

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

PRÉVISION DE SÉRIE ISSUE DE L'AGRÉGATION CONTEMPORAINE :  
COMPARAISON ENTRE L'APPROCHE AGRÉGÉE ET DÉSAGRÉGÉE

MÉMOIRE  
PRÉSENTÉ  
COMME EXIGENCE PARTIELLE  
DE LA MAÎTRISE EN SCIENCE ÉCONOMIQUE

PAR  
PHILIPPE MARCIL

Juillet 2007

UNIVERSITÉ DU QUÉBEC À MONTRÉAL  
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.01-2006). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

## REMERCIEMENT

Je voudrais remercier Alain Guay, Frédérick Demers et Charles-Olivier Fiola de m'avoir aidé dans la rédaction de mon mémoire.

## TABLE DES MATIÈRES

REMERCIEMENT .....	ii
LISTE DES TABLEAUX .....	v
RÉSUMÉ .....	vi
INTRODUCTION .....	1
CHAPITRE I	
REVUE DE LA LITTÉRATURE	
1.1 Résumé historique et développement du champ d'études .....	3
1.2 Griliches et Grunfeld: une histoire de corrélation .....	4
1.3 Lutkepohl: moins est parfois mieux que trop .....	6
1.4 Hendry et Hubric : une question d'information .....	8
1.5 Granger et Newbold : les implications de l'agrégation de processus autorégressifs	9
1.6 Ce que l'on doit accomplir pour avancer la littérature .....	10
CHAPITRE II	
ANALYSE THÉORIQUE	
2.1 Élaboration du modèle .....	12
2.2 Comparaison des prévisions avec des processus connus .....	13
2.3 Comparaison des prévisions avec des processus inconnus et estimés .....	20
2.4 Nombre de retards variables entre le modèle agrégé et désagrégé .....	27
2.5 Quelles sont les innovations de notre modèle théorique? .....	30
CHAPITRE III	
SIMULATION	
3.1 Les liens entre les simulations, le modèle théorique et la littérature .....	31
3.2 Sélection des retards et caractéristiques générales des simulations .....	32

3.3 Simulations de processus à deux variables : le modèle de Grunfeld et Griliches ...	32
3.4 Simulation de processus désagrégés à trois variables .....	37
3.5 Simulations de processus agrégés identiques et l'impact des racines communes....	42
3.6 Les résultats des simulations appuient-ils notre modèle théorique? .....	44
CONCLUSION .....	46
RÉFÉRENCES .....	48

## LISTE DES TABLEAUX

Tableau		Page
3.1	Comparaison des erreurs de prévisions pour deux AR(2) indépendants .....	34
3.2	Comparaison des erreurs de prévisions pour des VAR(2) à deux variables ..	36
3.3	Comparaison des erreurs de prévisions pour trois AR(2) indépendants .....	39
3.4	Comparaison des erreurs de prévisions pour des VAR(2) à trois variables ..	41
3.5	Comparaison des erreurs de prévisions pour des VAR(2) à quatre variables	44

## RÉSUMÉ

Nous étudions la différence des erreurs de prévisions entre une représentation agrégée et désagrégée lorsque l'on s'intéresse à la prévision de la série agrégée. Pour ce faire, nous développons un modèle théorique qui étudie les erreurs de prévisions d'une représentation agrégée et désagrégée en unissant les théories de différents auteurs dans un même cadre analytique. Le modèle que nous avons développé démontre que la présence de racine commune est essentielle, mais est une condition non suffisante pour que le modèle agrégé ait une erreur de prévision plus faible que le modèle désagrégé. De plus, les simulations démontrent que la présence de racines communes aide à faire diminuer l'écart des erreurs de prévisions entre une représentation agrégée et désagrégée.

## INTRODUCTION

Les prévisions de séries macroéconomiques issues de l'agrégation de variables contemporaines sont essentielles au bon fonctionnement des institutions économiques et politiques. L'amélioration des prévisions de ces séries est donc d'un grand intérêt. Mais sous quelles conditions doit-on utiliser un modèle agrégé pour faire une prévision d'une série agrégée?

Plusieurs chercheurs ont essayé de répondre à cette question. Les modèles développés jusqu'à présent (Rose, 1977; Hendry et Hubric, 2006; Pesaran, 2003) sont cependant incapables d'expliquer les résultats empiriques divergents qui, selon les séries étudiées, démontrent des gains où des pertes à utiliser la désagrégation à des fins de prévision. La condition pour qu'une représentation d'un modèle agrégé soit le meilleur indicateur, selon la théorie d'agrégation classique, est que les paramètres du processus générateur de données d'un modèle linéaire sont égaux. Mais les études empiriques où le modèle agrégé est un meilleur indicateur que le modèle désagrégé ne répondent pas toujours à ce critère.

Même les modèles théoriques plus nuancés, comme celui de Lutkepohl (1984), sont incapables d'expliquer les divergences des résultats des applications empiriques.

Nous allons donc essayer de développer dans ce mémoire un modèle théorique qui peut expliquer les résultats empiriques divergents. Pour ce faire, nous nous intéressons à l'importance des racines autorégressives communes et leurs impacts sur les erreurs de prévisions du modèle agrégé car elles n'ont pas été étudiées en profondeur dans la littérature.

Pour ce faire, nous allons examiner les travaux de plusieurs auteurs (Grunfeld et Griliches, 1960; Lutkepohl, 1984; Hendry et Hubric, 2006; Granger et Newblod, 1977). Nous allons ensuite reprendre le développement de Rose (1977) et Reinsel (1980) et poursuivre leur travail en utilisant les modèles théoriques analysés dans notre revue de la littérature alors que nous unirons ces différents modèles théoriques dans un même cadre analytique. Finalement, nous allons effectué des simulations pour s'assurer de la validité des résultats



théoriques et aussi explorer certaine des implications du modèle théorie de Grunfeld et Griliches (1960).

## CHAPITRE I

### REVUE DE LA LITTÉRATURE

#### 1.1 Résumé historique et développement du champ d'études

La désagrégation contemporaine a été le sujet d'une vaste littérature qui débute avec le travail de Theil (1954). Ce dernier étudiait la disparité entre les paramètres estimés d'équations agrégés et désagrégés et il constata que les paramètres agrégés étaient biaisés. Le biais d'agrégation fut étudié par plusieurs auteurs, donc notamment Rose (1977) et Pesaran, Pierse et Kumar (1989). D'autres auteurs ont cependant cherché des failles à son travail, donc notamment Grunfeld et Griliches (1960) et Lutkepohl (1984).

Les résultats des études empiriques qui utilisent la désagrégation ne donnent cependant raison à aucun des deux camps. La majorité des études constate une réduction de l'erreur quadratique de prévision par rapport à une représentation agrégée. D'excellents exemples sont les travaux de Demers et Dechamplain (2005), Demers et Dupuis (2005) et Tobias et Zellner (2000). Cependant, certains chercheurs ont été confrontés avec des résultats négatifs. Notamment Hubrich (2005), Zotteri, Kalchschmidt et Caniato(2005) et Benalal, Dei Hoyo, Landau, Roman et Skudelny (2004). Ces résultats plus négatifs sont en ligne avec les idées de Lutkepohl (1984) et de Grunfeld et Griliches (1960). Il existerait donc des processus générateur de données pour lesquels l'agrégation est préférable.

Mais est-ce que les travaux de Grunfeld et Griliches et de Lutkepohl peuvent expliquer les échecs de la désagrégation à améliorer les prévisions qu'ont rencontré Hubrich (2005), Zotteri, Kalchschmidt et Caniato(2005) et Benalal, Del Hoyo, Landau, Roman et Skudelny (2004)?

Pour répondre à ces questions, il nous faut d'abord étudier le travail de Grunfeld et Griliches (1960) et Lutkepohl (1984) qui donnent des explications théoriques aux problèmes potentiels de la désagrégation. Nous examinons ensuite le travail d'Hendry et Hubric (2006)

qui ont développé de nouvelles méthodologies empiriques concernant les modèles désagrégés et une théorie basée sur l'information disponible pour la prévision. Finalement, nous étudions la théorie de Granger et Newbold (1977) sur l'agrégation de processus ARMA et démontrons son potentiel malheureusement inutilisé. Nous terminons ce tour d'horizon avec les implications de la littérature sur ce travail.

## 1.2 Griliches et Grunfeld : une histoire de corrélation

Griliches et Grunfeld ont surtout étudié l'estimé des régressions agrégées et désagrégées dans leur article « Is Aggregation Necessarily Bad » (1960). Ils s'intéressent surtout au  $R^2$  qu'ils utilisent comme critère d'évaluation. La première section de leur travail compare différentes études longitudinales : une sur l'investissement de 8 firmes individuelles et l'autre sur la demande d'engrais de 9 groupes d'États Américains, dont les  $R^2$  des modèles agrégés sont plus élevés que ceux des modèles désagrégés. Ils élaborent deux hypothèses et deux modèles théoriques qui peuvent expliquer ce phénomène.

Leur première hypothèse est qu'il existe un gain de synchronisation si l'agrégation s'effectue entre des variables donc la corrélation est élevée. Plus la corrélation entre les variables désagrégées est forte, plus la différence entre le  $R^2$  de la représentation agrégée et de la représentation désagrégée est grande. Grunfeld et Griliches démontrent cette hypothèse en établissant un modèle simple :

$$y_i = \beta_i x_i + \mu_i \quad (1.2.1)$$

et qui s'agrège et donne

$$Y = BX + \mu \quad (1.2.2)$$

où  $Y = \sum_{i=1}^m y_i$  et  $X = \sum_{i=1}^m x_i$ . Sous les hypothèses que les coefficients sont égaux,

$\beta_i = \beta_j = B$ , et que la variance et la covariance entre les variables désagrégées sont les mêmes,  $var(x_i) = var(x_j)$  et  $cov(x_i, x_j) = cov(x_g, x_k)$  pour  $i \neq g$  et  $j \neq k$ . Il en est aussi de même pour la variance et covariance des erreurs,  $var(\mu_i) = var(\mu_j)$  et  $cov(\mu_i, \mu_j) = cov(\mu_g, \mu_k)$  pour  $i \neq g$  et  $j \neq k$ . Sous ces conditions, Grunfeld et Griliches démontrent que :

$$R_a/R_i = \frac{B^2 + \text{var}(\mu)/\text{var}(X)}{B^2 + [\text{var}(\mu) * \text{corr}(\mu_i, \mu_j) / \text{var}(X) * \text{corr}(x_i, x_j)]} \quad (1.2.3)$$

où  $R_a$  est le  $R^2$  de l'équation agrégée,  $R_i$  est le  $R^2$  d'une équation désagrégée. Par conséquent, le modèle agrégé peut avoir plus de pouvoir explicatif qu'une équation désagrégée si la corrélation des erreurs  $\text{corr}(\mu_i, \mu_j)$  est plus petite que la corrélation des variables indépendantes  $\text{corr}(x_i, x_j)$ . Donc, pour qu'il existe un gain à l'agrégation, il faut que les variables désagrégées possèdent soit corréllées entre elles.

Leur seconde hypothèse est que l'on ne connaît pas suffisamment les équations désagrégées pour les spécifier correctement, mais que l'équation agrégée permet de limiter les problèmes de spécification des équations désagrégées. Il existe alors un bénéfice à agréger. La démonstration effectuée par Grunfeld et Griliches pour démontrer ce phénomène est peu convaincante, car ils utilisent comme modèle :

$$y_i = \beta_i x_i + \beta X + \mu_i \quad (1.2.4)$$

où  $X = \sum x_i$ . Le modèle désagrégé mal spécifié est :

$$y_i = \beta_i x_i + \mu_i \quad (1.2.5)$$

Puisque  $X$  est la somme des  $x_i$ , nous pouvons intuitivement déduire que le modèle agrégé :

$$Y = B X + \epsilon \quad (1.2.6)$$

est un meilleur que le modèle désagrégé estimé. Donc, si les variables désagrégées sont affectées par la variable agrégée et non les mouvements d'une sous composante, le modèle agrégé pourrait potentiellement être supérieur. Malheureusement, Grunfeld et Griliches n'étudient qu'un type d'erreur de spécification et l'on ne peut donc pas généraliser leur analyse à toutes les erreurs de spécifications que l'on puisse imaginer. En résumé, Grunfeld et Griliches ont démontré que :

- La corrélation entre les variables désagrégées pouvait, sous certaines conditions, causer un gain à l'agrégation.
- Certains processus générateurs de données qui sont mal spécifiés au niveau désagrégé peuvent être avantagés par l'agrégation.

