement par une hausse de l'inflation et n'exerce aucun effet sur la production et l'emploi. Ainsi, la courbe de Phillips permet à Friedman d'établir un lien entre sa théorie de court terme du revenu monétaire et la dichotomie classique de long terme entre les variables monétaires et les variables réelles (Desai, 1981). Ce résultat est obtenu en postulant, d'une part, l'existence d'un taux de chômage naturel uniquement déterminé par les facteurs réels de l'offre et indépendant des facteurs monétaires. D'autre part, Friedman impose une élasticité unitaire du salaire nominal à l'inflation anticipée dans l'équation (75), correspondant à l'absence d'illusion monétaire à long terme et à une courbe de Phillips qui devient verticale à cet horizon temporel (Friedman, 1975).

Pour Desai (1981), l'interprétation proposée par Friedman de la courbe de Phillips n'apporte aucun élément théorique fondamentalement différent de l'analyse keynésienne. La principale différence entre Friedman et les keynésiens porte sur le degré de l'illusion monétaire. En effet, pour les keynésiens, la valeur de l'élasticité du salaire nominal au prix dans l'équation (75) est strictement inférieure à l'unité. Tobin (1972) considère que Friedman ne fait qu'adopter l'arbitrage fourni par la courbe de Phillips keynésien, mais avec la disparition de cet arbitrage à long terme.

2. Politique monétaire monétariste

A partir de sa nouvelle interprétation de la courbe de Phillips, Friedman place l'argument de l'inflation non anticipée au cœur de la justification de l'arbitrage de court terme entre l'inflation et le chômage. Les autorités monétaires peuvent user de la surprise d'une hausse de la masse monétaire pour provoquer une baisse temporaire du chômage. Mais, l'accélération de l'inflation et l'effet nul à long terme sur l'emploi qui résultent du processus d'une expansion monétaire répétée finit par décourager les autorités monétaires de poursuivre une telle politique. L'enseignement général qui est tirée par Friedman de son analyse de la courbe de Phillips, c'est que les autorités monétaires peuvent contrôler l'évolution des variables monétaires, mais qu'elles ne peuvent utiliser cette capacité de contrôle pour stabiliser les grandeurs réelles (Friedman, 1968).

De fait, Friedman déplace l'objectif de la politique économique de celui de lutte contre le chômage vers celui de la stabilité des prix. La stabilité des prix assure un meilleur fonctionnement du système économique et favorise l'activité des producteurs et des consommateurs qui ne risquent pas d'être surpris par des variations inattendus des prix. En adoptant un taux de croissance modéré de la masse monétaire, la politique monétaire peut éviter de provoquer des périodes d'accélération de l'inflation et garantir des niveaux de prix et de taux d'intérêt nominaux faibles à long terme.

On retrouve alors les mêmes recommandations de politiques économiques formulées par Friedman dans ses travaux de la fin des années 1950 jusqu'au début des années 1960. Dans la section 2 ci-dessus, nous avons montré le poids des arguments empiriques pour la formulation de ces politiques. La nouvelle lecture de la courbe de Phillips permet à Friedman de fournir

un fondement théorique à la décomposition du revenu nominal entre production et prix et à l'origine monétaire de l'inflation, mais l'ampleur des arguments empiriques est toujours présente.

En effet, Friedman (1968) considère que l'ampleur et la variabilité du délai d'influence de l'offre de monnaie sur les prix rendent difficile la prévision exacte de l'effet de la politique monétaire sur les prix. Ainsi, en raison de cette difficulté d'appréciation, toute tentative de contrôle direct des prix risque d'introduire encore plus de perturbations dans l'activité économique. De plus, l'importance du délai séparant les actions des autorités monétaires de leurs effets doit les inciter à éviter les changements brusques et forts de politique face à des événements imprévus.

Pour atteindre l'objectif de stabilité des prix, Friedman recommande l'adoption d'une règle de croissance régulière d'un agrégat monétaire se situant entre 3% et 5%. Cette règle est justifiée par la combinaison de l'argument relatif à la variabilité et la longueur des délais d'action de la monnaie avec celui relatif aux observations historiques montrant une correspondance entre les périodes de stabilité de la croissance de la masse monétaire et de stabilité de l'activité économique.

Conclusion

Ce chapitre a permis de montrer qu'avant la publication de la nouvelle lecture friedmanienne de la courbe de Phillips en 1968, la proposition centrale des monétaristes du lien causale entre la monnaie et le revenu s'appuie exclusivement sur des arguments empiriques. Il met également en évidence le rôle central de l'identité de Fisher dans le corpus théorique des monétaristes, pour qui la monnaie exerce d'abord une influence sur le niveau général des prix avant le taux d'intérêt nominal.

La nouvelle lecture de la courbe de Phillips permet au monétariste d'expliquer la dynamique de la décomposition du revenu nominal entre la production et les prix le long d'un équilibre walrasien. Elle confirme l'existence d'un arbitrage entre l'inflation et le chômage, mais uniquement à court terme. De même, la reformulation de cette courbe fonde théoriquement les recommandations de politique monétaire formulées à la fin des 1950 concernant l'adoption d'une règle de croissance constante de la masse monétaire pour assurer l'objectif de stabilité des prix.

Chapitre 3

L'Evaluation de la Politique Monétaire par les Modèles VAR : l'Empirie Face à l'Impératif Théorique

Introduction

Ce chapitre expose les difficultés économétriques et les problèmes méthodologiques que rencontre l'approche des modèles VAR (Vectoriel AutoRegressif) pour l'évaluation de la politique monétaire. Ces modèles ont été introduits par Sims dans un contexte de vifs débats entre Tobin (1970) et Friedman (1970) sur le mode d'articulation entre théorie et économétrie pour la justification de la causalité entre la monnaie et le revenu. Dès son article de 1958, Friedman s'appuie sur l'observation de corrélation et de décalages temporels entre ces variables pour justifier l'influence causale de la monnaie sur le revenu monétaire. Tobin rejette la justification de la causalité par l'antécédence temporelle et développe un modèle théorique keynésien qui génère les décalages observés par Friedman, mais où la monnaie ne joue aucun rôle causal sur l'activité économique.

La position de Tobin s'inscrit dans la tradition de modélisation de la Commission Cowles, selon laquelle la relation causale est un concept théorique et non pas empirique (Leamer, 1985 ; Zellner, 1979 ; Cooley et LeRoy, 1985). Les keynésiens préfèrent l'estimation des paramètres d'un modèle dans le cadre d'équations structurelles simultanées et insistent sur l'importance des *a priori* théoriques pour interpréter les corrélations observées entre les données (Desai, 1981). Ils conçoivent l'exogénéité comme une forme de restriction imposée sur les paramètres, nécessaire à l'identification de la forme structurelle du modèle.

C'est dans ce contexte que Sims (1972) propose de soumettre les hypothèses d'exogénéité de la monnaie et de la présence d'une causalité entre celle-ci et le revenu à des tests économétriques directs et précis s'appuyant sur le test de causalité de Granger. L'intérêt de ce concept de G-causalité est, selon Sims, d'offrir la possibilité de tester aisément et directement la présence d'une causalité unidirectionnelle, qu'il identifie avec une hypothèse de stricte exogénéité. Les limites du caractère a-théorique du concept de G-causalité ont été largement soulignées (Zellner, 1979; Pierce, 1977; Leamer, 1985; Cooley et LeRoy, 1985) et de vives critiques ont porté sur la fragilité et le manque de robustesse des résultats des tests de G-causalité et de S-causalité (Feige et Pearce, 1979).

L'introduction de la modélisation VAR a permis à Sims (1980, a, b), suite à sa célèbre critique, d'étendre l'analyse de la causalité au sens de Granger à un vecteur de plus de deux

variables. Dans cette critique, Sims reproche aux modèles structurels de la Commission Cowles de comporter trop d'hypothèses théoriques non testées empiriquement et présente les modèles VAR comme une alternative à ces modèles structurels. Mais, la sensibilité des résultats des VAR utilisés par Sims aux choix des spécifications retenus a été mise en évidence par plusieurs auteurs (Eichenbaum et Singleton, 1986; Runkle, 1987; Spencer, 1989)

De plus, une transformation sur le système VAR s'est avérée nécessaire pour pouvoir interpréter les innovations canoniques comme des chocs exogènes de politique monétaire dans le cadre d'un VAR dit structurel (VARs). Ainsi, l'évolution de la modélisation VAR conduit à une situation paradoxale où, pour justifier les restrictions d'identification imposées aux innovations contemporaines, Sims se réfère à la notion de chaîne causale introduite par Wold (1954). La volonté de la méthode des VAR de s'affranchir de l'*a priori* théorique s'est alors révélée illusoire (Pagan, 1987). De plus, l'analyse des réponses impulsionnelle dans le cadre des modèles VAR fournit, selon Hoover (2009), un bon exemple de ce que Cartwright (2007) considère comme un contrefactuel "imposteur". Dans un VAR structurel, l'intervention de la Banque centrale est perçue comme un choc sur un système stable, alors que l'on sait pertinemment que la politique monétaire n'est pas conduite à travers des chocs.

Par ailleurs, l'assimilation des corrélations conditionnelles avec un ordre causal dans le cas de l'approche VAR ne peut se justifier, selon Cooley et LeRoy (1985), que sous une hypothèse de prédétermination qui est non testable en l'absence de restrictions *a priori* dérivées de la théorie. En l'absence de restrictions sur-identifiantes, les VARs sont tous observationnellement équivalents (Hoover, 2004 ; Hoover et Demiralp, 2003), puisqu'ils ont la même forme réduite qui fait l'objet d'une estimation empirique explicite. L'indétermination empirique de la causalité dans un VAR, liée au problème de l'équivalence observationnelle (Basman, 1965), impose l'adoption d'un schéma d'identification sur la base d'*a priori* théorique pour identifier les chocs de politique monétaire. Ceci constitue un cas extrême du problème de la sous-détermination de la théorie par les données soulevé par la thèse de Duhem-Quine (Duhem, 1906; Quine, 1951).

Section 1. Modèles VAR et causalité monnaie-revenu

A la suite du débat qui a opposé Friedman (1970) et Tobin (1970) sur le sens de la causalité entre la monnaie et le revenu, Sims (1972) utilise le test de causalité de Granger et confirme le résultat des monétaristes. Suite à sa célèbre critique des modèles structurels de la Commission Cowles, Sims (1980 a, b) introduit les modèles VAR et étend le concept de G-causalité à un vecteur de plus de deux variables. L'application de la méthode des VAR à un vecteur composé de la monnaie, du revenu, des taux d'intérêt et des prix permet à Sims de créer une structure causale plus complexe, à la faveur de la conception keynésienne de l'endogénéité de la monnaie (Paulré, 2007). Ainsi, les effets sur le revenu qui étaient auparavant attribués à la monnaie sont maintenant essentiellement expliqués par le taux d'intérêt. Toutefois, les modèles VAR affichent une forte sensibilité aux choix de spécification et exigent l'imposition d'un ordre causal, *a priori*, pour pouvoir interpréter les innovations canoniques comme des chocs exogènes.

Sous-section 1. G-causalité et S-causalité

1. Exogénéité de la monnaie et G-causalité

Le débat théorique sur la causalité entre la monnaie et le revenu qui a opposé Tobin et Friedman au début des années 1970 a mis en évidence l'écart profond qui existe entre les keynésiens et les monétaristes sur le mode d'articulation entre théorie et économétrie. Rappelons que Friedman (1958) s'appuie sur l'observation d'une corrélation et de décalages temporels entre le cycle économique et la variation de la masse monétaire pour justifier l'influence causale de la monnaie sur le revenu monétaire. Il utilise cet argument pour recommander l'adoption d'une politique monétaire passive, s'appuyant sur une règle de croissance constante de la masse monétaire, et il met en garde contre l'utilisation de la politique monétaire pour la stabilisation conjoncturelle. Tobin (1970) rejette la justification de la causalité par l'antécédence temporelle et il développe un modèle théorique keynésien qui génère les décalages observés par Friedman, mais où la monnaie ne joue aucun rôle causal sur l'activité économique.

La position de Tobin s'inscrit dans la tradition de la modélisation de la Commission Cowles, selon laquelle la relation causale est un concept théorique et non pas empirique (Leamer, 1985 ; Zellner, 1979 ; Cooley et LeRoy, 1985). Les keynésiens préfèrent l'estimation des paramètres d'un modèle dans le cadre d'équations structurelles simultanées et insistent sur l'importance des *a priori* théoriques pour interpréter les corrélations observées entre les données (Desai, 1981). Jusqu'au début des années 1970, les procédures de test de la Commission Cowles sont dominantes et les monétaristes n'ont pas réussi à offrir un cadre alternatif de tests statistiques, plus rigoureux que l'observation d'une corrélation, pour défendre et justifier leur hypothèse d'une influence causale de la monnaie sur le revenu.

C'est dans ce contexte que Sims (1972) propose de soumettre les hypothèses d'exogénéité de la monnaie et de la présence d'une causalité entre celle-ci et le revenu à des tests économétriques directs et précis. Il admet que l'observation d'une association positive entre la monnaie et le revenu ne peut être tenue pour une preuve de la causalité. Le phénomène observé peut très bien être le résultat d'une réaction passive et retardée de la monnaie aux fluctuations de l'activité économique. En s'appuyant sur une variante de la définition de la causalité de Granger (1969), Sims (1972) développe un test de causalité qui lui permet de confirmer l'hypothèse monétariste de l'influence causale de la monnaie sur le revenu et de rejeter les résultats du modèle proposé par Tobin (1970). Pour Sims (1972), l'intérêt supposé de ce concept de G-causalité est qu'il offre la possibilité de tester aisément et directement la présence d'une causalité unidirectionnelle qu'il identifie à une hypothèse d'exogénéité.

En raison de l'importance que revêt l'explication causale de la relation entre deux variables économiques, la définition opérationnelle de la causalité proposée par Granger (1969) dans le cadre des séries chronologiques a suscité un grand intérêt aussi bien pour l'économétrie théorique que pour les praticiens (Feige et Pearce, 1979). Cette définition de la causalité sera à la base de nouvelles méthodes de tests économétriques et au cœur des débats des années 1970 et 1980 sur les procédures de tests des propositions théoriques (Desai, 1981).

Le critère proposé par Granger pour déterminer la causalité s'appuie sur l'analyse des relations dynamiques existantes entre les séries chronologiques, mais sans référence à aucune restriction théorique *a priori* (Desai, 1981). La définition de la causalité se réfère uniquement à une condition de minimalité de la variance σ^2 de l'erreur de prévision ε_t . Ainsi, une variable Y "cause" une variable X si l'utilisation de toute l'information disponible U permet

d'améliorer la prédiction de X par rapport à la prédiction qui ne tient pas compte de Y (Granger, 1969), autrement dit, si :

$$\sigma^2(X|\bar{U}) < \sigma^2(X|\bar{U} - \bar{Y})$$

où les notations $\sigma^2(X|\bar{U})$ et $\sigma^2(X|\bar{U} - \bar{Y})$ désignent, respectivement, les variances des erreurs de prévision $\varepsilon_t(X|\bar{U})$ et $\varepsilon_t(X|\bar{U} - \bar{Y})$ définies à l'instant t par :

$$\varepsilon_t(X|\bar{U}) = X_t - (\hat{X}_t|\bar{U})$$

$$\varepsilon_t(X|\bar{U} - \bar{Y}) = X_t - (\hat{X}_t|\bar{U} - \bar{Y})$$

Avec $(\hat{X}_t|\bar{U})$ et $(\hat{X}_t|\bar{U} - \bar{Y})$ les prédicteurs des moindres carrés de X_t conditionnels, respectivement, à l'information passée \bar{U} de U et à l'information passée de U hormis les valeurs passées \bar{Y} de Y .

En cherchant à utiliser toute l'information disponible, la définition de la causalité ci-dessus comporte, selon Granger, un aspect "irréel" (Granger, 1969). Granger propose alors de réviser cette définition en retenant le sous-ensemble d'information contenant les séries "pertinentes" pour l'analyse causale. Lorsque par exemple l'ensemble d'information U se réduit aux deux séries X et Y , la condition " Y cause X " se réécrit :

$$\sigma^2(X|\bar{X}, \bar{Y}) < \sigma^2(X|\bar{X})$$

En choisissant d'utiliser la variance comme mesure de la prédiction, Granger impose aux séries chronologiques d'être stationnaires avant de subir le test de causalité (Granger, 1969). Lorsque la variable n'est pas stationnaire, la mesure de sa variance va dépendre du temps et l'existence d'une relation causale sera alors altérée au cours du temps. Comme la majorité des variables économiques ne sont pas stationnaires, la mise en œuvre de ce test dépend du choix d'une technique particulière pour les rendre stationnaires (Zellner, 1979 ; Paulré, 1985).

Afin de rendre cette définition de la causalité plus pratique et testable, Granger impose également une restriction sur la forme du prédicteur qu'il écrit sous la forme linéaire suivante :

$$(\hat{X}_t|\bar{X}, \bar{Y}) = \sum_{j \geq 1} a_j X_{t-j} + \sum_{j \geq 1} b_j Y_{t-j}$$

où les paramètres a_j et b_j sont fixés de façon à minimiser l'expression $\sigma^2(X|\bar{X},\bar{Y})$.

Ainsi, une variable constitue une cause au sens de Granger si elle améliore la prédiction d'une autre variable. Sous ces conditions simplificatrices, le modèle causal soumis par Granger au test de la condition "Y cause X" s'écrit sous la forme :

$$X_t = \sum_{1 \leq j \leq m} a_j X_{t-j} + \sum_{1 \leq j \leq m} b_j Y_{t-j} + \varepsilon_t \quad (1)$$

Granger envisage également le cas de la causalité bidirectionnelle due à l'existence d'une rétroaction de "X cause Y" en plus de celle de "Y cause X" :

$$Y_t = \sum_{1 \leq j \leq m} c_j X_{t-j} + \sum_{1 \leq j \leq m} d_j Y_{t-j} + \eta_t \quad (2)$$

Les processus ε_t et η_t représentent des bruits blancs, les variables X_t et Y_t sont supposées stationnaires. Granger montre que la définition de la condition "Y cause X" est équivalente à l'existence d'au moins un paramètre non nul b_j . De même, la condition de "X cause Y" est équivalente à l'existence d'au moins un paramètre non nul c_j .

Les deux équations (1) et (2) peuvent se réécrire sous la forme de la représentation autorégressive condensée suivante :

$$X_t = a(L)X_t + b(L)Y_t + \varepsilon_t \quad (3)$$

$$Y_t = c(L)X_t + d(L)Y_t + \eta_t \quad (4)$$

avec $a(L)$, $b(L)$, $c(L)$ et $d(L)$ des polynômes de l'opérateur retard L défini par $L^k X_t = X_{t-k}$. L'hypothèse nulle d'absence de causalité de Y vers X est équivalente à la condition que le polynôme $b(L)$ est identiquement nul dans l'équation (3). Dans ce cas, la variance minimale de l'erreur de prévision de X_t en utilisant le passé de X_t est égale à σ_ε^2 et cette variance ne peut être réduite par l'utilisation du passé de Y_t (Granger, 1969).

Pour Paulré (1985), Granger est soucieux de fournir une définition opérationnelle de la causalité qui puisse offrir un cadre pratique de diagnostic et de test pour la causalité. Mais, cet

objectif est réalisé au prix de l'imposition de plusieurs hypothèses simplifications et restrictions par rapport à la définition générale initiale de la causalité, dont, notamment, une sélection *a priori* des variables pertinentes pour le phénomène étudié

2. La S-Causalité de la monnaie-revenu

Sims (1972) n'applique pas directement le test de Granger au cas des variables du revenu et de la monnaie, mais il développe un nouveau test qui s'appuie sur la régression de la variable endogène sur les valeurs passées, présentes et futures de la variable exogène. Pour cela, Sims (1972) procède à une adaptation de la définition de la G-causalité au cas où deux séries chronologiques X et Y , supposées stationnaires, sont écrites sous la forme d'une représentation moyenne mobile de Wold (1954) :

$$X_t = P_{11}(L)\varepsilon_t + P_{12}(L)\eta_t \quad (5)$$

$$Y_t = P_{21}(L)\varepsilon_t + P_{22}(L)\eta_t \quad (6)$$

où ε_t et η_t sont des processus bruits blancs, les $P_{ij}(L), i, j = 1, 2$, sont des polynômes de l'opérateur retard L défini par $L^k X_t = X_{t-k}$. On a, par exemple, $P_{12}(L)\eta_t = \sum_0^{+\infty} P_{12k}\eta_{t-k}$. Sous cette formulation, Sims montre que la condition "Y ne cause pas X" au sens de Granger est équivalente à la condition $P_{11}(L)$ est nul ou $P_{12}(L)$ est nul. Dans le cas par exemple où $P_{12}(L)$ est nul, la relation (5) se réduit à :

$$X_t = P_{11}(L)\varepsilon_t$$

ou encore :

$$\varepsilon_t = P_{11}^{-1}(L)X_t$$

Si on porte dans l'équation (6) cette expression de ε_t , on obtient (Desai, 1981; Feige et Pearce, 1979) une représentation de Y_t en fonction des valeurs présentes et passées de X_t :

$$Y_t = P_{21}(L)P_{11}^{-1}(L)X_t + P_{22}(L)\eta_t \quad (7)$$

A partir de ce résultat, Sims (1972) montre que l'hypothèse nulle correspondant à l'absence de causalité au sens de Granger de Y vers X est équivalente à la condition de nullité des coefficients des valeurs futures de X dans la régression de Y sur les valeurs présentes, passées et futures de X . Autrement dit, le test de la causalité de Sims consiste à estimer un modèle de la forme :

$$Y_t = \sum_{-n_1}^{n_2} \alpha_j X_{t-j} + \varepsilon_t \quad (8)$$

et de tester l'hypothèse de nullité de tous les coefficients correspondant à des valeurs futures de X , c'est-à-dire que tous les $\alpha_j = 0$ pour $-n_1 \leq j < 0$, où n_1 et n_2 sont des entiers positifs (Feige et Pearce, 1979). Le rejet de l'hypothèse nulle implique que Y *G-cause* X .

Sims (1972) applique cette méthode au revenu nominal Y et à la masse monétaire M sur des données trimestrielles pour les Etats-Unis pour la période 1947-1969. Il utilise la masse monétaire M sous deux formes, celle de la base monétaire (la monnaie fiduciaire plus les réserves) et celle correspondant à l'agrégat M_1 (la monnaie fiduciaire plus les dépôts).

La vérification de l'hypothèse de nullité de certains des coefficients α_j s'appuie sur un test de Fisher qui suppose que les résidus ε_t ne sont pas autocorrélés. Or, cette dernière condition n'est en général pas vérifiée comme le montre, par exemple, la forme du terme de l'erreur dans l'équation (7). Pour résoudre ce problème, Sims applique un filtre particulier de la forme $(1 - 1,5L + 0,562L^2)$ aux deux variables Y et M exprimées en logarithme et notées y et m , dans la suite. Puis, il estime par la méthode des moindres carrés ordinaires les deux équations suivantes, avec un nombre de retards fixé à 8 et celui des futures fixé à 4 :

$$y_t = \sum_{j=0}^8 \alpha_j m_{t-j} + \sum_{j=1}^4 \beta_j m_{t+j} + \varepsilon_t \quad (9)$$

$$m_t = \sum_{j=0}^8 \gamma_j y_{t-j} + \sum_{j=1}^4 \delta_j y_{t+j} + \eta_t \quad (10)$$

Les résultats de Sims montrent que les valeurs futures du revenu sont significativement non nulles dans la régression de la masse monétaire (équation 10) et confirment l'existence d'une causalité de la monnaie vers le revenu. En revanche, les valeurs futures de la masse monétaire

ne sont pas significatives dans la régression du revenu (équation 9), ce qui constitue, d'après ce test, une preuve de l'absence d'une causalité inverse du revenu vers la monnaie.

Ces résultats constituent pour Sims (1972) la preuve de la réponse passive du revenu aux mouvements de la masse monétaire. Ils permettent, selon lui, le rejet de l'hypothèse opposée selon laquelle la monnaie réagirait de façon passive aux fluctuations de l'activité économique sans exercer une influence sur le revenu.

De plus, Sims considère que la G-causalité unidirectionnelle de la monnaie vers le revenu est équivalente à l'exogénéité de la monnaie. Plus généralement, il estime que la vérification de cette condition est nécessaire pour assurer la cohérence de l'estimation d'une relation dynamique qui intègre des variables autres que les retards de la variable endogène. Le résultat de l'exogénéité de la monnaie constitue, d'après Sims, un argument contre DeLeeuw et Kalchbrenner (1969) qui soutiennent que la base monétaire est endogène à l'activité économique et qui critiquent son traitement exogène dans la forme réduite du modèle de la Réserve Fédérale de St. Louis de Anderson et Jordan (1968). Feige et Pearce (1979) remarquent, toutefois, que l'estimation et l'interprétation des résultats de l'équation de Anderson et Jordan s'appuient sur l'hypothèse d'exogénéité des dépenses publiques sans aucun test de causalité.

Les conclusions de Sims à propos du sens de causalité entre la monnaie et le revenu ont été confirmées par le travail de Elliot (1975), mais elles ont été également contredites par les études de Bisignano (1974) et de Ireland et Morgan (1976). Cette opposition entre les conclusions des tests de causalité empiriques portant sur le même phénomène jette un doute sur la robustesse de ces résultats et soulève la question de la sensibilité de ces tests à certains choix empiriques retenus (Feige et Pearce, 1979).

Etant équivalent au concept de causalité de Granger, le test de S-causalité est soumis aux mêmes critiques concernant l'application d'un filtre *a priori* sur les données et le choix d'un nombre arbitraire de retards à inclure dans les estimations. L'utilisation d'un filtre a pour seule fonction d'éviter l'autocorrélation des résidus qui permet d'appliquer les tests de Fisher (Feige et Pearce, 1979). Toutefois, en dépit de l'application d'un filtre, Sims reconnaît que les résidus issus de la régression du revenu sur M_1 ne forment pas un bruit blanc au seuil de 5%, ce qui

montre que l'application d'un filtre particulier ne garantit pas toujours l'élimination de l'autocorrélation des erreurs.

Pour vérifier la sensibilité des résultats des estimations au type de filtre retenu, Feige et Pearce (1979) réestiment les équations de Sims (1972) en appliquant un filtre de différence première et un filtre ARIMA, pour des données, respectivement, ajustées et non ajustées pour la saisonnalité. Rappelons que l'étude de Sims est menée dans un contexte de données corrigées des variations saisonnières. Les résultats obtenus par Feige et Pearce, sur des données ajustées, à l'aide de la différence première confirment l'existence d'une S-causalité unidirectionnelle de la monnaie vers le revenu. En revanche, la régression de M_1 sur le revenu et celle du revenu sur la base monétaire affichent la présence d'une autocorrélation des résidus mesurée par la Q-statistique, ce qui entoure d'incertitude les F-statistiques utilisées. L'utilisation du filtre ARIMA conduit à une absence de S-causalité entre les deux variables. De même, l'exploitation de données non ajustées de la saisonnalité montre l'absence de S-causalité.

Feige et Pearce (1979) procèdent également à une réestimation des équations de Sims en utilisant des longueurs différentes pour les retards passés et futures et aboutissent à des valeurs des tests de Fisher fortement différentes. Même si ces grands écarts entre les tests ne permettent pas de rejeter les conclusions de Sims, ils sont le signe de la forte sensibilité de cette causalité au choix de la longueur des retards. Le choix d'un nombre de retards inapproprié pourrait conduire à l'omission d'un retard pertinent, ce qui risque de générer de l'autocorrélation dans les résidus de la régression, affecter les valeurs des coefficients et biaiser ainsi les tests d'hypothèse.

Afin d'examiner la robustesse de l'inférence causale au type de test utilisé, Feige et Pearce (1979) appliquent les tests de causalité de Granger et de Sims aux mêmes données et sur la même période que celles retenues dans l'étude de Sims (1972). Les auteurs vérifient l'existence d'une G-causalité entre le revenu et la masse monétaire mesurée, comme dans le cas de l'étude de Sims, par M_1 et par la base monétaire, respectivement. En utilisant les mêmes notations que ci-dessus, la mise en œuvre du test de Granger exige l'estimation des deux équations suivantes :

$$y_t = \sum_{j=1}^n \alpha_j y_{t-j} + \sum_{j=1}^k \beta_j m_{t-j} + \varepsilon_t$$

$$m_t = \sum_{j=1}^n \gamma_j m_{t-j} + \sum_{j=1}^k \delta_j y_{t-j} + \eta_t$$

Les tests sont menés sur des données corrigées et non corrigées des variations saisonnières et pour une valeur de k égale à 4 et des valeurs de n égales, respectivement, à 4 et à 8. Parmi les seize régressions effectuées, une seule estimation confirme l'existence d'une G-causalité allant de la base monétaire vers le revenu, alors que toutes les autres estimations montrent que le revenu ne G-cause pas la monnaie et la monnaie ne G-cause pas le revenu. L'utilisation de données non ajustées pour la saisonnalité conduit même à rejeter l'hypothèse de la G-causalité de la monnaie vers le revenu.

Les conclusions différentes auxquelles conduisent les applications des tests de G-causalité et de S-causalité, pourtant équivalents selon Sims, aux mêmes données soulèvent la question de la portée de l'explication causale de l'économétrie des séries chronologiques. Les résultats de Feige et Pearce (1979) jettent également un doute sur l'utilité de ces procédures pour tester les hypothèses d'exogénéité postulées par la théorie économique.

Sous-section 2. Critique de Sims et ses implications

1. Causalité et analyse impulsionnelle

Christopher Sims (1980 a) interprète la méthode *d'identification* des modèles structurels issues des travaux de la *Commission Cowles* comme une tentative d'établir une correspondance entre le modèle et la réalité. Toutefois, il rejette la distinction, qui est à la base de l'estimation des formes réduites de ces modèles, entre variables endogènes et variables exogènes sur une base théorique et introduit la modélisation VAR (Vecteur Auto-Régressif) comme une alternative à la modélisation structurelle de la Commission Cowles. Il reproche à ces modèles économétriques de comporter trop d'hypothèses théoriques non testées empiriquement, telles que le choix des variables exogènes, l'exclusion de certaines variables des équations du modèle ou encore la nature des décalages temporels dans le cas des modèles dynamiques. Ces restrictions de spécification, sans lesquelles selon Sims (1980 a) la forme structurelle du modèle ne serait pas identifiable, ne paraissent pas "crédibles" car elles sont

Avec : $Y_{t-j} = (y_{1,t-j}, y_{2,t-j}, \dots, y_{n,t-j})'$, $\theta^0 = (\theta_1^0, \theta_2^0, \dots, \theta_n^0)'$ un vecteur $n \times 1$ de termes constants, $\varepsilon_t = (\varepsilon_{1,t}, \varepsilon_{2,t}, \dots, \varepsilon_{n,t})'$ un vecteur de bruits blancs de matrice de variance-covariance $V(\varepsilon_t) = \Sigma$ de taille $n \times n$ et $\theta^1, \theta^2, \dots, \theta^p$ une suite de matrices carrées de taille $n \times n$ ayant la forme pour chaque $j = 1, \dots, p$:

$$\theta^j = \begin{pmatrix} \theta_{11}^j & \theta_{12}^j & \dots & \theta_{1n}^j \\ \theta_{21}^j & \theta_{22}^j & \dots & \theta_{2n}^j \\ \dots & \dots & \dots & \dots \\ \theta_{n1}^j & \theta_{n2}^j & \dots & \theta_{nn}^j \end{pmatrix}$$

où les θ_{ik}^j , pour $i, k = 1, \dots, n$, sont les paramètres estimés du modèle. Sims (1980 a) utilise le terme innovation pour qualifier le résidu de chaque variable du modèle VAR dans le sens où il constitue la composante "nouvelle" qui est non prévisible à partir des valeurs passées de cette variable.

Les deux seules contraintes introduites dans ce type de modélisation portent sur le choix des n variables y_1, y_2, \dots, y_n et le nombre de retards p , qui est le même pour toutes ces variables. En revanche, cette modélisation n'impose aucune contrainte *a priori* d'exclusion d'une variable ou de l'un de ses retards d'une équation du système VAR. De même, le modèle n'impose aucune hypothèse d'exogénéité *a priori* sur certaines variables.

Pour apprécier l'effet causal d'une variable du VAR sur les autres, Sims (1978) introduit les méthodes de l'analyse impulsionnelle et de la décomposition de la variance de l'erreur de prévision. Ces techniques cherchent à déterminer l'influence de l'innovation liée à une variable sur les autres variables à partir d'une réécriture du VAR sous la forme d'une représentation moyenne mobile. En effet, étant donné que toutes les variables d'un modèle VAR sont endogènes, il n'est pas possible d'effectuer un choc sur l'une de ces variables pour évaluer son influence sur les autres endogènes.

Dans son étude sur la causalité entre la masse monétaire et le revenu dans un cadre multivarié, Sims (1980 b) établit une équivalence entre l'étude de la G-causalité de la monnaie vers le revenu et l'analyse de la contribution des chocs qui affectent la masse monétaire dans la variance du revenu. En particulier, cette technique devrait permettre, d'après Sims, de tester la

thèse du courant des Nouveaux Classiques selon laquelle ce sont des chocs non anticipés sur la masse monétaire qui expliquent sa non-neutralité.

En utilisant l'opérateur retard L , le processus VAR donné par l'équation (11) peut se mettre sous la forme :

$$\Phi(L)Y_t = \varepsilon_t \quad (12)$$

avec $\Phi(L) = I - \Theta(L)$ et $\Theta(L) = \sum_{j=0}^p \theta^j L^j$ sont des matrices de polynômes en L , I étant la matrice unité.

D'après le théorème de Wold (1954), lorsque le processus Y_t est stationnaire, le polynôme $\Phi(L)$ peut être inversé, ce qui permet d'exprimer Y_t en fonction des valeurs présentes et passées des innovations :

$$Y_t = \Phi(L)^{-1} \varepsilon_t \quad (13)$$

ou encore :

$$Y_t = \varphi^0 + \sum_{j=0}^{+\infty} \varphi^j \varepsilon_{t-j} \quad (14)$$

Cette transformation permet d'écrire le VAR de l'équation (11) sous la forme d'une représentation moyenne mobile vectorielle infinie $VMA(\infty)$, dite représentation canonique (Sims, 1978). Dans ce cas, le modèle VAR est interprété comme une représentation de chocs aléatoires qui résultent des innovations canoniques ε_t .

Toutefois, la représentation $VMA(\infty)$ du VAR donnée par l'équation (14) présente une limite fondamentale pour l'examen de l'effet causal d'un choc sur une innovation ε_{it} (Sims, 1978). En effet, si l'innovation ε_{it} est corrélée avec une autre innovation ε_{kt} , alors un choc sur la première sera nécessairement accompagné par un choc sur la seconde. Ainsi, il ne sera pas possible de calculer la contribution de l'innovation de chaque variable y_{it} dans la variance du vecteur Y_t en présence d'une corrélation contemporaine entre les composantes de ε_t . Dans ces conditions, une analyse impulsionnelle ne peut être menée de façon isolée sur une innovation particulière que si la matrice de variance-covariance $V(\varepsilon_t) = \Sigma$ est diagonale.

Pour contourner ces difficultés, Sims (1978) propose, en faisant référence aux chaînes causales de Wold, d'appliquer une transformation orthogonale G au vecteur des innovations canoniques ε_t :

$$\mu_t = G\varepsilon_t$$

Parmi les transformations possibles, Sims retient la sous-famille des matrices G triangulaires inférieures. Si de plus, on normalise la variance de μ_t à l'unité en divisant ses composantes par leur écart-type, alors on peut montrer que la matrice de variance-covariance $V(\mu_t)$ est égale à la matrice identité I . Les éléments de μ_t sont interprétés comme des chocs structurels et l'équation (12) devient :

$$G\Phi(L)Y_t = G\varepsilon_t = \mu_t \quad (15)$$

Ce modèle est appelé un VAR structurel (VARs) car il intègre des restrictions identifiantes sur les variables contemporaines. En posant $\psi^j = \varphi^j G^{-1}$, l'analyse impulsionnelle peut se faire à partir de la réécriture suivante de l'équation (14) :

$$Y_t = \varphi^0 + \sum_{j=0}^{+\infty} \psi^j \mu_{t-j} \quad (16)$$

A chaque horizon temporel j , la colonne k de la matrice ψ^j constitue la réponse du système, après j périodes, à un choc indépendant et normalisé sur l'innovation de la composante k du vecteur $Y = (y_1, y_2, \dots, y_n)'$. Soit y_{kt} cette composante k de Y_t dont la représentation VMA s'écrit :

$$y_{kt} = \sum_{h=0}^{\infty} (\psi_{k1}^h \mu_{1,t-h} + \psi_{k2}^h \mu_{2,t-h} + \dots + \psi_{kn}^h \mu_{n,t-h})$$

où les ψ_{ki}^h sont les éléments de la ligne k de la matrice ψ^h qui sont interprétés comme des multiplicateurs dynamiques à l'horizon h . Dans ce cas, l'effet d'un choc structurel μ_{it} à un horizon temporel j sur la variable y_{kt} est égal à ψ_{ki}^j .