Ces résultats sont essentiels. Le premier point est plus près d'une condition dont l'impact a souvent été ignoré par les chercheurs, mais qui sera étudiée dans ce travail à l'aide d'un VAR. Le second point ne s'applique que dans le cas spécial décrit par Grunfeld et Griliches puisque les auteurs n'ont considéré qu'un type d'erreur de spécification : l'oubli de la variable agrégée dans la représentation agrégée. Par conséquent, on ne peut attribuer les erreurs de spécifications comme une source de disparité entre les modèles agrégés et désagrégés avec certitude. Les analyses empiriques de Grunfeld et Griliches démontrent que le problème de spécification qu'ils ont identifié peut être rencontré dans les données et qu'il est possiblement important. Nous avons cependant besoin d'une théorie plus générale pour expliquer les inconvénients de la désagrégation, car Grunfeld et Griliches n'en expliquent que deux aspects.

### 1.3 Lutkepohl : moins est parfois mieux que trop

Lutkepohl (1984) soulève l'idée que « la variabilité des modèles induite par la sélection de la spécification et l'estimation des paramètres peut rendre préférable de prévoir l'agrégat directement ». Lutkepohl base son analyse sur l'erreur quadratique de prévision moyenne de l'agrégat faite à l'aide de modèles agrégés et désagrégés. Il appuie son idée sur les résultats de Yamamoto (1981) et de Reinsel (1980) qui stipulent que l'erreur quadratique de prévision augmente plus il y a de paramètres à estimer dans l'équation. Lutkepohl considère le modèle suivant :

$$\begin{bmatrix} 1+\alpha L & 0 \\ 0 & 1+\alpha L \end{bmatrix} \begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} = \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix} \quad (1.3.1)$$

où  $L$  est un opérateur de retard. Lutkepohl écrit l'équation 1.3.1 aussi comme :

$$(1+\alpha L) \begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} = U \quad (1.3.2)$$

où  $y_t = x_{1,t} + x_{2,t}$ . L'erreur de prévision pour une période à l'avance,  $\hat{\Sigma}_x(h=1)$ , basée sur les résultats de Reinsel (1980) sera de :

$$\hat{\Sigma}_x(1) = \left(1 + \frac{kp}{T}\right) \Sigma_U \quad (1.3.3)$$

où  $k$ , le nombre de variable désagrégé dans l'équation, est égale à 2 et  $p$ , le nombre de retard du modèle désagrégé, est égale à 1. Lutkepohl démontre que l'erreur de prévision du modèle agrégé est :

$$\hat{\Sigma}_y(1) = \left(1 + \frac{p}{T}\right) \Sigma_U \quad (1.3.4)$$

Il s'intéresse à la différence des deux prévisions, qui peut se définir comme :

$$\hat{\Delta}(1) = \hat{\Sigma}_y(1) - A \hat{\Sigma}_x(1) A' \quad (1.3.5)$$

où  $A$  est une matrice d'agrégation. En remplaçant Lutkepohl obtient :

$$\hat{\Delta}(1) = \left(1 + \frac{p}{T}\right) \Sigma_U - \left(1 + \frac{kp}{T}\right) A \Sigma_U A' \quad (1.3.6)$$

En remplaçant  $k=2$  et  $p=1$ , Lutkepohl trouve :

$$\hat{\Delta}(1) = \left(\frac{-1}{T}\right) A \Sigma_U A' \quad (1.3.7)$$

Sous l'hypothèse que les coefficients autorégressifs sont égaux, le modèle désagrégé sera un meilleur indicateur que le modèle agrégé, car il existe une pénalité, en petit échantillon, quand on estime davantage de paramètres. Cependant, le modèle théorique de Lutkepohl n'étudie pas le cas où les paramètres des équations désagrégées divergent.

Les simulations de Lutkepohl sur différents modèles désagrégés à deux variables démontrent que sa théorie est vraie lorsque les paramètres des équations désagrégées sont égaux. Le modèle agrégé est alors le meilleur indicateur. Mais, s'il existe des divergences dans les coefficients des modèles désagrégés, la représentation désagrégée a une erreur de prévision plus faible que le modèle agrégé.

De plus, les études empiriques de Lutkepohl sur trois composantes de l'investissement et de la consommation confirment les résultats de ses simulations. Si les équations désagrégées ont des divergences de signes, la représentation désagrégée est le meilleur indicateur.

En résumé, Lutkepohl a identifié un élément essentiel de la désagrégation : il existe un coût à utiliser davantage de variables en prévision hors échantillons. Cependant, Lutkepohl n'a pas considéré qu'il existe aussi un gain à utiliser de l'information désagrégée, ce qu'Hendry et Hubric ont bien compris.

#### 1.4 Hendry et Hubric : une question d'information

Hendry et Hubric (2006) étudient la différence des prévisions d'un modèle agrégé et désagrégé par rapport aux ensembles d'information utilisés par chaque modèle. Par exemple, si on considère deux ensembles d'information,  $I$  et  $i$  où  $i \subset I$ . Hendry et Hubric considèrent que :

$$E(Y_t|I) = f(I) \neq E(Y_t) \quad (1.4.1)$$

$I$  contient donc de l'information qui aide à prévoir  $Y_t$ , on peut donc écrire :

$$E(Y_t|I) = f(I) + v_t \quad (1.4.2)$$

Ils considèrent maintenant l'ensemble d'information  $i$  :

$$E(Y_t|i) = g(i) \neq E(Y_t) \quad (1.4.3)$$

$$E(Y_t|i) = g(i) + \epsilon_t \quad (1.4.4)$$

$$\epsilon_t = Y_t - g(i) = f(I) - g(i) + v_t = w + v_t \quad (1.4.5)$$

où  $w = f(I) - g(i)$ . En comparant l'erreur quadratique de prévision moyenne sous l'hypothèse que  $E(wv_t) = 0$  Hendry et Hubric trouvent :

$$E(\epsilon_t \epsilon_t') = E[(v_t + w) \cdot (v_t + w)'] \quad (1.4.6)$$

$$E(\epsilon_t \epsilon_t') = E(v_t v_t') + E(wv_t) + E(v_t w) + E(ww) \quad (1.4.7)$$

$$E(\epsilon_t \epsilon_t') = E(v_t v_t') + E(ww) \quad (1.4.8)$$

Donc, utiliser un ensemble d'information plus petit fait en sorte que les erreurs de prévisions du modèle restreint soient plus élevées que celles du modèle à information complète. Cette idée est à la base des gains des modèles désagrégés qui utilisent un ensemble d'information plus grand que les modèles agrégés.

Hendry et Hubric supposent qu'il existe des problèmes de spécification quand les modèles agrégés ont des erreurs de prévisions plus faibles que les modèles désagrégés et discutent du gain à ajouter des variables désagrégées aux modèles agrégés et vice versa dans l'objectif d'accroître l'information disponible des modèles. Ils se rapprochent ainsi du travail de Grunfeld et Griliches, mais avec une base théorique plus convaincante.

Hendry et Hubric essaient d'appliquer cette idée au travail d'Hubric sur l'inflation à l'aide d'un VAR qui contient les variables désagrégées et la variable agrégée. Ils découvrent rapidement que le modèle agrégé est un meilleur modèle que le VAR sauf si les variables de

ce dernier sont sélectionnées à l'aide d'une méthodologie de modélisation du générale au spécifique qui permet d'éliminer les variables qui n'ajoutent pas d'information. Ils découvrent ainsi qu'un VAR composé de la variable agrégée et certaines variables désagrégées est un meilleur indicateur que la représentation agrégée.

L'analyse d'Hendry et Hubric explique bien les gains de l'agrégation avec un modèle simple d'information. Cependant, ils ne prennent pas en compte la théorie de Lutkepohl et ne peuvent alors expliquer pourquoi un modèle agrégé peut être plus efficient qu'un modèle désagrégé sans recourir à des problèmes de spécification, ce qui est une approche dangereuse si l'on considère que la démonstration de Grunfeld et Griliches ne s'applique qu'à un seul type d'erreur de spécification. Le travail d'Hendry et Hubric complète le travail de Lutkepohl, mais ces auteurs sous-estiment l'impact de l'agrégation sur les processus générateurs de données.

### 1.5 Granger et Newbold : les implications de l'agrégation de processus autorégressifs

Granger et Newbold (1977) ont développé des modèles mathématiques sur l'agrégation de processus AR et démontrent la règle suivante :

$$X_t \sim AR(p) \text{ et } Y_t \sim AR(q)$$

$$\text{si } Z_t = X_t + Y_t$$

$$\text{alors } Z_t \sim ARMA(a, b)$$

$$\text{où } a \leq p+q \quad b \leq \max(p, q)$$

Ce résultat implique que les processus agrégés auront presque toujours autant de paramètres à estimer qu'un processus désagrégé ce qui a des implications importantes par rapport à la théorie de Lutkepohl et d'Hendry et Hubric. Cependant, les deux auteurs ajoutent qu'il est possible qu'il existe une racine commune entre différents processus autorégressifs. Par exemple, les deux processus suivants;

$$(1-\alpha\beta)X_t = \epsilon_t \quad X_t \sim AR(1) \quad (1.5.1)$$

$$(1-\alpha\beta)(1-\theta\beta)Y_t = u_t \quad Y_t \sim AR(2) \quad (1.5.2)$$

partagent une racine autorégressive commune et leur somme  $Z_t$  peut s'écrire comme :

$$(1-\alpha\beta)(1-\theta\beta)Z_t = (1-\theta\beta)\epsilon_t + u_t \quad Z_t \sim ARMA(2,1) \quad (1.5.3)$$



La règle d'agrégation de deux processus AR avec  $k$  racines communes s'écrit comme suit :

$$\begin{aligned} X_t &\sim AR(p) \text{ et } Y_t \sim AR(q) \\ \text{si } Z_t &= X_t + Y_t, \\ \text{alors } Z_t &\sim ARMA(a, b) \\ \text{où } a &\leq p + q - k \quad b \leq \max(p - k, q - k) \end{aligned}$$

Par conséquent, on se trouve dans une situation où la représentation agrégée a moins de paramètres à estimer que la représentation désagrégée. Le travail de Granger et Newbold sera donc incorporé dans le prochain chapitre.

Le travail de Granger et Newbold est plus tard appliqué sur des VAR par Lutkepohl qui obtient des résultats similaires.

La théorie développée par Granger et Newbold a toujours été prise en considération par les auteurs vus précédemment, mais l'impact de leur modèle n'a jamais été directement utilisé dans les modèles théoriques développés dans ce champ d'étude. Les implications de leur travail ont été brièvement étudiées par Lutkepohl (1984) à l'aide de simulation, mais n'ont pas été ajoutées à son modèle théorique.

#### 1.6 Ce que l'on doit accomplir pour avancer la littérature

Hendry et Hubric ont démontré clairement qu'il existe un gain d'information à utiliser un modèle désagrégé et que par conséquent, l'erreur quadratique moyenne de prévision d'une représentation agrégée sera plus élevée. En contrepartie, le travail de Lutkepohl a aussi démontré qu'il existe une pénalité à estimer un modèle avec davantage de paramètres. Pour obtenir une image complète du problème, il faut donc unir les deux théories dans un seul cadre.

De plus, il serait intéressant d'étudier les problèmes soulevés par Grunfeld et Griliches, notamment la condition de l'existence de corrélation entre les variables désagrégées. Cependant, les questions qu'ils soulèvent dépassent le cadre théorique de ce travail, mais seront tout de même étudiées dans le cadre des simulations au chapitre 3.

Finalement, la théorie de Granger et Newbold doit être intégrée au travail de Lutkepohl et d'Hendry et Hubric. On crée ainsi un modèle qui englobe toute la théorie existante sur l'agrégation contemporaine. Il est en effet possible que la théorie de Granger et Newbold

puisse expliquer les déboires d'Hubrich (2005) et de Benalal, Del Hoyo, Landau, Roman et Skudelny (2004).

## CHAPITRE II

### ANALYSE THÉORIQUE

Ce chapitre propose un modèle théorique basé sur le travail de Lutkepohl et de Granger et Newbold avec des rapprochements avec le travail d'Hendry et Hubric. L'objectif est d'unir le travail de ces trois auteurs dans un seul et même cadre théorique.

Pour ce faire, nous cherchons à comparer les prévisions de modèle agrégé et désagrégé. Nous débuterons par spécifier un modèle général. Ensuite, nous calculerons l'erreur quadratique moyenne des prévisions sous différentes hypothèses en utilisant les pénalités d'estimations calculées par Reinsel. Les résultats obtenus sont comparés au travail d'Hendry et Hubric. Nous prenons ensuite en considération la théorie de Granger et Newbold et l'impact des racines communes sur l'erreur quadratique moyenne des prévisions du modèle agrégé et désagrégé. Nous terminons ce chapitre en résumant les implications de notre modèle théorique.

#### 2.1 Élaboration du modèle

Nous considérons le modèle suivant:

$$Y_t = Y_{t-1}G_1 + Y_{t-2}G_2 + \dots + Y_{t-p}G_p + \epsilon_t \quad (2.1.1)$$

où  $Y_t$  est de dimension  $t \times I$  et  $G_p$  est un scalaire. L'espérance des erreurs est nulle  $E(\epsilon_t) = 0$ , la matrice de variance-covariance des erreurs se définit comme  $E(\epsilon_t \epsilon_t') = \Sigma_\epsilon$  et  $E(\epsilon_t \epsilon_{t-i}') = 0$ .

L'équation (2.1.1) peut se réécrire de la manière suivante :

$$Y_t - Y_{t-1}G_1 - Y_{t-2}G_2 - \dots - Y_{t-p}G_p = \epsilon_t \quad (2.1.2)$$

Il est alors possible de simplifier cette équation en utilisant la fonction  $G(L)$ , celle-ci étant composée des paramètres du modèle et d'opérateurs de retard  $L$ .

$$G(L) = I - LG_1 - L^2G_2 - \dots - L^pG_p \quad (2.1.3)$$

$$G(L)Y_t = \epsilon_t \quad (2.1.4)$$

L'équation (2.1.4) peut être représentée par la représentation de Wold suivante :

$$Y_t = C(L)\epsilon_t \quad (2.1.5)$$

où  $C(L)$  est un processus de moyenne mobile infini. Cette représentation se révèle beaucoup plus aisée à manipuler.

Les vecteurs qui composent  $Y_t$  proviennent de la somme de  $m$  variables désagrégées

$$Y_t = \sum_{j=0}^m y_{j,t} \quad (2.1.6)$$

où l'ensemble des variables désagrégées peut être représenté par un VAR tel que:

$$\begin{matrix} y_t & = & y_{t-1} & g_1 & + & y_{t-2} & g_2 & + \dots & + & y_{t-p} & g_p & + & u_t \\ m \times 1 & & m \times 1 & m \times m & & m \times 1 & m \times m & & & m \times 1 & m \times m & & m \times 1 \end{matrix}$$

$$g(L) y_t = u_t \quad (2.1.7)$$

Chacune des variables désagrégées peut être transformée sous la forme de la représentation de Wold suivante :

$$y_t = c(L)u_t \quad (2.1.8)$$

De plus, l'espérance des erreurs est  $E(u_t) = 0$ , la matrice de variance-covariance des erreurs se définit comme  $E(u_t u_t') = \Omega$  et les erreurs ne sont pas autocorrélées  $E(u_t u_{t-i}') = 0$ .

Pour faciliter le développement mathématique, nous utilisons la forme matricielle suivante:

$$Y_t = A y_t \quad (2.1.9)$$

La matrice  $A$  effectue le lien entre les séries agrégées et désagrégées et n'est qu'un simple vecteur colonne composé de 1.

## 2.2 Comparaison des prévisions avec des processus connus

Nous commençons notre analyse théorique en reprenant la démonstration de Rose (1977) et Tiao et Guttman (1980). Ensuite, nous comparerons le résultat obtenu avec le modèle d'information d'Hendry et Hubric (2006) pour ainsi relier les deux modèles.

La prévision optimale des modèles agrégés et désagrégés,  $E_t(Y_{t+h})$  et  $E_t(y_{t+h})$  pour  $h$  périodes à l'avance se détermine par la fonction suivante:

$$E_t(Y_{t+h}) = \sum_{j=0}^{\infty} \Phi_{j+h} \epsilon_{t-j} \quad (2.2.1)$$

$$E_t(y_{t+h}) = \sum_{j=0}^{\infty} \phi_{j+h} u_{t-j} \quad (2.2.2)$$

où  $\Phi$  et  $\phi$  sont les coefficients de prévisions de  $Y_t$  et  $y_t$ . En pratique les mêmes coefficients s'utilisent pour décrire le modèle et effectuer des prévisions, mais il est plus facile de les distinguer aux fins de l'analyse. Cette distinction est validée par les simulations de Reinsel (1980).

L'erreur quadratique moyenne des prévisions est la mesure d'efficacité utilisée dans cette analyse. Comme Rose (1977), on s'intéresse à l'erreur de prévision du modèle agrégée qui se définit comme suit :

$$Q_h = E\{(Y_{t+h} - E_t(Y_{t+h}))(Y_{t+h} - E_t(Y_{t+h}))\} \quad (2.2.3)$$

$$\text{où } Y_{t+h} = C(L)\epsilon_{t+h} \quad (2.2.4)$$

$$Y_{t+h} = \sum_{j=0}^{\infty} C_j \epsilon_{t+h-j} \quad (2.2.5)$$

Si les éléments de l'équation 2.2.3 sont remplacés par les équations 2.2.1 et 2.2.5, nous obtenons :

$$Q_h = E\left\{\left(\sum_{j=0}^{\infty} C_j \epsilon_{t+h-j} - \sum_{j=0}^{\infty} \Phi_{j+h} \epsilon_{t-j}\right)\left(\sum_{j=0}^{\infty} C_j \epsilon_{t+h-j} - \sum_{j=0}^{\infty} \Phi_{j+h} \epsilon_{t-j}\right)\right\} \quad (2.2.6)$$

Le premier terme de l'équation peut être séparé en deux sommations,

$$Q_h = E\left\{\left(\sum_{j=0}^{h-1} C_j \epsilon_{t+h-j} + \sum_{j=0}^{\infty} C_{j+h} \epsilon_{t-j} - \sum_{j=0}^{\infty} \Phi_{j+h} \epsilon_{t-j}\right)\left(\sum_{j=0}^{h-1} C_j \epsilon_{t+h-j} + \sum_{j=0}^{\infty} C_{j+h} \epsilon_{t-j} - \sum_{j=0}^{\infty} \Phi_{j+h} \epsilon_{t-j}\right)\right\} \quad (2.2.7)$$

$$Q_h = E\left\{\left(\sum_{j=0}^{h-1} C_j \epsilon_{t+h-j} + \sum_{j=0}^{\infty} (C_{j+h} - \Phi_{j+h}) \epsilon_{t-j}\right)\left(\sum_{j=0}^{h-1} C_j \epsilon_{t+h-j} + \sum_{j=0}^{\infty} (C_{j+h} - \Phi_{j+h}) \epsilon_{t-j}\right)\right\} \quad (2.2.8)$$

En développant l'équation 2.2.8, nous avons

$$\begin{aligned} Q_h = & E\left(\sum_{j=0}^{h-1} C_j \epsilon_{t+h-j}\right)\left(\sum_{j=0}^{h-1} C_j \epsilon_{t+h-j}\right) + 2E\left(\sum_{j=0}^{h-1} C_j \epsilon_{t+h-j}\right)\left(\sum_{j=0}^{\infty} (C_{j+h} - \Phi_{j+h}) \epsilon_{t-j}\right) \\ & + E\left(\sum_{j=0}^{\infty} (C_{j+h} - \Phi_{j+h}) \epsilon_{t-j}\right)\left(\sum_{j=0}^{\infty} (C_{j+h} - \Phi_{j+h}) \epsilon_{t-j}\right) \quad (2.2.9) \end{aligned}$$

Sous l'hypothèse que les erreurs sont non corrélées,  $E(\epsilon_t \epsilon_{t-i}') = 0$ , le terme du centre peut être éliminé, alors 2.2.9 devient :

$$Q_h = E\left(\sum_{j=0}^{h-1} C_j \epsilon_{t+h-j}\right) \left(\sum_{j=0}^{h-1} C_j \epsilon_{t+h-j}\right)' + E\left(\sum_{j=0}^{\infty} (C_{j+h} - \Phi_{j+h}) \epsilon_{t-j}\right) \left(\sum_{j=0}^{\infty} (C_{j+h} - \Phi_{j+h}) \epsilon_{t-j}\right)'. \quad (2.2.10)$$

Les transformations effectuées jusqu'à maintenant ont permis de décortiquer l'erreur quadratique moyenne de prévision du modèle en deux parties. Le premier terme  $E\left(\sum_{j=0}^{h-1} C_j \epsilon_{t+h-j}\right) \left(\sum_{j=0}^{h-1} C_j \epsilon_{t+h-j}\right)'$  est l'erreur de prévision qui dépend seulement des paramètres du vrai modèle. Le second terme est l'erreur de prévision associée à l'utilisation de paramètre autre que ceux du vrai processus

On peut distribuer les termes de l'équation 2.2.10 et sachant que les erreurs ne sont pas autocorrélées les produits croisés s'annulent. Nous retrouvons alors les formules suivantes :

$$Q_h = E\left(\sum_{j=0}^{h-1} C_j \epsilon_{t+h-j} \epsilon_{t+h-j}' C_j'\right) + E\left(\sum_{j=0}^{\infty} (C_{j+h} - \Phi_{j+h}) \epsilon_{t-j} \epsilon_{t-j}' (C_{j+h} - \Phi_{j+h})'\right) \quad (2.2.11)$$

$$Q_h = \sum_{j=0}^{h-1} C_j E(\epsilon_{t+h-j} \epsilon_{t+h-j}') C_j' + \sum_{j=0}^{\infty} (C_{j+h} - \Phi_{j+h}) E(\epsilon_{t-j} \epsilon_{t-j}') (C_{j+h} - \Phi_{j+h})' \quad (2.2.12)$$

$$Q_h = \sum_{j=0}^{h-1} C_j \Sigma C_j' + \sum_{j=0}^{\infty} (C_{j+h} - \Phi_{j+h}) \Sigma (C_{j+h} - \Phi_{j+h})' \quad (2.2.13)$$

De toute évidence, les coefficients optimaux pour la prévision nécessitent que  $\Phi_{j+h} = C_{j+h}$ . Les meilleurs paramètres pour la prévision sont égaux à ceux du vrai processus pour que l'erreur quadratique moyenne de prévision soit minimale. Par conséquent, l'erreur quadratique moyenne de prévision pour le modèle agrégé devient :

$$Q_h = \sum_{j=0}^{h-1} C_j \Sigma C_j'. \quad (2.2.14)$$

comme dans la démonstration de Rose(1977). L'erreur quadratique moyenne des séries agrégées peut aussi se définir comme étant la somme des prévisions désagrégées. Donc, en prenant l'équation 2.2.3 et en y insérant la relation 2.1.9, on obtient le point de départ de Rose (1977) :

$$q_h = E\left\{\left(A y_{t+h} - A E_t(y_{t+h})\right) \left(A y_{t+h} - A E_t(y_{t+h})\right)'\right\} \quad (2.2.15)$$

$$\text{où } E_t(y_{t,h}) = \sum_{j=0}^{\infty} \phi_{j+h} u_{t-j} \quad \text{et} \quad y_{t+h} = \sum_{j=0}^{\infty} c_j u_{t+h-j}. \quad (2.2.16)$$