Une fois la représentation récursive (16) est imposée au modèle VAR, il est également possible de procéder à la décomposition de la variance de l'erreur de prévision de Y sous la

forme des contributions des différents chocs structurels. Par exemple, l'erreur de prévision de y_{kt} à un horizon temporel j s'écrit :

$$y_{k,t+j} - E(y_{k,t+j}) = \sum_{h=0}^{j-1} (\psi_{k1}^h \mu_{1,t+j-h} + \psi_{k2}^h \mu_{2,t+j-h} + \dots + \psi_{kn}^h \mu_{n,t+j-h})$$

Les $\mu_{i,t+j-h}$ étant non corrélés, on montre que la variance de cette erreur de prévision est donnée par :

$$E(y_{k,t+j} - E(y_{k,t+j}))^2 = \sum_{i=1}^n [(\psi_{ki}^0)^2 + (\psi_{ki}^1)^2 + \dots + (\psi_{ki}^{j-1})^2]$$

Pour chaque indice i , la somme $(\psi_{ki}^0)^2 + (\psi_{ki}^1)^2 + \dots + (\psi_{ki}^{j-1})^2$ représente la variance de l'erreur de prévision⁵ de la variable y_{kt} expliquée uniquement par le choc μ_{it} à l'horizon j . La contribution de ce choc à la variance totale de l'erreur de prévision de la variable y_{kt} est donnée par la formule :

$$\frac{(\psi_{ki}^0)^2 + (\psi_{ki}^1)^2 + \dots + (\psi_{ki}^{j-1})^2}{\sum_{i=1}^n [(\psi_{ki}^0)^2 + (\psi_{ki}^1)^2 + \dots + (\psi_{ki}^{j-1})^2]}$$

Cette méthode est mise en œuvre par Sims (1980 a, 1980 b) pour évaluer la contribution des chocs monétaires à la volatilité de la production ou des prix.

2. La robustesse de la causalité monnaie-revenu en question

L'introduction des modèles VAR a permis d'étendre l'analyse de la causalité à la Granger à un vecteur de plus de deux variables (Paulré, 2007). Sims (1980 b) utilise cette modélisation pour étudier les relations dynamiques entre quatre variables, à savoir le taux d'intérêt, la masse monétaire, le niveau des prix et la production. Il estime un premier modèle VAR qui intègre uniquement trois variables, la production, le prix et le stock de monnaie. Dans cette modélisation, Sims (1980 b) montre que les innovations du stock de monnaie représentent une

⁵ Par abus de langage, on parle parfois de variance de la variable.

contribution importante dans la variance de la production⁶, et que la variance de la monnaie est presque entièrement expliquée par ses propres innovations. Ce comportement de la monnaie est interprété par Sims comme une G-cause et lui permet de confirmer son premier résultat (Sims, 1972), à savoir que la monnaie G-cause la revenu. Sims vérifie également avec succès l'hypothèse de nullité des 12 coefficients correspondant aux retards de la production et du prix dans l'équation de régression de la masse monétaire (Sims, 1980 b).

Toutefois, l'application de la méthode des VAR à un vecteur composé de la monnaie, du revenu, des taux d'intérêt et des prix permet à Sims (1980 b) de créer une structure causale plus complexe, à la faveur de la conception keynésienne de l'endogénéité de la monnaie. Ainsi, les effets des chocs sur le revenu qui étaient auparavant attribués à la monnaie sont maintenant essentiellement expliqués par le taux d'intérêt (Paulré, 2007). Selon Sims, ce résultat jette un doute sur la pertinence de l'explication monétariste et confirme la thèse du rôle passif de la monnaie.

Dans les deux cas précédents, les modèles VAR sont estimés sur la période 1948-1978 pour des données mensuelles américaines. Les variables sont exprimées en logarithmes et chaque équation contient douze retards pour chacune des variables en plus d'une constante. Sims estime les coefficients de chaque équation séparément par la méthode des moindres carrés ordinaires. On présente une illustration de la démarche de Sims adoptée pour l'estimation et l'identification du second modèle VAR. Soient r , m , p et y les logarithmes, respectivement, du taux d'intérêt, de la monnaie, du prix et de la production. Le VAR intégrant ces quatre variables s'écrit sous la forme :

$$\begin{aligned}
 r_t &= \theta_r^0 + \sum_{j=1}^{12} \theta_{rr}^j r_{t-j} + \sum_{j=1}^{12} \theta_{rm}^j m_{t-j} + \sum_{j=1}^{12} \theta_{rp}^j p_{t-j} + \sum_{j=1}^{12} \theta_{ry}^j y_{t-j} + \varepsilon_{rt} \\
 m_t &= \theta_m^0 + \sum_{j=1}^{12} \theta_{mr}^j r_{t-j} + \sum_{j=1}^{12} \theta_{mm}^j m_{t-j} + \sum_{j=1}^{12} \theta_{mp}^j p_{t-j} + \sum_{j=1}^{12} \theta_{my}^j y_{t-j} + \varepsilon_{mt} \\
 p_t &= \theta_p^0 + \sum_{j=1}^{12} \theta_{pr}^j r_{t-j} + \sum_{j=1}^{12} \theta_{pm}^j m_{t-j} + \sum_{j=1}^{12} \theta_{pp}^j p_{t-j} + \sum_{j=1}^{12} \theta_{py}^j y_{t-j} + \varepsilon_{pt} \\
 y_t &= \theta_y^0 + \sum_{j=1}^{12} \theta_{yr}^j r_{t-j} + \sum_{j=1}^{12} \theta_{ym}^j m_{t-j} + \sum_{j=1}^{12} \theta_{yp}^j p_{t-j} + \sum_{j=1}^{12} \theta_{yy}^j y_{t-j} + \varepsilon_{yt}
 \end{aligned}$$

⁶ Plus précisément, il s'agit de la variance de l'erreur de prévision de ces variables.

où $\theta_r^0, \theta_m^0, \theta_p^0, \theta_y^0$ sont des constantes, $\varepsilon_{rt}, \varepsilon_{mt}, \varepsilon_{pt}, \varepsilon_{yt}$ sont les termes d'erreurs qui constituent des chocs sur les variables du VAR, et les θ_{ik}^j sont les paramètres à estimer du modèle et qui sont au nombre de 48 pour chaque équation.

Le calcul la matrice des variances-covariances Σ du vecteur ε_t des termes d'erreurs estimés montre que ceux-ci sont corrélés. Pour pouvoir interpréter causalement et de façon indépendante des interventions sur les innovations ε_{kt} et appliquer les méthodes de la réponse impulsionnelle et de la décomposition de la variance de l'erreur de prévision, Sims retient le schéma d'identification récursif suivant :

$$\varepsilon_{rt} = \mu_{rt} \quad (17)$$

$$\varepsilon_{mt} = \lambda_{rm} \cdot \mu_{rt} + \mu_{mt} \quad (18)$$

$$\varepsilon_{pt} = \lambda_{rp} \cdot \mu_{rt} + \lambda_{mp} \cdot \mu_{mt} + \mu_{pt} \quad (19)$$

$$\varepsilon_{yt} = \lambda_{ry} \cdot \mu_{rt} + \lambda_{my} \cdot \mu_{mt} + \lambda_{py} \cdot \mu_{pt} + \mu_{yt} \quad (20)$$

Avec les notations de la section précédente, cela revient à poser :

$$\varepsilon_t = G^{-1} \mu_t$$

où

$$G^{-1} = \begin{pmatrix} 1,0,0,0 \\ \lambda_{rm},1,0,0 \\ \lambda_{rp},\lambda_{mp},1,0 \\ \lambda_{ry},\lambda_{my},\lambda_{py},1 \end{pmatrix}$$

Cet ordre causal permet de relier, à travers les coefficients λ_{ik} , les innovations canoniques ε_{kt} à des chocs théoriques non observés μ_{kt} sur les variables du modèle dont la matrice des variances-covariances est égale à la matrice identité. Cette transformation lève l'indétermination qui caractérise l'effet d'un choc sur les innovations canoniques et permet d'identifier précisément le canal de transmission d'une variation des μ_{kt} (Todd, 1990).

La matrice de passage entre le vecteur ε_t et le vecteur μ_t a une forme triangulaire inférieure. Le choc canonique sur le taux d'intérêt ε_{rt} dépend uniquement du choc μ_{rt} et il est indépendant des autres chocs structurels $\mu_{mt}, \mu_{pt}, \mu_{yt}$, qui de ce fait n'exerceront aucun effet instantané sur

le taux d'intérêt dans le modèle VAR. A cet effet, le choc sur le taux d'intérêt μ_{rt} peut être considéré ici comme une variable de décision des autorités monétaires. En particulier, ce schéma d'identification implique que toutes les fluctuations communes à ε_{rt} et à ε_{mt} résultent uniquement de μ_{rt} . De façon récursive, le choc structurel μ_{rt} exerce un effet immédiat sur toutes les variables du VAR. Le choc μ_{mt} exerce un effet instantané sur la monnaie, le prix et la production. Le choc μ_{pt} a un effet immédiat sur le prix et la production et, enfin, le choc μ_{yt} n'exerce un effet instantané que sur la production.

Le choc sur la masse monétaire ε_{mt} est décomposé entre une composante $\lambda_{rm} \cdot \mu_{rt}$ qui reflète la réponse de la monnaie à un choc sur le taux d'intérêt, et un choc μ_{mt} qui est propre à la masse monétaire. De la même façon, le choc sur le prix ε_{pt} comporte une première composante $\lambda_{rp} \cdot \mu_{rt}$ qui reflète la réponse du prix à un choc sur le taux d'intérêt, une seconde composantes $\lambda_{mp} \cdot \mu_{mt}$ qui représente la réponse du prix à un choc sur la masse monétaire, et un troisième choc μ_{pt} qui est propre au prix. Finalement, le choc sur la production ε_{yt} intègre des réponses aux chocs sur les trois variables précédentes $\lambda_{ry} \cdot \mu_{rt}, \lambda_{my} \cdot \mu_{mt}, \lambda_{py} \cdot \mu_{pt}$.

L'utilisation de ce schéma d'identification permet à Sims de montrer que le choc μ_{mt} sur la masse monétaire joue un rôle faible dans l'explication de la production. A un horizon temporel de 48 mois, seulement 4% de la variance de l'erreur de prévision de la production est expliquée par le choc sur la monnaie (Sims, 1980 b). En revanche, le choc sur le taux d'intérêt μ_{rt} contribue à hauteur de 30% à la variance de l'erreur de prévision de la production. Etant conscient de la sensibilité des résultats de sa méthode au choix de l'ordre causal retenu,

Sims vérifie si l'utilisation d'une seconde identification permet de confirmer le rôle non causal de la monnaie sur la production. En effet, Sims a toujours insisté sur la nécessité de tester différents choix d'orthogonalisation triangulaire des innovations (Sims, 1978). Dans le cas présent, Sims se soucie de savoir si le rôle premier accordé au choc sur le taux d'intérêt dans la chaîne causale ci-dessus est responsable du résultat trouvé. Pour cela, il procède à l'inversion du rôle de ce choc avec celui de la monnaie dans les équations (17) et (18) tout en préservant ceux des équations (19) et (20). Dans ce nouveau schéma d'identification, la

décision de politique monétaire se transmet de la masse monétaire vers le taux d'intérêt, puis au prix et enfin à la production.

Les résultats fournis par cette nouvelle chaîne causale sont similaires à ceux trouvés dans la première identification. La part de la variance de l'erreur de prévision de la production qui est expliquée par le choc sur la monnaie n'excède pas 10%, alors qu'elle se situe dans la même proportion pour le choc du taux d'intérêt que dans la précédente simulation. Pour Sims, contrairement à la thèse monétariste, les fluctuations de la production ne semblent pas provenir d'un choc sur l'offre de monnaie, qu'il soit lié directement au stock de monnaie ou qu'il soit transmis par le taux d'intérêt.

Sims (1980 b) fournit également une interprétation alternative à la corrélation statistique mise en évidence par Friedman et Schwartz (1963) entre le revenu et la monnaie. L'examen des réponses impulsionnelles, à plusieurs horizons temporels, montre que le stock de monnaie et la production réagissent à la hausse et à la baisse à peu près de la même façon suite à un choc sur le taux d'intérêt. De plus, 56% de la variance de l'erreur de prévision de la masse monétaire est expliquée, à un horizon de 48 mois, par le choc sur le taux d'intérêt. Ces constats constituent, selon Sims, la preuve d'une réponse commune du stock de monnaie et de la production à des chocs sur le taux d'intérêt.

La méthode des modèles VAR développée par Sims comme une alternative à la modélisation structurelle de la Commission Cowles a fait l'objet à deux types de critiques. Le premier type de critiques, qui sera présenté dans la seconde suivante, est d'ordre méthodologique et porte sur l'incapacité de l'approche des modèles VAR à fournir des résultats qui puissent être interprétés comme étant structurels. Le second type de critiques concerne la robustesse et la sensibilité des résultats des VAR aux choix de spécifications retenus, en particulier il cherche à prouver que le rôle faible joué par la monnaie dans l'explication des fluctuations de la production résulte en fait d'une mauvaise spécification du VAR. Par exemple, les travaux de Eichenbaum et Singleton (1986), Runkle (1987) et Spencer (1989) montrent que la réponse impulsionnelle et la décomposition de la variance sont très sensibles à une modification de la spécification du VAR retenu par Sims. Ces auteurs étudient des spécifications alternatives de la représentation VAR retenue par Sims qui présentent une différence en ce qui concerne l'ordre des variables dans la chaîne causale, la longueur des retards dans le VAR, la méthode

pour stationnariser les séries, la fréquence de mesure des variables, l'indicateur de mesure des variables et la période d'estimation.

En réestimant le VAR de Sims sur des données trimestrielles, Runkle (1987) trouve une part de la variance de la production expliquée par le choc sur la monnaie qui atteint 28%. De plus, ce rôle important de la monnaie dans l'explication de la variance de la production est supérieur à celui fourni par le choc sur le taux d'intérêt qui n'est que de 27%. La fréquence des données utilisées pour l'estimation d'un VAR rend également ses résultats fragiles et peu précis. En effet, la probabilité de présence d'une corrélation instantanée entre les séries économiques augmente lorsque le degré d'agrégation temporelle de ces données s'accroît (Spencer, 1989). De ce fait, les résultats d'un VAR utilisant des données semestrielles ou annuelles sont beaucoup plus sensibles à l'ordre causal retenu que ceux obtenus à partir de données mensuelles ou trimestrielles.

L'influence du choix de l'ordre causal sur le résultat est d'autant plus important que la corrélation instantanée entre les innovations canoniques est forte (Spencer, 1989). L'invariance des résultats obtenus par Sims au choix de l'ordre causal est uniquement valable dans le cas des données mensuelles. Dans le cas de l'utilisation de données trimestrielles ou semestrielles, l'effet du choc de la masse monétaire sur la production est au moins deux à trois fois supérieur lorsque la monnaie se situe avant la production dans la chaîne causale que dans le cas contraire.

Section 2. Les VAR structurels et le retour obligé à la théorie

L'approche des modèles VAR a été incapable de fournir des résultats qui puissent être interprétés comme étant structurels sans imposer des restrictions d'identification aux innovations. Sans ces restrictions, les modèles VAR structurels (VARs) sont tous observationnellement équivalents (Basman, 1965) et, de fait, sont soumis au problème de l'indétermination empirique de la causalité lié au problème de la sous-détermination de la théorie par les données (Duhem, 1906; Quine, 1965). Par ailleurs, Hoover (2009) relève que l'utilisation d'un VAR pour l'évaluation de la politique monétaire constitue ce que Cartwright (2007) considère comme un contrefactuel "imposteur".

Sous-section 1. Causalité et exogénéité : questions de méthode

1. Le caractère a-théorique de la G-causalité

Pour Sims (1972), la définition de la G-causalité constitue l'analogie dans le cas dynamique des "chaînes causales" de Wold utilisées dans les modèles économétriques statiques. Mais, contrairement à la G-causalité qui est facilement testable, Sims estime que la structure de l'ordre causale de Wold n'est généralement pas testable dans un cadre multivarié en l'absence de restrictions théoriques sur-identifiantes (Sims, 1972). L'intérêt du concept de G-causalité, selon Sims (1972), est qu'il offre la possibilité de tester aisément et directement la présence d'une causalité unidirectionnelle de la monnaie vers le revenu.

Les modèles empiriques doivent constituer, d'après Sims, l'outil par lequel les théories sont confrontées à la réalité, et de fait les unes par rapport aux autres (Sims, 1980 a). Il estime (Sims, 1982) que les modèles structurels de la Commission Cowles peuvent être utilisés à des fins d'évaluation de la politique économique à condition de soumettre les hypothèses d'exogénéité des variables instruments aux tests de Granger ou de Sims. L'introduction des notions de G-causalité et de S-causalité est supposée offrir un cadre pour tester statistiquement l'hypothèse de l'exogénéité qui est assimilée à une absence de G-Causalité. Ainsi, la monnaie est considérée comme exogène par rapport au revenu car celui-ci ne G-cause pas la monnaie.

Certes, les VAR permettent d'étudier l'antécédence temporelle d'une variable par rapport à une autre, mais cela est loin de constituer une preuve pour la causalité. Comme le souligne Zellner (1979), en omettant de mentionner explicitement l'importance de la théorie et des lois économiques dans la définition de la causalité, Granger (1969) donne l'impression que celle-ci peut être obtenue à partir d'un critère purement statistique. En effet, ce dernier adopte le point de vue d'une relation temporelle asymétrique entre la cause et l'effet et ne fait aucune référence à une loi ou une théorie économique. Selon Granger (1969), l'écoulement du temps joue un rôle central dans cette définition puisqu'elle renferme l'idée de la succession de l'effet à la cause. Granger (1969) va plus loin en affirmant que le concept de causalité est de peu d'utilité dans la pratique si le temps n'est pas introduit dans l'analyse. D'après Paulré (1985), Simon (1953) rejette l'idée d'associer l'ordre causal avec la succession chronologique uniquement dans le but de faire apparaître une asymétrie entre la cause et l'effet.

La définition de la G-causalité, qui remonte à Wiener, est inhabituelle dans le sens où elle est intégrée dans un critère particulier de confirmation, à savoir la variance de l'erreur de prévision de l'estimateur des moindres carrés (Zellner, 1979). Cette forme de restriction, qui n'apparaît pas dans les concepts de causalité chez des auteurs comme Simon, Feigl ou Basmann, réduit le champ d'application de la définition de Granger. Ces auteurs n'identifient pas l'existence d'une asymétrie entre la cause et l'effet, comme celle retenue par Granger, nécessairement avec la succession chronologique de l'effet à la cause.

Zellner (1979) relève l'importance du concept de causalité en Economie et en Econométrie et note que ce concept ne peut être associé à une relation temporelle ou réduit à la minimisation de la variance de prévision d'un ensemble d'informations. Il adopte le concept fourni par Feigl (1953) selon lequel la causalité est définie en termes de prédictabilité par rapport à une loi. Selon cette définition, une prédictabilité sans la référence à une loi, ou à une théorie, ne serait pas une causalité. Relier la prédictabilité à une loi ou à un ensemble de lois est crucial pour évaluer les différents tests de causalité qui sont développés dans la littérature économétrique.

Pierce (1977) précise que le test de Granger s'articule autour d'une spécification empirique entre des séries chronologiques et que les données sont utilisées pour révéler l'existence ou l'absence de certaines formes d'interdépendances. Les relations examinées ne sont pas spécifiées conformément à une théorie précise, ce qui risque de conduire à retenir des formes incorrectes des variables ou d'exclure des variables pertinentes (Zellner, 1979). D'ailleurs,

Granger (1979) admet la possibilité de l'inférence d'une causalité fallacieuse ou fictive dans le cas de l'omission de variables importantes. A cet effet, Sims (1972) reconnaît que la méthode du test de Granger constitue bien une version sophistiquée du principe du "*post hoc ergo propter hoc*" qui fût rejeté par Tobin (1970) dans sa célèbre critique contre Milton Friedman.

Desai (1981) note la différence entre la définition de la G-causalité et la définition de la causalité chez les auteurs de la Commission Cowles, quoique les deux définitions impliquent l'estimation des paramètres d'un modèle. Dans un cas, il s'agit de minimiser l'erreur de prévision, alors que la causalité est définie dans le second cas en termes de validité des restrictions identifiantes de façon à assurer l'unicité de la transmission d'une variation des variables prédéterminées aux variables endogènes. De plus, Desai estime que la définition de la G-causalité n'intègre pas la vision de la Commission Cowles sur la simultanéité où il existe une rétroaction instantanée entre les deux variables X et Y . Même si la définition de la G-causalité suggère la possibilité d'une causalité instantanée, Granger considère que cette situation n'est qu'apparente et qu'elle disparaît si les variables sont mesurées avec une fréquence plus élevée.

Dans le prolongement du concept de G-causalité, dans l'approche VAR la propriété d'exogénéité et la direction de la causalité sont supposées constituer également des questions empiriques, à l'opposé du programme de la Commission Cowles qui traite ces concepts comme des données *a priori* et non testables (Leamer, 1985). Pour Cooley et LeRoy (1985), les modèles VAR sont davantage intéressés par les caractéristiques probabilistes des données que par les propriétés structurelles des paramètres. Mais, l'un des grands avantages de la recherche sur les modèles VAR, c'est l'attention grandissante qui fût accordée, selon Leamer (1985), aux trois concepts importants, quoique controversés, de causalité, exogénéité et structural.

La méthode des modèles VAR ne peut pas être utilisée, d'après Leamer (1985), pour effectuer une inférence causale à partir de données non expérimentales sans référence à une théorie. Lorsqu'une intervention sur le modèle ne peut être directement testée, une interprétation causale des corrélations observées exige la définition d'une structure théorique qui puisse servir pour la justification d'une modification des données comme si elles étaient générées dans un cadre expérimental. A cet effet, Leamer estime que la méthode des modèles VAR ne

peut pas constituer un substitut adéquat au programme de la Commission Cowles pour établir des inférences de la causalité.

Les auteurs de la Commission Cowles ont, en effet, proposé un cadre de travail qui suppose que les données sont engendrées par un modèle à équations simultanées, avec le choix d'un ensemble de variables exogènes et une connaissance *a priori* sur la direction de la causalité. Leamer (Leamer, 1985) précise que dans cette tradition, l'exogénéité est reliée à des restrictions sur les valeurs de certains paramètres. Selon Paulré (2007), Haavelmo et Frisch soulignent l'importance de l'invariance des relations du modèle, supposées représentées des mécanismes fondamentaux, à l'intervention extérieure et de l'intérêt des expérimentations hypothétiques pour l'interprétation causale.

Pour Cooley et LeRoy (1985), la causalité est en rapport avec l'intervention, qui nécessite une propriété d'invariance, laquelle n'est satisfaite que sous l'hypothèse d'exogénéité, dans le sens où une variable x est une cause précédant une variable y si et seulement si x est exogène et y est endogène. Une variable exogène est déterminée en dehors du modèle et il est envisagé de réaliser une expérience hypothétique qui consiste à évaluer l'effet de la modification de cette exogène sur les variables endogènes du modèle, en maintenant toutes les autres variables exogènes fixes. La notion d'exogénéité comprend l'idée d'intervention sur la variable exogène et elle implique également l'idée d'invariance sous intervention (Cooley et LeRoy, 1985).

L'hypothèse de *ceteris paribus* est retenue et cette restriction doit être définie sans ambiguïté, c'est-à-dire que toutes les autres variables exogènes restent fixes, de façon à ce que l'intervention hypothétique soit clairement définie. Pour que l'invariance exigée sous intervention ait un sens sans ambiguïté dans tous les contextes, particulièrement dans les grands systèmes, l'hypothèse que toutes les variables exogènes soient non corrélées est requise (Cooley et LeRoy, 1985).

2. Exogénéité et identification

Hoover (2001) rappelle que le débat au sein de la Commission Cowles sur la question de la causalité a connu un glissement progressif vers le problème de l'*identification*. Parmi les raisons qui expliquent cette situation, Hoover cite l'équivalence établie par Simon (1953)

entre la causalité et les conditions d'identification de la forme structurelle dont les paramètres sont déterminés, moyennant certaines restrictions théoriques, par l'estimation de la forme réduite. Depuis, le problème central de l'économétrie a été, selon Hoover (Hoover, 2009), celui de l'*identification*, c'est-à-dire, comment retrouver les paramètres du processus générateur des données à partir de l'observation des variables.

La question d'identification renvoie également au problème du biais de simultanéité dont souffrent les modèles à équations simultanées (Pirotte, 2004). En effet, Haavelmo (1943 a) a démontré l'inadéquation de la technique des moindres carrés pour l'estimation des paramètres d'un système à équations simultanées et que les estimateurs obtenus pour l'estimation de l'une des équations du système à l'aide de cette méthode sont biaisés et non convergents. La solution apportée par les chercheurs de la Commission Cowles à ce problème consiste en l'adjonction d'un nombre suffisant de restrictions, qui doivent être formulées par la théorie économique, sur les paramètres de la forme structurelle du modèle de façon à retrouver ces paramètres à partir de ceux estimés dans la forme réduite (Pirotte, 2004).

Dans ce cadre, les économistes de la Commission Cowles conçoivent l'exogénéité comme une forme de restriction sur les paramètres nécessaire à l'identification de la forme structurelle. En l'absence d'autres restrictions *a priori* issues de la théorie, l'hypothèse de l'exogénéité est sous-identifiée ou exactement identifiée, et donc non testable (Cooley et LeRoy, 1985). Dans le cas d'un ensemble de restrictions sur-identifiantes sur les paramètres, celles-ci peuvent être testées, et donc si l'hypothèse d'exogénéité est intégrée avec ces restrictions, un test joint sur ces restrictions et l'exogénéité peut être mené (Wu, 1973, Hausman 1978).

Dans un modèle statique où deux variables sont déterminées simultanément, aucune des deux ne peut être considérée comme exogène, ni comme causant l'autre. Il est généralement inapproprié de chercher à s'enquérir de l'effet d'un changement dans une variable endogène sur une autre endogène car le résultat de l'expérience qui consiste en la variation de l'endogène est ambigu (Cooley LeRoy, 1985). Pour illustrer cette situation, considérons le système générateur des données suivant entre la masse monétaire m et le revenu y :

$$m = \theta.y + \varepsilon_1 \quad (21)$$

$$y = \gamma.m + \varepsilon_2 \quad (22)$$

avec θ et γ des paramètres inconnus, ε_1 et ε_2 des perturbations non corrélées. Dans ce système, ε_1 est exogène relativement à m et ε_2 est exogène relativement à y . D'après ce qui précède, on ne peut pas dire que m ou y soit exogène l'une par rapport à l'autre, et ce, en dépit de leur corrélation conditionnelle. Ainsi, on ne peut pas interpréter l'une des deux variables comme causant l'autre.

On peut également montrer à partir de cet exemple que l'hypothèse de non-corrélation entre deux variables est une condition nécessaire, mais non suffisante pour l'exogénéité. Pour cela, définissons l'équation (23) suivante à partir d'une combinaison des équations (21) et (22) :

$$y = \delta \cdot m + \eta \quad (23)$$

où $\delta = \frac{\text{cov}(m, y)}{\text{var}(m)}$ et η est corrélée avec les variables exogènes ε_1 et ε_2 .

Le fait que η soit non corrélée avec m ne justifie pas son traitement comme une variable exogène. L'effet d'une variation de η est ambigu, car on ne sait pas si cette variation résulte d'un changement dans ε_1 , dans ε_2 ou dans les deux. Par exemple, une restriction *a priori* $\theta = 0$ assure l'exogénéité de m relativement à y et la corrélation conditionnelle de m relativement à y peut alors être identifiée avec une relation causale.

Dans le cas dynamique, et en recourant au concept de G-causalité, Sims (1972) cherche à tester statistiquement, sans restrictions *a priori*, le principe monétariste de l'exogénéité du stock de la monnaie contre l'idée keynésienne de l'endogénéité d'une composante importante de la masse monétaire. La mise en évidence d'une G-causalité unidirectionnelle qui va de la monnaie vers le revenu, sans rétroaction de ce dernier, est interprétée par Sims comme la preuve de l'exogénéité de la monnaie et de son influence causale sur le revenu (Sims, 1980 b). Ainsi, il qualifie de stricte exogénéité de la monnaie la condition que y_t ne G-cause pas m_t , c'est-à-dire que $\pi_{12} = 0$ dans le modèle VAR suivant :

$$m_t = \pi_{11} \cdot m_{t-1} + \pi_{12} \cdot y_{t-1} + u_{mt} \quad (24)$$

$$y_t = \pi_{21} \cdot m_{t-1} + \pi_{22} \cdot y_{t-1} + u_{yt} \quad (25)$$

où m_t et y_t expriment, respectivement, la masse monétaire et le revenu et les π_{jk} sont des paramètres à estimer.

Toutefois, Cooley et LeRoy (1985) considèrent que le concept d'exogénéité qui est pertinent pour l'analyse des interventions de la politique monétaire est celui de la *prédétermination* et non celui de la stricte exogénéité. Afin d'établir la relation entre la G-causalité, la *prédétermination* et la stricte exogénéité, supposons que le processus générateur des données dont est issue la forme réduite composée des équations (24) et (25) s'exprime sous la forme du système suivant :

$$m_t = \theta \cdot y_t + \beta_{11} \cdot m_{t-1} + \beta_{12} \cdot y_{t-1} + \varepsilon_{1t} \quad (26)$$

$$y_t = \gamma \cdot m_t + \beta_{21} \cdot m_{t-1} + \beta_{22} \cdot y_{t-1} + \varepsilon_{2t} \quad (27)$$

où θ , γ et les β_{jk} sont des paramètres inconnus. Les variables ε_{1t} et ε_{2t} sont sériellement et simultanément non corrélées. En effet, il n'existe *a priori* aucune raison pour exclure la possibilité de l'existence d'une relation instantanée et simultanée entre la masse monétaire et le revenu. Comme le note Hoover (2006), la simultanété rend difficile l'inférence causale, mais elle n'exclut pas la possibilité d'un ordre causal.

Dans ce dernier système, m_t et y_t sont des variables endogènes, et ε_{1t} et ε_{2t} sont exogènes. La variable m_t est dite *prédéterminée* relativement à y_t si $\theta = 0$. Dans ce cas, m_t est exogène par rapport à y_t et une intervention sur m_t est associée, sans ambiguïté, directement et uniquement à un changement dans ε_{1t} . L'effet d'une intervention d'un montant de Δm_t sur y_t est égal à $\gamma \Delta m_t$.

La relation entre le paramètre π_{12} de la G-causalité exprimée dans la forme réduite (24)-(25) et le paramètre θ de la *prédétermination* imposée dans le modèle structurel (26)-(27) est donnée par la formule :

$$\pi_{12} = \frac{\theta \beta_{22} + \beta_{12}}{1 - \theta \gamma} \quad (28)$$

Il apparaît ainsi que la non-causalité au sens de Granger n'est ni nécessaire ni suffisante pour la *prédétermination*. De fait, le test de Granger n'est pas pertinent pour savoir si une interprétation causale de la corrélation conditionnelle est justifiée. L'assimilation des corrélations conditionnelles avec un ordre causal dans le cas de l'approche VAR ne peut se justifier, selon Cooley et LeRoy (1985), que sous une hypothèse de *prédétermination* qui est

non testable en l'absence de restrictions *a priori* dérivées de la théorie. Le mérite qui était accordé aux tests de Granger et de Sims pour tester les théories apparaît comme une erreur.

De la même façon, la non causalité au sens de Granger ne peut pas être identifiée avec la stricte exogénéité qui est définie par les deux conditions suivantes :

$$\theta = 0 \text{ et } \beta_{12} = 0$$

La stricte exogénéité implique la non causalité de Granger, mais l'inverse n'est pas vrai. Plusieurs modèles avec des restrictions théoriques peuvent générer la non-causalité de Granger mais non la stricte exogénéité. De plus, le concept de la stricte exogénéité de m_t n'est pas nécessaire pour rendre non ambiguë l'effet sur y_t d'une intervention sur m_t .

La stricte exogénéité intervient lorsqu'on s'intéresse à une intervention non seulement sur une seule réalisation de la variable m_t , mais sur toute une trajectoire de cette variable. L'effet d'un tel changement dépend de la valeur du paramètre β_{12} dans l'équation (26). Dans le cas de la stricte exogénéité de m_t , c'est-à-dire, $\theta = 0$ et $\beta_{12} = 0$, l'effet sur y_t d'une seule réalisation de m_t est équivalent à celui d'une intervention sur toute la trajectoire de cette variable. En revanche, la non nullité du paramètre β_{12} autorise une accumulation des effets d'une intervention sur m_t à travers le terme $\beta_{12} \cdot y_{t-1}$. La stricte exogénéité assure ainsi uniquement l'équivalence entre les effets des deux types d'interventions.

Cooley et LeRoy (1985) estiment que ces concepts de *prédétermination* et de stricte exogénéité sont équivalents, dans le cas des modèles linéaires, aux concepts d'exogénéité faible et d'exogénéité forte qui ont été introduits par Engle, Hendry et Richard (1983) dans le cas des modèles non linéaires.

Sous-section 2. L'indétermination empirique de la causalité

1. Du paradoxe de la référence à Wold au VAR "contrefactuel imposteur"

Selon Sims (Sims, 1980 b), il existe une équivalence entre l'étude de la G-causalité et l'analyse de la contribution des chocs qui affectent une variable dans la variance des autres variables du modèle VAR. Sims exploite ce résultat (voir la section 1 ci-dessus) pour étendre l'analyse de la causalité à la Granger à un vecteur de plus de deux variables et il introduit les techniques de l'analyse impulsionnelle et de la décomposition de la variance de l'erreur de prévision à partir de la réécriture du modèle VAR sous la forme d'une représentation moyenne mobile infinie.

Mais, la présence d'une corrélation contemporaine entre les composantes du vecteur des innovations canoniques ε_t ne permet pas d'interpréter causalement et de façon indépendante des interventions sur les composantes ε_{it} de ces innovations. Pour dépasser ce problème, Sims impose un schéma d'identification récursif, faisant référence aux chaînes causales de Wold, qui lui permet d'établir un ordre causal entre les innovations canoniques ε_{it} et des chocs structurels théoriques non observés μ_{kt} . Dans ces conditions, l'évaluation de la politique monétaire se fait par action sur les termes aléatoires μ_{kt} , interprétés comme des chocs structurels exogènes, dans le cadre du modèle VAR structurel.

Par conséquent, la volonté de la méthode des modèles VAR de s'affranchir de l'*a priori* théorique s'est révélée illusoire (Pagan, 1987). En effet, l'évolution de la modélisation VAR conduit à une situation paradoxale où, pour justifier les restrictions d'identification utilisées pour permettre une interprétation causale de l'intervention sur les variables du VAR, Sims se réfère à la notion de chaîne causale introduite par Wold dans le cadre de ses travaux au sein de la Commission Cowles (Cooley et LeRoy, 1985; Paulré, 2007).

La préférence pour l'écriture d'un modèle sous une forme récursive est justifiée par Wold (1954) par la capacité d'une telle spécification à décrire correctement la succession des enchainements causaux observés ou supposés. Cet auteur estime qu'en l'absence d'une interprétation causale claire du modèle, celui-ci ne peut pas répondre à l'exigence explicative

des phénomènes qu'il privilégie par rapport à une approche descriptive. Pour cela, et par souci de transparence, chaque équation doit avoir une interprétation causale autonome et décrire un mécanisme particulier, indépendamment des autres équations du modèle.

Les causalités instantanées et les rétroactions sont exclues car, pour Wold, il existe toujours un intervalle de temps qui s'écoule entre le déclenchement d'une cause et l'apparition de son effet. Cette asymétrie entre la cause et l'effet conduit Wold à imposer un ordre pour les variables endogènes et un ordre pour les équations qui reflètent le sens de la causalité. La résolution du modèle se fait de façon unidirectionnelle puisqu'une même variable ne peut être à la fois une cause et un effet d'une autre variable. De la même façon, Strotz (1960) soutient que si une interprétation causale d'un système interdépendant est possible, alors ce système doit être mis sous une forme récursive.

Il faut rappeler que la position de Wold n'est pas partagée par tous les experts de la Commission Cowles. En particulier, Haavelmo et Koopmans n'excluent pas la possibilité de l'existence d'une causalité instantanée et d'une rétroaction entre les variables du modèle (Desai, 1981 ; Paulré, 1985). Selon Zellner (1979), Basmann (1963) reproche à Wold et à Strotz d'omettre plusieurs caractéristiques contenues dans d'autres définitions de la notion de causalité. Il cite, par exemple, le cas des équations structurelles dans un système interdépendant non nécessairement triangulaire et qui peuvent avoir une interprétation causale.

Les débats sur la causalité impliquent généralement deux questions principales, à savoir 1) la définition de la causalité et 2) la relation entre le concept de causalité et la forme du modèle économétrique (Zellner, 1979). Toutefois, ces débats s'accompagnent par la présence d'une multitude de définitions de la causalité, souvent incompatibles les unes par rapport aux autres. D'après Zellner (1979), Wold et Strotz (1960) soutiennent que personne n'a le monopole de la définition de la causalité et considèrent que le véritable enjeu est celui de fournir une explication à ce concept, mais que cette explication n'a pas besoin d'être unique.

Mais, ce choix de Wold et de Strotz représente une position *a priori* sur la nature de la causalité qui est non fondée méthodologiquement et qui ne justifie pas l'exclusion des autres formes de causalité. Par exemple, les définitions de la causalité de Basmann, Feigl et Jeffreys s'appliquent à l'ensemble des modèles économétriques, qu'ils soient écrits sous une forme

réursive, triangulaire ou interdépendante (Zellner, 1979). Ces modèles sont causaux dans un sens déductif et logique et n'exigent pas que chaque équation puisse être interprétée causalement. Imposer à chaque équation du modèle d'avoir une interprétation causale est une restriction *a priori* sur la forme du modèle qui n'a pas de justification méthodologique.