Ainsi, l'équation 2.2.15 peut se reformuler de la manière suivante selon Rose(1977) :

$$q_h = E \left\{ \left( A \sum_{j=0}^{\infty} c_j u_{t+h-j} - A \sum_{j=0}^{\infty} \phi_{j+h} u_{t-j} \right) \left( A \sum_{j=0}^{\infty} c_j u_{t+h-j} - A \sum_{j=0}^{\infty} \phi_{j+h} u_{t-j} \right) \right\}. \quad (2.2.17)$$

Si nous divisons la première sommation et qu'ensuite nous la regroupons avec la seconde sommation nous permet d'écrire :

$$q_h = E \left\{ \left( A \sum_{j=0}^{h-1} c_j u_{t+h-j} + A \sum_{j=0}^{\infty} (c_{j+h} - \phi_{j+h}) u_{t-j} \right) \left( A \sum_{j=0}^{h-1} c_j u_{t+h-j} + A \sum_{j=0}^{\infty} (c_{j+h} - \phi_{j+h}) u_{t-j} \right) \right\}. \quad (2.2.18)$$

Si les erreurs sont non corrélées,  $E(u_j u'_{j-h}) = 0 \forall h \geq 1$ , alors l'équation 2.2.18 peut atteindre une forme similaire à l'équation 2.2.12.

$$q_h = \sum_{j=0}^{h-1} A c_j E(u_{t+h-j} u'_{t+h-j}) c_j A' + \sum_{j=0}^{\infty} A (c_{j+h} - \phi_{j+h}) E(u_{t-j} u'_{t-j}) (c_{j+h} - \phi_{j+h}) A'. \quad (2.2.19)$$

Finalement, si nous considérons que  $\phi_{j+h} = c_{j+h}$  et que  $E(u_t u'_t) = \Omega$  nous trouvons que

$$q_h = \sum_{j=0}^{h-1} A c_j \Omega c_j A'. \quad (2.2.21)$$

comme l'a démontré Rose(1977). Maintenant, il faut déterminer si les prévisions désagrégées sont meilleures que les prévisions agrégées en tenant compte de l'hypothèse que les valeurs de  $C(L)$  et  $c(L)$  sont connues. Pour se faire, il faut calculer la différence entre l'erreur quadratique moyenne des prévisions des modèles agrégées et désagrégées que Rose(1977) définit comme.

$$\Delta Q_h = Q_h - q_h. \quad (2.2.21)$$

Si  $\Delta Q_h$  est positif, alors il est préférable d'utiliser le modèle désagrégé pour la prévision et vice versa. En remplaçant  $Q_h$  et  $q_h$  de 2.2.21 par les valeurs des équations 2.2.14 et 2.2.21, Rose(1977) déduit que

$$\Delta Q_h = \sum_{j=0}^{h-1} C_j \Sigma C_j - \sum_{j=0}^{h-1} A c_j \Omega c_j A'. \quad (2.2.22)$$

Pour simplifier la comparaison entre les deux erreurs quadratiques moyennes, la prévision d'une période à l'avance est utilisée.

$$\Delta Q_{h=1} = Q_{h=1} - q_{h=1} = C_0 \Sigma C_0 - A c_0 \Omega c_0 A'. \quad (2.2.23)$$

Puisque le théorème de Wold spécifie que  $c(0) = I_{mk}$  et  $C(0) = I_k$ . Alors, nous pouvons réécrire 2.2.23 comme

$$\Delta Q_1 = Q_1 - q_1 = \Sigma - A\Omega A' \quad (2.2.24)$$

ce qui est le résultat obtenu par Rose(1977) et Lutkepohl(1984). En suivant le travail de ces deux auteurs, on peut poursuivre la démonstration. En effet, selon la définition des erreurs on sait que

$$Q_1 = \Sigma = E(\epsilon_t \epsilon_t'). \quad (2.2.23)$$

En utilisant l'équation 2.1.9 et en remplaçant les variations de  $Y_t$  et  $y_t$  par les équations 2.1.5 et 2.1.8, Rose(1977) et Lutkepohl(1984) montre que

$$C(L)\epsilon_t = Ac(L)u_t \quad (2.2.24)$$

qui peut être reformulé de la manière suivante:

$$\epsilon_t = C^{-1}(L)Ac(L)u_t = \frac{Ac(L)}{C(L)}u_t. \quad (2.2.25)$$

Définissons maintenant  $\chi(L)$  qui est égal à  $\frac{c(L)}{C(L)}$ . Nous trouvons alors comme

Rose(1977) :

$$\epsilon_t = A\chi(L)u_t = \sum_{i=0}^{\infty} A\chi_i u_{t-i} = u_t + A\chi_1 u_{t-1} + A\chi_2 u_{t-2} + \dots + A\chi_{\infty} u_{t-\infty}.$$

Si nous substituons 2.2.25 dans 2.2.23 on trouve comme Rose(1977) que:

$$Q_1 = \Sigma = E(\epsilon_t \epsilon_t') = E[(Au_t + A\chi_1 u_{t-1} + \dots + A\chi_{\infty} u_{t-\infty})(Au_t + A\chi_1 u_{t-1} + \dots + A\chi_{\infty} u_{t-\infty})'] \quad (2.2.26)$$

$$Q_1 = E[(Au_t + A\chi_1 u_{t-1} + \dots + A\chi_{\infty} u_{t-\infty}) \sum_{i=0}^{\infty} u_{t-i}' \chi_i' A']$$

$$Q_1 = E[Au_t \sum_{i=0}^{\infty} u_{t-i}' \chi_i' A' + A\chi_1 u_{t-1} \sum_{i=0}^{\infty} u_{t-i}' \chi_i' A' + \dots + A\chi_{\infty} u_{t-\infty} \sum_{i=0}^{\infty} u_{t-i}' \chi_i' A']$$

$$Q_1 = E[Au_t \sum_{i=0}^{\infty} u_{t-i}' \chi_i' A'] + E[A\chi_1 u_{t-1} \sum_{i=0}^{\infty} u_{t-i}' \chi_i' A'] + \dots + E[A\chi_{\infty} u_{t-\infty} \sum_{i=0}^{\infty} u_{t-i}' \chi_i' A'] \quad (2.2.27)$$

Si les erreurs sont non corrélées,  $E(u_j u_{j-h}') = 0 \forall h \geq 1$ , alors l'équation (2.2.27) se simplifie à :

$$Q_1 = AE[u_t u_t'] A' + A\chi_1 E[u_{t-1} u_{t-1}'] \chi_1' A' + \dots + A\chi_{\infty} E[u_{t-\infty} u_{t-\infty}'] \chi_{\infty}' A'^{-1}$$

$$Q_1 = A\sigma A' + A\chi_1 \sigma \chi_1' A' + \dots + A\chi_{\infty} \sigma \chi_{\infty}' A'$$

$$Q_1 = \sum_{i=0}^{\infty} A\chi_i \sigma \chi_i' A' \quad (2.2.28)$$



Ce résultat est similaire au résultat théorique de Rose(1977). Nous pouvons réécrire l'équation 2.2.28 de la façon suivante:

$$Q_1 = A\sigma A' + \sum_{i=1}^{\infty} AX_i\sigma X_i'A' = A\sigma A' + W \quad (2.2.29)$$

où  $W = \sum_{i=1}^{\infty} AX_i\sigma X_i'A'$ . Nous pouvons faire un rapprochement entre l'équation 2.2.29 et le modèle d'information d'Hendry et Hubric(2006). Ils ont démontré que l'erreur de prévision d'un modèle qui avait moins d'information  $E(ee')$  pouvait s'écrire comme :

$$E(ee') = E(uu') + E(ww') \quad (2.2.30)$$

où  $E(uu')$  est l'erreur de prévision d'un modèle à information complète et  $E(ww')$  est l'erreur de prévision incombé par le manque d'information. Nous pouvons aisément faire un parallèle avec l'équation 2.2.29 et ainsi trouver une nouvelle signification au résultat de Rose (1977). L'erreur de prévision du modèle à information complète est alors donnée par  $A\sigma A'$  et  $W$  représente donc le coût supplémentaire incubé par l'utilisation d'un modèle agrégé.

En insérant 2.2.29 dans l'équation 2.2.24, nous nous retrouvons avec une nouvelle formulation :

$$\Delta Q_1 = Q_1 - q_1 = A\sigma A' + W - A\sigma A'. \quad (2.2.31)$$

Finalement, nous pouvons déterminer que :

$$\Delta Q_1 = W = \sum_{i=1}^{\infty} AX_i\sigma X_i'A'. \quad (2.2.32)$$

La différence entre les erreurs quadratiques de prévisions est positive et le modèle désagrégé se révèle ainsi le meilleur outil de prévision comme le prévoit le modèle de Rose(1977) et d'Hendry et Hubric(2006). Notons toutefois que cette affirmation n'est valide que lorsque les valeurs de  $C(L)$  et  $c(L)$  sont connues.

Par ailleurs, plusieurs auteurs (Rose, 1977; Clark, 2000; Hendry et Hubric, 2006 ; Pesaran, 2003) ont démontré que la différence entre l'erreur de prévision d'une représentation agrégée et désagrégée est égale à zéro si :

$$Ac(L) = C(L)A. \quad (2.2.33)$$

Cette condition est différente pour les VAR, comme le spécifient Clark (2000) et Khon (1982). Cette équation implique que la somme des coefficients du VAR désagrégé soit égale aux paramètres du modèle agrégé. Nous pouvons aisément le démontrer en suivant le raisonnement de Khon (1982):

$$A \cdot c(L) = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix} \cdot \begin{bmatrix} c(L)_{1,1} & c(L)_{1,2} & \dots & c(L)_{1,m} \\ c(L)_{2,1} & c(L)_{2,2} & \dots & c(L)_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ c(L)_{m,1} & c(L)_{m,2} & \dots & c(L)_{m,m} \end{bmatrix}$$

$$A \cdot c(L) = \begin{bmatrix} \sum_{j=1}^m c(L)_{j,1} & \sum_{j=1}^m c(L)_{j,2} & \dots & \sum_{j=1}^m c(L)_{j,m} \end{bmatrix}$$

on a aussi que  $C(L) \cdot A = \begin{bmatrix} C(L) & C(L) & \dots & C(L) \end{bmatrix}$  et nous pouvons donc écrire :

$$C(L) = \sum_{j=1}^m c(L)_{j,i} \forall i=1 \dots m \quad (2.2.34)$$

L'équation 2.2.34 sera fort utile lors des simulations, car elle permettra de tester l'importance des racines communes. Cette condition est simplifiée s'il n'y a aucun lien entre les variables désagrégées, car nous pouvons alors écrire :

$$A \cdot c(L) = \begin{bmatrix} c(L)_1 & c(L)_2 & \dots & c(L)_m \end{bmatrix} = \begin{bmatrix} C(L) & C(L) & \dots & C(L) \end{bmatrix}$$

ce qui implique que chaque équation désagrégée indépendante suit le même processus autorégressif que le modèle agrégé. Pour la présente démonstration, l'équation 2.2.33 implique que :

$$Q_1 = C^{-1}(L) A c(L) \sigma c'(L) A' C^{-1}(L) = C^{-1}(L) C(L) A \sigma A' C(L) C^{-1}(L)$$

$$Q_1 = A \sigma A' \quad (2.2.35)$$

ce qui est le résultat obtenu par Rose (1977) et Lukepohl (1984). En insérant ce nouveau résultat dans l'équation 2.2.24, nous découvrons comme ces deux auteurs que

$$\Delta Q_1 = Q_1 - q_1 = A \sigma A' - A \sigma A' = 0 \quad (2.2.34)$$

Donc, si la somme des coefficients des modèles désagrégés est égale aux coefficients du modèle agrégé, ce dernier prévoit aussi bien que le modèle désagrégé selon la relation décrite par Rose (1977). Il n'existe alors aucun gain à utiliser le modèle désagrégé.

Le travail de Rose (1977) et de plusieurs autres auteurs (Clark, 2000; Hendry et Hubric, 2006 ; Pesaran, 2003) s'arrête cependant ici. La condition pour que le modèle agrégé ait une meilleure erreur de prévision que le modèle désagrégé est très stricte pour ces chercheurs.