Par ailleurs, la définition de la causalité retenue par Wold et Strotz (1960) introduit, selon Zellner (1979), l'idée du contrôle des variables "*z est cause de y si, par hypothèse, il est ou il serait possible, en contrôlant z de contrôler indirectement y. Une relation causale est donc asymétrique par essence*". Dans son article de 1966, Wold présente la causalité comme étant la "*relation entre un stimulus et sa réponse dans les expérimentations contrôlées réelles ou fictives*" (Paulré, 1985). Cette définition exclut les situations où les variables ne peuvent pas être contrôlées. Zellner considère que Strotz et Wold exigent que le monde réel soit récursif en imposant une restriction *a priori* sur la forme d'un modèle économique.

Plus récemment, Cartwright (2006, 2007) a également évoqué la pluralité des théories de la causalité qui, en général, ont peu de contenu en commun. Dans ce contexte, chaque théorie présente des limites et fait l'objet de contre-exemples fournis par les théories alternatives. Souvent, le traitement de la causalité est valable dans un cas particulier, mais pas pour un autre cas. Cette situation reflète l'existence de différents systèmes causals et que l'affirmation d'une relation causale change en fonction des caractéristiques de ces systèmes.

Ainsi, il n'existe pas des théories concurrentes de la causalité, mais plutôt une pluralité de théories, chacune adaptée à une situation particulière parmi différents cas de systèmes de lois causales. Pour Cartwright, cela pose un problème de méthode, car il faut savoir quelle méthode est appropriée pour découvrir quel type de causalité.

Plus fondamentalement, elle postule que les théories de la causalité sont orientées, soit vers la caractérisation des causes, soit vers leur utilisation, mais pas les deux à la fois. Le problème central à la base de cette thèse est l'absence de connexion entre les méthodes de test et de caractérisation des propositions causales et les modes de justification de l'utilisation des structures causales établies. Cartwright rappelle que l'idée selon laquelle la connaissance des causes devrait permettre d'agir sur le monde est une idée dominante parmi les théories de la causalité.

La combinaison de cette idée avec les techniques d'inférence disponibles suggère que la représentation d'un ensemble de relations devrait permettre d'effectuer des inférences qui autoriseraient l'utilisation de ces causes comme des stratégies pour produire des effets. Cartwright (2006) précise que la représentation d'une relation comme étant causale n'offre pas, en général, des possibilités d'inférence à propos de son utilisation.

Cette critique de Cartwright s'adresse, particulièrement, à la théorie de la causalité de Woodward (2003) qui défend la causalité en termes de manipulation contrefactuelle et où justement cette connexion n'est pas possible. Woodward considère en effet, que la référence à des lois n'est pas indispensable pour doter l'économie d'une capacité explicative. Il adopte une interprétation interventionniste de la causalité qui s'appuie sur l'idée d'une manipulation expérimentale idéalisée.

La causalité au sens de Woodward est caractérisée par deux conditions. La première est une condition de modularité qui stipule que si une variable X est une cause d'une variable Y, alors X peut varier indépendamment de toutes les autres sources de variations de Y. La seconde est une condition d'invariance qui exige que la relation fonctionnelle entre les variables X et Y reste stable lorsque X varie en maintenant fixes les autres causes de Y. Cartwright juge que les conditions imposées dans cette définition sont trop fortes et trop restrictives par rapport aux situations prévues de mises en œuvre des politiques étudiées. La théorie de la causalité de Woodward fournit un contrefactuel pour tester ou établir des hypothèses causales, mais qui correspond rarement à la structure causale où on souhaite l'utiliser (Hoover, 2009).

Cartwright justifie son opposition à l'approche retenue par Woodward pour évaluer les relations causales à partir de la dissociation des liens de causalité par le fait que c'est le système intact et non pas le système déformé qui est nécessaire pour l'évaluation de la politique économique. Elle considère que le contrefactuel qui acquiert la justification empirique n'est pas celui qui devrait être utilisé dans une analyse contrefactuelle de la politique économique.

De façon générale, elle stipule que les techniques appropriées pour la caractérisation et le test de la causalité ne sont pas celles qui sont appropriées pour l'utilisation de cette causalité dans l'analyse contrefactuelle. Dans ces conditions, le décideur de la politique économique ne peut pas avoir une confiance dans le contrefactuel que Woodward cherche à évaluer, car, lors de la

mise en œuvre d'une politique on ne peut pas faire varier une cause indépendamment des autres causes. Pour ces raisons, Cartwright qualifie ce type de contrefactuel de '*contrefactuels imposteurs*'.

L'analyse des réponses impulsionnelle fournit, selon Hoover (2009), un bon exemple de ce que Cartwright considère comme un *contrefactuel imposteur*. Rappelons que dans le cas de l'étude de Sims, une fois le VAR structurel est estimé, il est utilisé pour répondre à la question contrefactuelle qui consiste à augmenter la masse monétaire, par surprise, de x% pendant une période et de la ramener à sa valeur initiale à la période suivante. Cette hausse est considérée comme un choc monétaire, effectué par action sur le terme d'erreur, et la réaction des variables du modèle à ce choc est appelée fonction de réponse impulsionnelle.

Dans ces conditions, l'intervention de la Banque centrale est perçue comme un choc sur un système stable, alors que l'on sait pertinemment que la politique monétaire n'est pas conduite à travers des chocs. Ainsi, la fonction de réponse impulsionnelle ne peut pas reproduire les effets d'un changement systématique de la politique monétaire. En effet, il n'est pas possible dans le cadre d'un VAR structurel de forcer une variable à évoluer dans une direction donnée, car cela constituerait une violation de l'hypothèse du caractère aléatoire des termes d'erreurs du modèle.

A cet effet, pour Cartwright (2006), l'intervention sur les innovations ne peut pas être associée aux changements dans les paramètres de la politique monétaire que l'on souhaite manipuler. Un changement systématique dans la politique monétaire ne peut être analysé dans ce cadre tout en supposant simultanément que la structure causale du modèle, qui inclut les termes aléatoires, est la fois stable et qu'elle se modifie. D'ailleurs, Sims (1999) reconnaît que des chocs aléatoires sur les termes d'erreurs ne peuvent pas fournir des réponses adéquates à des questions contrefactuelles, particulièrement lorsque la politique monétaire est caractérisée par l'adoption d'une règle (Hoover, 2009).

2. La sous-détermination de la théorie par l'observation

L'écriture d'un modèle sous une forme récursive ou triangulaire revient à imposer des restrictions de nullité sur certains paramètres du modèle. Au niveau empirique, cette forme de

restriction pourrait être justifiée si le modèle récursif pouvait fournir de meilleures prévisions qu'un modèle interdépendant. Or, Basmann (1965) a démontré que les deux formes de modèle sont *observationnellement équivalentes* et, de ce fait, il dénie tout rôle au choix de la spécification du modèle récursif dans la réalisation d'un meilleur ajustement aux observations relativement à un modèle interdépendant. Avec les mêmes restrictions, les formes réduites associées aux deux types de modèle peuvent générer les mêmes prévisions. Comme le rappelle Paulré (1985), Wold a lui-même reconnu l'équivalence de la forme récursive avec certaines formes non récursives, remettant ainsi en cause la supériorité statistique des chaînes causales.

En plus des limites du caractère restrictif de la forme récursive de Wold, la propagation des chocs sur les innovations dépend de la nature de l'ordre causal récursif retenu, dans le sens où un changement de cet ordre causal peut conduire à des causalités opposées (voir la section 1 ci-dessus). Le choix d'un ordre causal instantané ne peut pas être déterminé statistiquement à partir des données, mais la justification de ce choix exige le recours à des considérations théoriques préalables. Il y a autant d'ordres causals possibles que de permutations possibles des éléments du vecteur Y_t dans le modèle VAR structurel donné par l'équation (15) :

$$G\Phi(L)Y_t = \mu_t \quad (15)$$

En l'absence de restrictions sur-identifiantes, ces VARS sont tous *observationnellement équivalents* (Hoover, 2004 ; Hoover et Demiralp, 2003), puisqu'ils ont la même forme réduite d'origine, qui fait l'objet d'une estimation empirique explicite, donnée par l'équation (12) :

$$\Phi(L)Y_t = \varepsilon_t \quad (12)$$

Pour Cooley et LeRoy (1985), l'invariance du résultat d'un VAR au choix de l'ordre causal dans la forme récursive ne justifie pas une interprétation causale de l'intervention sur les innovations orthogonalisées. Ils considèrent que, comme pour le cas de la G-causalité, en l'absence d'une hypothèse de *prédétermination*, l'interprétation des fonctions des réponses impulsionnelles en termes d'un ordre causal est complètement injustifiée, même si les innovations orthogonalisées sont mutuellement non corrélées.

Rappelons, à cet effet, que la condition de non G-causalité exprimée par la nullité du paramètre π_{12} , donné par la relation (28) :

$$\pi_{12} = \frac{\theta\beta_{22} + \beta_{12}}{1 - \theta\gamma} \quad (28)$$

est compatible avec une infinité des paramètres θ et γ du processus générateur des données :

$$m_t = \theta \cdot y_t + \beta_{11} \cdot m_{t-1} + \beta_{12} \cdot y_{t-1} + \varepsilon_{1t} \quad (26)$$

$$y_t = \gamma \cdot m_t + \beta_{21} \cdot m_{t-1} + \beta_{22} \cdot y_{t-1} + \varepsilon_{2t} \quad (27)$$

Tous ces modèles structurels ont la même forme réduite constituée des équations (24) et (25) et sont donc tous *observationnellement équivalents* :

$$m_t = \pi_{11} \cdot m_{t-1} + \pi_{12} \cdot y_{t-1} + u_{mt} \quad (24)$$

$$y_t = \pi_{21} \cdot m_{t-1} + \pi_{22} \cdot y_{t-1} + u_{yt} \quad (25)$$

Autrement dit, en l'absence d'une hypothèse de *prédétermination*, l'estimation des paramètres π_{jk} ne permet pas de fixer de façon unique les valeurs des paramètres du modèle structurel, ce qui nous ramène au problème de l'identification.

Pour montrer l'importance de la *prédétermination* également dans le cas de l'analyse des fonctions impulsionnelles, Cooley et LeRoy (1985) considèrent la transformation qui consiste à remplacer l'équation (25) par elle-même moins l'équation (24) multipliée par le coefficient

$$\delta = \frac{\sigma_{my}}{\sigma_{mm}}, \text{ où } \sigma_{mm} = \frac{\sigma_1^2 + \theta^2 \sigma_2^2}{(1 - \theta\gamma)^2} \text{ est la variance du terme d'erreur } u_{mt} \text{ et } \sigma_{my} = \frac{\gamma\sigma_1^2 + \theta\sigma_2^2}{(1 - \theta\gamma)^2}$$

est la covariance de u_{mt} et u_{yt} . σ_1^2 et σ_2^2 représentent les variances de ε_{1t} et ε_{2t} , respectivement. On obtient alors le modèle suivant :

$$m_t = \rho_{11} \cdot m_{t-1} + \rho_{12} \cdot y_{t-1} + \eta_{mt} \quad (29)$$

$$y_t = \delta \cdot m_t + \rho_{21} \cdot m_{t-1} + \rho_{22} \cdot y_{t-1} + \eta_{yt} \quad (30)$$

où η_{mt} et η_{yt} sont les innovations orthogonalisées :

$$\eta_{mt} = u_{mt}$$

$$\eta_{yt} = u_{yt} - \delta \cdot u_{mt}$$

En termes des erreurs originelles ε_{1t} et ε_{2t} , les innovations orthogonalisées sont données par :

$$\eta_{mt} = \frac{(\varepsilon_{1t} + \theta\varepsilon_{2t})}{(1 - \theta\gamma)}$$

$$\eta_{yt} = \frac{(\gamma\varepsilon_{1t} + \varepsilon_{2t})}{(1 - \theta\gamma)} - \frac{(\gamma\sigma_1^2 + \theta\sigma_2^2)(\varepsilon_{1t} + \theta\varepsilon_{2t})}{(1 - \theta\gamma)(\sigma_1^2 + \theta\sigma_2^2)}$$

En dépit du fait que les innovations orthogonalisées η_{mt} et η_{yt} sont sériellement et simultanément non corrélées, elles ne peuvent pas être traitées comme des variables exogènes pour l'analyse des fonctions impulsionnelles. En effet, le résultat d'une action sur η_{mt} ou sur η_{yt} ne sera bien défini que lorsque l'on précise si l'origine de la variation dans ces innovations est due à ε_{1t} ou à ε_{2t} .

Les innovations orthogonalisées η_{mt} et η_{yt} sont perçues par Cooley et LeRoy comme des constructions théoriques arbitraires dont l'unique intérêt est d'offrir une description des caractéristiques probabilistiques des variables observées. Dans la même veine, Pagan (1987) soulève le problème de l'absence de signification économique des innovations orthogonalisées. Comme le soulignent Cooley et LeRoy, il existe une infinité de transformations pour re-normaliser le modèle VAR initial, qui sont toutes '*observationnellement équivalentes*' puisqu'elles génèrent la même distribution de probabilité pour les variables observées. Les systèmes formés des équations (29) et (30) sont tous *observationnellement équivalents* car ils ont la forme réduite constituée des équations (24) et (25) et de plus la détermination de leurs paramètres ne permet pas de fixer de façon unique ceux du modèle représenté par les équations (26) et (27).

Sans référence à une théorie, le choix par le modélisateur d'un cas particulier de re-normalisation sera fait uniquement par souci de "commodité" (Cooley et Leroy, 1985). Il paraît ainsi qu'une hypothèse de *prédétermination* soit nécessaire, par exemple $\theta = 0$, pour pouvoir interpréter les innovations orthogonalisées comme des variables exogènes et aboutir à une distinction entre des modèles tous '*observationnellement équivalentes*'. Dans ce cas, on aura $\delta = \gamma$, $\eta_{mt} = \varepsilon_{1t}$ et $\eta_{yt} = \varepsilon_{2t}$, ce qui nous ramène au processus générateur des données formé des équations (26) et (27), mais avec de plus la restriction $\theta = 0$.

Les économistes de la Commission Cowles étant soucieux de la distinction entre causalité et corrélation, ils introduisent une différence claire entre les propriétés d'exogénéité et celles de corrélation. Cette distinction entre corrélation et causalité implique que la présence ou l'absence d'exogénéité ne peut être inférée à partir des données. Dans cette perspective, la présence d'une corrélation entre la monnaie et le revenu est compatible à la fois avec l'hypothèse que la monnaie est exogène par rapport au revenu, ou que le revenu est exogène par rapport à la monnaie, ou aucune des deux n'est déterminée par l'autre, mais qu'elles constituent des fonctions d'une troisième variable qui est exogène (Cooley et LeRoy, 1985).

Pour cela, les modèles VAR se sont révélés de peu d'intérêt pour l'évaluation de l'effet d'une intervention sur une variable de politique monétaire à travers l'utilisation de la G-causalité ou de l'analyse impulsionnelle. L'indétermination empirique de la causalité dans un VAR, liée au problème de *l'équivalence observationnelle*, impose l'adoption d'un schéma d'identification sur la base d'*a priori* théorique pour identifier les chocs de politique monétaire.

Pierce (1977) relève que l'analyse statistique est d'une portée limitée pour la détermination de certaines relations économiques, car des propositions incompatibles peuvent toujours être vérifiées par l'utilisation des mêmes données. Il est alors impératif de disposer d'une information *a priori* concernant ces relations. L'observation d'une corrélation empirique ne permet pas de savoir s'il s'agit d'une relation causale pure exprimée par l'équation théorique ou tout simplement d'un mélange combinant les effets de plusieurs relations (Meidinger, 1994).

Ceci nous renvoie encore une fois à la question de l'identification qui constitue un cas extrême du problème de la sous-détermination de la théorie par les données soulevé par la *thèse de Duhem-Quine* (Duhem, 1906 ; Quine, 1951). Cette thèse est également connue sous le nom de *holisme de la confirmation*, selon lequel l'observation des données ne permet pas d'imposer une seule théorie, mais qu'il existe toujours des théories différentes qui sont toutes compatibles avec ces données (Esfeld, 2006).

Même si, en général, la direction de la régression doit respecter le sens de la causalité, il n'y a rien dans le contenu propre des données qui révèle le vrai sens de la causalité (Hoover, 2006). Par exemple, en raison du problème de *l'équivalence observationnelle*, il n'y a aucun moyen

qui puisse permettre de faire la discrimination ou le choix, uniquement à partir des données, entre le modèle composé des équations (31) et (32) :

$$m = \varepsilon_m \quad (31)$$

$$y = \gamma.m + \varepsilon_y \quad (32)$$

et le modèle constitué des équations (33) et (34) :

$$m = \tau.y + \eta_m \quad (33)$$

$$y = \eta_y \quad (34)$$

Les deux formes de modèles peuvent décrire les variables observées m , y et de plus ils possèdent la même forme réduite. Mais, les deux modèles offrent des interprétations différentes quant à l'effet des interventions sur ces variables. Par exemple, le dernier modèle pourrait être déduit du premier en posant $\tau = \frac{\gamma \cdot \text{var}(\varepsilon_m)}{\gamma^2 \text{var}(\varepsilon_m) + \text{var}(\varepsilon_y)}$, avec les variables

$\eta_y = \varepsilon_y + \gamma \cdot \varepsilon_m$ et $\eta_m = (1 - \tau \cdot \gamma) \cdot \varepsilon_m - \tau \cdot \varepsilon_y$ qui sont non corrélées. Symétriquement, on aurait pu partir du modèle formé des équations (33) et (34) et en déduire de la même façon les paramètres du modèle composé des équations (31) et (32) (Hoover, 2006).

La sélection par exemple du modèle formé des équations (31) et (32) comme modèle structurel ne peut se faire que sur la base d'*a priori* théorique, ce qui permet de spécifier le paramètre γ comme paramètre structural et ε_m et ε_y comme des variables exogènes (Cooley et LeRoy, 1985). Mais, dans ce cas, les variables η_m et η_y ne peuvent pas être considérées comme exogènes. Lorsqu'ils sont inscrits dans une vision non structurale, du type de celle des modèles VAR, les deux modèles précédents sont équivalents, alors qu'ils deviennent différents s'ils sont interprétés dans une perspective structurale, c'est-à-dire qu'ils sont destinés à évaluer les effets des interventions sur les variables exogènes.

Des versions différentes, mais *observationnellement équivalentes* d'un modèle, offrent des interprétations causales différentes. L'exploitation d'un VAR comme un modèle structural dans une perspective causale, pour l'évaluation de l'effet d'une intervention, exige une justification théorique pour les restrictions sur les paramètres utilisées dans les hypothèses de triangularisation ou d'orthogonalisation des innovations.

Conclusion

Ce chapitre a permis de mettre en évidence les limites économétriques et méthodologiques que rencontrent le concept de G-causalité et les modèles VAR pour l'évaluation de la politique monétaire. La volonté de Sims (1972) de tester aisément et directement, à l'aide de ce concept, la présence d'une causalité unidirectionnelle s'est heurtée à d'importantes limites théoriques. De même, l'interprétation des innovations canoniques comme des chocs exogènes de politique monétaire exige d'imposer des restrictions d'identification sur la structure des modèles VAR, qui, paradoxalement, sont nés de la volonté d'éviter l'*a priori* théorique.

En l'absence d'un schéma d'identification sur la base d'*a priori* théorique pour identifier les chocs de politique monétaire, les modèles VAR se heurtent au problème de l'équivalence observationnelle (Basman, 1965, Hoover, 2004). Cette indétermination empirique de la causalité constitue un cas extrême du problème de la sous-détermination de la "théorie" par "l'empirie" tel qu'initialement soulevé par la thèse de Duhem-Quine. De plus, l'utilisation d'un modèle VAR pour l'évaluation de l'intervention de la Banque centrale fournit, selon Hoover (2009), un exemple de ce que Cartwright (2007) qualifie un contrefactuel "*imposteur*".

Deuxième Partie

**Du Retour Forcé de la Théorie
à sa Primauté**

Introduction de la Deuxième Partie

Cette seconde partie s'attache à confirmer, d'une part, l'obligation de la référence à la théorie, mise en évidence dans le chapitre 3, et, d'autre part, montrer la primauté de la théorie sur l'empirie pour l'étude de la politique monétaire et que l'économétrie ne peut pas être décisive dans la remise en cause de la théorie. Les monétaristes ont bien cherché à minimiser l'importance de leur écart théorique avec les keynésiens et à déplacer le débat sur le terrain empirique. Pourtant, le volume impressionnant des travaux empiriques qui a été produit par chacune des deux parties n'a pas permis de trancher le débat. A y regarder de près, ce débat cache une forte différence de vision méthodologique quant à l'utilité d'un modèle économétrique.

Les keynésiens insistent sur la cohérence théorique et la portée explicative du modèle théorique structurel, alors que pour les monétaristes c'est la capacité prédictive de la forme réduite estimée à partir des données observées qui compte. A cet effet, les keynésiens établissent une nette distinction entre le modèle théorique et le modèle estimé à l'aide de techniques économétriques à partir des données observées. En revanche, pour les monétaristes cette distinction n'est pas claire et ne leur semble pas pertinente.

Le développement des Modèles à Correction d'Erreurs dans la tradition de la *London School of Economics (LES)* (Hendry, 1995) et des modèles VAR cointégrés (Engle et Granger, 1987) a permis de renouveler l'analyse des propositions monétaristes en mettant en évidence les insuffisances des techniques économétriques traditionnelles (Granger et Newbold, 1974; Nelsson et Plosser, 1982). Mais, ces modèles dynamiques se heurtent également au problème de l'identification qui exige le recours à des *a priori* théoriques.

En effet, les liens entre les propositions de cointégration, les notions d'équilibre de long terme et de déséquilibre de court terme sont rarement interprétés dans le cadre d'un modèle théorique rigoureux et complètement spécifié. Or, pour Faust et Leeper (1994), l'identification d'un modèle par l'imposition de contraintes peut s'avérer non fructueuse lorsque la théorie économique ne fournit pas suffisamment de restrictions sur les variables d'intérêt ou n'établit pas de distinction claire entre les dynamiques de court terme et de long terme. Dans le même sens, Faust et Whiteman (1997) relèvent l'absence d'un critère d'arbitrage dans la démarche de la *London School of Economics (LES)* en présence de conflit

entre le principe théorique et l'ajustement aux données, sinon une subordination de la théorie à l'économétrie.

Parallèlement au problème de l'identification, la critique de Lucas (1976) constitue la seconde critique fondamentale à laquelle fait face l'utilisation de l'économétrie pour l'étude de la politique monétaire. Malgré l'absence d'une confirmation empirique concluante (Ericsson, Irons, 1995), cette critique a eu un impact dévastateur sur la recherche en modélisation économétrique s'inscrivant dans la tradition de la Commission Cowles ainsi que sur celle portant sur les modèles VAR. Lucas (1980, 1986) adopte une nouvelle posture épistémologique considérant le modèle théorique comme une "fiction" ou une "imitation" de l'économie et non pas un ensemble de propositions sur le comportement d'une économie réelle. En raison de la préséance de la théorie, cette critique a révolutionné la conception et le rôle d'un modèle, a remis en cause les méthodes économétriques traditionnelles et a conduit à un renouveau de la politique monétaire dans le cadre du programme de recherche des Nouveaux Classiques. Lucas (1977) s'oppose à la vision traditionnelle de l'équilibre où celui-ci est réalisé uniquement à long terme et défend l'idée d'une explication du comportement du cycle en termes de discipline de l'équilibre.

La vulnérabilité de la méthode des VAR à la critique de Lucas a été mise en évidence par Benati et Surico (2009) qui ont établi la supériorité d'un modèle DSGE par rapport à un modèle VAR structurel (SVAR) pour l'évaluation de l'effet d'un changement de règle de politique monétaire. Les modèles DSGE (Dynamic Stochastic General Equilibrium), qui constituent les modèles de base de la théorie de la Nouvelle Synthèse, sont fortement influencés par la méthodologie lucasienne et s'inscrivent dans la continuité des modèles RBC (Taouil, 2011). Benati et Surico estiment que l'utilisation des SVAR ne permet pas toujours de bien identifier un changement de règle de politique monétaire. Cet échec est la conséquence directe des restrictions inter-équations imposées par l'hypothèse des anticipations rationnelles, tel que cela a été initialement soulevé par la critique de Sargent (1979).

Chapitre 4

**Tentatives Économétriques de Conciliation
entre Théorie et Empirie**

Introduction

Le débat des années 1960 et 1970 entre les monétaristes et les keynésiens se caractérise par une forte différence de vision méthodologique quant à l'utilité d'un modèle économétrique. Alors que les keynésiens insistent sur la cohérence théorique et la portée explicative du modèle théorique, pour les monétaristes c'est la capacité prédictive du modèle estimé à partir des données observées qui compte. A cet effet, les keynésiens établissent une nette distinction entre le modèle théorique et le modèle estimé à l'aide de techniques économétriques à partir des données observées. En revanche, pour les monétaristes cette distinction n'est pas claire et ne leur semble pas pertinente.

La controverse sur les méthodes d'estimation directes ou à travers la forme réduite d'un modèle structurel illustre parfaitement le profond désaccord entre les keynésiens et les monétaristes sur la méthode adéquate pour l'estimation des effets de la politique monétaire ou budgétaire sur l'activité économique. La principale faiblesse de la méthode des monétaristes est l'impossibilité de définir le modèle structurel à partir duquel leur forme réduite est dérivée. Or, c'est le modèle structurel qui permet de préciser les restrictions sur les coefficients des variables indépendantes à travers les identités, les retards les variables omises, ... etc. En l'absence de ce modèle, il n'est pas possible de comprendre la nature des relations causales entre les variables endogènes et exogènes.

Les investigations économétriques sur la demande de monnaie, qui se sont poursuivies dans le cadre des Modèles à Correction d'Erreur (MCE), ont montré que les élasticités de cette fonction sont très sensibles à une mauvaise spécification de la dynamique des variables d'intérêt. Le MCE développé par Hendry et Mizon (1978) permet d'interpréter la demande de monnaie comme les fluctuations d'une dynamique de court terme autour d'un équilibre de long terme stable. En effet, la méthode des modèles à correction d'erreur, défendue par Hendry, cherche à concilier entre une relation d'équilibre de long terme entre la demande de monnaie et ses déterminants, dictée par la théorie économique, et une spécification purement économétrique de la dynamique de court terme. Toutefois, la spécification d'une relation économique sous la forme d'un MCE est conditionnée par l'existence d'une relation de cointégration entre les variables.

L'économétrie des séries non stationnaires a permis de renouveler l'analyse des propositions monétaristes en mettant en évidence les insuffisances des techniques économétriques traditionnelles. La théorie des variables cointégrées découle d'un rapprochement entre l'économétrie dynamique selon la méthode de Hendry et la théorie des séries temporelles de Box et Jenkins (1976). A cet effet, la cointégration est formalisée par Engle et Granger (1987) en termes d'un modèle ARIMA qui s'écrit sous la forme d'un MCE à la Hendry, mais sans postuler *a priori* la relation de long terme qui est dérivée à partir des propriétés stochastiques des variables en étudiant leur comportement tendanciel. Les études portant sur l'estimation d'une relation de long terme pour la demande de monnaie utilisant les tests de cointégration aboutissent souvent à des résultats contrastés qui, de plus, dépendent de l'agrégat monétaire et du type de taux d'intérêt retenus (Hendry et Ericsson, 1990; Hafer et Jansen, 1991).

Les liens entre les propositions de cointégration, les notions d'équilibre de long terme et de déséquilibre de court terme sont rarement interprétés dans le cadre d'un modèle théorique rigoureux et complètement spécifié. De fait, l'interprétation des modèles de cointégration en termes de comportement économique devient souvent difficile et elle se focalise le plus souvent sur les relations de long terme et néglige les fluctuations de court terme. Toutefois, Cooley et Hansen (1993) ont développé un modèle monétaire qui montre que les fluctuations de court terme renseignent mieux sur le comportement de l'économie que les relations de cointégration. En particulier, ce modèle ne permet pas d'interpréter la relation de long terme entre le revenu, la monnaie, le prix et le taux d'intérêt comme une fonction de demande de monnaie.

De plus, les contraintes imposées pour l'identification et l'interprétation des relations de long terme détériorent en général les qualités d'ajustement de la dynamique de court terme. En effet, lorsque le nombre de vecteurs de cointégration est supérieur à un, les modèles VAR cointégrés présentent un problème d'identification qui exige le recours à des *a priori* théoriques. Pour Faust et Leeper (1994), l'identification d'un modèle par l'imposition de contraintes peut s'avérer non fructueuse lorsque la théorie économique ne fournit pas suffisamment de restrictions sur les variables d'intérêt ou n'établit pas de distinction claire entre les dynamiques de court terme et de long terme. Ces auteurs montrent que l'imposition de contraintes identifiantes de long terme dans un VAR à la Sims pour évaluer la neutralité de la monnaie sur la production ne permet pas en général d'obtenir des inférences fiables sur la dynamique de court terme. Les modèles VAR intègrent implicitement des restrictions de court

terme qui ne sont, en général, pas compatibles avec les contraintes de court terme que pourraient imposer la théorie. Dans le même sens, Faust et Whiteman (1997) relèvent l'absence d'un critère d'arbitrage dans la démarche de la *London School of Economics (LES)* en présence de conflit entre le principe théorique et l'ajustement aux données, sinon une subordination de la théorie à l'économétrie.

Monfort et Rabemanajara (1990) et Hendry et Mizon (1993) ont montré que l'hypothèse d'exogénéité faible, introduite par Engle, Hendry et Richard (1983), permet d'effectuer le passage entre la modélisation structurelle et la méthode des modèles VAR, mais sans résoudre le problème de l'identification de ces derniers. Dans le cas d'un modèle VAR cointégré bivarié, l'hypothèse d'exogénéité faible est testable uniquement en présence de contrainte identifiante sur le vecteur de cointégration. En revanche, en dimension supérieure, l'hypothèse d'exogénéité faible ne peut plus être directement vérifiée et devrait être postulée *a priori*.

Section 1 : Le critère prédictif des monétaristes

Les monétaristes appuient leurs propositions centrales de stabilité de la demande de monnaie et d'efficacité supérieure de la politique monétaire par rapport à la politique budgétaire sur des hypothèses empiriques, et non théoriques, dont les méthodes d'élaboration sont plus que défectueuses. En effet, la majorité des études économétriques ont montré la sensibilité de la demande de monnaie au taux d'intérêt et les limites de l'utilisation du coefficient de corrélation pour comparer deux modèles économétriques alternatifs.

Sous-section 1 : Le statut des paramètres du modèle

1. Le poids de l'empirie chez les monétaristes

La plus grande partie des débats des années 1960 entre monétaristes et keynésiens porte sur des résultats économétriques, puisque la majorité des propositions qui forment le noyau du monétarisme, publiés entre 1956 et 1963, s'appuient surtout sur des considérations empiriques plutôt que sur des constructions théoriques (Desai, 1981). En effet, avant la publication en 1968 de sa nouvelle lecture de la courbe de Phillips, Friedman n'avait pas encore forgé une théorie qui permet d'expliquer de façon complète l'influence de la masse monétaire sur les prix et la production.

Les principaux résultats des investigations empiriques présentés comme preuve de la supériorité des propositions monétaristes portent sur la stabilité de la demande de monnaie à long terme, l'exogénéité de l'offre de monnaie et son influence causale sur le revenu et les prix, la variabilité des décalages à court terme entre la croissance de la monnaie et celle de la production et des prix et la force du multiplicateur monétaire relativement au multiplicateur budgétaire (Desai, 1981). Les monétaristes tirent trois conclusions importantes de ces recherches empiriques. Premièrement, l'efficacité supérieure de la politique monétaire par rapport à la politique budgétaire, et, deuxièmement, la préférence d'une règle de croissance constante de l'offre de monnaie. La grande controverse et les vives réactions soulevées par ces travaux témoignent de l'importance qu'occupe désormais l'économétrie dans les débats économiques.

L'approche méthodologique adoptée par Friedman pour défendre ses positions privilégie le critère de confirmation des prévisions empiriques d'une proposition théorique par rapport à celui de la formalisation explicite des mécanismes de transmission d'une politique économique. Selon le chef de file des monétaristes, la pertinence d'une théorie se mesure d'abord par sa capacité à formuler des prévisions précises. Il abandonne la méthode des modèles structurels de la Commission Cowles et renoue avec celle du *National Bureau of Economic Research (NBER)* qui s'appuie sur une équation unique et les techniques des séries chronologiques. Friedman manifeste un souci de construction de régularités statistiques directes, mais à partir *d'a priori* théorique, puisqu'on ne peut observer des données qu'à partir d'un cadre théorique donné. L'absence de critères explicites et claires pour arbitrer entre des formulations économétriques concurrentes constitue l'une des raisons de la persistance de la controverse entre monétaristes et keynésiens. D'autant plus que les méthodes empiriques utilisées par les uns et par les autres ne sont pas neutres et ont beaucoup évolué avec le débat.

En soutenant que la différence entre les monétaristes et les keynésiens est surtout empirique plutôt que théorique, Friedman (1971) a réussi à placer les débats sur un terrain économétrique et à formuler ses critiques à partir d'éléments empiriques (Desai, 1981). Par exemple, il a réussi à déplacer le débat sur la théorie quantitative vers celui de la stabilité empirique de la demande de monnaie pour expliquer le rôle de la monnaie dans l'activité économique. Friedman estime que le désaccord sur le comportement de la demande de monnaie s'explique par le fait que l'ajustement aux perturbations monétaires prend beaucoup de temps et que celles-ci affectent plusieurs agrégats.

Mais, au-delà des clivages théoriques importants mis en évidence dans les chapitres 1 et 2, il convient de rappeler qu'en l'absence d'un accord sur la méthode empirique à utiliser, il est impossible de discriminer entre deux propositions théoriques concurrentes. A ce niveau, les monétaristes et les keynésiens ne retiennent pas les mêmes méthodes empiriques pour tester la capacité d'une théorie à expliquer/prédire les données observées. Par exemple, les deux courants ne sont pas d'accord sur la méthode statistique adéquate à utiliser pour identifier les relations causales entre, d'un côté, la monnaie, et, d'un autre côté, le revenu nominal ou les prix. Les monétaristes privilégient l'estimation de formes réduites et l'observation des corrélations alors que les keynésiens s'appuient sur des modèles à équations structurelles simultanées où la causalité est fixée sur la base d'un *a priori* théorique (voir le chapitre 3). A

partir de ce débat, Sims (1972) a introduit de nouvelles méthodes économétriques pour tester les hypothèses d'exogénéité, qui ont conduit par la suite à l'émergence des modèles VAR.

A cet effet, Desai (1981) considère que le cadre poppérien ne permet pas de comprendre la dynamique de ce débat et qu'il faut plutôt considérer que ce sont deux programmes de recherche, au sens de Lakatos, qui sont en présence. La version simplifiée de l'apport de Popper, qualifiée de "falsificationnisme naïf", consiste à soumettre aux tests empiriques les propositions théoriques. Lorsque la proposition théorique dit la même chose que la proposition empirique estimée, on dit alors que la théorie est "corroborée" par les faits. Sinon, la théorie est dite "réfutée" ou "falsifiée" par les données. Desai note que si les résultats de cette méthode de classification étaient clairs, on n'aurait pas assisté à d'aussi virulentes controverses entre les keynésiens et les monétaristes.

D'abord, les méthodes statistiques qui sont censées être neutres dans les tests ne le sont pas toujours. Il existe, en effet, plusieurs méthodes avec des complexités différentes et la réponse des tests dépend du choix de la méthode, de la nature de la proposition théorique et des données utilisées. Souvent, les méthodes économétriques évoluent beaucoup en conséquence du débat lui-même. Ensuite, les théories ne fournissent en général pas des hypothèses isolées et comportent un noyau d'hypothèses non testables, mais qui sont maintenues *a priori*. Enfin, les variables impliquées dans les propositions théoriques sont supposées être liées par des relations d'équilibre, statique ou dynamique, les unes aux autres. Or, les données utilisées pour tester ces propositions ne reflètent pas nécessairement des situations d'équilibre. De plus, les économistes introduisent une distinction entre le court terme et le long terme, où l'équilibre serait atteint uniquement à long terme sous l'hypothèse de la clause *ceteris paribus*. Comme on le verra dans la section 2 ci-dessous, la mesure de ces ajustements soulève d'importants problèmes.

En revanche, la présentation des théories comme des programmes de recherche au sens de Lakatos rend, selon Desai, la compréhension de la dynamique de la controverse plus facile à comprendre. Pour illustrer cette position, l'auteur se réfère à l'évolution du débat autour de la courbe de Phillips. Les tests économétriques menés sur la courbe de Phillips formulée par Lipsey ont montré que le coefficient de l'inflation est inférieur à 1. Ce résultat a remis en cause l'explication néoclassique du fonctionnement du marché du travail où les employés négocieraient le salaire réel et non nominal. Les monétaristes questionnent alors l'utilisation

dans cette équation de l'inflation mesurée et non celle anticipée. Toutefois, Friedman (1975) reconnaît l'échec de l'introduction des anticipations adaptatives à valider l'utilisation du salaire réel dans la relation de Phillips. Au lieu de rejeter l'hypothèse néoclassique d'absence d'illusion monétaire, la contribution de Lucas permet de la maintenir. La critique de l'hypothèse des anticipations adaptatives et son remplacement par celle des anticipations rationnelles permet alors de changer complètement la nature du débat (voir le chapitre 5).

Partant des travaux de Friedman comme économiste praticien, Hirsch et De Marchi (1984) estiment que l'approche de celui-ci coïncide avec celle de John Dewey (1938), dans le sens où il insiste sur la capacité prédictive d'un modèle, plutôt qu'avec celle de Karl Popper (1965). En effet, ils relèvent que pour Friedman une théorie n'est pas "falsifiable", et, qu'en même temps, ce dernier considère que si les implications d'une théorie correspondent aux données mieux qu'une seconde théorie, alors il choisira la première et questionnera la véracité des prémices de la seconde. Simultanément, Hirsch et De Marchi relèvent que les prévisions fournies par les modèles de Friedman sont mauvaises.

Par ailleurs, Friedman prétend qu'il est Marshallien et classe ses concurrents comme des Walrasiens. Cette opposition s'appuie sur deux traits distinctifs. D'un côté, par Marshalliens, Friedman entend les économistes qui dérivent leur théorie à partir d'un processus d'interaction entre l'observation des données et la construction théorique à partir d'une démarche en équilibre partiel. D'un autre côté, il considère que les auteurs Walrasiens privilégient les constructions abstraites et globales de l'économie.

2. Formes réduites versus formes structurelles et les restrictions a priori

On retrouve à la base des débats entre monétaristes et keynésiens des visions différentes pour la justification des restrictions sur les paramètres d'un modèle. En effet, les monétaristes n'hésitent pas à imposer des restrictions, non testées, sur certains paramètres de leurs équations empiriques, tels que le taux d'intérêt dans la demande de monnaie. Ainsi, Friedman (1956) postule une élasticité de la demande de monnaie par rapport au taux d'intérêt qui est faible à court terme et nulle à long terme.

Friedman et Meiselman publient en 1963 une étude économétrique, sur la période 1897-1958, utilisant des données annuelles américaines relatives au revenu, à l'investissement et à l'offre de monnaie. Cette étude soutient que l'explication des variations du revenu nominal par la théorie quantitativiste a un pouvoir prédictif plus fort que l'explication fournie par le modèle keynésien. Le multiplicateur monétaire, calculé à partir d'une demande de monnaie, est jugé plus stable et plus fort que le multiplicateur de l'investissement keynésien, déduit du modèle IS-LM. Pour cela, Friedman et Meiselman (1963) imposent une restriction *a priori* de nullité du coefficient du taux d'intérêt dans leur équation utilisée pour le calcul du multiplicateur monétaire :

$$C_t = \alpha_0 + \alpha_1.M_t + u_{1t} \quad (1)$$

avec C_t la consommation des ménages, M_t le stock de monnaie, α_j des paramètres et u_{1t} un résidu. Or, une version générale de la demande de monnaie M_t^d s'exprime sous la forme suivante :

$$M_t^d = \delta_0 + \delta_1.Y_t + \delta_2.r_t + \varepsilon_t \quad (2)$$

où Y_t est le revenu national, δ_j des paramètres et ε_t un résidu. Sachant que la demande de monnaie est égale à son offre M_t , on déduit de l'équation (2) la courbe LM :

$$Y_t = \frac{\delta_0}{\delta_1} + \frac{1}{\delta_1}.M_t - \frac{\delta_2}{\delta_1}.r_t - \frac{1}{\delta_1}\varepsilon_t \quad (3)$$

La comparaison entre les équations (1) et (3) montre (à part le remplacement de C_t par Y_t) que Friedman et Meiselman ont contraint, sans aucun test de restriction, le coefficient δ_2 à 0, autrement dit, ils calculent un multiplicateur monétaire pour une courbe LM supposée être verticale (Desai, 1981). De même, Friedman et Meiselman retiennent le modèle suivant pour calculer le multiplicateur keynésien :

$$C_t = \beta_0 + \beta_1.D_t + u_{2t} \quad (4)$$

où D_t représente les dépenses autonomes qui comportent l'investissement net, les exportations nettes et les dépenses gouvernementales et qui vérifie l'identité :

$$Y_t = C_t + D_t \quad (5)$$

Or, ce modèle s'appuie également sur l'hypothèse d'une demande d'investissement qui est inélastique par rapport au taux d'intérêt, c'est-à-dire, une courbe IS verticale.

Ando et Modigliani (1965) procèdent à l'estimation d'une forme réduite déduite à partir d'un système d'équations simultanées qui sont données par un modèle IS-LM. Ces auteurs préfèrent partir d'un cadre général dans lequel la politique monétaire et budgétaire, données par les équations IS et LM, déterminent le revenu et le taux d'intérêt. En effet, les débats théoriques présentés dans les chapitres 1 et 2 montrent que ni les keynésiens, ni les monétaristes ne pensent que la courbe IS ou la courbe LM est verticale. Ainsi, au niveau empirique, Friedman et Meiselman (1963) retiennent un *a priori* que Friedman rejette au niveau théorique et qu'il ne teste pas économétriquement.

Les tests économétriques de Ando et Modigliani conduisent au rejet des restrictions empiriques retenues par Friedman et Meiselman comme étant non compatibles avec les données, ce qui implique que l'utilisation des formes réduites (1) et (4) est inappropriée car la condition d'indépendance de la variable explicative par rapport au résidu n'est pas assurée ici. En utilisant les techniques d'estimation des variables instrumentales, Ando et Modigliani montrent également que la spécification du multiplicateur d'investissement par Friedman et Meiselman souffre d'un biais de simultanéité.

Le rejet du modèle monétaire simple (1) et du modèle keynésien simple (4) confirme l'obligation de recourir à une forme structurelle qui puisse permettre de tenir compte des interdépendances et des rétroactions existantes entre les variables. Dans cette forme structurelle, la distinction entre variable endogène et variable exogène remplace la distinction entre variable indépendante et variable dépendante retenue dans une régression simple du type (1) et (4). Ces dernières relations ne peuvent conduire à une interprétation causale que lorsqu'elles sont spécifiées et estimées dans le cadre d'un système d'équations simultanées. Le débat autour du modèle de St. Louis de Andersen et Jordan (1968) a clairement mis en évidence l'opposition entre deux méthodologies économétriques relatives, respectivement,

aux avantages des grands modèles économétriques contre les petits modèles et aux estimations des formes structurelles contre celles des formes réduites.

Il existe de nombreuses études empiriques, y compris celles réalisées par des économistes monétaristes, qui ont montré l'importance du rôle du taux d'intérêt dans la fonction de demande de monnaie. Par exemple, Meltzer (1973) trouve une élasticité négative de -0,7 du taux d'intérêt sur les bons à 20 ans, en utilisant M1 et M2 pour l'économie américaine sur la période 1900-1958. De même, Brunner et Meltzer (1963) démontrent, en utilisant plusieurs combinaisons de variables de l'économie américaine sur une longue période, le rôle empirique important du taux d'intérêt dans la fonction de demande de monnaie. Dans ces conditions, Laidler (1970) considère que l'élasticité non nulle du taux d'intérêt dans la demande de monnaie est établie empiriquement sans aucun doute.

Sous-section 2 : Propositions monétaristes et limites des tests empiriques

1. Coefficient de corrélation et problèmes de spécification

La notion de corrélation entre séries chronologiques, basée sur la mesure du coefficient de corrélation R^2 , occupe une place insigne dans les argumentations empiriques des monétaristes. On la retrouve au cœur des débats entre les monétaristes et les keynésiens, en l'occurrence sur l'efficacité relative de la politique monétaire par rapport à la politique budgétaire et sur la causalité entre l'offre de monnaie et le revenu nominal.

Friedman et Meiselman (1963) considèrent que les effets de la politique monétaire sont plus rapides et plus fiables que ceux de la politique budgétaire et remettent en cause l'idée dominante à l'époque, à savoir que les dépenses d'investissement ou la politique budgétaire puissent exercer une influence significative sur les fluctuations conjoncturelles du cycle économique. La supériorité supposée de l'explication monétaire ne s'appuie pas ici sur des arguments théoriques, mais plutôt sur des considérations économétriques, en particulier le calcul de coefficients de corrélation à partir des résultats des régressions linéaires. La grande controverse soulevée par cet article entre ses auteurs et les keynésiens reflète le poids de l'économétrie dans les débats économiques.

Cette controverse a pour origine des confusions élémentaires sur l'utilisation des modèles de régressions linéaires simples (Desai, 1981), concernant en particulier l'hypothèse de non auto-corrélation des résidus et celle de l'indépendance de la variable explicative par rapport au terme d'erreur. En remarquant que, sur plusieurs périodes, le coefficient de corrélation de l'équation (1) est supérieur à celui de l'équation (4), les auteurs concluent à la supériorité du modèle monétariste par rapport au modèle keynésien. Rappelons que le coefficient de corrélation entre la consommation et la masse monétaire dans l'équation (1), par exemple, s'écrit :

$$r = \frac{\text{cov}(C_t, M_t)}{\sqrt{\text{var}(C_t) \cdot \text{var}(M_t)}} \quad (6)$$

En réaction à ce résultat, Ando et Modigliani (1965) relèvent la dimension biaisée du calcul du coefficient de corrélation dans l'équation (4) en raison de la non indépendance de la variable D_t par rapport au terme d'erreur u_{2t} , autrement dit, cette équation est mal spécifiée. Pour mettre en évidence ce problème de spécification, considérons les deux équations suivantes qui relient la consommation au revenu disponible Y_t^d et à l'épargne S_t des ménages:

$$C_t = \gamma_0 + \gamma_1 \cdot Y_t^d + v_t \quad (7)$$

$$Y_t^d = C_t + S_t \quad (8)$$

La substitution de (8) dans (7) donne :

$$C_t = \frac{\gamma_0}{(1 - \gamma_1)} + \frac{\gamma_1}{(1 - \gamma_1)} \cdot S + \frac{1}{(1 - \gamma_1)} v_t \quad (9)$$

Contrairement à l'équation (4), la variable explicative dans l'équation (9) est l'épargne des ménages S_t et non les dépenses autonomes D_t et les deux variables diffèrent par plusieurs éléments (les transferts nets reçus du reste du monde, les revenus nets des investissements, les taxes reçues par le gouvernement, ...) qu'on note par X_t . Dans ce cas, l'écriture de l'équation (4) serait valide seulement si on peut remplacer Y_t^d dans l'équation (7) par $Y_t^d + X_t$:

$$C_t = \gamma_0 + \gamma_1 \cdot (Y_t^d + X_t) + v_t \quad (7')$$

ce qui constitue clairement une mauvaise spécification de la fonction de consommation des ménages. Lorsque des restrictions de nullité sont imposées sans tests économétriques sur les paramètres d'une équation, il est légitime de douter de l'interprétation causale des autres coefficients estimés. En effet, les variables correspondantes à ces coefficients sont probablement corrélées avec les termes d'erreurs et ne peuvent, de ce fait, être considérées comme étant exogènes. Dans ces conditions, l'ampleur du coefficient de corrélation n'est d'aucune utilité pour comparer les performances des modèles keynésien et monétariste.

En conséquence de ce problème de spécification, Ando et Modigliani montrent que le coefficient de corrélation correspondant à l'équation (4) est fortement altéré à la baisse relativement à celui de l'équation (1) par l'inclusion de la période de la guerre 1942-1946 dans l'échantillon de l'estimation. De plus, ils notent que la variance des résidus constitue une meilleure mesure du pouvoir prédictif de l'équation que le coefficient de corrélation, surtout que ce dernier est sensible à la spécification de l'équation en niveau ou en différence première.

Desai (1981) fait remarquer que l'utilisation d'une spécification linéaire par, d'une part, Friedman et Meiselman, et, d'autre part, Ando et Modigliani, conduit à des multiplicateurs linéaires. L'adoption de spécifications log-linéaires par Morishima et Saito (1964) dans un modèle pour les Etats-Unis permet d'obtenir des multiplicateurs qui varient le long du cycle conformément à l'opinion de Keynes. Dans ce modèle, les courbes IS et LM sont non linéaires et le multiplicateur budgétaire est plus élevé que le multiplicateur monétaire.

Par ailleurs, Granger et Newbold (1974) montrent qu'une valeur élevée du coefficient de corrélation ne peut pas constituer une base solide pour comparer les performances d'équations estimées à partir du niveau des variables. Ils montrent qu'on peut trouver pour ce type d'équation des paramètres significatifs et un coefficient de corrélation élevé même si les variables sont générées par des processus indépendants sous forme de marche aléatoire.

Par exemple, Hendry a mis en évidence ce problème en estimant les deux équations ci-dessous sur des données trimestrielles du Royaume Uni entre 1964 et 1975 :

$$p = 0,02 + 0,73.m \quad , \quad R^2 = 0,984 \quad (10)$$

$$p = 10,9 - 3,2.C + 0,39.C^2, \quad R^2 = 0,982 \quad (11)$$

où p désigne le niveau logarithmique du prix à la consommation, m le niveau logarithmique de la masse monétaire $M3$ et C représente l'accumulation des précipitations pluviométriques. Les paramètres sont significatifs dans les deux équations et les coefficients de corrélation sont très élevés et proches. Mais, il serait absurde de conclure que la pluviométrie constitue une cause des prix de même importance que la monnaie.

Ce type de corrélation fallacieuse a pour origine la présence de tendance commune entre les deux variables de la régression. Dans ce cas, la valeur élevée de la corrélation ne peut être interprétée comme une présence de relation causale et il convient d'éliminer la tendance en procédant en général à la différenciation de ces variables. Autrement, la présence d'une forte autocorrélation des résidus dans la régression est le signe d'une mauvaise spécification de l'équation, et ce, quelle que soit la valeur du coefficient de corrélation.

Contrairement à la théorie économique qui s'intéresse uniquement au sens économique des variables, l'analyse des séries chronologique insiste sur l'importance de tenir compte des propriétés statistiques des variables. Celles-ci sont, en effet, considérées comme la réalisation statistique de processus générateurs de données dont il faut extraire le maximum d'informations de sorte à ce que le terme résiduel non expliqué soit la somme de bruits blancs.

2. Les problèmes de mesure

En général, les théories économiques ne précisent pas le contenu empirique des variables qu'elles cherchent à modéliser, ce qui rend compliquer les débats autour des tests économétriques. A cet égard, Mayer se pose la question de savoir si le théoricien doit se soucier de la facilité de mesure des variables de sa théorie (Mayer, 1975). Ceci a été le cas lors de la controverse entre les monétaristes et les keynésiens à propos de l'estimation de la fonction de demande de monnaie. En effet, la réalisation de tests empiriques exige au préalable de s'entendre sur la définition précise à donner, par exemple, à la masse monétaire et au taux d'intérêt.

Lorsqu'on parle de la masse monétaire, il faut préciser à quel agrégat on fait référence sachant qu'il en existe plusieurs et que, de plus, il existe plusieurs actifs proches de la monnaie au sens strict. S'agit-il de la monnaie fiduciaire représentée par l'agrégat M1, de l'agrégat M2 qui inclut, en plus de M1, les dépôts à vue bancaires ou de l'agrégat plus large M3 qui intègre en plus de M2 les dépôts à terme. L'agrégat M1 est généralement retenu par les théories qui insistent sur le motif de transaction de la monnaie. Mais, dès que la demande de monnaie intègre d'autres considérations que le moyen d'échange, comme le motif de spéculation ou le degré de substituabilité entre actifs, l'agrégat M1 n'est plus suffisant et il faut inclure les autres catégories d'agrégats (Laidler, 1970). En raison de ces difficultés, on rencontre plusieurs définitions de la monnaie qui sont testées dans les études empiriques.

La variable richesse ou patrimoine, introduite par Friedman (1956) dans sa fonction de demande de monnaie, pose encore plus de difficultés de mesure. Cette variable comporte, au-delà du patrimoine physique, également le capital humain. La variable patrimoine est généralement approximée dans les études empiriques par la valeur actualisée du revenu futur. Ce dernier est lui-même approximé par une moyenne pondérée de ses valeurs passées et présentes. Cette méthode de la moyenne pondérée est aussi retenue par les monétaristes pour la mesure du revenu permanent, utilisé dans la fonction de consommation des ménages, et pour la mesure de la valeur anticipée d'une variable, telle que celle du prix dans la relation de Phillips.

La célèbre controverse des années 1960 sur le statut du capital nous rappelle à quel point il est difficile de mesurer monétairement autant le capital physique que le capital humain. Le premier problème porte sur la difficulté d'agrégation des différents types de capital techniques, qui sont par nature hétérogènes, sans connaître leur prix. Or, c'est la formation des prix que les néo-classiques et plus tard les monétaristes cherchent à expliquer à travers la décomposition du revenu nominal entre production et prix. Le capital humain ne se prête pas non plus à une quantification aisée et il est généralement approximé par le nombre d'année d'étude, la nature de la qualification ou le diplôme.

Les études empiriques rencontrent également des difficultés pour trouver de longues séries sur les rendements des actifs autres que la monnaie. Le choix est généralement très limité et porte uniquement sur les rendements des obligations, sachant que celles-ci ne constituent pas nécessairement le meilleur substitut de la monnaie. Dans le cas où les rendements sur

plusieurs types d'actifs sont disponibles, il faut savoir lequel est le plus lié à la demande de monnaie. De plus, faut-il retenir des rendements et des taux d'intérêt de court terme ou de long terme? Ce choix dépend-il de la nature de l'agrégat monétaire retenu dans la relation à estimer?

Par ailleurs, la quantité de monnaie demandée n'étant pas observée, c'est la quantité offerte qui est utilisée dans les régressions économétriques. En raison du problème d'identification, ceci n'est possible qu'en supposant que l'équilibre sur le marché de la monnaie est réalisé. Pour cela, Laidler (1970) considère que les estimations des fonctions de demande de monnaie ne reflètent pas uniquement le comportement de la demande de monnaie.

Les études empiriques de la demande de monnaie utilisant les techniques de cointégration et les modèles à correction d'erreur (voir la section 2 ci-dessous) ont continué à soulever la question de la mesure "correcte" de la masse monétaire et du taux d'intérêt à utiliser (Hafer et Jansen, 1991). Hallman, Porter et Small (1989) et Gavin et Dewald (1989) trouvent que l'utilisation de la masse monétaire M2 fournit des relations de long terme plus fiables que M1. Hafer et Jansen (1991) montrent qu'il existe une relation de cointégration entre l'encaisse réelle utilisant l'agrégat M2, le revenu réel et le taux d'intérêt, mais qu'il n'en existe pas avec M1. En revanche, les études de Hafer (1984) et de Darby, Mascar et Marlow (1989) concluent à la pertinence de l'utilisation de l'agrégat M1.

Section 2. La cointégration et l'articulation entre court et long terme

L'examen des propositions empiriques des monétaristes s'est poursuivi dans le cadre des méthodes des modèles à correction d'erreur et de la cointégration, qui cherchent à tenir compte des propriétés statistiques des variables macroéconomiques. Le point commun entre ces approches est la volonté de concilier entre une relation d'équilibre de long terme et une dynamique purement économétrique de court terme. Toutefois, la volonté de tenir compte des propriétés stochastiques des variables économiques ne suffit pas pour fournir une spécification complète de ces modèles. A l'instar des modèles VAR à la Sims, la nécessité de recourir à des *a priori* théoriques pour identifier et postuler des hypothèses d'exogénéité continue également de se poser avec acuité pour les modèles de cointégration.

Sous-section 1 : La dynamique des variables monétaires

1. Demande de monnaie et modèle à correction d'erreur

Depuis les premiers travaux de Friedman, l'analyse de la politique monétaire a été concernée par les effets dynamiques des variations de l'offre de monnaie. Le débat sur la fonction de la demande de monnaie qui s'est poursuivi au cours des années 1970 et 1980 a porté sur la nécessité de tenir compte des propriétés d'autocorrélation des résidus et conduit au développement de nouvelles méthodes d'analyse dans le cadre des modèles à correction d'erreur. En effet, les élasticités de cette fonction relatives au taux d'intérêt, au revenu ou à l'inflation sont très sensibles à la spécification du terme d'erreur et à une mauvaise spécification de la dynamique des variables.

Ainsi, Hasche (1974) décrit le comportement de la demande de monnaie par un processus d'ajustement partiel :

$$\Delta \text{Log}(M_t) = (1 - \gamma) \cdot (\text{Log}(M_t^*) - \text{Log}(M_t)) + v_t \quad (12)$$

où la demande de monnaie M s'ajuste à une demande de monnaie désirée M^* , donnée par la relation :

$$\text{Log}(M_t^*) = \alpha_0 + \alpha_1 \text{Log}(Y_t) + \alpha_2 \text{Log}(P_t) + \alpha_3 \text{Log}(1 + r_t) + \varepsilon_t \quad (13)$$

avec M la production, P le niveau des prix, r le taux d'intérêt et Δ l'opérateur différence première. En remplaçant le niveau logarithmique des variables par des minuscules et en imposant $\alpha_2 = 1$, la condensation des deux relations précédentes donne :

$$\Delta(m_t - p_t) = \beta_1 \cdot \Delta y_t + \beta_2 \cdot \Delta(1 + r_t) + \beta_3 \cdot \Delta(m_{t-1} - p_{t-1}) + v_t \quad (14)$$

Hendry et Mizon (1978) soumettent à des tests statistiques la forme autorégressive (14) ci-dessus et la contrainte *a priori* $\alpha_2 = 1$, à partir de l'estimation de la forme dynamique plus générale suivante :

$$\phi(L) \cdot \psi(L) y_t = \phi(L) \cdot \theta(L) x_t + \mathcal{G}(L) e_t \quad (15)$$

où $\phi(L)$, $\psi(L)$, $\theta(L)$, $\mathcal{G}(L)$ sont des polynômes de l'opérateur retard L . L'équation autorégressive (14) de Hasche ne permet pas de faire apparaître la demande de monnaie de long terme, alors que l'équation (15) de Hendry et Mizon permet d'interpréter le modèle comme les fluctuations d'une dynamique de court terme autour d'un équilibre de long terme stable (Desai, 1981). Dans ce cadre, la spécification de la fonction de demande de monnaie de Hendry et Mizon prend la forme :

$$\begin{aligned} \Delta(m_t - p_t) = & \beta_0 + \beta_1 \cdot \Delta y_t + \beta_2 \cdot \Delta(1 + r_t) + \beta_3 \cdot \Delta(m_{t-1} - p_{t-1}) + \beta_4 \cdot \Delta p_t \\ & + \beta_5 \cdot (m_{t-1} - p_{t-1} - y_{t-1}) + \beta_6 \cdot (1 + r_{t-4}) + u_t \end{aligned} \quad (16)$$

Ici, les paramètres β_5 et β_6 jouent un rôle important puisqu'elles représentent des élasticités de long terme pour la demande de monnaie et le taux d'intérêt qui sont exprimés en niveau et non pas en différence première. Alors que Hasche a conclu à l'instabilité de la demande de monnaie, Hendry et Mizon ont mis en évidence, moyennant de nouvelles techniques économétriques, une demande de monnaie de long terme stable. Par ailleurs, Davidson, Hendry, Srba et Yeo (1978) ont également estimé, à l'aide des mêmes techniques utilisées par Hendry et Mizon, une fonction de consommation des ménages stable à long terme. A cet effet, Hendry (1995) estime que l'instabilité de la demande de monnaie est plus le résultat de mauvaises spécifications que d'un changement du comportement de la demande.

Les travaux présentés ci-dessus s'inscrivent dans la continuité de ceux développés par Sargan (1964) qui introduit la dynamique, sous la forme de processus autorégressif d'ordre un, dans

une application au salaire et au prix. Dans ce cadre, les coefficients des variables retardées ne sont plus contraints à zéro mais sont estimés librement. L'approche de Sargan (1964) établit un pont entre le rôle de la théorie dans le choix de la spécification et le traitement *ad hoc* de la dynamique dans le cadre de la méthode des séries chronologiques.

Le type de spécification retenu par les articles de référence de Hendry et Mizon (1978) et de Davidson et *ali.* (1978) a été popularisé par Hendry sous l'appellation de Modèle à Correction d'Erreurs (MCE). De façon plus générale, un MCE reliant une variable y_t à une variable x_t s'écrit sous la forme :

$$\Delta y_t = \mu + \sum_{i=1}^p \alpha_i \cdot \Delta y_{t-i} + \sum_{i=1}^q \beta_i \cdot \Delta x_{t-i} + \lambda \cdot (y_{t-1} - \delta \cdot x_{t-1}) + e_t \quad (17)$$

Par cette spécification, Hendry cherche à concilier entre des relations d'équilibre de long terme, dictées par la théorie économique, et des schémas dynamiques d'ajustement de court terme vers ces cibles de long terme. Cette démarche se justifie par le fait que la théorie économique se contente souvent de formuler des relations statiques associées à des états stationnaires de long terme de l'économie. Des spécifications purement économétriques sont alors utilisées pour modéliser le comportement dynamique de court terme.

Hendry et Ericsson (1991) applique la méthode du général au particulier pour développer une fonction de demande de monnaie qui englobe et rend compte des défauts de celle construite par Friedman et Schwartz (1982) pour le Royaume-Uni. Un modèle est déduit par transformations et réductions du Processus Générateur de Données (PGD) caractérisé par la densité jointe de la masse monétaire, du revenu, du taux d'intérêt et du prix. Le modèle conditionnel de la masse monétaire, en termes autorégressifs à retards échelonnés, est réduit à un modèle à correction d'erreurs. Hendry et Ericsson placent leur approche, qui cherche à exploiter l'information contenue dans les données en s'appuyant sur la théorie pour interpréter les résultats, en opposition de la méthode de Friedman et Schwartz qui visent à corroborer leur théorie par les résultats des régressions.

L'équation de demande de monnaie estimée par Friedman et Schwartz (1982) a la forme suivante :

$$(m - p - n) = \alpha_0 + \alpha_1 \cdot y + \alpha_2 \cdot g\tilde{y} + \alpha_3 \cdot i \quad (18)$$

où m , p , n , y désignent le logarithme de la masse monétaire, du prix, de la population et du revenu réel, $g\tilde{y}$ est le taux de croissance du revenu nominal et i est le niveau du taux d'intérêt. Toutes les variables sont exprimées en termes de moyennes. En effet, pour isoler des relations de long terme, ces auteurs utilisent des calculs de moyennes de variables annuelles le long des phases du cycle mesuré par le National Bureau of Economic Research (NBER) pour le Royaume-Uni sur la période 1871-1975.

Friedman et Schwartz cherchent à établir la stabilité de cette fonction de demande de monnaie sur l'ensemble de la période étudiée, mais ils ne procèdent à aucun test pour vérifier cette stabilité. Hendry et Ericsson (1991) réestiment la demande de monnaie (18) et montrent l'instabilité de ses paramètres en utilisant la statistique de Chow (1960). Cette demande de monnaie souffre de plusieurs problèmes de spécification et de nombreuses lacunes statistiques et les propositions de Friedman et Schwartz manquent de crédibilité en raison de leur incapacité à fournir des tests statistiques pertinents. La présence d'importantes mauvaises spécifications invalide leur inférence, en particulier leur demande de monnaie n'est pas stable.

Friedman et Schwartz évitent les régressions multiples et préfèrent, par souci de simplicité, recourir à des formes réduites et estimer des équations où sont introduites au plus deux variables explicatives. Ils regroupent ensuite les résultats de ces différentes estimations dans une équation finale. Hendry et Ericsson critiquent vivement cette méthodologie qu'ils considèrent comme statistiquement hautement déficiente. D'abord, il est erroné d'affirmer qu'un simple modèle permet de mieux comprendre, car le comportement qu'il décrit peut être trompeur à moins que ce modèle ne soit parfaitement cohérent avec les données. Ensuite, il n'est admis de regrouper dans une même équation les différents résultats que si leurs influences conditionnelles sont orthogonales. Puis, la difficulté d'interprétation des résultats des régressions multiples ne justifie pas l'ajustement de modèles simples qui soient mal spécifiés. Enfin, l'estimation des effets joints des variables après celle de leur effet individuel n'est possible que s'il n'y a pas d'interaction entre ces variables.

Hendry et Ericsson (1991) spécifient une relation de long terme de la demande de monnaie :

$$(m - p)_t^* = \alpha_0 + \alpha_1 \cdot y_t + \alpha_2 \cdot i_t + \alpha_3 \cdot (1 + \dot{p}_t) \quad (19)$$

où m , p , y désignent le logarithme de la masse monétaire, du prix, du revenu réel, i est le niveau du taux d'intérêt et \dot{p} est l'inflation. L'ajustement dynamique est fourni par le modèle à correction d'erreurs :

$$\begin{aligned} \Delta(m-p)_t = & \Gamma_0(L)\Delta(m-p)_{t-1} + \Gamma_1(L).\Delta p_t + \Gamma_2(L).\Delta y + \Gamma_3(L).\Delta i \\ & + \Gamma_4.((m-p)_{t-1}^* - (m-p)_{t-1}) + \varepsilon_t \end{aligned} \quad (20)$$

Les $\Gamma_i(L)$ sont des polynômes de l'opérateur retard L . Dans ce modèle, le paramètre Γ_4 associé à la relation de long terme doit être négatif. Hendry et Ericsson considèrent que, puisque la théorie ne précise pas la vitesse de réaction des agents à des changements dans leur environnement, les degrés des polynômes $\Gamma_i(L)$ doivent être déterminés par les données. Statistiquement, l'équation ci-dessus constitue une reparamétrisation d'un modèle autorégressif à retards échelonnés des variables en niveau.

La stabilité de la demande de monnaie exige la vérification de la constance des paramètres et de la même qualité d'ajustement sur les différentes sous-périodes. Hendry et Ericsson (1991) démontrent que l'équation (18) de demande de monnaie de Friedman et Schwartz ne vérifie aucune de ces hypothèses et, qu'en revanche, leur équation les vérifie et leur permet d'interpréter les coefficients comme ceux d'une demande de monnaie.

Toutefois, la spécification d'une relation économique entre deux variables sous la forme d'un modèle à correction d'erreur (MCE) est conditionnée par la stationnarité de la relation de long terme, autrement dit par la cointégration de ces deux variables (Malgrange, 1992).

2. Cointégration et relation économétrique de long terme

La théorie des variables cointégrées découle d'un rapprochement entre l'économétrie dynamique selon la méthode de Hendry et la théorie des séries temporelles de Box et Jenkins (1976) (Maurel, 1989). Engle et Granger (1987) formalisent la cointégration en termes de modèles ARIMA qui peuvent se mettre sous la forme d'un MCE à la Hendry, mais sans postuler *a priori* la relation de long terme. Celle-ci est, en effet, dérivée à partir des propriétés stochastiques des variables en étudiant leur comportement tendanciel. Ceci permet à la théorie

de la cointégration de formuler des hypothèses économétriquement testables sur les propriétés dynamiques des variables et de tenir explicitement compte de leur caractère non stationnaire (Maurel, 1989 ; Urbain, 1990).

Une variable Y est dite intégrée d'ordre d , notées $I(d)$, si sa différence à l'ordre d est stationnaire, c'est-à-dire de moyenne et de variance finies, et possède une représentation ARMA inversible. Deux séries Y et M sont dites cointégrées d'ordre (d,b) , qu'on note $CI(d,b)$, si elles sont $I(d)$ et s'il existe un couple d'entiers β_1 et β_2 tel que $\beta_1.Y + \beta_2.M$ est $I(d-b)$. Par souci de simplicité, on ne considère dans la suite que les variables intégrées à l'ordre 1.

Depuis les articles de Granger et Newbold (1974) et de Nelsson et Plosser (1982), les travaux empiriques accordent beaucoup d'attention, à travers les tests de racine unitaire de Dickey et Fuller (1979, 1981), à la nature déterministe ou stochastique de la non stationnarité des variables macroéconomiques. Lorsqu'une série se caractérise par une tendance stochastique et la présence d'une racine unitaire, cela implique l'accumulation des chocs aléatoires qui affectent le comportement de la série (Urbain, 1990).

Les tests statistiques sur l'ordre d'intégration et de cointégration sont devenus très utiles et incontournables dans la modélisation empirique. Ils ont permis de mettre en évidence que le revenu, la masse monétaire et les prix n'évoluent pas en général de façon indépendante au cours du temps dans les différents pays. Plus généralement, ces nouvelles techniques ont montré l'intérêt de tenir compte dans les estimations des tendances stochastiques communes aux variables considérées.

L'économétrie des séries non stationnaires a permis de renouveler l'analyse des propositions monétaristes en mettant en évidence les insuffisances des techniques économétriques traditionnelles. Par exemple, Engle et Granger (1987) montrent qu'il existe une relation de cointégration, autrement dit une relation stable à long terme, uniquement entre la masse monétaire mesurée par l'agrégat $M2$ et le revenu monétaire, alors que cette hypothèse est rejetée dans le cas des agrégats $M1$ et $M3$. De même, ils ont montré que la consommation des ménages et le revenu sont cointégrés, alors que les salaires et les prix ne le sont pas.

Pour Engle et Granger (1987), la bonne spécification d'un modèle, lorsque les variables sont non stationnaires et cointégrées, consiste à utiliser une forme à correction d'erreurs. A cet effet, le théorème de représentation de Granger (Engle et Granger, 1987) démontre l'équivalence entre un système cointégré et un modèle à correction d'erreur. Mais, la représentation à correction d'erreurs du système cointégré CI(1,1) formé de Y et M est bivariée et a une forme plus générale que la spécification de Hendry :

$$\Delta Y_t = \delta_{11}(L)\Delta Y_{t-1} + \delta_{12}(L)\Delta M_{t-1} - \alpha_1 \cdot (\beta_1 \cdot Y_{t-1} + \beta_2 \cdot M_{t-1}) + \mu(L)e_{1t} \quad (21)$$

$$\Delta M_t = \delta_{21}(L)\Delta Y_{t-1} + \delta_{22}(L)\Delta M_{t-1} - \alpha_2 \cdot (\beta_1 \cdot Y_{t-1} + \beta_2 \cdot M_{t-1}) + \mu(L)e_{2t} \quad (22)$$

La forme du modèle à correction d'erreur de Hendry correspond dans ce cas à l'une des deux équations (21) ou (22) précédentes avec $\mu(L) = 1$.

Le théorème de Granger établit une synthèse entre la méthodologie des MCE à la Hendry, qui cherche à concilier entre théorie économique et rigueur économétrique, et l'approche des modèles VAR. De plus, ce théorème montre que l'hypothèse de cointégration, qui fait intervenir des variables en niveau dans l'écriture du modèle, n'est pas compatible avec l'écriture d'un VAR en différence première qui serait dans ces conditions mal spécifié (Malgrange, 1992). Dans un cadre multivarié, un modèle à correction d'erreur dans la tradition de Hendry s'écrit sous la forme :

$$\Delta X_t = \mu + \sum_{i=1}^s \varphi_i \cdot \Delta X_{t-i} + \alpha \cdot z_{t-1} + \varepsilon_t \quad (23)$$

où $X_t = (x_{1t}, \dots, x_{nt})'$ et $z_t = \beta' \cdot X_t$ représente la relation de long terme déduite de la théorie économique. Le vecteur X_t , constitué de variables I(1), est dit cointégré s'il existe un vecteur β de cointégration tel que $\beta' \cdot X_t$ est I(0), autrement dit stationnaire. D'après le théorème de Granger (Engle et Granger, 1987), il y a une équivalence entre l'existence de r vecteurs de cointégration indépendants β_i et l'existence d'une représentation sous la forme d'un modèle à correction d'erreur de la variable cointégrée de rang r, X_t :

$$\Phi(L) \cdot \Delta X_t = \alpha \cdot \beta' \cdot X_{t-1} + \theta(L) \cdot e_t \quad (24)$$

avec β une matrice de dimension (n,r) formée des vecteurs β_i . La matrice α , de dimension (n,r) , représente la matrice des poids de la solution de long terme dans chaque équation du modèle. Lorsque r est égal à 1, le vecteur de cointégration β peut s'interpréter comme la solution de long terme du modèle. En revanche, si r est strictement supérieur à 1, alors l'interprétation en termes économiques de ce vecteur devient plus difficile.

La demande de monnaie a connu un changement au début des années 1970 et 1980, perçu comme le signe d'une instabilité dans cette relation (Hafer et Jansen, 1991). Laidler (1981) fait remarquer que cette fonction a connu de larges déviations au cours de la décennie 1970 et que la vitesse de circulation de la monnaie n'est pas constante, mais qu'elle est fonction du revenu, du prix et du taux de croissance du stock de la monnaie. Parmi les raisons avancées pour expliquer cette instabilité, on trouve le passage aux régimes de change flexible après 1973, les innovations introduites sur les marchés financiers, mais aussi les tentatives de contrôle de l'évolution de la masse monétaire sous l'impulsion des recommandations des monétaristes eux-mêmes (Laidler, 1980).

De nombreux travaux ont cherché à expliquer cette apparente instabilité, ce qui présuppose l'existence d'une relation d'équilibre entre la demande d'encaisses réelles et certaines variables explicatives. Si cette hypothèse est fautive, toute tentative d'estimation d'une fonction de demande de monnaie est invalide. Or, la stabilité de la demande de monnaie exige la constance des paramètres de la fonction sous-jacente. Selon Hendry et Ericsson (1990), cette constance des paramètres, qui a été remise en cause par cette instabilité de la masse monétaire, est importante pour l'inférence, la prévision et l'évaluation de la politique monétaire.