En résumé, lorsque les processus sont connus, il est toujours préférable d'utiliser le modèle désagrégé sauf si les coefficients des modèles agrégés et désagrégés sont égaux. Ces résultats sont conformes à la littérature. Il faut maintenant lier ces premiers résultats avec le modèle de Lutkepohl(1984) en tenant compte des problèmes de l'erreur de prévision provoqués par l'utilisation de modèles estimés (Reinsel, 1980).

### 2.3 Comparaison des prévisions avec des processus inconnus et estimés

Nous étudions maintenant les prévisions effectuées à l'aide de paramètres estimés en reprenant la démonstration de Reinsel (1980). Nous nous assurons ainsi que l'utilisation par Lutkepohl (1984) des erreurs quadratiques moyennes de prévisions asymptotiques pour la prévision de la série agrégée à l'aide d'une représentation désagrégée est valable. Nous reprenons ensuite la démonstration de Lutkepohl (1984) avec un modèle plus général tout en reliant nos résultats avec le modèle d'information d'Hendry et Hubric (2006).

Pour ce faire, nous cherchons à comparer l'erreur quadratique moyenne de prévision asymptotique, désigné par  $q_h^a$  pour le modèle désagrégé et  $Q_h^a$  pour le modèle agrégé. Nous commençons par chercher l'erreur quadratique moyenne de prévisions asymptotiques du modèle désagrégé qui s'écrit :

$$q_h^a = E \{ (A y_{t+h} - A E_t(\hat{y}_{t+h})) (A y_{t+h} - A E_t(\hat{y}_{t+h}))' \} \quad (2.3.1)$$

où  $\hat{y}_{t+h}$  est la prévision effectuée à partir d'un modèle estimé d'un échantillon de taille T. La démonstration suivante est inspirée des travaux effectués par Yamamoto (1980, 1981), Reinsel (1980) et Lewis et Reinsel (1985). La plus grande distinction entre leurs travaux et la démonstration suivante est l'introduction de la matrice  $A$  qui permet d'additionner les variables désagrégées.

Nous nous intéressons à l'ensemble des variables désagrégées décrites plus haut et qui peuvent être représentées par un VAR(p) tel que:

$$y_t = g_1 y_{t-1} + g_2 y_{t-2} + \dots + g_p y_{t-p} + u_t \quad (2.3.2)$$

Considérons maintenant  $Y_t$ , une matrice qui se compose des variables désagrégées  $y_t$  à  $y_{t-p+1}$  de sorte que

$$Y_t = \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p+1} \end{bmatrix} \quad (2.3.3)$$

et définissons les matrices des coefficients et des erreurs suivantes:

$$\Gamma = \begin{bmatrix} \gamma_1 & \gamma_2 & \cdots & \gamma_{t-p+2} & \gamma_{t-p+1} \\ I & 0 & \cdots & \cdots & 0 \\ 0 & I & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \cdots & \vdots \\ 0 & \vdots & \cdots & I & 0 \end{bmatrix} \quad \omega_t = \begin{bmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (2.3.4)$$

Cette représentation espace d'état est fort utile et nous permet de calculer  $Y_t$  en  $t+h$  comme:

$$Y_{t+h} = \Gamma^h Y_t + \sum_{j=0}^{h-1} \Gamma^j \omega_{t+h-j} \quad (2.3.5)$$

car l'on peut rapidement montrer que pour  $h = 2$  l'équation 2.3.5 s'écrit alors:

$$Y_{t+2} = \Gamma Y_{t+1} + \omega_{t+2} = \Gamma (\Gamma Y_t + \omega_{t+1}) + \omega_{t+2} = \Gamma^2 Y_t + \sum_{j=0}^1 \Gamma^j \omega_{t+2-j}$$

Nous introduisons maintenant comme Reinsel (1980) une matrice

$S_t = \begin{bmatrix} I & 0 & \cdots & 0 \end{bmatrix}$  qui permet de formuler les trois identités suivantes:

$$S Y_t = y_t \quad (2.3.6)$$

$$S \omega_t = u_t \quad (2.3.7)$$

$$S \Gamma_t = \begin{bmatrix} \gamma_1 & \gamma_2 & \cdots & \gamma_{t-p+2} & \gamma_{t-p+1} \end{bmatrix} \quad (2.3.8)$$

La matrice  $S$  nous permet donc de transformer la valeur dans l'équation 2.3.3 pour obtenir les données désagrégées  $y_{t+h}$ .

$$\begin{aligned}
 SY_{t+h} &= S\Gamma^h Y_t + \sum_{j=0}^{h-1} S\Gamma^j \omega_{t+h-j} \\
 y_{t+h} &= S\Gamma^h Y_t + \sum_{j=0}^{h-1} S\Gamma^j \omega_{t+h-j} \quad (2.3.9)
 \end{aligned}$$

L'équation 2.3.9 détermine la valeur de  $y_{t+h}$  par rapport à  $Y_t$  et d'une somme pondérée des erreurs  $\omega_t$ . Puisque  $\omega$  est inconnu, la prévision hors échantillon pour la période  $t+h$  est donnée par:

$$E_t(y_{t,h}) = S\Gamma^h Y_t \quad (2.3.10)$$

ce qui est le départ de l'analyse de Reinsel (1980). Mais contrairement à ce dernier, nous nous intéressons à la forme agrégée de l'équation 2.3.9 et 2.3.10 :

$$A y_{t+h} = A S \Gamma^h Y_t + \sum_{j=0}^{h-1} A S \Gamma^j \omega_{t+h-j} \quad (2.3.11)$$

$$A E_t(y_{t,h}) = A S \Gamma^h Y_t \quad (2.3.12)$$

Pour simplifier notre démonstration, nous posons que  $\alpha^h = A S \Gamma^h$ , alors nous pouvons réécrire l'équation 2.3.11 :

$$Y_{t+h} = \alpha^h Y_t + \sum_{j=0}^{h-1} \alpha^j \omega_{t+h-j}$$

Cette équation est similaire à celle obtenue par Reinsel (1980) et nous permet de poursuivre les étapes de sa démonstration. De la même façon, on peut réécrire la prévision de l'équation 2.3.12 comme:

$$A E_t(y_{t,h}) = E_t(Y_{t,h}) = \alpha^h Y_t \quad (2.3.13)$$

Pour simplifier notre démonstration comparativement à celle de Reinsel(1980) on s'intéresse à  $A S \Gamma^h$  si  $h=1$  :

$$\alpha = A S \Gamma$$

où  $\alpha$  est un vecteur ligne qui contient les coefficients des modèles désagrégés. On cherche maintenant l'estimateur de  $\alpha$  et sa loi. Si on reprend l'équation 2.3.2

$Y_{t'} = \Gamma \cdot Y_{t-1} + \omega_{t'}$  et que l'on prémultiplie par les matrices  $A S$  nous obtenons :

$$A S Y_{t'} = A S \Gamma \cdot Y_{t-1} + A S \omega_{t'} \quad (2.3.14)$$

$$Y_t = \alpha Y_{t-1} + A u_t \quad (2.3.15)$$

En supposant comme Reinsel (1980) que les termes d'erreurs suivent une loi normale, nous pouvons écrire la vraisemblance conditionnelle de  $Y_t$  comme:

$$Y_t | Y_{t-1} \dots \sim N(\alpha Y_{t-1}, A \Omega A')$$

ce qui permet d'écrire le logarithme de la vraisemblance de l'équation 2.3.15 de la façon suivante:

$$\begin{aligned} \mathcal{L}(\theta) = & -(Tk/2) \log(2 \cdot \pi) + T/2 \log |(A \sigma A')^{-1}| \\ & - 1/2 \sum_{t=1}^T [(Y_t - \alpha Y_{t-1})' (A \sigma A')^{-1} (Y_t - \alpha Y_{t-1})] \end{aligned}$$

où  $\theta = (\alpha', A \Omega A')$ . La dérivée de la fonction de vraisemblance pour  $\alpha$  est donnée par:

$$\partial \mathcal{L}(\theta) / \partial \alpha = \sum_{t=1}^T [Y_t' (A \sigma A')^{-1} Y_{t-1} - \alpha' Y_{t-1}' (A \sigma A')^{-1} Y_{t-1}] = 0$$

$$\partial^2 \mathcal{L}(\theta) / \partial \alpha \partial \alpha' = \sum_{t=1}^T [-Y_{t-1}' (A \sigma A')^{-1} Y_{t-1}]$$

En insérant la seconde dérivée dans la matrice de variance-covariance, on peut écrire la loi de  $\hat{\alpha}$  :

$$\sqrt{T}(\hat{\alpha} - \alpha) \sim N(0, A \Omega A' [E(Y_t Y_t')]^{-1}) \quad (2.3.16)$$

Cette loi permet de calculer l'erreur quadratique moyenne de prévision asymptotique de l'équation 2.3.1:

$$q_h^{\alpha} = E \{ (A y_{t+h} - A E_t(\hat{y}_{t+h})) (A y_{t+h} - A E_t(\hat{y}_{t+h}))' \} \quad (2.3.17)$$

Si nous insérons dans l'équation 2.3.17 les équations 2.3.11 et 2.3.13 avec un vecteur  $\alpha$  estimé selon la vraisemblance décrite plus haut on obtient comme Reinsel (1980):

$$\begin{aligned} q_h^{\alpha} = & E \{ [(\alpha^h Y_t + \sum_{j=0}^{h-1} \alpha^j \omega_{t+h-j} - \hat{\alpha}^h Y_t) [(\alpha^h Y_t + \sum_{j=0}^{h-1} \alpha^j \omega_{t+h-j} - \hat{\alpha}^h Y_t)'] \} \\ q_h^{\alpha} = & E \{ [(\alpha^h - \hat{\alpha}^h) Y_t + \sum_{j=0}^{h-1} \alpha^j \omega_{t,h-j}] [(\alpha^h - \hat{\alpha}^h) Y_t + \sum_{j=0}^{h-1} \alpha^j \omega_{t,h-j}]' \} \quad (2.3.18) \end{aligned}$$

Si nous développons l'équation 2.3.18 nous trouvons comme Reinsel (1980) que:

$$q_h^a = E\{(\alpha^h - \hat{\alpha}^h)Y_t Y_t' (\alpha^h - \hat{\alpha}^h) + 2(\alpha^h - \hat{\alpha}^h)Y_t \sum_{j=0}^{h-1} \omega_{t+h-j}(\alpha^j) + \sum_{j=0}^{h-1} \alpha^j \omega_{t+h-j} \omega_{t+h-j}'(\alpha^j)\}$$

Nous distribuons l'opérateur d'espérance et nous considérons la relation suivante:

$$E(\omega_{t+h-j}) = E \begin{bmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} E(u_t) \\ 0 \\ \vdots \\ 0 \end{bmatrix} = 0$$

nous pouvons alors reformuler l'équation 2.3.18:

$$q_h^a = E\{(\alpha^h - \hat{\alpha}^h)Y_t Y_t' (\alpha^h - \hat{\alpha}^h)\} + E\left\{\sum_{j=0}^{h-1} \alpha^j \omega_{t+h-j} \omega_{t+h-j}'(\alpha^j)\right\} \quad (2.3.19)$$

ce qui suit la démonstration de Reinsel(1980). Pour simplifier notre démonstration par rapport à celle de Reinsel (1980) nous allons remplacer le deuxième terme de l'équation 2.3.19. En effet, le terme  $E\{\sum_{j=0}^{h-1} \alpha^j \omega_{t+h-j} \omega_{t+h-j}'(\alpha^j)\}$  peut être remplacé par  $\sum_{j=0}^{h-1} A c_j \Omega c_j' A'$ . En effet, si  $\hat{\alpha} = \alpha$  l'équation 2.3.19 devient:

$$q_h^a = E\left\{\sum_{j=0}^{h-1} \alpha^j \omega_{t+h-j} \omega_{t+h-j}'(\alpha^j)\right\}$$

Puisque les coefficients sont connus, l'erreur quadratique moyenne de prévision asymptotique est alors similaire à l'erreur quadratique moyenne de prévision et par conséquent  $q_h^a = q_h$ .

On peut alors écrire que:

$$E\left\{\sum_{j=0}^{h-1} \alpha^j \omega_{t+h-j} \omega_{t+h-j}'(\alpha^j)\right\} = \sum_{j=0}^{h-1} A c_j \Omega c_j' A' \quad (2.3.20)$$

et l'équation 2.3.19 se reformule comme

$$q_h^a = E\{(\alpha^h - \hat{\alpha}^h)Y_t Y_t' (\alpha^h - \hat{\alpha}^h)\} + \sum_{j=0}^{h-1} A c_j \Omega c_j' A' \quad (2.3.21)$$

Pour simplifier la démonstration, nous considérons le cas où  $h=1$ . L'équation 2.3.21 est alors égale à :

$$q_1^a = E\{(\alpha - \hat{\alpha})Y_t Y_t' (\alpha - \hat{\alpha})\} + A \Omega A'$$

que nous pouvons réécrire comme étant :

$$q_1^a = E\{(\hat{\alpha} - \alpha)Y_t Y_t' (\hat{\alpha} - \alpha)\} + A \Omega A'$$

ce qui est forme simplifiée des résultats de Reinsel (1980). L'équation

$E\{(\hat{\alpha} - \alpha)Y_t Y_t' (\hat{\alpha} - \alpha)'\}$  est un scalaire, nous pouvons donc utiliser l'opérateur trace  $tr$  tout comme Reinsel (1980);

$$\begin{aligned} q_1^a &= tr(E\{(\hat{\alpha} - \alpha)Y_t Y_t' (\hat{\alpha} - \alpha)'\}) + A \Omega A' \\ q_1^a &= tr(E\{(\hat{\alpha} - \alpha)'(\hat{\alpha} - \alpha)Y_t Y_t'\}) + A \Omega A' \\ q_1^a &= tr(E_Y E\{(\hat{\alpha} - \alpha)'(\hat{\alpha} - \alpha)Y_t Y_t'\}) + A \Omega A' \\ q_1^a &= tr(E_Y \{E[(\hat{\alpha} - \alpha)'(\hat{\alpha} - \alpha)]Y_t Y_t'\}) + A \Omega A' \quad (2.3.22) \end{aligned}$$

Dans l'équation 2.3.22, on remplace la variance de  $\hat{\alpha}$  par l'équation 2.3.16 pour obtenir:

$$\begin{aligned} q_1^a &= tr(E_Y \{1/T A \Omega A' [E(Y_t Y_t')]^{-1} Y_t Y_t'\}) + A \Omega A' \\ q_1^a &= 1/T A \Omega A' tr(E_Y \{Y_t Y_t' [E(Y_t Y_t')]^{-1}\}) + A \Omega A' ; \end{aligned}$$

on a que  $\underbrace{E(Y_t Y_t')}_{(pm \times pm)} \underbrace{(E(Y_t Y_t'))^{-1}}_{(pm \times pm)} = \underbrace{I}_{pm \times pm}$ , ce qui nous permet d'écrire

$$\begin{aligned} q_1^a &= 1/T A \Omega A' tr(I_{pm}) + A \Omega A' \\ q_1^a &= pm/T A \Omega A' + A \Omega A' \\ q_1^a &= (1 + pm/T) A \Omega A' \quad (2.3.23) \end{aligned}$$

ce qui est conforme aux résultats obtenus par Reinsel (1980) à l'exception de la présence des matrices d'agrégation A. L'équation 2.3.23 est donc l'erreur de prévision effectuée à l'aide d'un modèle désagrégé de la série agrégée lorsque les coefficients sont estimés et que  $h=1$ . L'équation 2.3.23 est exactement l'erreur quadratique moyenne de prévision asymptotique trouvée par Yamamoto(1980,1981), Reinsel(1980) et Lewis et Reinsel(1985) à l'exception de la présence de la matrice d'agrégation A. Nous validons donc l'utilisation de ce résultat par Lutkepohl (1984). Pour notre démonstration, nous avons aussi besoin de l'erreur de prévision du modèle agrégé, donc quand  $m=1$ . L'erreur de prévision du modèle agrégé s'écrit donc comme:

$$Q_1^a = (1 + p/T) \Sigma \quad (2.3.24)$$

Abandonnons maintenant la démonstration de Reinsel (1980) pour continuer la démonstration de Lutkepohl (1984). Ce dernier avait déterminé que la différence des erreurs



quadratiques moyennes de prévisions asymptotiques entre le modèle agrégé et désagrégé pour  $h=1$  est donnée par:

$$\Delta Q_1^a = Q_1^a - q_1^a$$

$$\Delta Q_1^a = (1 + p/T)\Sigma - (1 + pm/T)A\Omega A' \quad (2.3.25)$$

Il avait conclu qu'il était préférable d'utiliser le modèle agrégé. Nous pouvons cependant pousser plus loin son analyse et innover dans ce champ d'études. En effet, si on s'inspire des résultats de Rose(1977) que nous avons repris plus haut et que l'on remplace  $\Sigma$  par l'équation 2.2.29 on a que:

$$\Delta Q_1^a = (1 + p/T)(A\Omega A' + W) - (1 + pm/T)A\Omega A'$$

$$\Delta Q_1^a = (1 + p/T)W + A\Omega A'(1 + p/T - 1 - mp/T)$$

$$\Delta Q_1^a = (1 + p/T)W + A\Omega A'(p/T - mp/T) \quad (2.3.26)$$

$$\Delta Q_1^a = (1 + p/T)W + p/T A\Omega A'(1 - m) \quad (2.3.27)$$

L'équation 2.3.27 que nous venons d'élaborer est composée d'un terme positif, que nous identifions, en nous basant sur la théorie d'Hendry et Hubric (2006), comme étant la perte de précision des prévisions lorsque l'on utilise un modèle agrégé :

$$(1 + p/T)W ,$$

et d'un terme négatif qui croît avec le nombre de variables désagrégées :

$$p/T A\Omega A'(1 - m) < 0 \quad \text{si } m > 1 .$$

Nous identifions ce second terme comme étant le gain du modèle agrégé lié à l'estimation d'un nombre de paramètres plus restreint par rapport aux modèles désagrégés.

Hendry et Hubric (2006) ont déterminé que la différence entre les erreurs d'un modèle agrégé à information incomplète  $e_t$  et d'un modèle désagrégé  $v_t$  est :

$$E(e_t e_t') - E(v_t v_t') = E(w_t w_t') .$$

On remarque que l'équation 2.3.27 comporte des éléments supplémentaires. D'abord, il a été déterminé à la section 2.2 que  $W = E(w_t w_t')$  et par conséquent on observe que si le modèle est estimé, le gain d'information à la désagrégation est conditionnelle à la taille de l'échantillon  $T$  et au nombre de variable du modèle agrégé  $p$ . Plus particulièrement, si l'échantillon est très grand, il n'y aura aucun gain d'information à utiliser un modèle

désagrégé. L'élément le plus important est l'ajout du terme négatif  $p/T A \Omega A'(1-m)$  qui reflète l'idée de Lutkepohl qu'il existe un coût à estimer davantage de variables. Par conséquent, le gain d'information et de précision trouvé par Hendry et Hubric peut être annulé par ce coût potentiel. L'équation 2.3.27 est donc la synthèse des deux théories.

Selon l'équation que nous avons développée, il sera alors préférable d'utiliser un modèle désagrégé si:

$$(1+p/T)W > -p/T A \Omega A'(1-m) .$$

Il faut donc que le gain d'information soit suffisamment important pour qu'elle efface le coût d'estimer davantage de variables. Cette inéquation est cruciale et permet d'expliquer pourquoi la désagrégation peut être bénéfique, mais non dans toutes les situations. Nous développons ainsi le début d'une réponse à l'un des problèmes communément rencontrés dans ce champ d'études.

En résumé, reprendre la théorie de Lutkepohl (1984) et Reinsel (1980) nous a permis d'établir les bases nécessaires à sa combinaison avec les approches d'Hendry et Hubric (2006) et de Rose (1977). Nous avons avancé la théorie dans ce champ d'études et nous sommes maintenant en mesure de pousser davantage le développement de la théorie en nous attachant maintenant au travail de Granger et Newbold (1977).

#### 2.4 Nombre de retards variables entre le modèle agrégé et désagrégé

Jusqu'à présent, nous avons considéré que le nombre de retards du modèle agrégé et des modèles désagrégés était égal. Cependant, la théorie développée par Granger et Newbold indique plutôt que cela ne sera jamais le cas. En effet, ces deux auteurs ont démontré que :

$$\text{si } X_t \sim AR(p) , Y_t \sim AR(q) \text{ et que } Z_t = X_t + Y_t,$$

$$\text{alors } Z_t \sim ARMA(r, s)$$

$$\text{où } r \leq p+q \text{ et } s \leq \max(p, q) .$$

Si on applique cette théorie à l'agrégation de  $m$  variables avec  $p$  retard, le modèle agrégé sera:

$$X_t \sim AR(p) \quad Z_t = \sum_{j=0}^m X_j$$

alors  $Z_t \sim ARMA(r, s)$

où  $r \leq mp$  et  $s \leq p$ .

Le nombre total de paramètres  $P$  à estimer sera de  $P \leq p(m+1)$ . Si nous insérons ce résultat dans l'équation 2.3.24 et 2.3.25 nous obtenons :

$$Q_1^o = (1 + (m+1)p/T) \Sigma \quad (2.4.1)$$

$$\Delta Q_1^o = (1 + (m+1)p/T) \Sigma - (1 + pm/T) A \sigma A' \quad (2.4.2).$$

Si nous développons l'équation 2.4.2 nous trouvons que :

$$\begin{aligned} \Delta Q_1^o &= (1 + mp/T + p/T) (A \sigma A' + W) - (1 + pm/T) A \sigma A' \\ \Delta Q_1^o &= (1 + mp/T + p/T) A \sigma A' + (1 + mp/T + p/T) W - (1 + pm/T) A \sigma A' \\ \Delta Q_1^o &= (p/T) A \sigma A' + (1 + mp/T + p/T) W \quad (2.4.3). \end{aligned}$$

L'équation 2.4.3 est toujours positive, car nous avons la perte d'information du modèle agrégé  $W$  et une pénalité reliée à l'estimation de paramètres supplémentaires par le modèle agrégé comparativement aux modèles désagrégés. Par conséquent, le modèle agrégé ne sera pas optimal. Les résultats de la section 2.3 sont donc invalidés.

Nous allons à présent prendre en considération la possibilité qu'il existe  $c$  racines autorégressives communes entre les  $m$  variables du modèle désagrégé. Si on extrapole les résultats de Granger et Newbold (1977) on trouve que :

$$\begin{aligned} Z_t &\sim ARMA(r, s) \quad \text{où } r \leq m(p-c) + c \text{ et } s \leq p-c \\ \text{alors } P &\leq (m+1)p - mc \end{aligned}$$

Si nous insérons ce résultat dans l'équation 3.3.38 et que l'on développe nous trouvons que :

$$\begin{aligned} \Delta Q_1^o &= (1 + (mp - mc + p)/T) \Sigma - (1 + pm/T) A \sigma A' \\ \Delta Q_1^o &= (1 + mp/T + (p - cm)/T) A \sigma A' + (1 + (mp + p - mc)/T) W - (1 + pm/T) A \sigma A' \\ \Delta Q_1^o &= (p - cm)/T A \sigma A' + (1 + (mp + p - mc)/T) W \quad (2.4.4) \end{aligned}$$

Si la condition suivante est vraie,

$$p < mc \quad (2.4.5)$$

il est alors possible que le modèle agrégé soit plus efficace que le modèle désagrégé, car la pénalité d'estimation sera positive. En effet, si  $p = mc - k$ , où  $k$  n'est qu'une constante, le premier terme peut s'écrire comme:

$$(p - cm)/T A \sigma A' = ([cm - k - cm]/T) A \sigma A' = -k/T A \sigma A'$$

Cette condition implique que s'il existe suffisamment de racines communes par rapport au nombre de retards du VAR désagrégé, la représentation agrégée estimera une quantité plus petite de paramètres que le VAR.