Hafer et Jansen (1991) s'attachent à vérifier la véracité de l'existence d'une relation d'équilibre pour la demande de monnaie en évaluant l'existence d'une relation de cointégration entre la demande d'encaisses réelles, le revenu réel et le taux d'intérêt. Ils emploient, à cette fin, les procédures des tests de cointégration développées par Johansen (1988) et Johansen et Juselius (1990). Ils montrent alors qu'il existe bien une relation de cointégration entre l'encaisse réelle utilisant l'agrégat M2, le revenu réel et le taux d'intérêt, mais que cette cointégration n'est pas vérifiée dans le cas de l'agrégat M1.

Les liens entre les propositions de cointégration, les notions d'équilibre de long terme et de déséquilibre de court terme sont rarement interprétées dans le cadre d'un modèle théorique rigoureux et complètement spécifié (Vredin, 1994). L'explication des mouvements de court terme dans une relation de cointégration par l'idée de déséquilibre par rapport à un équilibre de long terme s'appuie sur des arguments théoriques de coûts d'ajustement et de rigidité, mais ce lien n'est jamais clairement explicité. De fait, l'interprétation des modèles de cointégration en termes de comportement économique devient souvent difficile et elle se focalise le plus souvent sur les relations de long terme et néglige les fluctuations de court terme. De plus, les contraintes imposées pour l'identification et l'interprétation des relations de long terme détériorent en général les qualités d'ajustement de la dynamique de court terme.

Cooley et Hansen (1993) ont développé un modèle monétaire qui met clairement en évidence l'intérêt du recours à la théorie pour savoir vers quel aspect de l'analyse des données il faut orienter l'attention. En effet, leur modèle montre que les fluctuations de court terme renseignent mieux sur le comportement de l'économie que les relations de cointégration décrivant la dynamique de long terme. En particulier, ce modèle ne permet pas d'interpréter la relation de long terme entre le revenu, la monnaie, le prix et le taux d'intérêt comme une fonction de demande de monnaie.

Sous-section 2 : Validation empirique et exigence théorique

1. Le retour au problème de l'identification

Considérons un modèle VAR cointégré de la forme :

$$\Delta X_t = \Pi \cdot X_{t-1} + \zeta(L) \cdot \Delta X_{t-1} + \varepsilon_t \quad (25)$$

où $\Pi = \alpha \cdot \beta'$ est une matrice de dimension (n,n). Lorsqu'il n'existe qu'un seul vecteur de cointégration β , ce modèle ne pose aucun problème d'identification, puisqu'il suffit d'une seule restriction de normalisation sur les coefficients de ce vecteur. Toutefois, cette normalisation, qui doit être compatible avec les données, n'a souvent aucune interprétation économique. Par contre, si le nombre de vecteurs de cointégration r est strictement supérieur

1, alors il existe un problème d'identification. En effet, le modèle VAR cointégré ci-dessus est une forme réduite et on n'observe que les éléments de la matrice Π . La décomposition $\Pi = \alpha.\beta'$ n'est pas unique, puisqu'il est toujours possible de trouver une matrice inversible C de dimension (r,r) telle que :

$$\Pi = \alpha.\beta' = (\alpha.C^{-1}).(C.\beta) = \alpha^*.\beta^* \quad (26)$$

Sans restrictions supplémentaires, les éléments des matrices α et β ne sont pas identifiés. Les restrictions sur β portent sur la solution de long terme du modèle et exigent le recours à des *a priori* théoriques pour pouvoir les interpréter économiquement (Johansen, 1995). Les restrictions sur α sont interprétées en terme d'exogénéité (voir le point 2 ci-dessous). Bauwens et Lubrano (1996) ont montré que l'identification des matrices α et β nécessite d'imposer r^2 restrictions. De plus, la procédure d'identification de β doit tenir compte de l'ordre des variables dans le modèle puisque, dans certains cas, elle peut introduire des considérations d'exogénéité sur certaines variables.

Une méthode d'identification des relations de long terme dans le cadre d'un modèle VAR structurel est présentée par Blanchard et Quah (1989). Cette approche impose, sur la base d'*a priori* théorique, une chaîne causale sur la façon dont les chocs se propagent à long terme. Le modèle porte sur le taux de croissance de la production et le taux de chômage, où seul le PIB présente une racine unitaire. Puisqu'il y a une seule variable $I(1)$, il n'y a pas de cointégration dans ce modèle.

Pour illustrer cette distinction, considérons un modèle VAR structurel qui s'écrit sous la forme d'une moyenne mobile infinie :

$$X_t = C_0.\varepsilon_t + C_1.\varepsilon_{t-1} + C_2.\varepsilon_{t-2} + \dots \quad (27)$$

Nous avons vu dans le chapitre 3 que du fait que la matrice C_0 n'est pas contrainte à l'identité, les chocs contemporains ε_t peuvent avoir des effets croisés sur les composantes du vecteur X_t . La solution d'identification retenue par Sims (1980) consiste à imposer une structure récursive sur la matrice C_0 .

Blanchard et Quah introduisent une distinction entre les chocs de demande, qui sont supposés exercer une influence temporaire, et les chocs d'offre qui ont une influence durable. La présence de la racine unitaire dans le PIB contribue à expliquer la persistance des chocs d'offre et l'influence de court terme des chocs de demande est justifiée par la présence de rigidités nominales. Le schéma d'identification pour les chocs d'offre de long terme adopté dans ce modèle introduit une restriction sur la matrice C_1 . Les auteurs montrent que les chocs de demande exercent une influence temporaire sur le revenu réel et que les politiques de demande peuvent donc être utilisées pour la stabilisation conjoncturelle. Ce résultat va à l'encontre du courant classique qui privilégie le recours exclusif à des politiques d'offre.

Pour Faust et Leeper (1994), l'identification d'un modèle par l'imposition de contraintes peut s'avérer non fructueuse lorsque la théorie économique ne fournit pas suffisamment de restrictions sur les variables d'intérêt ou n'établit pas de distinction claire entre les dynamiques de court terme et de long terme (Faust, Leeper, 1994). La recherche de restrictions identifiantes supplémentaires dans le cadre de modèles VAR à la Sims a amené par exemple Blanchard et Quah (1989), King, Plosser, Stock et Watson (1991) et Shapiro et Watson (1988) à introduire des restrictions de neutralité de long terme de la monnaie sur la production.

Faust et Leeper (1994) montrent que l'imposition de contraintes identifiantes de long terme ne permet pas en général d'obtenir des inférences fiables sur la dynamique de court terme. Les modèles VAR intègrent implicitement des restrictions de court terme qui ne sont, en général, pas compatibles avec les contraintes de court terme que pourraient imposer la théorie (Faust et Leeper, 1994).

2. Modèle conditionnel et exogénéité

Dans leur critique de l'étude de Friedman et Schwartz (1982), Hendry et Ericsson (1991) relèvent que l'utilisation par ceux-ci de la masse monétaire comme régresseur dans plusieurs équations exige que celle-ci vérifie une hypothèse de faible exogénéité pour pouvoir valider les estimations et interpréter correctement les paramètres d'intérêt. De même, dans le modèle à correction d'erreur (20) ci-dessus, les régresseurs doivent vérifier l'hypothèse de faible

exogénéité, les paramètres doivent être constants et invariants aux changements historiques qui ont affecté le processus déterminant la distribution marginale.

Hendry et Ericsson (1991) prouvent, à cet effet, que les variables conditionnelles sont super-exogènes relativement aux paramètres de l'équation (20), autrement dit, ces paramètres restent constants même lorsque le PGD des variables conditionnelles change sur l'histoire. Le test de la super-exogénéité revient à s'assurer que les paramètres sont invariants aux changements des régimes monétaires qui ont caractérisés la période. Les auteurs exploitent, à cet effet, l'introduction simultanée du flottement de change et de la compétition et la régulation du contrôle du crédit en 1971, et l'utilisation du test de Chow leur permet de rejeter les hypothèses de constance et de super-exogénéité sur la période 1971-1975.

Les notions d'exogénéité faible et de super-exogénéité ont été introduites dans le célèbre article de Engle, Hendry et Richard (1983) en termes de densités jointes et de modèles conditionnels. Leur méthode permet, en partitionnant le vecteur de paramètres θ entre deux sous-ensembles φ et ψ et le vecteur X_t entre Y_t et Z_t , de faire apparaître le vecteur de paramètre d'intérêt φ uniquement dans la densité conditionnelle de Y_t par rapport à Z_t , $D_{Y|Z}(Y_t|Z_t, X_{t-1}^0, \varphi)$, et le vecteur ψ dans la densité marginale de Z_t , $D_Z(Z_t|X_{t-1}^0, \psi)$, selon la formule :

$$D_X(X_t|X_{t-1}^0, \theta) = D_{Y|Z}(Y_t|Z_t, X_{t-1}^0, \varphi) \cdot D_Z(Z_t|X_{t-1}^0, \psi) \quad (28)$$

où $D_X(X_t|X_{t-1}^0, \theta)$ est la distribution jointe de probabilité, indexée par le paramètre $\theta = (\varphi, \psi)$, du vecteur $X_t = (Y_t, Z_t)$ conditionnellement à son passée X_{t-1}^0 . Cette décomposition cherche à regrouper les invariants dans le modèle conditionnel de Y_t et la partie instable dans le modèle marginal de Z_t .

Le vecteur Z_t est dit faiblement exogène pour le paramètre φ si on peut négliger le processus marginal de Z_t pour l'inférence de φ . Autrement dit, la variable Z_t peut être considérée comme fixe et permettre l'inférence sur φ , sans causer de perte d'information, uniquement à partir du modèle conditionnel de Y_t . Le vecteur Z_t est dit super-exogène pour le paramètre

φ si Z_t est faiblement exogène et si φ reste invariant lorsque des modifications affectent le paramètre ψ .

Florens et Moucahrt (1985) montrent qu'une condition pour la faible exogénéité de la variable Z_t est que les deux paramètres φ et ψ soient libres en variation, c'est-à-dire qu'il n'y a pas de restriction entre les deux systèmes d'équations générées par les parties de la décomposition de la distribution jointe de probabilité. Monfort et Rabemanajara (1990) et Hendry et Mizon (1993) ont montré que l'hypothèse d'exogénéité faible permet d'effectuer le passage entre la modélisation structurelle et la méthode des modèles VAR, mais sans résoudre le problème de l'identification de ces derniers.

La notion d'exogénéité faible permet également de mettre en évidence le lien entre un modèle VAR cointégré bivarié et un modèle à correction d'erreur à la Hendry dans le cas où il existe une seule relation de cointégration. Reprenons pour cela le cas simple de la représentation à correction d'erreurs entre les deux variables Y et M présentée dans le point 2 de la sous-section 1 ci-dessus par les deux équations (21) et (22) :

$$\Delta Y_t = \delta_{11}(L)\Delta Y_{t-1} + \delta_{12}(L)\Delta M_{t-1} - \alpha_1 \cdot (\beta_1 Y_{t-1} + \beta_2 M_{t-1}) + \mu(L)e_{1t} \quad (21)$$

$$\Delta M_t = \delta_{21}(L)\Delta Y_{t-1} + \delta_{22}(L)\Delta M_{t-1} - \alpha_2 \cdot (\beta_1 Y_{t-1} + \beta_2 M_{t-1}) + \mu(L)e_{2t} \quad (22)$$

Ici, le problème de l'identification se réduit au choix d'une normalisation du vecteur $\beta = (\beta_1, \beta_2)$ dans la relation de cointégration $z_{t-1} = \beta_1 Y_{t-1} + \beta_2 M_{t-1}$, qui se résout en posant $\beta_1 = 1$ ou $\beta_2 = 1$.

Ce modèle, constitué d'un système de deux équations, peut être considéré comme représentant la distribution jointe de $(\Delta Y_t, \Delta M_t)$. Retenir le modèle à correction d'erreur constitué de la première équation par exemple équivaut à considérer la distribution conditionnelle de ΔY_t par rapport à ΔM_t , et négliger la distribution marginale de ΔM_t . Pour cela, il faut que les paramètres du modèle conditionnel et du modèle marginal soient libres en variation, autrement dit que ΔM_t soit exogène pour l'inférence de α et β (Johansen, 1992).

Une condition nécessaire et suffisante pour cette exogénéité de ΔM_t , correspond à la condition économétriquement testable de nullité du coefficient α_2 , c'est-à-dire que le terme à correcteur d'erreur apparaît dans le modèle conditionnel et pas dans le modèle marginal. Toutefois, ce test d'exogénéité présuppose que le modèle soit déjà identifié par des restrictions identifiantes qui sont, elles, non testables. De plus, en l'absence de la contrainte de cointégration, l'hypothèse d'exogénéité faible est non testable et devrait être postulée *a priori*.

Dans le cas général où le nombre de relations de cointégration r est strictement supérieur à 1, la relation entre un modèle VAR cointégré et un modèle à correction d'erreur devint plus difficile à établir. L'hypothèse d'exogénéité faible ne peut plus être directement vérifiée et il faut introduire des restrictions d'identification pour pouvoir interpréter économiquement les coefficients du modèle (Ericsson, 1995).

Conclusion

La volonté des monétaristes de concilier entre un cadre théorique s'appuyant sur la fonction de demande de monnaie et les résultats de leurs investigations empiriques s'est heurtée à d'importantes limites économétriques. En l'absence d'un modèle théorique dynamique rigoureux et spécifiant l'ensemble des comportements, le recours à l'économétrie est de peu d'utilité.

La même conclusion émerge des tentatives de la théorie de la cointégration et des Modèles à Correction d'erreur pour concilier entre les dynamiques de court et de long terme des variables de la politique monétaire. Ces nouvelles méthodes souffrent également d'importants problèmes d'identification et ne permettent pas de tester les hypothèses d'exogénéité sans *a priori* théorique.

Chapitre 5

La Critique de Lucas et ses Implications

Introduction

La modélisation macroéconométrique structurelle et les modèles VAR sont soumis à la critique de Lucas (1976) relative à l'invariance des paramètres d'un modèle économétrique aux changements de politique économique. Lucas (1977) défend une explication du comportement du cycle économique en termes de discipline de l'équilibre, où celle-ci s'articule autour des postulats du comportement optimisateur des agents et de l'apurement des marchés. Les comportements agrégés de l'économie sont expliqués à partir de la résolution d'un programme d'optimisation d'un agent représentatif à anticipations rationnelles sous contraintes (Sargent, 1982).

Le nouveau programme de recherche initié par Lucas et adopté par les nouveau Classiques s'appuie sur une vision épistémologique qui conçoit le modèle théorique comme une fiction ou une imitation de l'économie (Lucas, 1980, 1986). Kydland et Prescott (1982) ont procédé à la mise en œuvre du programme de Lucas pour la confrontation de la théorie avec les données à travers l'application de la technique du calibrage aux modèles des Cycles Réels (RBC). Cette technique est présentée comme une stratégie alternative aux méthodes économétriques traditionnelles pour le calcul des paramètres structurels de ces modèles

S'inscrivant dans la continuité des contributions de Lucas, Kydland et Prescott (1977) conçoivent l'équilibre économique comme un jeu stratégique entre le décideur public et les agents privés. Sous l'hypothèse d'anticipations rationnelles, une politique économique discrétionnaire, même temporellement cohérente, peut conduire à un équilibre non optimal. La contribution de Kydland et Prescott (1977) a permis de déplacer le débat sur l'efficacité de la politique monétaire vers celui du rôle des institutions en charge de celle-ci (Hairault et Langot, 2005).

Le jeu non coopératif entre le décideur public et les agents privés conduit à un biais inflationniste qui est d'autant plus élevé que le Gouvernement accorde un poids plus élevé à la réduction du chômage qu'à la lutte contre l'inflation Barro et Gordon (1983). La solution proposée pour éviter une perte de crédibilité dans la politique monétaire est de conférer la responsabilité de celle-ci à une Banque centrale indépendante avec un Gouverneur conservateur (Rogoff, 1985).

Le problème du financement du déficit public soulevé par les monétaristes dans le cadre du débat sur l'effet d'éviction a été analysé de façon plus rigoureuse par Barro (1974) dans le cadre du principe de l'équivalence ricardienne. Sargent et Wallace (1981) remettent en cause la pertinence des propositions monétaristes et considèrent que la manière avec laquelle la politique monétaire et la politique budgétaires sont coordonnées influence la capacité de la Banque centrale à contrôler l'inflation.

De fait, le *policy-mix*, initialement défini comme l'articulation optimale entre la politique budgétaire et la politique monétaire, est devenu celui de la coordination entre les autorités de politique économique. Pour Sargent et Wallace, la capacité d'engagement de la Banque centrale sur une règle monétaire lui permet de se comporter en meneur du jeu, dans le cadre d'un régime de politique monétaire dominante, et d'imposer sa décision et une discipline à l'autorité budgétaire.

Section 1. La critique de Lucas : une innovation cardinale

A la suite de la critique de Lucas (1976), l'Ecole des Nouveaux Classiques a développé un programme de recherche visant à doter l'analyse quantitative de la macroéconomie avec une nouvelle démarche de modélisation structurelle, en rupture totale avec la modélisation à équations simultanées de la Commission Cowles et les modèles VAR. Le raisonnement de Lucas s'inscrit dans un cadre d'analyse qui s'appuie sur le comportement dynamique optimisateur d'agents individuels à anticipations rationnelles, avec un retour à la discipline de l'équilibre sous l'égide du postulat de l'apurement des marchés. La vision épistémologique qui sous-tend cette approche conçoit le modèle comme une fiction et adopte pour cela la technique du calibrage pour fixer la valeur de ses paramètres.

Sous-section 1. La discipline de l'équilibre

1. Invariance des paramètres et anticipations rationnelles

Initialement, Lucas (1976) s'attaque aux modèles macroéconométriques qui s'inscrivent dans la tradition de la Commission Cowles et remet en cause l'usage de ces modèles pour l'évaluation des politiques économiques. Les modèles VAR sont également soumis à la critique de Lucas relative à l'invariance des paramètres des méthodes économétriques. En particulier, ces modèles sont incapables de rendre compte des ruptures structurelles dans les comportements des agents induites par la modification de la politique monétaire.

Lucas reproche à ce type de modèles de maintenir invariants les paramètres θ , estimés économétriquement à partir des données historiques, dans leurs fonctions de comportement de la forme :

$$y_{t+1} = F(y_t, x_t, \theta, \varepsilon_t) \quad (1)$$

où y_t est un vecteur de variables endogènes, x_t un vecteur de variables exogènes, ε_t un vecteur de chocs aléatoires et θ un vecteur de paramètres.

Pour montrer que la procédure d'évaluation des politiques économiques à l'aide des modèles économétriques est incompatible avec le maintien du paramètre θ constant, Lucas inscrit d'emblée son raisonnement dans un cadre d'analyse qui s'appuie sur le comportement dynamique optimisateur d'agents individuels. Il postule, d'une part, que la fonction F et le paramètre θ sont issus de règles de décision optimales des agents économiques. D'autre part, il suppose qu'une formulation correcte de ces règles de décision est tributaire de la connaissance par ces agents du comportement futur des variables les concernant. En particulier, Lucas considère qu'il serait malaisé de déterminer une règle de décision optimale d'un agent lorsque la politique économique, représentée par la variable exogène x_t , suit une évolution arbitraire. Il privilégie la représentation de la politique économique sous la forme d'une règle qui dépend des variables endogènes du modèle et d'un vecteur de paramètres λ représentant la nature de cette politique :

$$x_t = G(y_t, \lambda, \eta_t) \quad (2)$$

où η_t est un vecteur de perturbations. Sous ces hypothèses, maintenir l'hypothèse de stabilité du paramètre θ revient à supposer que les agents ne modifient pas leur opinion sur le comportement des chocs récurrents qui affectent le système, et de fait, n'adaptent pas leur règle de décision aux variations des grandeurs qu'ils cherchent à anticiper. Selon Sargent (1981), les règles de décisions ne peuvent pas rester invariantes en présence d'interventions de politique économique qui affectent les contraintes des agents. D'après ces arguments théoriques, les estimations économétriques historiques de ces règles ne peuvent fournir que de mauvaises prévisions sur les comportements dans un nouvel environnement.

A partir de ses considérations théoriques, Lucas estime que le paramètre θ est censé refléter la réaction des agents aux différents choix des politiques x_t , et, qu'en principe, ce paramètre doit varier si ces politiques changent. Pour cela, le modèle (1) doit plutôt s'écrire :

$$y_{t+1} = F(y_t, x_t, \theta(\lambda), \varepsilon_t) \quad (3)$$

De fait, les règles de décision optimales des agents économiques décrites par la structure du modèle économétrique (3) varient systématiquement sous l'effet d'un changement de la politique économique représentée par le paramètre λ à travers la modification simultanée de

la variable de politique x_t et du paramètre de comportement $\theta(\lambda)$. De plus, le changement de ce dernier est intimement lié à la façon dont la modification de la politique économique est mise en œuvre. Un modèle qui n'explique pas pourquoi les agents choisissent d'adopter telle ou telle règle de décision est sans aucune utilité pour l'évaluation des conséquences du changement d'une politique économique. En revanche, Lucas (1977) pense que ni les goûts ni la technologie ne varient de façon systématique suite à une modification de la politique économique. Un modèle d'équilibre de la forme :

$$y_{t+1} = F(y_t, x_t, \theta(\lambda), \mu, \varepsilon_t) \quad (4)$$

est défini pour prévoir comment les agents choisissent de réagir à de nouvelles situations, pour des goûts et une technologie donnés par un paramètre stable μ .

Si la politique économique est définie par des règles préalablement annoncées, il sera possible de prévoir la modification de $\theta(\lambda)$ suite à un changement de cette politique. En revanche, si le changement de la politique intervient de façon arbitraire et non annoncée, alors les agents économiques ne s'en rendront compte que graduellement, et de fait, la transformation de $\theta(\lambda)$ se fera de façon lente et non perceptible.

Toutefois, Lucas conditionne, en faisant référence à Knight (1921) et à Muth (1961), la capacité du modélisateur à prévoir les réactions des agents par la connaissance de ceux-ci du modèle utilisé et des chocs susceptibles de l'affecter. La référence à la connaissance du modèle renvoie au concept des *anticipations rationnelles* introduites et définies par Muth (1961) comme étant les prévisions qui sont fournies par ce qu'il appelle la théorie économique "*pertinente*". Si on désigne par Ω_{t-1} l'ensemble d'information disponible à la date $t-1$, l'anticipation rationnelle x_t^a correspondant à la variable x_t est donnée par l'espérance conditionnelle suivante :

$$x_t^a = E_t(x_t | \Omega_{t-1}) \quad (5)$$

L'écart entre la variable x_t et son anticipation rationnelle x_t^a diffère par un terme d'erreur aléatoire ε_t :

$$x_t - x_t^a = \varepsilon_t \quad (6)$$

Par ce concept, Muth cherche à souligner l'importance de la relation entre la conviction des individus et le comportement stochastique du modèle (Pirotte, 2004). En effet, ce concept tient compte de toute l'information disponible qui est susceptible d'influencer les variables du modèle. Autrement dit, le modèle en question doit convenablement décrire les caractéristiques de l'économie qui sont pertinentes pour la formation des anticipations des agents (Hoover, 1984). Dans ce cadre, l'hypothèse des *anticipations rationnelles* implique que, en l'absence d'éléments imprévisibles, les agents sont capables de prévoir, en moyenne et en fonction de l'information dont ils disposent, les résultats futurs du modèle théorique.

2. Optimisation intertemporelle et apurement des marchés

Dans l'un de ses articles fondateurs, Lucas (1977) critique Keynes (1936) d'avoir abandonné, dans sa Théorie générale, le postulat du comportement optimisateur des agents et le postulat de l'apurement des marchés, en cherchant à montrer l'existence du chômage involontaire. La démarche de Keynes a libéré des générations d'économistes de ce que Lucas (1977) qualifie de "*discipline imposée par la théorie de l'équilibre*", constituée des deux postulats précédents. De plus, c'est en raison de l'abandon de cette "discipline de l'équilibre" que Keynes (1936) a pu conclure qu'une théorie de l'équilibre au sens des auteurs classiques est inatteignable. De Vroey (2009) conçoit l'importance accordée par Lucas à la discipline de l'équilibre par la volonté de celui-ci de fixer sur une base théorique tous les paramètres, au sens large, du modèle. Sinon, un raisonnement en dehors de l'équilibre ouvre la voie à "l'arbitraire" en l'absence de règles précises et laisse une grande marge pour le modélisateur pour fixer les grandeurs du modèle.

Luca s'oppose à la vision traditionnelle de l'équilibre où celui-ci est réalisé uniquement à long terme (De Vroey, 2009). Il défend l'idée d'une explication du comportement du cycle en termes d'équilibre où les deux postulats précédents sont réalisés à chaque point du temps (Lucas et Sargent, 1979). La définition du cycle d'affaire, au sens de Lucas (Lucas, 1977), correspond aux fluctuations répétées du produit et de l'emploi comme réponse optimale d'un agent face à un mouvement des prix (Lucas, 1977). Cette réponse dépend de deux facteurs : a/ la manière avec laquelle l'agent interprète l'information contenue dans ce mouvement, et b/ sa

préférence concernant la substitution intertemporelle entre la consommation et le loisir (Lucas, 1977, p. 19). Lucas place cette hypothèse de substitution intertemporelle au cœur de son analyse et la considère comme étant à la base de tous les modèles dynamiques (Lucas, 1981). Comme le précise Hoover (Hoover, 1984), c'est la capacité d'optimisation des agents, en exploitant l'ensemble de l'information disponible, qui leur permet de maintenir en permanence leur situation d'équilibre. Les fluctuations économiques sont interprétées dans ce cadre comme la conséquence des changements non anticipés dans les variables qui affectent les décisions des agents (Lucas et Sargent, 1979).

L'une des implications radicales de cette thèse c'est que les variations conjoncturelles ne sont plus perçues comme la conséquence des dysfonctionnements d'une économie décentralisée, qui nécessitent une intervention de l'Etat pour y apporter des remèdes (De Vroey, 2010). Kydland et Prescott rejettent l'idée keynésienne selon laquelle la demande globale serait à l'origine des fluctuations économiques. Il serait contre-productif de chercher à stabiliser ces fluctuations puisqu'elles sont considérées comme étant optimales.

Sargent (1982) estime que le meilleur moyen d'éviter la critique de Lucas c'est d'interpréter le modèle (4) comme une fonction-objectif d'un *individu représentatif*, d'en déduire sa réaction face à un changement de politique économique ou d'autres contraintes, et d'utiliser ces résultats pour dériver les comportements agrégés de l'économie. L'établissement de relations structurelles dans ce modèle est présenté comme une justification du recours à un modèle à agent représentatif (De Vroey, 2010). De fait, la construction du modèle "*pertinent*" de l'économie exige d'expliquer les relations agrégées à partir de la résolution d'un programme d'optimisation d'un *agent représentatif* sous contraintes. Dans un cadre dynamique, le comportement des agents change lorsque les contraintes qu'ils subissent varient (Sargent 1981), car les changements dans la perception de leurs contraintes modifient les plans qui décrivent leur choix en fonction de l'information dont ils disposent (Sargent 1981).

Kydland et Prescott (1982) reprennent les exigences formulées par Lucas (*discipline de l'équilibre, anticipations rationnelles, substitution intertemporelle*), en plus de l'hypothèse de la concurrence parfaite, pour développer le premier modèle canonique des fluctuations économiques qui sera désigné par la suite de *modèle des Cycles Réels (RBC)*. Le projet des Nouveaux Classiques est d'expliquer les fluctuations du système économique comme des déviations par rapport à une tendance à partir du modèle de croissance néoclassique. Le cycle

est interprété comme le résultat du comportement optimisateur d'agents rationnels en réponse à des chocs exogènes de nature réelle, en particulier technologique, la prise en compte des chocs monétaires n'étant pas jugée nécessaire.

Sous-section 2. Une nouvelle posture épistémologique

1. Le statut formel non descriptif du modèle

La critique de Lucas a eu un impact dévastateur sur la recherche en modélisation économétrique s'inscrivant dans la tradition de la Commission Cowles ainsi que sur celle des modèles VAR. Les conséquences de cette critique ne se limitent pas uniquement à l'utilisation des modèles économétriques traditionnels pour l'évaluation de la politique économique. Mais, cette critique a révolutionné la conception et le rôle d'un modèle et elle a remis en cause les méthodes économétriques traditionnelles (Pirotte, 2004). Chez les Nouveaux Classiques, il devient essentiel pour les besoins d'identification de distinguer les paramètres de politique économique, des préférences et de la technologie des autres paramètres du modèle.

Dans ce nouveau programme de recherche, les ingrédients d'une macroéconomie complètement compatible avec la théorie microéconomique ont été développés à travers une réécriture de la macroéconomie et de la microéconomie dans un langage dynamique et stochastique (Lucas et Sargent, 1979 ; Lucas, 1980). Dans ce cadre, les Nouveaux classiques cherchent à développer des modèles structurels basés sur des comportements microéconomiques optimisateurs en référence au programme de recherche initié par Lucas.

Ce programme s'appuie sur une nouvelle vision épistémologique⁷, constituée de deux éléments essentiels, portant sur la nature d'un modèle théorique (De Vroey, 2009). D'un côté, Lucas abandonne la distinction entre la notion de théorie et de modèle dans le sens où les deux sont fusionnées. Cette position tranche radicalement avec la conception antérieure qui considérait la théorie comme un ensemble de propositions à propos de la "réalité", et que le modèle constitue une tentative pour exposer certaines implications de la théorie. D'un autre

⁷ Pour une analyse sur le point de vue épistémologique de Lucas, voir Vercelli (1991). Pour des lectures plus récentes sur cette position épistémologique, voir De Vroey (2011) et Taouil (2014 a).

côté, Lucas conçoit le modèle théorique comme un système parallèle ou une imitation de l'économie et non pas un ensemble de propositions sur le comportement d'une économie réelle. Dans cette perspective, *"un bon modèle ne sera pas plus réaliste qu'un modèle moins bon, mais il fournira une meilleure prédiction"* (Lucas, 1980 a). Pour Lucas (1987 b), un modèle doit être nécessairement abstrait et *"irrél"* pour qu'il puisse aider à résoudre des questions pratiques de politique économique. En raison de la difficulté d'effectuer des expérimentations sur les effets de telles politiques, un modèle est perçu comme une fiction ou une économie artificielle dont la manipulation permet de comprendre le fonctionnement de l'économie réelle (Lucas, 1981 b).

Dans le même sens, Lucas (1987 b) souligne que *"la recherche du réalisme d'un modèle économique pervertit son utilité potentielle à penser la réalité"*. Les principales hypothèses des modèles, comme celle des anticipations rationnelles ou la fiction de l'agent représentatif, constituent dans ces conditions des énoncés analytiques et des principes de modélisation plutôt que des assertions à propos de la réalité. A ce titre, la vision de Lucas rejoint celle de Friedman (1953) concernant la récusation du réalisme des hypothèses d'un modèle macroéconomique (Taouil, 2014 a). Les hypothèses les plus fécondes pour la construction théorique sont celles qui ont un faible contenu descriptif. En conséquence, une théorie ne peut être remise en cause en raison de l'absence de pertinence empirique de ses hypothèses (Taouil, 2014 b).

En dépit de son caractère abstrait, le modèle doit être testé, estime Lucas, par rapport aux données d'observation (Lucas, 1977, 1981, De Vroey, 2011). Ainsi, le test empirique de la théorie constitue une étape cruciale puisque le modèle s'appuie sur des axiomes abstraits qui sont nécessairement faux (Lucas, 1986). Dans cette perspective, Sargent (1981) juge que l'utilisation des techniques économétriques dynamiques traditionnelles doit évoluer afin de devenir cohérente avec le principe d'influence des contraintes sur les règles de décisions des agents. Cela entraîne d'importants ajustements dans la manière avec laquelle un modèle économétrique est formulé, identifié et estimé.

2. Calibrage versus tests économétriques

Ce sont Kydland et Prescott (1982) qui ont procédé à la mise en œuvre du programme de Lucas pour la confrontation de la théorie avec les données à travers l'application de la technique du calibrage aux modèles *RBC*. Cette technique est présentée comme une stratégie alternative aux méthodes économétriques traditionnelles pour le calcul des paramètres structurels de ces modèles. Les propriétés cycliques obtenues par le modèle sont conditionnelles à ces paramètres qui sont dérivés des règles de décisions optimales des agents fournies par la théorie. Le calibrage est réalisé à partir de données indépendantes de ces propriétés cycliques. Dans la plupart des cas, les valeurs des paramètres correspondent à des valeurs moyennes tirées de données d'enquêtes ou de données des comptes nationaux. Dans certains cas, ces valeurs sont estimées à partir de données individuelles.

En s'appuyant fortement sur la théorie, le calibrage reflète davantage une méthode pour l'établissement d'une correspondance entre le modèle contraint et les données observées de façon à reproduire certains faits stylisés (Cooley, 1997). Le calibrage n'a pas pour ambition de se substituer aux méthodes économétriques traditionnelles, et de ce fait il n'exclut pas le recours à ces méthodes pour l'estimation de certains paramètres du modèle. Mais, l'économétrie ne peut pas remettre en cause ces modèles, en exclure ou y adjoindre des variables sur une base purement empirique et modifier de la sorte la structure du modèle. La spécification des équations du modèle est l'affaire de la théorie économique et non de l'économétrie. Le degré de confiance en la réponse quantitative du modèle à une question dépend, selon Kydland et Prescott (1991), du degré de confiance en la théorie économique. Ainsi, la technique de mesure utilisée s'appuie plutôt sur des considérations théoriques que sur des critères statistiques qui rechercheraient la meilleure adéquation possible du modèle avec les données observées.

A cet effet, Faust et Whiteman (1997) relèvent deux questions fondamentales qui opposent profondément les approches des modèles *RBC*, des modèles *VAR* et des méthodes du général au particulier dans la tradition de la *London School of Economics* (Hendry, 1995). La première porte sur les caractéristiques des données à reproduire par le modèle et la seconde concerne l'ampleur des modifications théoriques à introduire dans la spécification empirique pour améliorer la qualité de l'ajustement.

Section 2. La cohérence temporelle des politiques économiques : un renouveau de l'économie politique

Partant d'une conception de l'équilibre économique comme un jeu stratégique entre les agents économiques, Kydland et Prescott (1977) établissent la supériorité d'une politique économique basée sur une règle par rapport à celle discrétionnaire. En introduisant les notions de crédibilité et de réputation, Barro et Gordon (1983) montrent que le décideur public est obligé de se conformer à une règle de politique monétaire pour éviter à l'économie d'être piégée dans un équilibre sous-optimal.

L'altruisme intergénérationnel adopté par Barro (1974), dans un cadre d'anticipations rationnelles, inscrit le comportement du Ménage et de l'Etat dans un horizon temporel infini et remet en cause les propositions monétaristes. La contrainte budgétaire intertemporelle de l'Etat rend le mode de financement des dépenses publiques le lieu d'une interdépendance entre les instruments des politiques économiques et soulève le problème de la stabilité dynamique de ces politiques (Sargent et Wallace, 1981).

Sous-section 1. L'incohérence temporelle de la politique discrétionnaire

1. Anticipations rationnelles et optimalité de la politique économique

Depuis son article de 1958, Friedman (Friedman, 1958) défend, sur la base d'arguments essentiellement empiriques, l'idée selon laquelle la règle de politique monétaire domine sa discrétion. C'est seulement en 1968 (Friedman, 1968) qu'il réussit à apporter la preuve théorique d'une telle proposition à partir d'une relecture de la courbe de Phillips sur la base de l'hypothèse des anticipations adaptatives. En inscrivant leur raisonnement dans le cadre d'une interaction stratégique entre acteurs privés et autorités publiques et en s'appuyant sur l'hypothèse des anticipations rationnelles, Kydland et Prescott (1977) confirment cette intuition de Friedman et l'étendent à toute politique économique en général (Hairault et Langot, 2005).

Kydland et Prescott inscrivent leur analyse dans le droit fil des contributions de Lucas et montrent que les interactions entre le décideur public et les agents privés peuvent conduire à

un équilibre sous-optimal. Pour montrer ce résultat, ils supposent que les décisions des agents à la période courante x_t dépendent de leurs décisions passées, des politiques économiques passées, présentes π_t et anticipées pour les périodes futures :

$$x_t = X_t(x_{t-j}, \dots, x_{t-1}, \pi_{t-j}, \dots, \pi_t, \pi_{t+1}, \dots, \pi_{t+k}) \quad (7)$$

Une séquence de politiques économiques $\pi = (\pi_{t-j}, \dots, \pi_t, \pi_{t+1}, \dots, \pi_{t+k})$ est optimale si elle permet de maximiser, sous la contrainte (7), une fonction-objectif sociale qui dépend des valeurs passées, présentes et anticipées des décisions des agents et des politiques économiques:

$$S_t(x_{t-j}, \dots, x_t, x_{t+1}, \dots, x_{t+k}, \pi_{t-j}, \dots, \pi_t, \pi_{t+1}, \dots, \pi_{t+k}) \quad (8)$$

La politique économique $\pi = (\pi_{t-j}, \dots, \pi_t, \pi_{t+1}, \dots, \pi_{t+k})$ est dite temporellement cohérente si, pour chaque période t , la politique π_t maximise la fonction-objectif (8) sous la contrainte (7) pour des valeurs données x_{t-j}, \dots, x_{t-1} des décisions passées des agents (Kydlan et Prescott, 1977). L'équilibre économique est considéré comme un jeu où les individus ont une bonne connaissance des différentes stratégies de l'Etat. Sous l'hypothèse d'anticipations rationnelles, les agents tiennent compte, de façon cohérente, dans leurs décisions des actions présentes et futures du décideur public.