Mais remplir la condition ne garantit pas que le modèle agrégé sera le meilleur indicateur. En effet, il existe toujours une perte d'information  $W$  à utiliser le modèle agrégé. Si nous reprenons l'exemple précédent, où  $p = mc - k$ , et que l'on remplace  $p$  dans le deuxième terme on trouve :

$$(1 + (mp + p - mc)/T) W = (1 + (mp - k)/T) W,$$

et l'on peut affirmer que remplir la condition 2.4.5 n'implique pas nécessairement que le modèle agrégé soit le meilleur indicateur, car il faut aussi que :

$$k/T A \sigma A' > (1 + (mp - k)/T) W.$$

Il est aussi intéressant d'examiner l'équation 3.3.40 dans le cas où toutes les racines sont communes, donc que  $p = c$ . Cependant, on ne fait pas l'hypothèse que  $A c(L) = C(L) A$  et ce faisant, on laisse la possibilité que les relations entre les variables désagrégées diffèrent. L'équation 2.4.4 devient alors :

$$\begin{aligned} \Delta Q_1^a &= ((p - pm)/T) A \sigma A' + (1 + p/T) W \\ \Delta Q_1^a &= (p(1 - m)/T) A \sigma A' + (1 + p/T) W \quad (2.4.6) \end{aligned}$$

Malgré que les processus autorégressifs sont exactement les mêmes, nous ne pouvons déterminer avec certitude lequel, d'une représentation désagrégée ou agrégée, est le meilleur indicateur.

Notre modèle théorique implique, contrairement aux résultats théoriques de Rose (1977) et de Lutkepohl (1984), qu'il est possible que la représentation désagrégée soit le meilleur indicateur même si les processus autorégressifs des variables désagrégées ne sont pas égaux. Nous démontrerons cette affirmation dans nos simulations au chapitre suivant et comparons ces résultats avec les simulations de Lutkepohl (1984) qui appuient aussi notre modèle théorique.

### 2.5 Quelles sont les innovations de notre modèle théorique?

En reprenant la démonstration de Rose (1977) et de Reinsel (1980) nous avons élaboré un modèle théorique avec une base solide en revalidant leurs résultats. De plus, nous avons profité de cette démonstration pour tisser des liens importants avec la théorie d'Hendry et Hubric (2006) et identifier le terme relié à la perte d'information du modèle agrégé. Grâce à l'idée de Lutkepohl (1984) nous avons fait progresser le modèle théorique de ce champ d'études et estimé avec plus de précision les avantages du modèle agrégé qui sont reliés à l'estimation d'un nombre plus restreint de paramètres. Finalement, à l'aide de la théorie de Granger et Newbold (1977), nous avons développé un modèle théorique plus nuancé.

Nous avons également démontré que la présence de racines autorégressives communes entre les variables désagrégées est nécessaire pour que le modèle agrégé soit le meilleur indicateur, mais qu'elle n'est pas suffisante. En effet, notre modèle théorique a démontré que même sous l'hypothèse que toutes les racines autorégressives sont communes, il est possible que le modèle désagrégé ait une erreur de prévision plus petite que le modèle agrégé.

Les conclusions de notre modèle théorique contredisent et dépassent la condition d'agrégation parfaite défendue par Rose (1977), Clark (2000), Pesaran (2003) et Hendry et Hubric (2006). Leur théorie stipule que seuls des paramètres autorégressifs identiques permettent aux modèles agrégés d'être un meilleur indicateur sauf s'il existe des erreurs de spécifications pour le modèle désagrégé. De plus, notre modèle est plus nuancé que celui de Lutkepohl (1984).

En pratique, la présence de racine autorégressive commune peut être difficile à identifier et ce phénomène devrait être relativement rare. Nous observons en effet que la majorité des études empiriques qui utilisent une représentation désagrégée voient leurs erreurs de prévisions s'améliorer. Seulement, certaines études sont confrontées avec un modèle agrégé qui est un meilleur indicateur que la représentation désagrégée. Par conséquent, les résultats des études empiriques existantes semblent supporter les résultats de notre modèle théorique et rejettent la condition d'agrégation parfaite.

Mais, le véritable test pour notre théorie est d'étudier le problème dans un environnement contrôlé. La prochaine étape consiste donc à vérifier que notre modèle théorique est valable à travers des simulations de processus autorégressifs.

## CHAPITRE III

### SIMULATION

Nous pouvons à présent valider le modèle théorique en effectuant des expériences de Monte-Carlo. La première section élabore en détail les objectifs des simulations. On s'intéresse ensuite à la sélection des retards des modèles agrégés et au nombre de simulation effectuée. Puis, nous simulons des processus générateurs de données désagrégées pour lesquelles nous estimons un modèle autorégressif désagrégé et agrégé. Nous commençons d'abord par trois AR(2) et trois VAR(2) à deux variables avec différentes combinaisons de racines communes. Nous simulons ensuite des VAR(2) et des AR(2) à trois variables pour analyser l'impact des racines communes relativement au nombre de variables désagrégées. Finalement, nous simulons deux VAR(2) à quatre variables pour étudier l'importance des racines communes lorsque l'on maintient le processus agrégé constant selon la définition de Clark (2000). Nous terminons ce chapitre par un résumé des résultats des simulations.

#### 3.1 Les liens entre les simulations, le modèle théorique et la littérature

Les simulations ont des objectifs multiples. Nous voulons d'abord vérifier certaines conclusions de la théorie développée au chapitre précédent. Le modèle théorique développé plus haut est incapable de déterminer si le modèle agrégé aura l'erreur de prévision plus faible que le modèle désagrégé même lorsque la condition  $p < mc$  est remplie. Les simulations permettront donc de savoir si la présence de racine commune peut être suffisante pour que le modèle agrégé soit le meilleur indicateur. De plus, notre modèle théorique implique que la présence de racine commune devrait normalement diminuer l'écart des erreurs de prévision entre la représentation agrégée et désagrégée et il est intéressant de vérifier ce résultat à l'aide de simulation.

Finalement, le chapitre précédent n'a pas abordé le travail de Grunfeld et Griliches (1960). Ces deux auteurs avaient déterminé qu'une corrélation positive entre les variables désagrégées était nécessaire pour que le modèle agrégé soit un meilleur indicateur que le

modèle désagrégé. Il est très aisé de vérifier les résultats de leur modèle théorique à l'aide de simulation, ce que Grunfeld et Griliches (1960) n'ont pas effectué. Pour ce faire, nous allons donc étudier l'impact sur l'écart des erreurs de prévisions entre le modèle agrégé et désagrégé lorsqu'il existe une corrélation entre les termes erreurs des variables désagrégées. De plus, nous étudierons l'importance des liens de causalité entre les variables désagrégées sur l'écart des erreurs de prévisions du modèle désagrégé et agrégé en comparant des  $AR(p)$  indépendants et un  $VAR(p)$  comme processus générateurs de données. Nous allons ainsi pouvoir vérifier l'intuition de Grunfeld et Griliches.

### 3.2 Sélection des retards et caractéristiques générales des simulations

Malgré que le modèle théorique que nous avons développé implique que nous estimons des processus  $ARMA(p,q)$ , nous avons décidé de n'estimer que des processus  $AR(p)$ . Cependant, l'ordre des processus  $AR(p)$  agrégés est déterminé par le critère d'Akaike pour s'assurer que les erreurs de la représentation agrégée soient un bruit blanc. Le nombre de retards maximal pour le modèle agrégé sera limité à 9 pour nos simulations. Finalement, l'ordre des processus désagrégés sera considéré comme connu pour les simulations.

Pour chaque expérience, on effectue 500 simulations pour un échantillon de 240 observations. De plus, 120 observations supplémentaires ont été simulées à la fin de l'échantillon d'estimation pour effectuer des prévisions hors échantillons. On s'assure ainsi de l'indépendance des coefficients estimés et des observations utilisées pour la prévision.

### 3.3 Simulations de processus à deux variables : le modèle de Grunfeld et Griliches

Dans cette section, nous étudions des processus générateurs de données désagrégées avec deux variables qui ont chacune deux racines autorégressives. Cette section s'intéresse surtout au travail de Grunfeld et Griliches (1960) car nous allons comparer les prévisions des représentations agrégées et désagrégées lorsqu'il y a de la corrélation entre les termes d'erreurs. Si la théorie de Grunfeld et Griliches (1960) est exacte, une corrélation positive entre les erreurs des deux modèles devrait mener à une diminution de la différence des erreurs de prévision entre la représentation agrégée et désagrégée. À l'inverse, une corrélation négative devrait augmenter l'écart de prévision entre les deux représentations.

Nous allons aussi comparer les prévisions de différent processus générateurs de données, soit deux AR(2) indépendants et un VAR(2) à deux variables. Selon Grunfeld et Griliches (1960), la différence entre les erreurs de prévision des deux représentations devrait être plus faible en présence d'un VAR(2).

Dans cette section, nous ne pouvons étudier la condition, car elle ne sera pas remplie dans cette section. En effet, si  $p=2$ ,  $m=2$  et qu'il y a une racine commune, donc que  $c=1$ , on aura  $mc=2*1=2$  et  $p=mc$ . Par conséquent, notre modèle théorique stipule que la présence d'une racine commune ne permet pas que l'erreur de prévision du modèle agrégé soit plus faible que l'erreur du modèle désagrégé. Nous voulons cependant vérifier si la présence de racine commune permet de diminuer l'écart entre les erreurs de prévision d'une représentation agrégée et désagrégée.

Nous allons d'abord simuler les trois processus générateurs de données suivantes. Ce sont des groupes de deux AR(2) indépendants. Le premier processus étudié est celui où toutes les racines autorégressives sont équivalentes :

$$(1+0.7L)(1-0.4L)y_t = \epsilon_t \quad (1+0.7L)(1-0.4L)x_t = \mu_t \quad (3.3.1)$$

$$\text{où } \epsilon_t \sim N(0,1) \text{ et } \mu_t \sim N(0,1) .$$

De plus, nous étudions le cas d'une racine autorégressive commune :

$$(1+0.7L)(1-0.4L)y_t = \epsilon_t \quad (1+0.7L)(1-0.6L)x_t = \mu_t \quad (3.3.2)$$

$$\text{où } \epsilon_t \sim N(0,1) \text{ et } \mu_t \sim N(0,1)$$

et le cas où toutes les racines diffèrent :

$$(1+0.7L)(1-0.4L)y_t = \epsilon_t \quad (1+0.5L)(1-0.6L)x_t = \mu_t \quad (3.3.3)$$

$$\text{où } \epsilon_t \sim N(0,1) \text{ et } \mu_t \sim N(0,1) .$$

Nous allons étudier ces processus dans trois cas particuliers. Dans le premier cas les erreurs des variables désagrégées ne sont pas corrélées entre elles,  $E(\epsilon_t, \mu_t) = 0$ . Dans le second cas, les erreurs sont corrélées positivement entre elles,  $E(\epsilon_t, \mu_t) = 0.5$ . Dans le dernier cas, les erreurs sont corrélées négativement,  $E(\epsilon_t, \mu_t) = -0.5$ .



Les résultats des prévisions de la série agrégée pour la représentation agrégée et désagrégée pour ces différents processus générateurs de données sont présentés dans le tableau 3.1.

Tableau 3.1

Comparaison des erreurs de prévisions pour deux AR(2) indépendants

Corrélation entre les erreurs	Erreurs quadratiques de prévisions moyennes		
	$E(\epsilon_i, \mu'_i)=0$	$E(\epsilon_i, \mu'_i)=0.5$	$E(\epsilon_i, \mu'_i)=-0.5$
<b>Si le pgd* est le modèle eq. 3.3.1 – avec 2 racines communes</b>			
$q_1^a$ : prévision avec 2 AR(2) indépendants	2.033	2.845	0.962
$Q_1^a$ : prévision avec un AR(p <sub>AIC**</sub> ) agrégé.	2.057	2.877	0.958
$\Delta Q_1^a = Q_1^a - q_1^a$	<b>0.024</b>	<b>0.032</b>	<b>-0.004</b>
<b>Si le pgd* est le modèle eq. 3.3.2 – avec 1 racine commune</b>			
$q_1^a$ : prévision avec 2 AR(2) indépendants	2.039	2.669	1.062
$Q_1^a$ : prévision avec un AR(p <sub>AIC**</sub> ) agrégé.	2.094	2.736	1.136
$\Delta Q_1^a = Q_1^a - q_1^a$	<b>0.055</b>	<b>0.067</b>	<b>0.074</b>
<b>Si le pgd* est le modèle eq. 3.3.3 – sans racine commune</b>			
$q_1^a$ : prévision avec 2 AR(2) indépendants	2.029	3.287	0.842
$Q_1^a$ : prévision avec un AR(p <sub>AIC**</sub> ) agrégé.	2.143	3.370	0.990
$\Delta Q_1^a = Q_1^a - q_1^a$	<b>0.114</b>	<b>0.083</b>	<b>0.148</b>

\*Processus générateur de données

\*\*Les retards sont choisis avec le critère d'Akaike pour le modèle agrégé et avec un maximum de 9 retards.