A partir de ce cadre d'analyse, Kydlan et Prescott montrent qu'une politique temporellement cohérente à la période t est optimale si les décisions des agents dépendent uniquement des politiques économiques présentes et passées et non de celles anticipées. Or, sous l'hypothèse des anticipations rationnelles, les agents prennent aujourd'hui des décisions en fonction des actions publiques futures. Il en résulte qu'une séquence de politique économique discrétionnaire, même temporellement cohérente, c'est-à-dire, optimale à chaque période compte tenu des informations disponibles, peut conduire à une solution non optimale si elle ne tient pas compte des anticipations des agents. Pour la même raison, une solution optimale ne correspond pas nécessairement à une décision temporellement cohérente à une période donnée. La raison de cet apparent paradoxe, selon Kydlan et Prescott, est qu'il s'agit d'un jeu entre agents économiques rationnels plutôt qu'un jeu contre la nature.

A partir de ce résultat, les auteurs montrent la supériorité d'une politique économique fondée sur une règle par rapport à un pouvoir discrétionnaire. Certes, ce dernier peut choisir la meilleure décision, mais uniquement à partir d'une situation donnée qui peut conduire à une solution sous optimale. Les actions de politique économique, non compatibles avec les objectifs des décideurs publics, sont jugées par les individus comme étant temporellement incohérentes et, de fait, ne sont pas intégrées dans la formation de leurs anticipations et dans leur prise de décision. Les problèmes d'incohérence temporelle exigent de s'intéresser aux règles de politique économique applicables à long terme. L'engagement du décideur public en faveur d'une règle peut être une source d'efficacité étant donné que les agents privés agissent en fonction des actions futures de ce décideur. Paradoxalement, comme le souligne, le raisonnement de Kydland et Prescott oblige les autorités publiques à se lier les mains par l'adoption d'un cadre institutionnel qui permet de limiter leurs marges de manœuvre et d'éviter les effets négatifs des stratégies discrétionnaires. En conséquence, Kydland et Prescott (1977) réfutent les politiques de stabilisation keynésiennes destinées à la régulation de l'activité économique.

2. Crédibilité et indépendance de la Banque centrale

En introduisant les notions de crédibilité et de réputation, Barro et Gordon (1983, a, b) montrent que le décideur public est obligé de se conformer à une règle de politique monétaire pour éviter à l'économie d'être piégée dans un équilibre sous-optimal. Comme gage de cette crédibilité, Rogoff (1985) propose de choisir, dans le cadre d'une Banque centrale indépendante vis-à-vis du pouvoir politique, un Gouverneur conservateur qui accorde un poids plus important que la société à la lutte contre l'inflation.

Le résultat de Kydland et Prescott (1977) a été appliqué dans le cadre du dilemme inflation-chômage pour montrer la prédominance d'une politique monétaire fondée sur une règle par rapport à une politique discrétionnaire (Barro et Gordon, 1983, a). Pour cela, on considère une courbe d'offre à la Lucas (1972), représentée par une courbe de Phillips augmentée des anticipations d'inflation, qui décrit le comportement des agents privés :

$$u_t = \lambda.(\pi_t^a - \pi_t) + u^* \quad (9)$$

où u_t est le taux de chômage courant, u^* le taux de chômage naturel, π_t le taux d'inflation courant, π_t^a le taux d'inflation anticipé et λ un paramètre positif. Les agents privés négocient et signent, en début de période, leurs contrats salariaux en termes nominaux en fonction de leur anticipation d'inflation π_t^a .

D'après la relation (9), le Gouvernement peut chercher à réduire le taux de chômage, de façon temporaire, en dessous de son niveau naturel en fixant, par une politique monétaire expansive, le taux d'inflation au-dessus de son niveau anticipé par les agents privés. Mais, il ne peut atteindre cet objectif qu'en usant de la surprise, car sinon, l'effet expansif de la politique monétaire sur l'emploi sera nul puisque les agents privés sont capables d'anticiper correctement le taux d'inflation. Ainsi, une politique monétaire discrétionnaire est inefficace, puisqu'elle ne fait qu'augmenter le niveau de l'inflation sans modifier le taux de chômage qui reste situé à son niveau naturel. Dans ce cas, le comportement optimal du Gouvernement consiste à adopter un taux d'inflation nul.

Conformément à l'analyse de Kydland et Prescott (1977), cette politique d'engagement à pratiquer une inflation nulle n'est pas temporellement cohérente. En effet, si les anticipations des agents sont formées sur la base d'une inflation nulle, la meilleure réponse du décideur public consiste à tricher et à augmenter le taux d'inflation pour réduire le taux de chômage. Or, les agents privés forment des anticipations rationnelles, réagissent de façon optimale à leur environnement et connaissent les préférences du décideur public qui cherche, en poursuivant un objectif de chômage et d'inflation, à minimiser le coût de l'inflation donné par la formule:

$$S(u_t, \pi_t) = \frac{1}{2}(u_t - \alpha.u^*)^2 + \frac{\beta}{2}.\pi_t^2 \quad (10)$$

Le premier terme de cette fonction mesure le coût positif, à court terme, de l'inflation non anticipée sur l'écart du taux de chômage effectif par rapport à un taux de chômage objectif $\alpha.u^*$ qui est fonction du taux de chômage naturel, où α est un paramètre positif inférieur à 1. Le second terme mesure le coût négatif de l'inflation, où β représente le poids que le Gouvernement accorde à l'inflation relativement au chômage.

L'action du décideur public est conditionnée par les anticipations des agents privés qui connaissent les objectifs de celui-ci et le comportement qu'il va adopter conformément aux équations (9) et (10). Le jeu non coopératif entre le Gouvernement et les agents privés, qui savent que celui-ci cherche à les bernier en augmentant le taux d'inflation, conduit à un équilibre sous-optimal avec un taux de chômage à son niveau naturel et un taux d'inflation positif donné par :

$$\pi_t = \frac{\lambda \cdot (1 - \alpha) \cdot u^*}{\beta} \quad (11)$$

Ce biais inflationniste est d'autant plus important que le poids accordé par le décideur public à la réduction du chômage est important, c'est-à-dire que β est petit, et que l'objectif du chômage est proche du taux naturel, c'est-à-dire que α est également petit.

Barro et Gordon (1983, b) ont proposé une solution au problème de la tentation de l'autorité publique à renier son engagement initial. Ils considèrent, de façon répétée, un jeu de politique monétaire qui tient compte du résultat du jeu antérieur. Les auteurs montrent que lorsque le Gouvernement s'engage à poursuivre une politique monétaire contrainte par une norme et que les agents privés croient en cette règle, le coût de l'inflation est inférieur à celui d'une politique discrétionnaire. Toutefois, lorsque les agents privés peuvent être systématiquement trompés en conservant des anticipations de faible inflation à chaque période, la tricherie du Gouvernement procure un optimum qui est supérieur à celui du respect de la règle.

Dans cette dernière situation, les agents privés remarquent que le décideur public n'a pas respecté la norme qu'il a annoncée à la période précédente et ils en déduisent qu'il a agi de façon discrétionnaire. En conséquence, le coût de la tricherie sera des anticipations d'inflations plus fortes pour le futur, ce qui permet aux agents privés de sanctionner le décideur public en perdant confiance dans ses engagements, car il les a induits en erreur. La solution discrétionnaire non coopérative pour les périodes suivantes constitue une punition pour l'abandon de la règle. Cette perte de crédibilité trouve son origine dans la tentation de tricherie du Gouvernement qui est supérieure à son incitation à la discipline. Ceci montre l'intérêt de faire des engagements contractuels entre les agents privés et le décideur public qui soient crédibles et qui puissent permettre à celui-ci de construire une réputation conforme aux anticipations de ceux-là et à leur capacité de mémoire.

Sous-section 2. La subordination de la politique budgétaire à la politique monétaire

1. Policy-mix et mode de financement des dépenses publiques

Les résultats du modèle économétrique de la Réserve Fédérale de Saint Louis développé par Andersen et Jordan (1968) ont été à l'origine d'un vif débat entre monétaristes et keynésiens sur l'efficacité relative de la politique monétaire et la politique budgétaire pour la stabilisation de l'activité et des prix. Ces résultats confirment l'étude économétrique initiée par Friedman et Meiselman (1963) et montrent, dans un cadre dynamique, la supériorité de la politique monétaire sur la politique budgétaire.

Au niveau théorique, les monétaristes insistent sur l'importance exclusive de la règle de croissance de la masse monétaire dans la lutte contre l'inflation et sur l'inefficacité à long terme de la politique budgétaire. Ils estiment que la politique budgétaire ou fiscale n'engendre que des effets temporaires puisque des effets d'éviction finissent par annuler l'impact initial (Friedman, 1972). Ces arguments conduisent à l'élimination des politiques économiques actives et, de fait, à l'abandon de toute idée de *policy-mix*. Mais, lorsque cela est nécessaire, ils préfèrent le recours à l'emprunt, qu'ils assimilent à un impôt différé, pour le financement du déficit public et ils refusent l'utilisation de la création monétaire.

En privilégiant le financement du déficit public par emprunt, les recommandations des monétaristes soulèvent, comme le relève Villieu (1998), la question de la soutenabilité de la politique monétaire dans les situations où la dette publique est élevée. En effet, une hausse du taux d'intérêt entraîne une augmentation du revenu disponible et de la consommation des Ménages qui peut être supérieure à la réduction de l'investissement. L'augmentation de la demande globale qui en résulte conduit à des pressions inflationnistes contraires à l'objectif de la politique monétaire restrictive initiale. Ainsi, lorsque la contrainte budgétaire intertemporelle de l'Etat est prise en compte, les propositions des monétaristes ne tiennent plus.

En effet, Blinder et Solow (1973) et Buiter et Tobin (1976) estiment que le respect de la contrainte budgétaire de l'Etat peut conduire à des effets boule de neige à travers l'alimentation de la charge de la dette par les paiements futurs des intérêts. Le recours à un financement monétaire peut constituer dans ce cas une solution au problème de la solvabilité

de l'Etat à long terme. Mais, dans ce genre de situation, le risque d'une utilisation exagérée de l'émission monétaire est de provoquer une dynamique inflationniste. Blinder et Solow (1973) et Buiter et Tobin (1976) considèrent qu'un financement monétaire du déficit public peut, à court terme, augmenter l'effet du multiplicateur en réduisant la pression sur le taux d'intérêt et allégeant l'effet d'éviction. Mais, ils admettent que le respect de la contrainte budgétaire de l'Etat à long terme conduit à l'égalité de la valeur du multiplicateur quel que soit le mode de financement du déficit.

L'équivalence, soulevée par les monétaristes, entre le financement aujourd'hui par emprunt d'un déficit public résultant d'une baisse d'impôts et son financement par des impôts futurs a été analysée de façon plus rigoureuse par Barro (1974) dans le cadre du principe de l'équivalence ricardienne. Sous l'hypothèse d'anticipations rationnelles, les consommateurs tournés vers le futur savent que la dette publique sera tôt ou tard financée par une augmentation des impôts. Dans leur décision de consommation, ils ne tiennent pas compte uniquement de leur revenu courant, mais également de leur revenu futur ainsi que de celui de leur descendance qui se perpétue indéfiniment. Cet altruisme intergénérationnel inscrit l'analyse de Barro du comportement du Ménage et de l'Etat dans un horizon temporel infini. Dans ce contexte, la réduction de l'impôt financée par emprunt laisse inchangée la consommation des ménages qui préfèrent épargner la partie correspondante de leur revenu disponible pour payer les impôts futurs. Il en résulte que la baisse de l'épargne publique est exactement compensée par une hausse de l'épargne privée de sorte que l'épargne globale reste inchangée et la baisse d'impôts n'exerce aucun effet sur la demande globale.

Il convient de rappeler que l'équivalence ricardienne concerne uniquement le financement par emprunt et par impôt, et non le financement monétaire des dépenses publiques. En particulier, l'effet sur la richesse des individus d'un financement du déficit public par une création monétaire n'est pas annulé par l'anticipation d'une augmentation future des impôts. Certes, cette richesse nominale peut être érodée par l'inflation résultant de l'expansion monétaire (McCallum, 1986). Le principe de l'équivalence ricardienne permet aux Ménages d'intégrer la contrainte budgétaire de l'Etat à la leur et offre une base d'analyse pour la comparaison des effets de la politique budgétaire et monétaire.

2. Le *policy-mix* comme jeu stratégique

Initialement concerné par la question de l'affectation optimale des instruments aux objectifs suite à l'article fondateur de Mundell (1962), le problème du *policy-mix*, défini comme l'articulation optimale entre politique budgétaire et politique monétaire, est devenu celui de la coordination entre les autorités de politique économique (Villieu, 1998). Suite à la contribution de Kydland et Prescott (1977) relative aux problèmes d'incohérence temporelle, les règles de politique économique sont étudiées dans une perspective de long terme dans le cadre d'un jeu stratégique entre agents privés, Banque centrale et Gouvernement. La contrainte budgétaire intertemporelle de l'Etat rend le mode de financement des dépenses publiques le lieu d'une interdépendance entre les instruments de politique économique et soulève le problème de la stabilité dynamique de ces politiques (Sargent et Wallace, 1981).

Dans cette perspective, Sargent et Wallace (1981) remettent en cause la pertinence des propositions monétaristes et considèrent que la manière avec laquelle la politique monétaire et la politique budgétaires sont coordonnées influence la capacité de la Banque centrale à contrôler l'inflation. Leur modèle s'appuie sur une contrainte budgétaire où le déficit public est financé soit par émission de titres soit par création monétaire (ou seignuriage) :

$$b_t = (1 + r_{t-1} - g_{t-1})b_{t-1} + d_t - (\pi_{t-1} + g_{t-1})m_t \quad (12)$$

où b_t , d_t et m_t représentent, respectivement, les ratios par rapport au revenu réel du stock de la dette publique, du solde primaire du budget de l'Etat et du stock de la monnaie, toutes ces variables exprimées en termes réels. Les variables r_t , g_t et π_t désignent, respectivement, le taux d'intérêt réel, le taux de croissance du revenu réel et le taux d'inflation.

En l'absence de coordination entre les autorités budgétaire et monétaire, l'indépendance de la Banque centrale la met en situation de conflit stratégique avec le Gouvernement où chacun poursuit son propre objectif, sans nécessairement tenir compte de la cohérence d'ensemble du *policy-mix*. Lorsque le taux d'intérêt devient supérieur au taux de croissance de l'économie, l'équation (12) met clairement en évidence le problème de la soutenabilité de la dette qui devient explosive à long terme. Si le Gouvernement refuse de limiter son déficit, la Banque centrale est contrainte, pour atténuer ce processus, de créer plus de monnaie pour monétiser une partie de la dette publique et augmenter le seignuriage qui est donné par le dernier terme

de l'équation (12). En effet, une politique monétaire restrictive visant à lutter contre l'inflation uniquement par le contrôle de la masse monétaire peut être insoutenable à long terme, car elle est incompatible avec cette contrainte budgétaire intertemporelle de l'Etat. Cette dernière configuration conduit à une instabilité de la contrainte budgétaire de l'Etat en raison des incertitudes quant à l'évolution future de la masse monétaire et du déficit public.

En revanche, si la Banque centrale réussit à imposer une rigueur monétaire, alors le Gouvernement doit contenir le déficit de son budget qu'il financera par emprunt et par un montant fixé de seignuriage. Pour Sargent et Wallace, la capacité d'engagement de la Banque centrale sur une règle monétaire lui permet de se comporter en meneur du jeu, dans le cadre d'un régime de politique monétaire dominante, et d'imposer sa décision et une discipline à l'autorité budgétaire. En excluant toute possibilité de financement monétaire du déficit public, l'autorité monétaire oblige le Gouvernement à fixer son déficit à un niveau compatible avec les sources de financement disponibles et se constitue une réputation qui lui permet d'assurer la stabilité des prix.

Les interdépendances stratégiques entre les autorités de politique monétaire et budgétaire posent de façon cruciale la question de la coordination et de la crédibilité du *policy-mix* (Villieu, 1998). La contrainte budgétaire de l'Etat constitue un lieu de conflit potentiel entre politique budgétaire et politique monétaire qui exige une coordination qui puisse permettre d'éviter l'incohérence temporelle de la politique économique. L'analyse de Sargent et Wallace souligne l'obligation pour la Banque centrale de tenir compte de la cohérence d'ensemble de tous les instruments de la politique économique pour assurer la stabilité des prix à long terme. En outre, cette analyse soulève le problème de la coordination entre la politique monétaire et la politique budgétaire dont la cohérence est déterminante pour la crédibilité du *policy-mix* dans son ensemble. Surtout que la présence d'un niveau élevé d'endettement public constitue une incitation forte pour pratiquer des surprises d'inflation (Calvo, 1988).

Conclusion

Suite à la critique de Lucas (1976), l'invariance des paramètres d'un modèle économétrique jette un doute sur l'utilité d'un tel modèle pour l'évaluation de la politique monétaire dans un environnement d'anticipations rationnelles. L'alternative proposée par Lucas (1977) est de décrire le comportement du cycle économique en termes de discipline de l'équilibre dans le cadre d'une vision épistémologique qui conçoit le modèle comme une imitation de l'économie (Lucas, 1980, 1986). La technique du calibrage est considérée comme une stratégie alternative aux méthodes économétriques traditionnelles pour fixer les paramètres du modèle (Kydland et Prescott, 1982).

Sous l'hypothèse des anticipations rationnelles, Kydland et Prescott (1977) analysent les interactions entre le décideur public et les agents privés dans le cadre d'un jeu stratégique, ce qui les conduit à déplacer le débat sur l'efficacité de la politique monétaire vers celui des institutions en charge de celle-ci. Dans cette perspective, le choix d'un Gouverneur conservateur à la tête d'une Banque centrale indépendante est retenu comme solution contre la perte de crédibilité en la politique monétaire (Barro et Gordon, 1983 ; Rogoff, 1985). La capacité d'engagement de l'autorité monétaire sur une règle lui permet de se comporter en meneur d'un *policy-mix* où la politique budgétaire serait subordonnée à la politique monétaire (Sargent et Wallace, 1981).

Chapitre 6

Les Modèles DSGE et l'Extension de la Méthodologie des Cycles Réels

Introduction

Les modèles DSGE (*Dynamic Stochastic General Equilibrium*) constituent les modèles de base de la théorie de la Nouvelle Synthèse (Goodfriend et King, 1997 ; Clarida, Gali et Gertler, 1999,) et forment, depuis le milieu des années 2000, la référence pour l'étude des conditions d'exercice de la politique monétaire par les Banques centrales (Woodford, 2003). Toutefois, en dépit de l'importance accordée au taux d'intérêt nominal en tant qu'instrument de la politique monétaire, les conditions d'équilibre de marché sont établies dans ces modèles indépendamment de toute référence à la monnaie (Taouil, 2014 a). En effet, en adoptant les principes de la discipline de l'équilibre, la conception de ces modèles est fortement influencée par le programme de recherche lucasien et s'inscrit dans la continuité de la méthodologie des modèles RBC.

De plus, la conception d'un modèle comme étant, selon Lucas, une représentation formelle et non descriptive de l'économie, conduit les adeptes des DSGE à privilégier le recours aux techniques du calibrage pour fixer les valeurs des paramètres. Les fortes restrictions simplificatrices, basées sur des *a priori* théoriques, sur la structure d'un DSGE ne permettent pas de tester ces modèles par les outils économétriques traditionnels. Simultanément, la préférence pour le calibrage connaît une évolution, pour des raisons de proximité conceptuelle, vers une combinaison avec les méthodes d'estimation bayésienne et les techniques d'inférence simulée.

L'explication de l'existence d'effets réels de la politique monétaire dans les modèles DSGE s'appuie sur l'hypothèse de rigidité à court terme des prix. Associée avec les imperfections du marché, cette rigidité joue un rôle central dans l'explication des déviations de court terme par rapport à l'équilibre concurrentiel de long terme. L'inefficacité de l'allocation des ressources à laquelle conduisent ces déviations, justifie l'utilisation de la politique monétaire pour contrer les pressions inflationnistes et les écarts au produit potentiel.

La maquette de base représentative d'un modèle DSGE s'appuie sur une réécriture de la relation IS et de la courbe de Phillips. La première décrit le comportement de la demande agrégée, dont le fondement théorique en équilibre général dynamique est constitué de l'équation d'Euler de la consommation. La version Nouvelle Keynésienne de la courbe de

Phillips résume le processus de formation des prix par les firmes en concurrence monopolistique. Elle est dérivée à partir d'un modèle de tarification optimale en concurrence imparfaite et d'une théorie de rigidités des prix (Roberts, 1997 ; Clarida et *ali.*, 1999 ; McCallum et Nelson, 1999 ; Svensson et Woodford, 2004 ; Woodford, 2003 ; Gianoni et Woodford, 2005). Ces deux équations sont complétées par une règle de Taylor qui représente un substitut à la courbe LM (Taylor, 1993, 1999) et permet aux autorités monétaires de cibler le niveau optimal de l'inflation. Elle est déduite de l'objectif de la Banque centrale sous les contraintes de la courbe de Phillips et de la demande agrégée.

Le cadre d'analyse d'un modèle DSGE accorde à la Banque centrale un rôle *leader* dans la conduite d'un *policy mix* s'articulant autour de la crédibilité monétaire et de la discipline budgétaire pour assurer la stabilité des prix. Ainsi, la représentation de la politique monétaire dans un modèle DSGE s'inscrit dans la continuité des travaux de Barro et Gordon (1983 a) qui ont montré la prédominance d'une politique monétaire fondée sur une règle par rapport à une politique discrétionnaire.

Dans le cadre du débat sur les causes de la "Grande Modération" aux Etats-Unis, Benati et Surico (2009) estiment que, contrairement aux modèles DSGE, l'utilisation des modèles VAR structurels (SVAR) ne permet pas toujours de bien identifier un changement de règle de politique monétaire. Cet échec des modèles SVAR n'est pas lié à des problèmes d'estimation comme le soutient Canova (2006), mais il est la conséquence directe des restrictions inter-équations imposées par l'hypothèse des anticipations rationnelles tel que cela a été initialement soulevé par la critique de Sargent (1979) à l'encontre de ces modèles.

Section 1. Les fondements des modèles DSGE

Les modèles DSGE (*Dynamic Stochastic General Equilibrium*) constituent la référence pour l'étude des conditions d'exercice de la politique monétaire par les Banques centrales (Woodford, 2003). Toutefois, la conception de ces modèles est fortement influencée par la vision lucasienne de modélisation et elle s'inscrit dans le prolongement de la méthodologie des modèles RBC. L'existence d'effets réels de la politique monétaire dans ces modèles est expliquée par la rigidité à court terme des prix. Les déviations par rapport à l'équilibre concurrentiel de long terme conduisent à une inefficience de l'allocation des ressources et justifient l'utilisation de la politique monétaire pour lutter contre l'inflation et réduire les écarts au produit potentiel.

Le noyau de la maquette d'un modèle DSGE est constitué par une relation IS et une courbe de Phillips. La première décrit le comportement de la demande agrégée et la seconde explique la formation des prix dans un contexte de concurrence monopolistique.

Sous-section 1. L'influence de la méthodologie Lucasienne

1. Une hégémonie de l'approche réelle

Les modèles DSGE (*Dynamic Stochastic General Equilibrium*) constituent les modèles de base de la théorie de la Nouvelle Synthèse Néoclassique (Goodfriend et King, 1997), qualifiée de Nouvelle Synthèse Néokeynésienne (Clarida, Gali et Gertler, 1999) ou encore de Nouveau Consensus Monétaire (Mishkin, 2008). Cette théorie résulte d'un rapprochement entre la Nouvelle Economie Classique, portée par les modèles des Cycles Réels (RBC), et la Nouvelle Economie Keynésienne qui s'appuie sur un cadre d'analyse de concurrence imparfaite et de rigidités nominales. Mais, comme le relève Taouil (2011), à y regarder de près on découvre que, fortement influencée par la méthodologie lucasienne, la Nouvelle synthèse constitue plutôt une absorption de la Nouvelle Economie Keynésienne par la Nouvelle Economie Classique (voir la première section du chapitre 5 pour une présentation plus détaillée sur la méthodologie lucasienne).

D'abord, les modèles DSGE s'inscrivent dans la continuité des modèles RBC, en adoptant une approche de l'équilibre général dynamique et stochastique qui analyse les comportements intertemporels des agents sous l'égide de la discipline de l'équilibre lucasienne, constituée des postulats du comportement optimisateur des agents et de l'apurement des marchés (Lucas, 1977). Ensuite, la fiction de l'agent représentatif est retenue comme principe de base pour dériver, conformément à la critique de Lucas (1976), le comportement des macro-variables à partir de choix microéconomiques d'agents à anticipations rationnelles qui sont sensibles aux décisions de politique économique. Enfin, la Nouvelle Synthèse adhère à l'impératif de l'adoption de règles de politiques économiques en accordant une priorité à l'objectif de la stabilité macroéconomique (Barro et Gordon, 1983, Kydland et Prescott, 1977).

Les équations d'un modèle DSGE expriment des déviations de court terme par rapport à l'équilibre concurrentiel de long terme. Ces fluctuations, qui résultent de l'introduction des imperfections et des rigidités à court terme du marché, sont considérées comme des réactions optimales des agents face aux perturbations de leur environnement (Taouil, 2014 b). En soumettant les rigidités et les imperfections de marché à la norme de rationalité optimisatrice, la Nouvelle Synthèse étend ainsi le cadre explicatif de la discipline de l'équilibre (Taouil, 2011). De fait, en raison de cette optimalité du comportement des agents, l'économie est à l'équilibre à court terme et à long terme, le premier étant sous-optimal au vue de la norme de l'optimum parétien associé à une économie parfaitement concurrentielle.

Le produit naturel de long terme de l'économie est déterminé par les conditions de l'offre, en l'occurrence, les préférences des agents et la technologie de production sous l'hypothèse de la pureté des ajustements concurrentiels. Les paramètres du modèle, reflétant ces préférences et cette technologie, sont des données invariantes et indépendantes des comportements des agents et de la politique économique. La fonction d'utilité intertemporelle, le taux d'intérêt réel et la passivité de la monnaie jouent un rôle central dans la dérivation des équations du modèle, reflétant ainsi le statut hégémonique de l'approche réelle.

En dépit de l'importance qu'accorde la Nouvelle Synthèse au taux d'intérêt nominal dans la conduite de la politique monétaire, la description des comportements du consommateur par des fonctions d'utilité et du producteur par des fonctions de production conduit cette théorie à évacuer tout rôle de la monnaie comme unité de compte. En effet, les conditions d'équilibre du marché sont établies dans des termes réels en fonction des prix relatifs indépendamment de

toute référence à la monnaie. En tant qu'instrument de la politique monétaire, le taux d'intérêt nominal n'intervient dans la demande agrégée qu'à travers la relation de Fisher qui relie le différentiel du taux d'intérêt réel avec l'inflation anticipée par le ménage représentatif. Celui procède à un arbitrage intertemporel entre la consommation présente et la consommation future en fonction du coût d'opportunité représenté par le taux d'intérêt réel. L'utilisation du taux d'intérêt nominal comme règle d'action de la politique monétaire pour le contrôle de l'inflation conduit à l'endogénéisation de l'offre de monnaie qui ne sert plus de variable de contrôle.

2. Combinaison du calibrage et de l'économétrie

La conception de la politique monétaire dans les modèles DSGE est fortement influencée par l'approche des modèles RBC. Les fortes restrictions simplificatrices sur la structure d'un DSGE (voir les parties ci-dessous) altèrent naturellement la qualité de l'ajustement du modèle aux données (Vo Phuong et *ali*, 2012). Lucas et Prescott considèrent que ces modèles sont "mal spécifiés" et "faux" et qu'ils seraient de toute façon rejetés s'ils sont testés par les outils économétriques traditionnels. Ainsi, l'échec des DSGE à reproduire les données ne peut pas justifier leur rejet comme étant faux. Dans cette optique, on ne peut pas opposer les modèles VAR aux modèles DSGE car ils ne se situent pas sur le même plan et n'ont pas la même visée. Un VAR ne peut être considéré comme étant supérieur à un DSGE tout simplement parce qu'il offre un meilleur ajustement aux données.

Dans la continuité de la tradition des Cycles Réels, les tenants des modèles DSGE s'abstiennent de l'utilisation d'une démarche descriptive et formelle de l'économétrie et privilégient le recours aux techniques du calibrage. Simultanément, on assiste à une évolution de l'évaluation des DSGE vers une réconciliation du calibrage, qui préserve toujours un rôle central dans la détermination des paramètres du DSGE, avec les méthodes d'estimation bayésiennes, voire les méthodes d'inférence simulée, telles que le bootstrap et l'inférence statistique indirecte (Pirrotte, 2004 ; Vo Phuong et *ali*, 2012).

La préférence pour l'estimation bayésienne est généralement expliquée par sa proximité avec la technique du calibrage, puisqu'elle s'appuie sur des densités de probabilités *a priori* permettant d'étudier des mécanismes non répétitifs. Pour cette raison, l'approche bayésienne

est qualifiée de subjective à l'opposé de l'approche fréquentiste qui est à la base de l'économétrie traditionnelle. Dans un cadre bayésien, les paramètres du modèle sont considérés comme des variables aléatoires, décrites par des distributions qui résument l'incertitude qui entoure les valeurs réelles de ces paramètres. Ces distributions de probabilités expriment les croyances *a priori* qu'on a sur les paramètres et elles sont indépendantes de tout examen des données observées.

Les informations dérivées d'un échantillon de données permettent de transformer, à travers le théorème de Bayes, les distributions *a priori* en des distributions de probabilité *a posteriori*. Si on note par y le vecteur des observations et par θ le vecteur des paramètres du modèle, alors on peut écrire :

$$p(\theta|y) = \frac{p(\theta) \times p(y|\theta)}{p(y)} \quad (1)$$

où $p(\theta|y)$ désigne la distribution *posterior* de θ , $p(\theta)$ le *prior* et $p(y|\theta)$ la fonction de vraisemblance. Si on considère que $p(y)$ est un terme de normalisation, alors on peut écrire :

$$p(\theta|y) \propto \text{information a priori} \times \text{information de l'échantillon} \quad (2)$$

où le signe \propto signifie "proportionnel à".

Le développement des méthodes de simulation de Monte Carlo par chaînes de Markov ont offert de nouvelles perspectives pour faciliter le passage d'une loi *a priori* à une loi *a posteriori* (Pirrotte, 2004). Le logiciel Dynare, largement utilisé dans la modélisation DSGE, contient des algorithmes de simulation, comme celui de Metropolis-Hastings, pour calculer de manière empirique ces distributions de probabilité.

Sous-section 2. Demande agrégée et courbe de Phillips

1. Rigidités et concurrence monopolistique

Les modèles DSGE cherchent à apporter une solution à l'absence de monnaie dans les modèles RBC et réintroduire les considérations de la politique monétaire. La rigidité à court terme des prix permet de justifier les effets réels positifs durables en termes de bien-être d'une politique monétaire expansionniste, remettant ainsi en cause le point de vue défendu par Friedman et Lucas sur l'inefficacité d'une politique monétaire active (Malgrange, 2006). En raison de cette rigidité, une baisse du taux d'intérêt nominal se traduit par une diminution du taux d'intérêt réel et une substitution de la consommation à l'épargne. L'augmentation de la demande globale qui en résulte se perpétue au-delà de la période courante en raison de la présence de délais d'ajustement des prix. Les défauts d'ajustement s'avèrent ainsi un préalable à la viabilité de la politique monétaire (Taouil, 2011). Toutefois, l'opportunité de l'arbitrage entre inflation et production est valable uniquement à court terme car la flexibilité des prix à long terme rend inefficace la politique monétaire à cet horizon.

L'explication du comportement dynamique des prix à court terme et de leur rigidité par des décalages dans leurs ajustements est fournie par trois théories concurrentes : le modèle de contrats imbriqués de Taylor (1977) dans lequel les salaires constituent la principale cause de modification des prix, le modèle d'ajustement dynamique optimal (Rotemberg, 1982) qui suppose que la vitesse d'ajustement des prix est choisie de façon optimale et le modèle de tarification échelonnée de Calvo (1983) pour qui les variations des prix sont aléatoires. Les hypothèses de coût d'ajustement sont également utilisées pour l'explication des retards dans l'ajustement des prix (Mankiew, 1985 ; Parkin, 1986).

L'hypothèse de concurrence imparfaite sur le marché des biens confère aux entreprises un pouvoir de marché qui peut entraîner un niveau plus élevé des prix et un niveau plus faible de la production qu'en concurrence parfaite (Wickens, 2006). Etant rigide à court terme, le prix de vente de la firme peut se situer au-dessus de son coût marginal de production et permettre à celle-ci d'élever sa production pour répondre à une demande accrue même lorsque ce coût augmente. En fixant un taux d'ajustement de leurs prix, les producteurs endogénéisent de façon différenciée ce qu'ils perçoivent comme degré optimal de flexibilité de ces prix entraînant par là un changement dans les prix relatifs.

Dixit et Stiglitz (1977) ont développé l'un des modèles les plus utilisés pour la fixation des prix optimaux à long terme dans le cadre d'une concurrence imparfaite entre firmes produisant des biens imparfaitement substituables. L'équation de base du modèle s'appuie sur une relation entre le prix de la production et son coût marginal moyennant un taux de marge constant. Le coût marginal est égal au rapport entre le taux de salaire nominal et la productivité marginal du travail ; et le taux de marge est déterminé en fonction de la substituabilité entre les biens dans les fonctions de préférence des agents pour la diversité.

Les rigidités nominales, considérées comme des réactions optimales à des chocs par des individus optimisateurs dans le cadre de concurrence imparfaite et de formation rationnelle des anticipations, obligent les firmes à ne pas ajuster automatiquement leurs prix en cas de variation de la demande. Etant données leurs contraintes, les agents ne révisent pas les plans de leurs actions qui sont par définition conformes à leurs anticipations. Les déviations engendrées par ces rigidités par rapport au sentier optimal correspondant à la flexibilité des prix conduisent à une allocation inefficente des ressources et justifie l'action de la politique monétaire pour résoudre les problèmes des tensions inflationnistes et des écarts au produit potentiel (Gaffard, 2012). Selon Malgrange (2006), ce cadre d'analyse de rigidité des prix et de concurrence imparfaite permet également d'expliquer la persistance sur une longue période des effets réels de chocs monétaires sur des séries de macro-variables.

En procédant à une réécriture du modèle IS/LM, la maquette de base représentative des modèles DSGE est constituée de trois équations résumant le comportement de trois agents économiques, les ménages, les firmes et la Banque centrale. Ces comportements sont résumés par des équations de demande, d'inflation et de taux d'intérêt. Les deux premières équations, issues de cette maquette sont présentées dans le point suivant et la troisième équation est décrite dans la deuxième section ci-dessous.

2. Relation IS et courbe de Phillips néo-keynésiennes

La courbe IS décrit le comportement de la demande agrégée, dont le fondement théorique en équilibre général dynamique est constitué de l'équation d'Euler de la consommation. Celle-ci est dérivée à partir de la maximisation de la valeur actualisée de la fonction d'utilité

intertemporelle du ménage représentatif dont le problème consiste en la maximisation de l'expression (Wickens, 2006) :

$$E_t \left[\sum_{s=0}^{\infty} \beta^s U(c_{t+s}, l_{t+s}) \right] \quad (3)$$

sous les contraintes :

$$y_t = c_t + i_t \quad (4)$$

$$y_t = F(k_t, n_t) \quad (5)$$

$$\Delta k_{t+1} = i_t - \delta \cdot k_t \quad (6)$$

$$n_t + l_t = 1 \quad (7)$$

L'équation (4) constitue l'identité du revenu national y_t avec c_t la consommation et i_t l'investissement, (5) est la fonction de production où k_t est le stock de capital au début de la période t et n_t est le temps de travail du ménage, l'équation (6) représente l'identité des inventaires avec δ le taux de dépréciation du capital à chaque période, (7) exprime le temps total, normalisé à l'unité, disponible pour le ménage, exprimé comme la somme du temps de travail et du temps de loisirs.

L'utilité future $U(c_{t+s}, l_{t+s})$ est moins valorisée que l'utilité courante $U(c_t, l_t)$ car elle est actualisée par le paramètre β compris entre 0 et 1 et qui est fonction du paramètre d'escompte θ :

$$\beta = \frac{1}{(1 + \theta)} \quad (8)$$

E_t désigne l'opérateur des anticipations conditionnelles à l'information disponible à la période t . Les conditions du premier ordre du lagrangien correspondant à ce programme d'optimisation, combinées avec la condition de transversalité :

$$\lim_{s \rightarrow \infty} \beta^s U'_{c,t+s} \cdot k_{t+s+1} = 0 \quad (9)$$

permettent de déterminer la condition d'optimalité intertemporelle du programme du consommateur, donnée par la relation d'Euler :

$$E_t \left[\beta \cdot \frac{U'_{c,t+1}}{U'_{c,t}} \cdot (1 + r_{t+1}) \right] = 1 \quad (10)$$

où $U'_{c,k}$ désigne la dérivée de l'utilité à la période k par rapport à la consommation et r_t est le rendement réel du capital.

En utilisant une approximation à l'ordre un de la dérivée de l'utilité en période $t+1$:

$$U'_{c,t+1} \approx U'_{c,t} + U'' \cdot \Delta c_{t+1} \quad (11)$$

et une formulation logarithmique de la consommation, la relation (10) peut être mise sous la forme :

$$E_t \Delta \ln(c_{t+1}) = \frac{1}{\sigma} (E_t r_{t+1} - \theta) \quad (12)$$

où $\sigma = -\frac{c_t U''}{U'}$ est le coefficient d'aversion relative pour le risque.

Pour introduire l'expression de la production y_t dans la formule (12) considérons une approximation log-linéaire de la contrainte des ressources (4) :

$$\ln(y_t) = \frac{c}{y} \ln(c_t) + \frac{k}{y} [\ln(k_{t+1}) - (1 - \delta) \ln(k_t)] \quad (13)$$

En remplaçant $\ln(c_t)$ dans l'équation (12) par sa valeur tirée de l'équation (13) et en utilisant une approximation du premier ordre on obtient :

$$\ln(y_t) = E_t \ln(y_{t+1}) - \frac{1}{\sigma} (E_t r_{t+1} - \theta) \quad (14)$$

Exprimée en termes de déviations par rapport à l'état stationnaire, sachant que θ exprime le taux d'intérêt réel à l'état stationnaire et en omettant les symboles "ln", l'équation (14) permet de retrouver la courbe IS :

$$y_t = E_t(y_{t+1}) - \alpha \cdot (i_t - E_t(\pi_{t+1})) + \varepsilon_{y,t} \quad (15)$$

où π_t est le taux d'inflation et i_t le taux d'intérêt nominal. Toutes les variables sont exprimées en déviation par rapport à leur niveau de long terme. $\varepsilon_{y,t}$ est un choc exogène de demande. Le paramètre α mesure la substitution intertemporelle entre la consommation et l'épargne et constitue le canal de transmission de la politique monétaire vers le produit. La courbe IS décrit le comportement de consommation des ménages en reliant la production courante positivement à la production anticipée (effet de richesse) et négativement au taux d'intérêt réel (effet de substitution). Le comportement d'optimisation intertemporelle du consommateur est affecté par la stratégie de prix des entreprises et les décisions de politique monétaire de la Banque centrale (Gaffard, 2012).

La deuxième équation décrit le comportement de fixation des prix par les firmes, en concurrence monopolistique, à travers une version Nouvelle Keynésienne de la courbe de Phillips. Celle-ci est dérivée à partir d'un modèle de tarification optimale en concurrence imparfaite et d'une théorie de rigidités des prix (Roberts, 1997 ; Clarida et *ali.*, 1999 ; McCallum et Nelson, 1999 ; Svensson et Woodford, 2004 ; Woodford, 2003 ; Gianoni et Woodford, 2005).

La relation de Phillips relie positivement le taux d'inflation courant au taux d'inflation future anticipé, à la production et à un choc d'offre exogène $\varepsilon_{\pi,t}$ sur les coûts de production. La fonction d'offre agrégée est alors donnée par :

$$\pi_t = \beta \cdot E_t(\pi_{t+1}) + \gamma \cdot y_t + \varepsilon_{\pi,t} \quad (16)$$

où γ mesure la vitesse d'ajustement des prix à la variation de la demande ou encore le degré d'arbitrage entre les déviations de la production et celles de l'inflation. Sa valeur est déterminante pour l'action de la politique monétaire sur l'inflation. Le paramètre β est un taux

d'escompte compris entre 0 et 1 qui détermine ici l'influence sur l'inflation courante de la valeur anticipée de l'inflation.

Section 2. Evaluation de la politique monétaire : suprématie des DSGE sur les VAR

La règle de Taylor constitue la troisième équation, en plus des courbes IS et de Phillips, constitutive de la maquette d'un modèle DSGE. Cette règle est déduite de la fonction-objectif de la Banque centrale et lui permet de cibler le niveau optimal de l'inflation. La conception de la politique monétaire dans un modèle DSGE défend l'idée de la prédominance de la règle sur la discrétion (Barro et Gordon, 1983 a).

Benati et Surico (2009) montrent, dans le cadre du débat sur les causes de la "Grande Modération" aux Etats-Unis, qu'un modèle DSGE permet de mieux identifier un changement de règle de politique monétaire qu'un modèle VAR structurel. Cette impuissance des VAR s'explique par l'existence de restrictions inter-équations imposées par l'hypothèse des anticipations rationnelles Sargent (1979).

Sous-section 1. Règle de politique monétaire et objectif de stabilité des prix

1. Régime à politique monétaire dominante

Les modèles DSGE constituent une référence de base pour l'étude des conditions d'exercice de la politique monétaire par les Banques centrales (Woodford, 2003). En érigeant la stabilité des prix au rang d'objectif ultime, ce cadre d'analyse accorde à la Banque centrale un rôle *leader* dans la conduite d'un *policy mix* s'articulant autour de la crédibilité monétaire et de la discipline budgétaire afin d'éviter l'incohérence temporelle de la politique économique (Barro et Gordon, 1983 ; Kydland et Prescott, 1979 ; Sargent et Wallace, 1981). L'intégration de la rationalité des anticipations implique que les interactions entre les agents privés, le Gouvernement et la Banque centrale sont conçues dans le cadre d'un jeu stratégique qui permet d'identifier les instruments de la politique monétaire comme des paramètres qui reflètent des règles d'action (Taouil, 2011).

Dans ce régime à politique monétaire dominante, la politique budgétaire est subordonnée aux décisions de la Banque centrale et le respect des règles de politique économique est censé favoriser le maintien d'une basse inflation (Sargent et Wallace, 1981). L'exclusion de toute possibilité de financement monétaire du déficit oblige le Gouvernement à respecter sa contrainte budgétaire intertemporelle et permet à l'autorité monétaire de se forger une réputation autour de la stabilité des prix. Selon Rogoff (1985), la crédibilité d'un tel régime est conditionnée par le choix à la tête d'une Banque centrale indépendante d'un Gouverneur conservateur qui accorde plus de poids que la société à la lutte contre l'inflation. Dans le prolongement de l'analyse de Kydland et Prescott (1977), il est recommandé aux autorités monétaires d'adopter un cadre institutionnel qui permet d'éviter les tentations des mesures discrétionnaires et les effets pervers des politiques économiques temporellement incohérentes.

Il apparaît ainsi que l'emprunte des travaux des Nouveaux Classiques, présentés dans le chapitre 5, et de la synthèse des débats sur la question de la règle *versus* discrétion sont clairement présentes dans la Nouvelle synthèse. La représentation de la politique monétaire dans un modèle DSGE s'inscrit dans la continuité des travaux de Barro et Gordon (1983 a) qui ont montré, à partir d'une courbe d'offre à la Lucas (1972), la prédominance d'une politique monétaire fondée sur une règle par rapport à une politique discrétionnaire dans le cadre de l'arbitrage inflation-chômage. Le comportement de la Banque centrale, résumé par une règle de Taylor, cherche à influencer l'évolution à court terme de la demande agrégée dans un contexte de rigidité des prix.

Si à long terme, le marché retrouve sa force régulatrice, la présence d'altérations de marché contraignantes à court terme justifie l'opportunité d'une politique monétaire visant à réduire l'impact de ces frictions. En manipulant son instrument de taux d'intérêt nominal, la Banque centrale cherche à assurer la convergence de la demande agrégée et de l'inflation vers leurs valeurs d'équilibre de long terme. La crédibilité de l'engagement des autorités monétaires est censée envoyer aux agents des signaux sur l'évolution future de ces variables de manière à ancrer leurs anticipations sur une stabilité macroéconomique.

2. Règle de Taylor et ciblage de l'inflation

La troisième et dernière équation de la maquette de base d'un modèle DSGE représente un substitut à la courbe LM et résume le comportement de la Banque centrale à travers une relation de Taylor (1993, 1999) qui s'ajoute à la relation IS et la courbe de Phillips (voir la première section ci-dessus). Cette relation est constituée d'une règle optimale de taux d'intérêt des autorités monétaires qui leur permet de cibler le niveau optimal de l'inflation. Elle est déduite de l'objectif de la Banque centrale sous les contraintes de la courbe de Phillips et de la demande agrégée décrites ci-dessus. La fonction-objectif de la politique monétaire porte sur le programme de maximisation de la fonction de bien-être social, définie à l'aide de la fonction d'utilité intertemporelle du consommateur représentatif. Woodford (2003) a montré que ce programme peut être approximé par la minimisation de la fonction quadratique intertemporelle suivante :

$$E_t \sum_{s=0}^{\infty} \beta^s [\pi_{t+s}^2 + \delta \cdot y_{t+s}^2] \quad (17)$$

où l'inflation et la production sont mesurées en écarts par rapport à leur cible respective et δ exprime la pondération qu'accorde l'autorité monétaire à l'objectif de la production relativement à celui de l'inflation. Une forme courante de la règle de Taylor relie l'écart du taux d'intérêt nominal de court terme par rapport à son niveau naturel positivement à l'écart du taux d'inflation par rapport à sa cible et à l'output gap ainsi qu'à un choc exogène $\varepsilon_{i,t}$ qui rend compte des entorses inattendues à cette règle :

$$\dot{i}_t = \varphi_{\pi} \cdot \pi_t + \varphi_y \cdot y_t + \varepsilon_{i,t}$$

L'ajustement du taux d'intérêt se fait de façon endogène par la Banque centrale en fonction des déviations de l'inflation et de la production et il dépend des préférences de celle-ci pour la stabilité des prix φ_{π} ou de l'activité φ_y .

Dans le cas de ciblage de l'inflation, la Banque centrale utilise l'action du taux d'intérêt nominal sur la production pour contrôler l'inflation. L'efficacité de la politique monétaire dépend de l'intensité du lien entre le taux d'intérêt et la production, à travers le paramètre α , et entre la production et l'inflation, à travers le paramètre γ (Wickens, 2006). De plus, les

anticipations jouent un rôle fondamental dans l'exercice de la politique monétaire étant donné que les fonctions de demande et d'offre ainsi que les préférences de la Banque centrale sont tournées vers le futur. Dans ce cadre, le comportement des agents tient compte non seulement de la politique monétaire actuelle ou passée, mais également de son évolution future. L'ancrage des anticipations sur la cible de l'inflation permet à la Banque centrale d'amortir les chocs de demande et d'offre qui risquent d'écarter à court terme l'économie de son sentier d'équilibre de long terme en raison des imperfections de marché et de la présence de rigidités (Taouil, 2011).

Sous-section 2. La Grande modération : l'échec explicatif des VAR

1. Règle monétaire versus hypothèse de la "chance"

A partir du milieu des années 1980, plusieurs économies industrialisées ont connu une forte baisse simultanée de la volatilité de la croissance économique et de l'inflation, après ses niveaux élevés des années 1970 (Blanchard et Simon, 2001, Stock et Watson, 2003). Selon Blanchard et Simon (2001), la volatilité de la production américaine, mesurée par son écart-type, après 1984 a été divisée par 3 comparativement à son niveau antérieur. Les facteurs avancés par la littérature économique pour expliquer cette stabilité macroéconomique, qualifiée par Bernanke (2004) de "Grande Modération", sont généralement regroupés autour de deux principaux axes. D'un côté, la thèse des effets de la réduction des chocs, qualifiée d'hypothèse de la "chance" (appelée "good luck" dans la littérature anglo-saxonne). D'un autre côté, l'importance des changements structurels au niveau de l'économie et de la politique monétaire suivie, regroupés sous le nom de "bonne politique" (ou "good policy").

L'explication par l'hypothèse de la "bonne politique" a été initialement avancée par Clarida, Gali et Gertler (2000) et reprise dans d'autres études comme celles de Cogley et Sargent (2001, 2005) ou Boivin et Giannoni (2002). Pour ces auteurs, la politique monétaire américaine ne fût pas suffisamment efficace dans sa lutte contre l'inflation durant les années 1970, et elle serait devenue plus agressive dès le début des années 1980. Ce changement de comportement de la Banque centrale américaine dans la lutte contre l'inflation serait alors la cause de la stabilité macroéconomique enregistrée aux Etats-Unis. Dans cette perspective,

Bernanke (2004) et Gali et Gambetti (2008) estiment que l'adoption d'une politique monétaire centrée sur l'ancrage des anticipations introduit une discipline dans le comportement de fixation des prix et des salaires et contribue à réduire la fréquence et l'ampleur des chocs.

Les tenants de l'hypothèse de la "*chance*" considèrent que le phénomène de la Grande Modération s'explique exclusivement par la baisse de la fréquence et de l'amplitude des chocs macroéconomiques exogènes qui ont affecté les économies concernées, notamment ceux des prix des produits pétroliers. La justification de cette hypothèse s'appuie sur la méthode des modèles VAR structurels et elle est soutenue par des auteurs comme Stock et Watson (2002), Sims et Zha (2006) et Canova, Gambetti et Pappa (2006).

De plus, les défenseurs de l'hypothèse de la "*chance*" cherchent à dénier tout rôle à la politique monétaire dans l'avènement de la Grande Modération. En s'appuyant sur un VAR structurel bayésien, Canova et Gambetti (2006) montrent que la fonction de réaction de la Banque centrale américaine a été relativement stable et qu'il n'y a eu ainsi aucun changement dans le régime de la politique monétaire durant la période en question. Dans le même sens, Leeper et Zha (2003) considèrent qu'il n'y a pas eu de changements significatifs dans la règle de politique monétaire et que la propagation des chocs monétaires a été relativement stable.

Les adeptes du "*good policy*" rejettent ces conclusions et s'interrogent sur la capacité de la méthode des VAR structurels à rendre compte des causes de la Grande Modération. En s'appuyant sur un modèle DSGE pour l'économie américaine, Lubik et Schorfheid (2004), d'une part, et Boivin et Giannoni (2006), d'autre part, confirment l'importance de la modification de la politique monétaire dans l'explication de la Grande Modération. Benati et Surico (2009) estiment que, contrairement aux modèles DSGE, l'utilisation des VAR structurels ne permet pas toujours de bien identifier un changement de règle de politique monétaire. En conséquence, l'utilisation des modèles VAR sous-estime l'importance du rôle joué par la politique monétaire et privilégie l'effet des chocs macroéconomiques.

Benati et Surico (2009) estiment que la différence des résultats obtenus par les deux approches peut être expliquée par des différences méthodologiques. Leur thèse centrale porte sur l'incapacité des modèles VAR structurels, en raison de leur nature même, à identifier un changement de règle de politique monétaire à partir de données simulées à l'aide d'un modèle DSGE.

2. VAR et restrictions inter-équations

Benati et Surico procèdent à une évaluation de la méthode des modèles VAR à partir d'une expérience contrôlée réalisée à l'aide d'un modèle DSGE. Le modèle VAR est estimé à partir des simulations du DSGE, considéré comme un processus générateur de données. Il s'agit d'un modèle DSGE standard à la Smets et Wouters (2003) constitué des trois équations suivantes :

$$\begin{aligned}i_t &= \rho.i_{t-1} + (1 - \rho).(\varphi_\pi.\pi_t + \varphi_y.y_t) + \varepsilon_{i,t} \\ \pi_t &= (1 - \beta).\pi_{t-1} + \beta.E_t(\pi_{t+1}) + \gamma.y_t + \varepsilon_{\pi,t} \\ y_t &= (1 - \lambda).y_{t-1} + \lambda.E_t(y_{t+1}) - \alpha.(i_t - E_t(\pi_{t+1})) + \varepsilon_{y,t}\end{aligned}$$

où y_t est la production, π_t le taux d'inflation et i_t le taux d'intérêt. E_t est l'opérateur d'anticipation conditionnelle à l'information en période t . En intégrant les valeurs retardées des variables endogènes, ces relations comportent des formulations dans une optique prospective et rétrospective (*forward looking* et *backward looking*). Les paramètres ont la même signification que dans les équations présentées ci-dessus. Le paramètre ρ constitue un coefficient de pondération entre la valeur passée et la valeur courante du taux d'intérêt. Le paramètre β est un taux d'escompte compris entre 0 et 1 qui détermine également ici l'influence sur l'inflation courante de la pondération entre sa valeur anticipée et sa persistance mesurée par sa valeur retardée. Le paramètre λ détermine la pondération entre l'influence de la valeur anticipée de la production et la persistance de sa valeur retardée.

La calibration et l'estimation des paramètres du modèle, pour l'économie américaine, sont réalisées par des méthodes bayésiennes pour les périodes avant et après octobre 1979. Le choix de cette date est justifié dans la littérature économique comme constituant la fin de la période des politiques de désinflation menées par Volcker. L'hypothèse fondamentale de travail adoptée consiste à supposer que la période de la Grande Modération aux Etats-Unis est la conséquence exclusive du passage d'un régime de politique monétaire passive à un régime de politique monétaire active mise en place après octobre 1979. Les auteurs retiennent ainsi l'hypothèse de Clarida et *al.* (2001) selon laquelle un comportement passif de la Banque centrale qui n'ajusterai pas son taux d'intérêt de façon suffisamment réactive aux variations

de l'inflation conduit à de fortes fluctuations économiques, désignées dans la littérature par les "taches solaires".

Sous l'effet unique du changement de la politique monétaire, la simulation du modèle DSGE reproduit convenablement la transition, en termes de réduction de la volatilité des variables d'intérêt, de l'économie américaine vers la période de la Grande Modération (Benati, Surico, 2009). Toutefois, les résultats fournis par le modèle VAR estimé à partir des données engendrées par le modèle structurel sont en complet désaccord avec ce résultat. En effet, la méthode des VAR structurels s'appuyant sur ce PDG tend à expliquer le passage à la période de la Grande Modération par l'hypothèse de la "chance". Or, par construction, l'origine des changements dans le processus générateur de données (PGD) est indépendante à la fois des variations de la volatilité des chocs structurels et des modifications dans les autres paramètres structurels du modèles qui sont non liés à la politique monétaire.

Cet échec des modèles SVAR n'est pas lié, selon Benati (2010), à des problèmes d'estimation comme le soutient Canova (2006), mais il est la conséquence directe des restrictions inter-équations imposées par l'hypothèse des anticipations rationnelles tel qu'il a été initialement soulevé par la critique de Sargent (1979) à l'encontre des modèles SVAR. La vulnérabilité des modèles VAR à la critique de Lucas (1976), relative aux méthodes économétriques d'évaluation des politiques économiques, que mentionne Sargent (1979) est la conséquence de la présence de ces restrictions inter-équations. Autrement dit, les modèles VAR ne permettent pas d'évaluer les impacts économiques consécutifs à un changement dans la règle dont disposent les autorités monétaires pour fixer le comportement des variables de politique monétaire (Sargent, 1979).

Pour illustrer ces éléments, Benati (2010) considère deux règles de Taylor distinctes caractérisant un modèle DSGE, chacune identifiée par les trois paramètres ρ , φ_π et φ_y correspondants :

$$\text{Taylor}^1 \equiv (\rho^1, \varphi_\pi^1, \varphi_y^1)$$

$$\text{Taylor}^2 \equiv (\rho^2, \varphi_\pi^2, \varphi_y^2)$$

Combinées avec les autres équations du modèle DSGE, ces deux relations conduisent à deux représentations VAR différentes correspondant au régime avant la Grande modération (régime 1) et au régime de la Grande Modération (régime 2) :

$$\begin{aligned} \text{Taylor}^1 &\rightarrow \text{DSGE}^1 \rightarrow \text{VAR}^1 \rightarrow \text{SVAR}^1 \rightarrow \text{Règle Monétaire}^1 \\ \text{Taylor}^2 &\rightarrow \text{DSGE}^2 \rightarrow \text{VAR}^2 \rightarrow \text{SVAR}^2 \rightarrow \text{Règle Monétaire}^2 \end{aligned}$$

La Règle Monétaire¹ et la Règle Monétaire² expriment les équations du taux d'intérêt dans les représentations SVAR¹ et SVAR², respectivement.

On considère que la représentation VAR structurel (SVAR) du modèle DSGE s'écrit sous la forme suivante :

$$A_0 X_t = A_1 X_{t-1} + \dots + A_p X_{t-p} + \varepsilon_t$$

où $X_t = (i_t, \pi_t, y_t)'$ le vecteur des variables endogènes, $\varepsilon_t = (\varepsilon_{i,t}, \varepsilon_t)'$ constitue le vecteur des chocs structurels dans le modèle VAR, avec $\varepsilon_{i,t}$ le choc de politique monétaire dans la règle de Taylor et ε_t le vecteur des autres chocs structurels. Posons :

$$A_0 = \begin{pmatrix} A_0^i(\theta) \\ A_0^j(\theta) \end{pmatrix}, A_1 = \begin{pmatrix} A_1^i(\theta) \\ A_1^j(\theta) \end{pmatrix}, \dots, A_p = \begin{pmatrix} A_p^i(\theta) \\ A_p^j(\theta) \end{pmatrix}$$

où $\theta = (\rho, \varphi_\pi, \varphi_y)$ est le vecteur des paramètres de la règle de politique monétaire, $A_j^i(\theta)$ représente la première ligne de la matrice correspondant à l'équation du taux d'intérêt dans le modèle VAR et $A_j^j(\theta)$ désigne la matrice des autres équations du SVAR, pour $j = 0, \dots, p$.

La présence du vecteur θ dans toutes les équations est l'expression des restrictions inter-équations qu'implique l'hypothèse des anticipations rationnelles sur la solution d'un modèle d'équilibre général. La justification de ces restrictions ne trouve pas son origine dans des considérations économétriques, mais résulte de l'articulation qu'impose la théorie entre la formation des anticipations et les décisions d'agents optimisateurs (Lucas, Sargent, 1979). Dans ces conditions, il n'est pas approprié de substituer l'équation d'une nouvelle règle de

politique monétaire dans le modèle tout en maintenant inchangées les autres équations. En principe, toutes les autres équations doivent également changer en raison des restrictions inter-équations qui sont imposées par la théorie économique sur les paramètres du modèle.

En effet, deux vecteurs de paramètres $\theta^1 = (\rho^1, \varphi_\pi^1, \varphi_y^1)$ et $\theta^2 = (\rho^2, \varphi_\pi^2, \varphi_y^2)$ correspondant à deux règles de politique monétaire distinctes sont associés aux deux représentations SVAR suivantes :

$$\begin{aligned} A_0(\theta_1)X_t &= A_1(\theta_1)X_{t-1} + \dots + A_p(\theta_1)X_{t-p} + \varepsilon_t \\ A_0(\theta_2)X_t &= A_1(\theta_2)X_{t-1} + \dots + A_p(\theta_2)X_{t-p} + \varepsilon_t \end{aligned}$$

Lorsqu'on impose la règle de politique monétaire du régime 2 dans le SVAR correspondant au régime 1 (ou inversement), on obtient la formulation suivante :

$$\begin{pmatrix} A_0^i(\theta_2) \\ A_0^i(\theta_1) \end{pmatrix} X_t = \begin{pmatrix} A_1^i(\theta_2) \\ A_1^i(\theta_1) \end{pmatrix} X_{t-1} + \dots + \begin{pmatrix} A_p^i(\theta_2) \\ A_p^i(\theta_1) \end{pmatrix} X_{t-p} + \varepsilon_t$$

Le SVAR ainsi construit génère une augmentation de la volatilité de la production et de l'inflation alors que le PGD simule une baisse de cette volatilité (Benati, Surico, 2009). Il est clair que ce VARS ne peut reproduire les résultats simulés par le modèle DSGE que si le vecteur des paramètres de la politique monétaire θ^1 n'apparaît pas dans les équations du SVAR, qui ne décrivent pas le comportement de la politique monétaire.

De l'aveu même de Canova (2006), la critique de Benati et Surico va au-delà de la critique soulevée par Chari et *ali.* (2005) concernant l'estimation biaisée par un VARS des réponses simulées à l'aide d'un modèle de cycles réels ou des précédentes critiques adressées à l'encontre des restrictions traditionnelles, sans référence à la théorie, utilisées pour identifier les chocs structurels dans les modèles VAR (Cooley et Leroy, 1985, Faust et Leeper, 1997 et Cooley et Dweyer, 1998). La présente critique est d'ordre méthodologique et elle est liée à la nature même des modèles VAR structurels (Benati et Surisco, 2009). Elle est valable même lorsque des restrictions identifiantes appropriées sont imposées et elle montre les limites de l'utilisation de la méthode VAR, sans référence à un modèle structurel, pour l'évaluation des politiques économiques.

Conclusion

Les modèles DSGE de la Nouvelle Synthèse s'inscrivent dans la continuité du programme de recherche des Nouveaux Classiques et de la vision épistémologique de Lucas. En substituant une règle de Taylor à la courbe LM dans un cadre de rigidité des prix et d'imperfection de marché, ces modèles constituent une référence pour l'évaluation de la politique monétaire depuis le milieu des années 2000. La détermination des paramètres de ces modèles s'appuie sur l'utilisation jointe des techniques du calibrage et des méthodes d'estimation bayésienne et d'inférence simulée.

Le rôle central de l'hypothèse des anticipations rationnelles dans ces modèles se traduit par des restrictions inter-équations dont les modèles VAR ne tiennent pas compte (Sargent, 1979). Dans le cadre du débat sur les causes de la "Grande Modération", Benati et Surico (2009) montrent que les modèles DSGE permettent de mieux rendre compte d'un changement dans la règle de politique monétaire qu'un modèle VAR structurel.

Conclusion générale

Les années 1960 et 1970 ont connu d'importants débats et de vives controverses sur le meilleur moyen d'établir une relation causale entre la monnaie et le revenu. En l'absence d'un cadre théorique rigoureux, les monétaristes tentent d'abord de déplacer leur débat avec les auteurs de la synthèse sur un terrain empirique. Leur attaque de la modélisation structurelle de la Commission Cowless s'appuie sur les ambiguïtés du modèle IS-LM et sur les divergences entre le cadre théorique de ce modèle et les versions estimées.

Les keynésiens critiquent le recours des monétaristes à des formes réduites pour établir la supériorité de la politique monétaire sur la politique budgétaire et rejettent l'utilisation par ceux-ci des décalages temporels pour prouver la causalité de la monnaie sur le revenu. A l'opposé, les monétaristes reprochent aux keynésiens d'introduire une version *ad hoc* de la courbe de Phillips dans leur modèle. La nouvelle lecture théorique par Friedman (1968) de cette courbe permet aux monétaristes d'expliquer l'influence causale de la masse monétaire sur le revenu nominal et la décomposition de celui-ci entre la production et les prix.

Sims (1972, 1980 a) remet en cause l'utilisation d'*a priori* théorique pour l'inférence de la causalité et de l'exogénéité. Toutefois, sa volonté de s'appuyer exclusivement sur l'économétrie, à travers le concept de G-causalité et les modèles VAR, pour étudier la causalité entre la monnaie et ses déterminants s'est heurtée à d'importantes limites théoriques et méthodologiques. D'un côté, l'évaluation de la politique monétaire à l'aide d'un VAR impose de choisir des restrictions d'identification sur la base d'*a priori* théorique pour éviter le problème de l'équivalence observationnelle. En effet, étant des formes réduites, les modèles VAR font face au problème de l'indétermination empirique de la causalité qui constitue un cas extrême de la sous-détermination de la "théorie" par "l'empirie", soulevée par la thèse de Duhem-Quine. D'un autre côté, Hoover (2009) qualifie, dans la continuité de des travaux de Cartwright (2007), les modèles VAR de contrefactuels "imposteurs".

Les tentatives de conciliation entre "théorie" et "économétrie" ne manquent pas de soulever d'importantes difficultés. En l'absence d'un modèle théorique dynamique rigoureux et spécifiant l'ensemble des comportements, les résultats économétriques des monétaristes ont été jugés hautement défectueux. Les méthodes développées par les auteurs de la *London School of Economics (LES)* et les tenants de l'approche de la cointégration pour concilier

entre les dynamiques de court et de long terme dans l'analyse de la politique monétaire souffrent d'importants problèmes d'identification. De plus, l'absence d'un critère d'arbitrage en présence de conflit entre le principe théorique et l'ajustement aux données conduit à une subordination de la théorie à l'économétrie (Faust et Whiteman, 1997).

La critique de Lucas (1976) met en évidence l'incapacité des modèles économétriques de rendre compte des ruptures structurelles dans les comportements des agents économiques induites par un changement de politique économique. Simultanément, Lucas (1977, 1980, 1986) développe une nouvelle vision épistémologique, considérant le modèle comme une fiction et où le cycle économique est décrit en termes de discipline de l'équilibre. Dans cette perspective, les Nouveaux Classiques recourent au calibrage, au détriment de l'économétrie, pour calculer les paramètres du modèle (Kydland et Prescott, 1982). Ce nouveau programme de recherche recommande la subordination de la politique budgétaire à une politique monétaire menée par une Banque centrale indépendante (Kydland et Prescott, 1977 ; Barro et Gordon, 1983 ; Sargent et Wallace, 1981).

Les modèles DSGE de la Nouvelle Synthèse constituent une extension des modèles des Cycles Réels (RBC) développés par les Nouveaux Classiques. La politique monétaire y est représentée par une règle de Taylor qui permet à la Banque centrale d'agir sur la demande agrégée dans un cadre de rigidité des prix à court terme. En raison des restrictions inter-équations induites par l'hypothèse des anticipations rationnelles (Sargent, 1979), Benati et Surico (2009) montrent la suprématie des modèles DSGE sur les modèles VAR pour l'identification d'un changement de règle de politique monétaire.

Toutefois, les modèles DSGE font l'objet de plusieurs critiques, surtout leur incapacité à rendre compte des événements liés à la crise financière de 2008. Les questionnements à propos de cette limite portent généralement sur l'irréalisme des hypothèses de ces modèles, notamment, celles des anticipations rationnelles, de l'agent représentatif ou encore de l'efficience des marchés financiers (Buitier, 2009 ; Cartapanis, 2010 ; Kay, 2012). Dès l'introduction de cette thèse, nous avons relevé le caractère contreproductif de ce type de critiques qui sont incapables de remettre en cause la théorie de la Nouvelle Synthèse (Taouil, 2014, b).

Mais, d'autres critiques insistent plutôt sur l'absence ou l'inutilité de la monnaie dans ces modèles censés servir à l'étude des conditions d'exercice de la politique monétaire par les Banques centrales (Woodford, 2003). En effet, ces modèles n'intègrent pas les fonctions de la monnaie comme moyen d'échange, réserve de valeur ou mode de financement, mais la monnaie est plutôt imposée dans la fonction d'utilité comme unité de compte. L'intégration de la monnaie dans un modèle néoclassique est une question récurrente en macroéconomie monétaire et qui oppose depuis longtemps les keynésiens aux néoclassiques.

Par ailleurs, on assiste, depuis les années 1960, à un va-et-vient permanent entre le développement de l'économétrie et les débats théoriques sur la politique monétaire. En dépit des limites auxquelles elle fait face, l'économétrie a souvent pris le pas sur la théorie et a permis d'approfondir la connaissance empirique des phénomènes monétaires. Toutefois, en accordant le primat à la théorie, la révolution des Nouveaux Classiques a relégué l'économétrie au second plan. Les modèles DSGE, qui sont fortement influencés par le programme de recherche des Nouveaux Classiques, s'appuient de plus en plus sur les estimations bayésiennes présentées comme une troisième voie entre le calibrage et l'économétrie des séries chronologiques.

Vo Phuong et *ali.* (2012) soutiennent que le choix de la meilleure méthode d'évaluation des performances empiriques d'un modèle DSGE reste une question ouverte. Dans cette perspective, on assiste à un recours croissant de ces modèles à des méthodes économétriques alternatives, telles que les méthodes d'inférence simulée, le bootstrap ou encore l'inférence statistique indirecte. Sachant que les modèles DSGE sont mal spécifiés, les adeptes de l'économétrie proposent de mesurer la "proximité" de ces modèles avec les données. L'objectif est d'établir une classification parmi ces modèles, qui permettrait d'apporter une réponse au questionnement de Canova (1994), "*étant donné que ces modèles sont faux, dans quelle mesure sont-ils vrais ?*".

A cet effet, Cooley (1997) note que l'idée de mesurer la "distance" entre ces modèles artificiels et les données est largement acceptée. Ce qui est refusé par les tenants du calibrage, c'est de rejeter ces modèles sur la base de cette mesure. Cette tendance inaugure de l'ouverture d'une seconde ère de conciliation entre "théorie" et "empirie" et ouvre la voie à un éventuel retour en force de l'économétrie dans la modélisation économétrique.

Références bibliographiques

Andersen, L. et Jordan, J. (1968) “Monetary and Fiscal Actions: A Test of Their Relative Importance in Economic Stabilization”, *Federal Reserve Bank of St. Louis Review*, 50, 11-24.

Ando, A. et Modigliani, F. (1965) “Velocity and the Investment Multiplier”, *American Economic Review*, Septembre, 693-728.

Artus, A., Deleau, M. et Malgrange, P. (1986) *Modélisation Macroéconomique*, Economica.

Barro, R. (1974) “Are Government Bonds Net Wealth? ”, *Journal of Political Economy*, vol. 82, n° 6, décembre, 1095-1118.

Barro, R. et Gordon, D. (1983) “Rules, Discretion and Reputation in a Model of Monetary Policy”, *Journal of Monetary Economics*, vol. 12.

Basman, R. L. (1963) “The Causal Interpretation of Non-Triangular Systems of Economic Relations”, *Econometrica*, 31: 439-48.

Basman, R.L. (1965) “A Note on the Statistical Testability of 'Explicit Causal Chain' against the Class of 'Interdependant' Models”, *Journal of the American Statistical Association*, 60.

Benati, L. (2010) “Are Policy Counterfactuals Based on Structural VAR Reliable”, European Central Bank, Working Paper Series, n° 1188, May.

Benati, L. et Surico, P. (2009) “VAR Analysis and the Great Moderation”, *American Economic Review*, 99(4).

Bernanke, B. (2004) “The Great Moderation”, Remarks at the meetings of the Eastern Economic Association, Washington, DC, February 20.

Bisignano, J. (1974) “Money, Income and Causality: Another Look”, unpublished paper, Federal Reserve Bank of San Francisco.

Blinder, A. et Solow, R. (1973) "Does Fiscal Policy Matter? ", *Journal of Public Economics*, vol. 2, 319-337.

Boivin, J. Giannoni, M. (2002) "Assessing Changes in the Monetary Transmission Mechanism : A VAR Approach", *Economic Policy Review*, Federal Reserve Bank of New York, no. 1, 97-111.

Boivin, J. Giannoni, M. (2006) "Has Monetary Policy Become More Effective?", *The Review of Economics and Statistics*, 88, 445-462.

Bordo M. et Schwartz, A. (2004) "IS-LM and Monetarism", in *The IS-LM Model: Its Rise, Fall and Strange Persistence*, De Vroey M. et Hoover K. (ed.)? Duke University Press.

Brown, W. (1940) "International Gold Standard Reinterpreted", Columbia University Press, New York.

Buiter, W. (2009) "The infortunate uselessness of most 'state of the art' academy monetary economics", *Financial Times*, 3 mars.

Buiter, W. et Tobin, J. (1976) "Long Run Effects of Fiscal and Monetary Policy on Aggregate Demand" in J. Stein (éd.), *Monetarism*, North Holland, p. 273-309.

Calvo, G. (1983) "Staggered Contracts in a Utility-Maximizing Framework", *Journal of Monetary Economics*, 12, septembre.

Calvo, G. (1988) "Servicing the public debt : the role of expectations", *American Economic Review*, vol. 78, 179-194.

Canova, F. (2006) "You Can Use VARs for Structural Analyses. A Comment to VARs and the Great Moderation", sur <http://www.crei.cat/people/canova/papers.html>.

Canova, F., Gambetti, L. et Pappa, E. (2006) "The Structural Dynamics of U.S. Output and Inflation: What Explains the Changes?" *Journal of Money, Credit, and Banking*, 40(2-3), 369—388.

Cartapanis, A. (2010) "Les économistes et la crise", in Ch. De Boissieu et B. Jacquillot (eds.) *A quoi servent les économistes*, Paris : Descartes et Cie, Presses universitaires de France.

Cartwright, N. (2006) "Where is the Theory in Our "Theories" of Causality?", *Journal of Philosophy*, Vol. 103, No. 2, 55-66.

Cartwright, N. (2007) *Hunting Causes and Using Them*, Approaches in Philosophy and Economics, Cambridge, Cambridge University Press.

Chari, V.V., Kehoe, P.J. et McGrattan, E.R. (2005) "A Critique of Structural VARs Using Real Business Cycle Theory", Working Paper 631, Federal Reserve Bank of Minneapolis.

Clarida, R., Gali, J. et Gertler, M. (1999) "The Science of Monetary Policy: A New-Keynesian Perspective", *Journal of Economic Literature*, vol. 37, n° 4, December, p.1661-1707.

Clarida, R., Gali, J. et Gertler, M. (2000) "Monetary policy rules and macroeconomic stability: Evidence and some theory", *Quarterly Journal of Economics*, CXV, 147-180.

Cogley, T. et Sargent, T. (2001) "Evolving post-world war II U.S. inflation dynamics", NBER Macroeconomic Annual, 16, 331-373.

Cogley, T. et Sargent, T., J. (2005) "Drifts and volatilities: Monetary policies and outcomes in the post WWII U.S. ", *Review of Economic Dynamics*, 8, 262-302.

Cooley, T. (1997) "Calibrated Models", University of Rochester.

Cooley, T. et Dwyer, M. (1998) "Business Cycle Analysis without much Theory: A Look at Structural VARs", *Journal of Econometrics*, 83, 57-88.

Cooley, T. et LeRoy, S. (1985) "A-theoretical macroeconomics : A critique", *Journal of Monetary Economics*, 16, 283-308.

Davidson, P. (1972) "A Keynesian View of Friedman's Theoretical Framework for Monetary Analysis", *Journal of Political Economy*, Vol. 80, N° 5, 864-882.

Davidson, J., Hendry, D., Srba, F. et Yeo, S. (1978) "Econometric Modelling of the Aggregate Time-Series Relationship between Consumers' Expenditure and Income in the United Kingdom", *Economic Journal*, 88, 661-92.

Deleplace, G. (1999) *Histoire de la Pensée Economique*, Dunod.

DeLeeuw, F. et Kalchbrenner, J. (1969) "Monetary and Fiscal Actions : a Test of their Relative Stability - Comment", *Federal Reserve Bank St. Louis Review*, 51: 6-11.

Desai, M. (1981) *Testing Monetarism*, Frances Pinter (Publishers) Ltd., London.

Desquilbet, J.B. et Villieu, P. (1998) "La théorie du *policy-mix* : un bilan critique", *Revue d'économie financière*, N°45, pp. 31-62.

De Vroey, M. (2000) "IS-LM a la Hicks versus IS-LM a la Modigliani", *History of Political Economy*, Volume 32, N° 2, 293-316.

De Vroey, M. (2009) *Keynes, Lucas: d'une macroéconomie à l'autre*, Dalloz.

De Vroey, M. (2010) "La crise de 2008: Quel effet de retour sur la théorie macroéconomique?", *Reflets et Perspectives*, XLIX.

De Vroey, M. (2011) "Lucas on the relationship between theory and Ideology", *Economics*, No. 4, *Economics: The Open-Access, Open-Assessment E-Journal*, Vol. 5, Iss. 4, pp. 1-39.

De Vroey, M. et Malgrange, P. (2006) "La théorie et la modélisation macroéconomiques, d'hier à aujourd'hui", Document de Travail n° 33, Paris-Jourdain Sciences Economiques.

De Vroey, M. et Malgrange, P. (2012) ‘‘Klein et l’émérgence de la modélisation macroéconomique’’, *Economie et Statistique*, n° 451-453.

Dewey, J. (1938) *Logic : The Theory of Inquiry*, New York: Henry Bolt.

Duhem, P. (1914) *La théorie physique. Son objet - sa structure*, Paris, Vrin. 2ème édition.

Dixit, A. et Stiglitz, J. (1977) ‘‘Monopolistic Competition and Optimum Product Diversity’’, *American Economic Review*, 67: 297-308.

Eichenbaum, M. et Singleton, K.J. (1986) ‘‘Do equilibrium real business cycle theories explain postwar U.S. business cycles?’’ NBER Macroeconomics, ed. Stanley Fischer, 1: 91-135. Cambridge, Mass.: MIT Press.

Elliot, J. (1975) ‘‘The Influence of Monetary and Fiscal Actions on Total Spending: The St. Louis Total Spending Equation Revisited’’, *Journal of Money, Credit and Banking*, 7.

Engle, R. Hendry, D. et Richard, J.F. (1983) ‘‘Exogeneity’’, *Econometrica*, 51, 277-304.

Ericsson, N. et Irons, J. (1995) ‘‘Lucas Critique in Practice: Theory without measurement’’, in *Macroeconometrics : Developments, Tensions, and Prospects*, K. D. Hoover (ed.), Boston, Kluwer Academic Press.

Esfeld, M. (2006) *Philosophie des sciences: une introduction*, Presses polytechniques et universitaires romandes.

Evans, M. Klein, L. et Schink, G. (1967) *The Wharton Econometric Forecasting Model*, Wharton School of Finance and Commerce Studies in Quantitative Economics, no. 2. Philadelphia: Univ. Pennsylvania Press.

Faust, J. et Leeper, E. (1997) ‘‘Do Long Run Restrictions Really Identify Anything?’’, *Journal of Business and Economic Statistics*, 15, 345-353.

Faust, J., et Whiteman, C. (1997) "General-to-specific procedures for fitting a data-admissible, theory-inspired, congruent, parsimonious, encompassing, weakly-exogenous, identified, structural model of the DGP: A translation and critique", Federal Reserve Board, Washington, n° 576.

Feigl, H. (1953) "Notes on Causality", in Herbert Feigl and Mary Brodbeck, editors. *Readings in the Philosophy of Science*. New York: Appleton-Century-Crofts, pp. 408-418.

Feige, E.L. et Pearce, D.K. (1979) "The Casual Causal Relationship Between Money and Income: Some Caveats for Time Series Analysis", *The Review of Economics and Statistics*, Vol. 61, No. 4

Fève, P. et Grégoir, S. (2002) "L'économétrie de la politique économique", *Annales d'Economie et de Statistique*, n° 67-68.

Florens, J et Mouchart, M. (1985) "Conditioning in Dynamic Models", *Journal of Time Series Analysis*, vol 6, n° 1, 15-34.

Friedman, M. (1956) "The Quantity Theory of Money: a Restatement", In *Studies in the Quantity Theory of Money*, edited by Milton Friedman. Chicago: Univ. Chicago Press.

Friedman, M. (1957) "The Permanent Income Hypothesis: Comment", *American Economic Review*, 48, 990-991

Friedman, M. (1958) "The Supply of Money and Changes in Prices and Output", In *The Relationship of Prices to Economic Stability and Growth*, pp. 241-56. U.S. Congress, Joint Economic Committee, Compendium.

Friedman, M. (1959) "The Demand for Money : Some Theoretical and Empirical Results", *Journal of Political Economy*, 67 : 327-351.

Friedman, M. (1961) "The Lag in Effect of Monetary Policy", *Journal of Political Economy*, Vol. 69, N° 5, 447-466.

Friedman, M. (1966) "Interest Rates and the Demand for Money", *Journal Law and Economics*, 9, 71-85.

Friedman, M. (1968) "The Role of Monetary Policy", *American Economic Review*, 58:1-17.

Friedman, M. (1969) *The Optimum Quantity of Money and Other Essays*, Chicago: Aldine.

Friedman, M. (1970 a) "A Theoretical Framework for Monetary Analysis", *Journal of Political Economy*, 78: 193-238.

Friedman, M. (1970 b) "Comment on Tobin", *Quarterly Journal of Economy*, volume 84, n° 2, 318-327.

Friedman, M. (1971 a) "A Monetary Theory of Nominal Income", *Journal of Political Economy*, 79: 323-337.

Friedman, M. (1971 b) "A Theoretical Framework for Monetary Analysis", Occasional Paper 112, New York : NBER.

Friedman, M. (1972) "Comments on the Critics", *Journal of Political Economy*, Vol. 80, No. 5, 906-950.

Friedman, M. (1975) "Unemployment versus Inflation? An Evaluation of the Phillips Curve", Institute of Economic Affairs.

Friedman, M. et Meiselman, D. (1963) "The Relative Stability of Monetary Velocity and the Investment Multiplier in the United States, 1897-1958", In *Stabilization Policies*, Commission on Money and Credit. Englewood Cliffs, N.J.: Prentice-Hall, 1963.

Friedman, M. et Schwartz, A. (1963 a) "Money and Business Cycles", *Review Economics and Statistics*, pp. 32-64.

Friedman, M. et Schwartz, A. (1963 b) *A Monetary History of the United States, 1867-1960*, Princeton, N.J.: Princeton Univ. Press.

Friedman, M. et Schwartz, A. (1982) *Monetary Trends in the United States and United Kingdom: Their Relation to Income, Prices, and Interest Rates, 1867-1975*, University of Chicago Press.

Gaffard, J.L. (2012) "Crise de la Théorie et Crise de la Politique Economique: des Modèles d'Equilibre Général Stochastique aux Modèles de Dynamique hors de l'Equilibre", *Revue de l'OFCE*, 10.

Gali, J. et Gambetti, L. (2009) "On the Sources of the Great Moderation", *American Economic Review*, Macroeconomics, vol. 1, n° 1, January.

Giannoni, M.P. et Woodford, W. (2005) "Optimal Inflation Targeting Rules", In *The Inflation Targeting Debate*, edited by B. S. Bernanke and M. Woodford. Chicago: University of Chicago Press.

Goodfriend, M. et King, R. (1997) "The New Neoclassical Synthesis and the Role of Monetary Policy", Federal Bank of Richmond, Working paper Series, No.98-05.

Granger, C. (1969) "Investigating causal relations by econometric models and cross-spectral methods", *Econometrica*, vol. 37 (3), pp. 424-438.

Granger, C. et Newbold, P. (1974) "Spurious Regressions in Econometrics", *Journal of Econometrics*, 26, 1045-1066.

Gupta, S. (1964) "Expected Rate of Change of Prices and Rates of Interest", Ph.D. dissertation, Univ. Chicago.

Haavelmo, T. (1943) "The statistical implications of a system of simultaneous equations", *Econometrica*, vol. 11.

Haavelmo, T. (1944) "The probability approach in econometrics", Supplement to *Econometrica*, 12.

Hairault, J. O. et Langot, F. (2005). ‘‘F. Kydland Et E. Prescott : Prix Nobel d’Economie 2004’’, *Revue d’économie politique*, N°1, Vol. 115, p. 65 à 83.

Hausman, J.A. (1978) ‘‘Specification tests in econometrics’’, *Econometrica* 49.1251-1271.

Hendry, D. (1995) ‘‘Econometrics and Business Cycle Empirics’’, *The Economic Journal*, Vol. 105, No. 433, pp. 1622-1636.

Hendry, D. et Ericsson, N. (1991 a) ‘‘Modeling the demand for narrow money in the United Kingdom and the United States’’, *European Economic Review*, 35, 4, p. 833-881.

Hendry, D. et Ericsson, N. (1991 b) ‘‘An Econometric Analysis of U.K. Money Demand in Monetary Trends in the United States and the United Kingdom by Milton Friedman and Anna J. Schwartz’’, *The American Economic Review*, Vol. 81, No. 1, pp. 8-38.

Hendry, D. et Mizon, G. (1978) ‘‘Serial correlation as a convenient simplification, not a nuisance: a comment on a study of the demand for money by the Bank of England’’, *Economic Journal*, 88, 549-563.

Hendry, D et Mizon, G. (1993) ‘‘Evaluating dynamic econometric models by encompassing the VAR’’, in *Models, Methods and Applications of Econometrics* (ed. P. C. B. Phillips), pp. 272-300. Oxford: Basil Blackwell.

Hicks, J. (1936) ‘‘Mr Keynes theory of employment’’, *Economic journal*, 46, p. 238-253.

Hicks, J. (1937) ‘‘Mr. Keynes and the ‘Classics’: a Suggested Interpretation’’, *Econometrica*, Vol. 5, n°2, p. 147-159.

Hirsch, A. et De Marchi, N. (1984) ‘‘Methodology: A Comment on Frazer and Boland’’, *The American Economic Review*, Vol. 74, No. 4, pp. 782-788.

Hoover, K. D. (1984) ‘‘ Two Types of Monetarism’’, *Journal of Economic literature*, vol. XXII, Mars.

Hoover, K. D. (2001) *Causality in Macroeconomics*, Cambridge : Cambridge University Press.

Hoover, K. D. (2004) "Automatic Inference of the Contemporaneous Causal Order of a System of Equations ", Departments of Economics, University of California.

Hoover, K. D. (2006) "Causality in Economics and Econometrics", Departments of Economics and Philosophy, Duke University.

Hoover, K. D. (2009) "Counterfactuals and Causal Structure", Department of Economics and Department of Philosophy, Duke University.

Hoover, K. D. et Demiralp, S. (2003) "Searching for the Causal Structure of a Vector Autoregression ", Working papers, University of California, Department of Economics.

Ireland, M. et Frederick, M. (1976) "In Search of the Causal Relationship between Money and Income", paper given at the meetings of the Southern Economic Association, Atlanta.

Kareken, J. et Solow, S. (1963) "Lags in Monetary Policy", Commission on Money and Credit, Stabilization Policies (Englewood Cliffs, NJ.: Prentice-Hall, Inc.), 14-25.

Kay, J. (2012) "The map is not the territory: An essay on the state of economics", *Voprosy Ekonomiki*, vol. 5.

Keane M. et Runkle D. (1990) "Testing the rationality of price forecasts: from panel data", *American Economic Review*, vol. 80.

Keynes, J.M. (1936) *Théorie générale de l'emploi, de l'intérêt et de la monnaie*, Edition 1969, Paris, Bibliothèque Scientifique, Payot.

Keynes, J.M. (1937) "The general theory of employment", *The Quarterly Journal of Economics*, vol.51, pp.209-223. Traduction française, *Revue Française d'Économie*, 1990, vol.5, n°4, pp.141-156.

Klein, L. (1966) *The Keynesian Revolution*, New York: Macmillan, 1947. 2d ed. New York: Macmillan, 1966.

Klein L.R. et Goldberger A. (1955) *An Econometric Model of the United States, 1929-1952*, Amsterdam : North-Holland.

Knight, F. (1921) *Risk, Uncertainty and Profit*, Washington, DC: Beardbooks.

Knight, F. (1941) "The Business Cycle, Interest, and Money: A Methodological Approach", *Review of Economic Statistics*, 23, 53-67.

Koopmans, T. (1947) "Measurement without Theory", *Review of Economic Statistics*, 29, 161-172.

Koopmans, T. (1950) "Statistical Inference in Dynamic Economic Models", Cowles Commission Monograph 10, New York: Wiley.

Kydland, F. et Prescott, E. (1977) "Rules rather than Discretion : The Inconsistency of Optimal Plans", *Journal of Political Economy*, 85, 473-190.

Kydland, F. et Prescott, E. (1982) "Time to Build and Aggregate Fluctuations", *Econometrica*, November, 50(6), 1345-1370.

Kydland, F. et Prescott, E. (1991) "The econometrics of the general equilibrium approach to business cycles", *Scandinavian Journal of Economics*, vol. 93, n°2.

Kydland, F. et Prescott, E. (1996) "The Computational Experiment : an Econometric Tool", *Journal of Economic Perspectives*, 10, 69-85.

Laidler, D. (1970) *The Demand for Money: Theories and Evidence*, Scranton, Pa.: International Textbook.

Laidler, D. (1980) "The demand for money in the United States: yet again", ed. K. Brunner and A. H. Meltzer, Carnegie-Rochester Conference Series, The State of Macroeconomics. Amsterdam: North Holland.

Laidler, D. (1981) "Monetarism: An Interpretation and an Assessment", *The Economic Journal*, Vol. 91, N° 361, p. 1-28.

Leamer, E. (1985) "Vector Autoregressions for Causal Inference?", in K. Brunner et A. Meltzer (eds.), Understanding Monetary Regimes, Carnegie Rochester Conference Series, Volume 22.

Leeper, E. et Zha, T. (2003) "Small Policy Interventions", *Journal of Monetary Economics*, 50, 1673-1700.

Lipsey, R.G. (1960) "The Relation Between Unemployment and the Rate of Change of Money Wage Rate in the United Kingdom 1862-1957: A Further Analysis", *Economica*, 27 : 1-31.

Lovell, M. (1986) "Tests of rational expectations hypothesis", *American Economic Review*, vol. 76.

Lubik, A. et Schorfheide, F. (2004) "Testing for Indeterminacy: An Application to U.S. Monetary Policy", *American Economic Review*, 94(1), 190-217.

Lucas, R. (1972) "Expectations and the Neutrality of Money", *Journal of Economic Theory*, vol. 4, avril.

Lucas, R. (1976) "Econometric Policy Evaluation : A Critique", Carnegie Rochester Conference Series on Public Policy, 1, 19-46.

Lucas, R. (1977) "Understanding Business Cycles", in K. Brunner et A. Meltzer (eds.), Stabilization of the Domestic and International Economy, Amsterdam : Carnegie Rochester Conference Series, North-Holland.

Lucas, R. (1980) “Methods and Problems in Business Cycle Theory”, *Journal of Money, Credit and Banking*, vol. 12, novembre.

Lucas, R. (1981) “Tobin and Monetarism: A Review Article”, *Journal of Economic Literature*, 19 (2): 558–567.

Lucas, R. (1982) “Studies in business cycles Theory”, MIT Press, Cambridge.

Lucas, R. (1986) “Adaptive Behavior and Economic Theory”, *Journal of Business*, 59 (4): 401-426.

Lucas, R. (1987), *Models of Business Cycles*, New York, NY : Basil Blackwell.

Lucas, R. et Sargent T. (1979) “After Keynesian Macroeconomics”, *Federal Reserve Bank Minneapolis Quarterly Review*, 3.

Mankiw, N. (1985) “Small menu costs and large business cycles: A macroeconomic model of monopoly”, *The Quarterly Journal of Economics*, Vol. 100, No. 2, pp. 529-537.

Maurel, F. (1989) “Modèles à correction d'erreur : l'apport de la théorie de la cointégration”, *Economie et Prévision*, n° 88-89, 105-125.

Mayer, T. (1975) “The structure of monetarism”, *Kredit und Kapital*, 8, p. 191–215.

McCallum, B. (1986) “Monetary versus Fiscal Policy Effects: A Review of the Debate”, Carnegie-Mellon University and National Bureau of Economic Research.

McCallum, B. et Nelson, E. (2000) “Timeless perspective versus discretionary monetary policy in forward-looking models”, *NBER Working Paper*, n°. 7915.

Meidinger, C. (1994) *Science Economique: Questions de Méthode*, Vuibert.

Mints, L. (1945) *A History of Banking Theory*, Chicago: University of Chicago Press.

Mishkin, F. (2008) “Comment Réagir aux Bulles des Prix d’Actifs”, *Revue de la stabilité financière*, n° 12, octobre.

Montfort, A. et Rabemananjara, R. (1990) “From a VAR model to a structural model, with an application to the wage-price spiral”, *Journal of Econometrics*, vol 5, pp 203-227.

Morishima, M. et Saito, M. (1964) “A dynamic analysis of the American economy, 1902-1952”, *International Economic Review*, p. 125164.

Mundell, R. (1962), “The Appropriate Use of Monetary and fiscal Policy under Fixed Exchange Rates”, *IMF Staff Papers*, vol. 9, p. 70-79.

Muth, J.F. (1961) “Rational Expectation and the Theory of Price Movements”, *Econometrica*, 29, p. 315-335.

Ohlin, B. (1937) “Alternatives Theories of the Rate of Interest”, *Economic Journal*.

Pagan, A. (1987) “Three Econometric Methodologies: A Critical Appraisal”, *Journal of Economic Surveys*, 1, 2-24.

Park, Y.C. (1972) “Some Current Issues on the Transmission Process of Monetary Policy”, *Staff Papers, International Monetary Fund*, Vol. 19, No. 1, pp. 1-45.

Parkin M. [1986], “The output inflation Trade off when prices and Costly to change”, *Journal of Political Economy*, vol. 94, n°1.

Patinkin, D. (1969) “The Chicago Tradition, The Quantity Theory, And Friedman”, *Journal of Money, Credit and Banking*, Vol. 1, No. 1, pp. 46-70.

Patinkin, (1972) “Friedman on the Quantity Theory and Keynesian”, *Economics Journal of Political Economy*, Vol. 80, No. 5, pp. 883-905

Paulré, B. (1985) *La causalité en économie: signification et portée de la modélisation structurelle*, Presses universitaires de Lyon.

Paulré, B. (2007) "La causalité en économie: mise en perspective et approches opérationnelles", in *Leçons de Philosophie Economique, Tome 3, "Science Economique et Philosophie des Sciences"*, Leroux, A., Livet, P.

Pierce, D.A. (1977) "Relationships-and the Lack Thereof-between Economic Time Series with Special Reference to Money and Interest Rates", *Journal of the American Statistical Association*, 72.

Pirotte, A. (2004) *L'Econométrie: des Origines aux Développements Récents*, CNRS éditions.

Popper, K. (1965) *Conjectures and Refutations: The Growth of Scientific Knowledge*, London: Routledge and Kegan Paul.

Rogoff, K. (1985) "The Optimal Degree of Commitment to an Intermediate Monetary Target", *Quarterly Journal of Economics*, vol. 100, p. 1169-1189.

Roberts, J. M. (1997) "Is Inflation Sticky? ", *Journal of Monetary Economics*, 39, 173-96.

Robertson, D. (1940) "M. Keynes and the Rate of Interest", *Essays in Monetary Theory*, London.

Rotemberg, J. (1982) "Sticky Prices in the United States", *Journal of Political Economy*, Vol. 90, No. 6, pp. 1187-1211.

Runkle, E. (1987) "Vector autoregressions and reality", *Journal of Business and Economic Statistics*, 5 (October): 437-42.

Quine, W. V. O. (1951) "Two dogmas of empiricism", *Philosophical Review*, 60, pp. 20-43.

Sargent, T. (1979) "Estimating Vector Autoregressions Using Methods Not Based on Explicit Economic Theories", *Federal Reserve Bank of Minneapolis Quarterly Review*, 3, 8-15.

- Sargent, T. (1981) "Interpreting economic time series", *The Journal of Political Economy*.
- Sargent, T. (1982) "Beyond Demand and Supply Curves in Macroeconomics", *American Economic Review*, vol. 72, n°2, May.
- Sargent T. et Wallace N. (1981) "Some Unpleasant Monetarist Arithmetics", *Federal Reserve Bank of Minneapolis Quarterly Review*, 5, p. 1-17.
- Simon, H. A. (1953) "Causal Order and Identifiability", in Hood and Koopmans editors, *Studies in Econometric Method*, Cowles Commission Monograph 14. New York: Wiley.
- Simons, H. (1948) *Economic Policy for a Free Society*, Chicago: University of Chicago Press.
- Sims, C. (1972) "Money, Income, and Causality", *American Economic Review*, vol. 62 (4), pp. 540-552.
- Sims, C. (1978) "Small Econometric Models of the US and West Germany Without Prior Restrictions", Discussion Paper No. 78 - 105, Center for Economic Research, Department of Economics, University of Minnesota.
- Sims, C. (1980 a) "Macroeconomics and Reality", *Econometrica*, 48 (1), 1-48.
- Sims, C. (1980 b) "Comparison of Interwar and Postwar Cycles : Monetarism Reconsidered", Working Paper n° 430, NBER.
- Sims, C. (1986) "Are Forecasting Models Usable for Policy Analysis?", *Minneapolis Federal Reserve Bank Quarterly Review*, vol. 10 (1), pp. 2-16.
- Sims, C. (1996) "Macroeconomics and Methodology", *Journal of Economic Perspectives*, 10, 105-120.
- Sims, C. (1999) "Drift and Breaks in Monetary Policy", Princeton University.

Sims, C., Goldfeld, S. et Sachs, J. (1982) "Policy Analysis with Econometric Models", *Brookings Papers on Economic Activity*, n°1.

Sims, C. et Zha, T. (2006) "Were there regime switches in US monetary policy?", *American Economic Review*, 96, 54-81.

Smets, F. et Wouters, R. (2003) "An Estimated Stochastic Dynamic General Equilibrium Model for the Euro Area", *Journal of the European Economic Association*, 1 (5).

Spencer, D. (1989) "Does money matter? The robustness of evidence from vector autoregressions", *Journal of Money, Credit, and Banking*, 21 (November): 442-54.

Stock, J. et Watson, M. (2003) "Understanding changes in Business Cycle dynamics", NBER Working Paper, No. 9859.

Strotz, R. (1960) "Interdependence as a Specification Error", *Econometrica*, Vol. 28, No. 2, pp. 428-442.

Strotz, R. et Wold, H. (1960) "Recursive vs. Nonrecursive Systems: An Attempt at Synthesis", *Econometrica*, Vol. 28, No. 2, pp. 417-427.

Svensson, L. et Woodford, M. (2004) "Implementing Optimal Policy through Inflation-Forecast Targeting", In *The Inflation Targeting Debate*, edited by B. S. Bernanke and M. Woodford. Chicago: University of Chicago Press.

Taouil, R. (2011) "La Nouvelle Synthèse et la Théorie des Cycles Réels", *Economie Appliquée*, Tome LXIV, n° 1, 75-102.

Taouil, R. (2014 a) "R. E. Lucas et les fondements des modèles macroéconomiques", Centre de Recherche en Economie de Grenoble, Université de Grenoble, à paraître.

Taouil, R. (2014 b) "La macroéconomie est-elle en crise?", Centre de Recherche en Economie de Grenoble, Université de Grenoble, à paraître.

Taylor, J. (1993) "Discretion versus Policy Rules in Practice", *Carnegie Rochester Conference Series on Public Policy*, 39, p.195-214.

Taylor, J. (1999) *Monetary Policy Rules*, Chicago: The University of Chicago Press.

Tobin, J. (1970) "Money and income: *Post hoc ergo propter hoc?*", *Quarterly Journal of Economics* (82) 301-317.

Tobin, J. (1972) "Friedman's Theoretical Framework", *Journal of Political Economy*, Vol. 80, No. 5, pp. 852-863.

Todd, R. (1990). "Vector Autoregression Evidence on Monetarism: Another Look at the Robustness Debate", *Federal Reserve Bank of Minneapolis Quarterly Review*, Vol. 14, No. 2.

Vercelli, A. (1991) *Methodological Foundations of Macroeconomics: Keynes and Lucas*, Cambridge: Cambridge University Press.

Viner, J. (1937) *Studies in the Theory of International Trade*, New York: Harper and Brothers.

Vining, R. (1949) "Koopmans on the Choice of Variables to be Studied and the Methods of Measurement", *The Review of Economics and Statistics*, Vol. 31, No. 2, p. 76-86.

Vo Phuong M. L., Meenagh, D., Minford, P. et Wickens, M. (2012) "Testing DSGE models by indirect inference and other methods : Some Monte Carlo experiments", *Cardiff Economics Working Papers*.

Wickens, M. (2006) *Macroeconomics Theory : A Dynamic General Equilibrium Approach*, Princeton University Press.

Wold, H. (1954) "Causality and Econometrics", *Econometrica*, 22, pp. 114.

Woodward, James B. (2003) *Making Things Happen: A Theory of Causal Explanation*, Oxford University Press.

Woodford, M. (2003) *Interest and Prices: Foundations of a Theory of Monetary Policy*, Princeton, NJ: Princeton University Press.

Wu, D.M. (1973) "Alternative tests of independence between stochastic regressions and disturbances", *Econometrica*, 41, 733-750.

Zellner, A. (1979) "Causality in Econometrics", in K. Brunner and A. H. Meltzer (eds.), *Three Aspects of Policy and Policymaking: Knowledge, Data, and Institutions* Carnegie-Rochester Series on Public Policy 10 (1979), 9-54.

Table des Matières

| | |
|---|----|
| <i>Résumé</i> | 6 |
| <i>Abstract</i> | 8 |
| REMERCIEMENTS | 10 |
| SOMMAIRE | 14 |
| INTRODUCTION GENERALE | 18 |
| | |
| PREMIERE PARTIE : LA MODELISATION STRUCTURELLE REMISE EN CAUSE PAR LES MONETARISTES ET LES MODELES VAR | 34 |
| | |
| CHAPITRE 1 : LA CRITIQUE MONETARISTE DE LA POLITIQUE MONETAIRE KEYNESIENNE | 38 |
| INTRODUCTION | 40 |
| | |
| SECTION 1 : IS-LM COMME CADRE DE LA MODELISATION STRUCTURELLE | 42 |
| | |
| Sous-section 1 : Modèles macroéconométriques et souci de validation empirique | 42 |
| 1. <i>Le modèle économétrique de Klein- Goldberger</i> | 42 |
| 2. <i>Les versions monétaire et à prix fixé de IS-LM</i> | 45 |
| | |
| Sous-section 2 : L’attaque monétariste des fondements keynésiens de la synthèse | 50 |
| 1. <i>Demande effective et autonomie de l’investissement et l’épargne</i> | 50 |
| 2. <i>Nature monétaire du taux d’intérêt et récursivité du modèle</i> | 53 |

| | |
|--|-----------|
| SECTION 2 : CANAL DU TAUX D'INTERET VERSUS CANAL DES PRIX | 56 |
| Sous-section 1 : La transmission de la politique monétaire | 56 |
| 1. <i>Courbe LM et multiplicateur monétaire.....</i> | 56 |
| 2. <i>Choix de portefeuille et canal monétariste</i> | 59 |
| Sous-section 2: La critique monétariste | 62 |
| 1. <i>Fixité des prix et trappe à liquidité</i> | 62 |
| 2. <i>Effet d'éviction et prééminence de la politique monétaire.....</i> | 64 |
| CONCLUSION..... | 70 |
| | |
| CHAPITRE 2 : LES LIMITES THEORIQUES DE LA POLITIQUE MONETAIRE MONETARISTE..... | 72 |
| | |
| INTRODUCTION..... | 74 |
| | |
| SECTION 1 : DEMANDE DE MONNAIE ET COMPOSITION DU PORTEFEUILLE | 76 |
| | |
| Sous-section 1 : Demande de monnaie monétariste et préférence pour la liquidité | 76 |
| 1. <i>Théorie quantitative et demande de monnaie</i> | 76 |
| 2. <i>L'influence keynésienne</i> | 81 |
| | |
| Sous-section 2 : Un modèle dynamique du revenu..... | 84 |
| 1. <i>Demande de monnaie et relation de Fisher</i> | 84 |
| 2. <i>Le problème de la dynamique du revenu</i> | 88 |
| | |
| SECTION 2 : DECALAGES TEMPORELS ET CAUSALITE MONNAIE-REVENU : L'ARGUMENT DE LA CAUSALITE INVERSEE | 93 |
| | |
| Sous-section 1 : Théorie du revenu permanent et stabilité de la demande de monnaie .. | 93 |
| 1. <i>Revenu permanent et décalages temporels</i> | 93 |
| 2. <i>Incohérence théorique de l'explication de la vitesse-revenu.....</i> | 96 |

| | |
|---|------------|
| Sous-section 2 : Limites théoriques de la causalité monnaie-revenu..... | 99 |
| 1. <i>De la causalité empirique monnaie-revenu</i> | 99 |
| 2. <i>... à l'a priori théorique</i> | 101 |
| | |
| SECTION 3 : LA RELECTURE MONETARISTE DE LA COURBE DE PHILLIPS | 108 |
| | |
| Sous-section 1 : De la rigidité du salaire à la courbe de Phillips dans IS-LM | 108 |
| 1. <i>Chômage involontaire et efficacité de la politique monétaire</i> | 108 |
| 2. <i>La base empirique de l'arbitrage inflation-chômage</i> | 113 |
| | |
| Sous-section 2 : Objectif de stabilité des prix et règle monétariste | 116 |
| 1. <i>Courbe de Phillips verticale à long terme</i> | 116 |
| 2. <i>Politique monétaire monétariste</i> | 120 |
| | |
| CONCLUSION..... | 122 |
| | |
| CHAPITRE 3 : L'EVALUATION DE LA POLITIQUE MONETAIRE PAR LES MODELES VAR : L'EMPIRIE FACE A L'IMPERATIF THEORIQUE | 124 |
| | |
| INTRODUCTION..... | 126 |
| | |
| SECTION 1. MODELES VAR ET CAUSALITE MONNAIE-REVENU..... | 128 |
| | |
| Sous-section 1. G-causalité et S-causalité..... | 128 |
| 1. <i>Exogénéité de la monnaie et G-causalité.....</i> | 128 |
| 2. <i>La S-Causalité de la monnaie-revenu</i> | 132 |
| | |
| Sous-section 2. Critique de Sims et ses implications | 136 |
| 1. <i>Causalité et analyse impulsionnelle.....</i> | 136 |
| 2. <i>La robustesse de la causalité monnaie-revenu en question</i> | 141 |

| | |
|--|------------|
| SECTION 2. LES VAR STRUCTURELS ET LE RETOUR OBLIGE A LA THEORIE | 147 |
| Sous-section 1. Causalité et exogénéité : questions de méthode | 147 |
| 1. <i>Le caractère a-théorique de la G-causalité</i> | 147 |
| 2. <i>Exogénéité et identification</i> | 150 |
| Sous-section 2. L'indétermination empirique de la causalité | 155 |
| 1. <i>Du paradoxe de la référence à Wold au VAR "contrefactuel imposteur"</i> | 155 |
| 2. <i>La sous-détermination de la théorie par l'observation</i> | 159 |
| CONCLUSION | 165 |
| | |
| DEUXIEME PARTIE : DU RETOUR FORCE DE LA THEORIE A SA PRIMAUTE | 166 |
| | |
| CHAPITRE 4 : TENTATIVES ECONOMETRIQUES DE CONCILIATION ENTRE THEORIE ET EMPIRIE | 170 |
| INTRODUCTION | 172 |
| SECTION 1 : LE CRITERE PREDICTIF DES MONETARISTES | 175 |
| Sous-section 1 : Le statut des paramètres du modèle | 175 |
| 1. <i>Le poids de l'empirie chez les monétaristes</i> | 175 |
| 2. <i>Formes réduites versus formes structurelles et les restrictions a priori</i> | 178 |
| Sous-section 2 : Propositions monétaristes et limites des tests empiriques | 181 |
| 1. <i>Coefficient de corrélation et problèmes de spécification</i> | 181 |
| 2. <i>Les problèmes de mesure</i> | 184 |

| | |
|---|------------|
| SECTION 2. LA COINTEGRATION ET L'ARTICULATION ENTRE COURT ET LONG TERME | 187 |
| Sous-section 1 : La dynamique des variables monétaires..... | 187 |
| 1. <i>Demande de monnaie et modèle à correction d'erreur</i> | <i>187</i> |
| 2. <i>Cointégration et relation économétrique de long terme</i> | <i>191</i> |
| Sous-section 2 : Validation empirique et exigence théorique..... | 195 |
| 1. <i>Le retour au problème de l'identification</i> | <i>195</i> |
| 2. <i>Modèle conditionnel et exogénéité.....</i> | <i>197</i> |
| CONCLUSION..... | 201 |
| | |
| CHAPITRE 5 : LA CRITIQUE DE LUCAS ET SES IMPLICATIONS..... | 202 |
| | |
| INTRODUCTION..... | 204 |
| | |
| SECTION 1. LA CRITIQUE DE LUCAS : UNE INNOVATION CARDINALE..... | 206 |
| | |
| Sous-section 1. La discipline de l'équilibre | 206 |
| 1. <i>Invariance des paramètres et anticipations rationnelles</i> | <i>206</i> |
| 2. <i>Optimisation intertemporelle et apurement des marchés</i> | <i>209</i> |
| | |
| Sous-section 2. Une nouvelle posture épistémologique | 211 |
| 1. <i>Le statut formel non descriptif du modèle.....</i> | <i>211</i> |
| 2. <i>Calibrage versus tests économétriques.....</i> | <i>213</i> |
| | |
| SECTION 2. LA COHERENCE TEMPORELLE DES POLITIQUES ECONOMIQUES : UN RENOUVEAU DE L'ECONOMIE POLITIQUE | 214 |
| | |
| Sous-section 1. L'incohérence temporelle de la politique discrétionnaire | 214 |
| 1. <i>Anticipations rationnelles et optimalité de la politique économique.....</i> | <i>214</i> |
| 2. <i>Crédibilité et indépendance de la Banque centrale</i> | <i>216</i> |

| | |
|--|------------|
| Sous-section 2. La subordination de la politique budgétaire à la politique monétaire.. | 219 |
| 1. <i>Policy-mix et mode de financement des dépenses publiques</i> | 219 |
| 2. <i>Le policy-mix comme jeu stratégique</i> | 221 |
| CONCLUSION | 223 |
| | |
| CHAPITRE 6 : LES MODELES DSGE ET L'EXTENSION DE LA METHODOLOGIE DES CYCLES REELS | 224 |
| | |
| INTRODUCTION | 226 |
| | |
| SECTION 1. LES FONDEMENTS DES MODELES DSGE | 228 |
| | |
| Sous-section 1. L'influence de la méthodologie Lucasienne | 228 |
| 1. <i>Une hégémonie de l'approche réelle</i> | 228 |
| 2. <i>Combinaison du calibrage et de l'économétrie</i> | 230 |
| | |
| Sous-section 2. Demande agrégée et courbe de Phillips | 232 |
| 1. <i>Rigidités et concurrence monopolistique</i> | 232 |
| 2. <i>Relation IS et courbe de Phillips néo-keynésiennes</i> | 233 |
| | |
| SECTION 2. EVALUATION DE LA POLITIQUE MONETAIRE : SUPREMATIE DES DSGE SUR LES VAR | 238 |
| | |
| Sous-section 1. Règle de politique monétaire et objectif de stabilité des prix | 238 |
| 1. <i>Régime à politique monétaire dominante</i> | 238 |
| 2. <i>Règle de Taylor et ciblage de l'inflation</i> | 240 |
| | |
| Sous-section 2. La Grande modération : l'échec explicatif des VAR | 241 |
| 1. <i>Règle monétaire versus hypothèse de la "chance"</i> | 241 |
| 2. <i>VAR et restrictions inter-équations</i> | 243 |
| | |
| CONCLUSION | 247 |

CONCLUSION GENERALE 248

REFERENCES BIBLIOGRAPHIQUES..... 252

TABLE DES MATIERES 272