Quand toutes les racines sont communes, il existe une légère différence entre les erreurs de prévisions du modèle agrégé et désagrégé. Cette différence est possiblement attribuable à l'utilisation du critère d'Akaike pour le modèle agrégé alors que nous considérons le nombre de retards connu au niveau désagrégé. Sinon, le modèle désagrégé est toujours le meilleur outil de prévision, même en présence de racine autorégressive commune. La présence d'une racine commune (l'équation 3.3.2) semble diminuer l'écart entre les erreurs de prévisions du processus agrégé et désagrégé comparativement au modèle où toutes les racines diffèrent (équation 3.3.3). Un résultat qui est en ligne avec nos attentes et qui confirme notre modèle théorique.

Pour des AR(2) indépendants, les résultats ne semblent pas diverger de façon constante en présence de corrélation positive ou négative entre les erreurs des variables désagrégées. Par exemple, on observe que la présence de corrélation positive fait diminuer l'écart entre les prévisions lorsque toutes les racines diffèrent. Cependant, elle fait augmenter l'écart lorsque les paramètres sont égaux. Par conséquent, nous ne pouvons valider ou invalider le modèle théorique de Grunfeld et Griliches.

Nous simulons maintenant des VAR(2) à deux variables qui suivent les processus générateurs de données suivants:

$$\begin{bmatrix} (1+0.7L)(1-0.4L) & 0.2L-0.2L^2 \\ 0.2L-0.2L^2 & (1+0.7L)(1-0.4L) \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} \epsilon_t \\ \mu_t \end{bmatrix} \quad (3.3.4)$$

où  $\epsilon_t \sim N(0,1)$  et  $\mu_t \sim N(0,1)$

$$\begin{bmatrix} (1+0.7L)(1-0.4L) & 0.2L-0.2L^2 \\ 0.2L-0.2L^2 & (1+0.7L)(1-0.6L) \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} \epsilon_t \\ \mu_t \end{bmatrix} \quad (3.3.5)$$

où  $\epsilon_t \sim N(0,1)$  et  $\mu_t \sim N(0,1)$

$$\begin{bmatrix} (1+0.7L)(1-0.4L) & 0.2L-0.2L^2 \\ 0.2L-0.2L^2 & (1+0.5L)(1-0.6L) \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} \epsilon_t \\ \mu_t \end{bmatrix} \quad (3.3.6)$$

où  $\epsilon_t \sim N(0,1)$  et  $\mu_t \sim N(0,1)$

Les résultats des simulations pour ces trois modèles sont présentés dans le tableau 3.2.

Tableau 3.2

Comparaison des erreurs de prévisions pour des VAR(2) à deux variables

Corrélation entre les erreurs	Erreurs quadratiques de prévisions moyennes		
	$E(\epsilon_t, \mu_t) = 0$	$E(\epsilon_t, \mu_t) = 0.5$	$E(\epsilon_t, \mu_t) = -0.5$
<b>Si le pgd* est le modèle eq. 3.3.4 – avec 2 racines communes</b>			
$q_1^a$ : prévision avec un VAR(2)	2.048	2.610	0.919
$Q_1^a$ : prévision avec un AR(p <sub>AIC**</sub> ) agrégé.	2.049	2.595	0.917
$\Delta Q_1^a = Q_1^a - q_1^a$	<b>0.001</b>	<b>-0.015</b>	<b>-0.002</b>
<b>Si le pgd* est le modèle eq. 3.3.5 – avec 1 racine commune</b>			
$q_1^a$ : prévision avec un VAR(2)	2.046	2.607	0.919
$Q_1^a$ : prévision avec un AR(p <sub>AIC**</sub> ) agrégé.	2.070	2.623	0.966
$\Delta Q_1^a = Q_1^a - q_1^a$	<b>0.024</b>	<b>0.016</b>	<b>0.057</b>
<b>Si le pgd* est le modèle eq. 3.3.6 – sans racine commune</b>			
$q_1^a$ : prévision avec un VAR(2)	2.047	2.888	0.990
$Q_1^a$ : prévision avec un AR(p <sub>AIC**</sub> ) agrégé.	2.135	2.918	1.119
$\Delta Q_1^a = Q_1^a - q_1^a$	<b>0.088</b>	<b>0.030</b>	<b>0.129</b>

\*Processus générateur de données

\*\*Les retards sont choisis avec le critère d'Akaike pour le modèle agrégé et avec un maximum de 9 retards.

Lorsque qu'il n'y a pas de corrélation entre les erreurs des variables désagrégées, les résultats sont très similaires à ce que l'on a vu précédemment. Le VAR(2) est le meilleur modèle sauf si les racines autorégressives sont identiques, alors la représentation agrégée et désagrégée a des erreurs de prévisions similaires. Sinon, le modèle agrégé est le plus efficient. Encore une fois, la présence d'une racine autorégressive commune diminue l'écart des erreurs de prévisions entre le modèle agrégé et désagrégé comparativement au modèle ou toutes les racines diffèrent. Ces résultats demeurent en ligne avec notre modèle théorique.

Contrairement aux simulations précédentes, nous observons que la présence d'une corrélation positive entre les erreurs des variables désagrégées diminue l'écart des erreurs de prévisions moyennes de façon constante par rapport aux processus sans corrélation des erreurs. Inversement, la présence de corrélation négative entre les variables désagrégées implique un écart des erreurs de prévisions plus grand par rapport aux processus sans corrélation entre les erreurs. Le modèle théorique de Grunfeld et Griliches est donc validé par ces résultats puisque la corrélation positive entre les termes d'erreurs a un effet positif sur les gains d'agrégation et inversement pour la corrélation négative.

Si nous comparons le tableau 3.1 et 3.2, nous remarquons que la différence entre les erreurs quadratiques de prévisions moyennes des modèles agrégés et désagrégés est plus faible si les processus générateurs de données sont un VAR(2). Nous pouvons expliquer ce résultat en revenant à la théorie de Grunfeld et Griliches (1960) qui stipule qu'il doit exister de la corrélation entre les variables désagrégées pour que l'agrégation soit efficace. Puisqu'il n'existe aucun lien entre les AR(2) indépendants, le travail de Grunfeld et Griliches voudrait que le modèle désagrégé soit toujours le meilleur modèle. Donc, les résultats de nos simulations semblent confirmer leur travail.

Les résultats de nos premières simulations sont jusqu'à présent conformes à notre modèle théorique. En effet, la présence de racine commune semble diminuer l'écart des erreurs de prévisions entre la représentation agrégée et désagrégée. De plus, nos résultats appuient les travaux de Grunfeld et Griliches (1960), car ils démontrent que la présence de corrélation positive entre les termes d'erreur peut diminuer l'écart des prévisions entre le modèle agrégé et désagrégé. De plus, l'écart des erreurs de prévisions entre les deux représentations diminue lorsque le processus désagrégé est un VAR.

#### 3.4 Simulation de processus désagrégés à trois variables

Nous nous intéressons maintenant à la condition  $p < mc$  qui est nécessaire pour que la représentation agrégée ait une erreur de prévision plus faible que la représentation désagrégée. Notre modèle théorique a déterminé que cette condition n'est pas suffisante pour que la représentation agrégée soit le meilleur indicateur et nous voulons donc vérifier notre affirmation.

Puisque l'on veut que la condition  $p < mc$  soit remplie, nous devons simuler des VAR(2) à trois variables et avec une racine commune. Nous avons alors que  $p=2$ ,  $m=3$  et  $c=1$  ce qui implique effectivement que  $p < mc$ .

De plus, nous allons à nouveau comparé des processus AR(p) indépendants et des VAR(p) et ainsi vérifié si l'augmentation du nombre de variables augmente l'écart des erreurs de prévisions entre la représentation agrégée et désagrégée.

Nous allons d'abord simuler trois AR(2) indépendants. Le premier processus simulé est une représentation où toutes les racines sont communes:

$$\begin{aligned}(1+0.7L)(1-0.4L)y_t &= \epsilon_t \\ (1+0.7L)(1-0.4L)x_t &= \mu_t \\ (1+0.7L)(1-0.4L)z_t &= \omega_t\end{aligned}\quad (3.4.1)$$

où  $\epsilon_t \sim N(0,1)$ ,  $\mu_t \sim N(0,1)$  et  $\omega_t \sim N(0,1)$ .

Nous étudions le cas d'une racine autorégressive commune:

$$\begin{aligned}(1+0.7L)(1-0.4L)y_t &= \epsilon_t \\ (1+0.7L)(1-0.5L)x_t &= \mu_t \\ (1+0.7L)(1-0.6L)z_t &= \omega_t\end{aligned}\quad (3.4.2)$$

où  $\epsilon_t \sim N(0,1)$ ,  $\mu_t \sim N(0,1)$  et  $\omega_t \sim N(0,1)$ .

et le processus générateur de données où les racines diffèrent :

$$\begin{aligned}(1+0.7L)(1-0.4L)y_t &= \epsilon_t \\ (1+0.6L)(1-0.5L)x_t &= \mu_t \\ (1+0.5L)(1-0.6L)z_t &= \omega_t\end{aligned}\quad (3.4.3)$$

où  $\epsilon_t \sim N(0,1)$ ,  $\mu_t \sim N(0,1)$  et  $\omega_t \sim N(0,1)$ .

Les résultats des prévisions de la série agrégée pour la représentation agrégée et désagrégée lors des simulations se trouvent dans le tableau 3.3 pour ces différents processus générateurs de données.

Tableau 3.3

Comparaison des erreurs de prévisions pour trois AR(2) indépendants

	<b>Erreurs quadratiques de prévisions moyennes</b>
<b>Si le pgd* est le modèle eq. 3.4.1 – avec 2 racines communes</b>	
$q_1^a$ : prévision avec 3 AR(2) indépendants	3.041
$Q_1^a$ : prévision avec un AR(p <sub>AIC**</sub> ) agrégé.	3.070
$\Delta Q_1^a = Q_1^a - q_1^a$	<b>0.029</b>
<b>Si le pgd* est le modèle eq. 3.4.2 – avec 1 racine commune</b>	
$q_1^a$ : prévision avec 3 AR(2) indépendants	3.028
$Q_1^a$ : prévision avec un AR(p <sub>AIC**</sub> ) agrégé.	3.104
$\Delta Q_1^a = Q_1^a - q_1^a$	<b>0.076</b>
<b>Si le pgd* est le modèle eq. 3.4.3 – sans racine commune</b>	
$q_1^a$ : prévision avec 3 AR(2) indépendants	3.040
$Q_1^a$ : prévision avec un AR(p <sub>AIC**</sub> ) agrégé.	3.155
$\Delta Q_1^a = Q_1^a - q_1^a$	<b>0.115</b>

\*Processus générateur de données

\*\*Les retards sont choisis avec le critère d'Akaike pour le modèle agrégé  
et avec un maximum de 9 retards.

L'augmentation du nombre de variables n'a pas changé les résultats par rapport à la section précédente. Remplir la condition d'agrégation ne semble pas avoir un impact important. Le modèle désagrégé est toujours le meilleur outil de prévision, même en présence de racine autorégressive commune. Nous observons encore que la présence de racine

commune diminue l'écart des erreurs de prévision entre la représentation agrégée et désagrégée.

Nous allons maintenant simuler les VAR(2) à trois variables suivants:

$$\begin{bmatrix} (1+0.7L)(1-0.4L) & 0.2L-0.2L^2 & 0.2L-0.2L^2 \\ 0.2L-0.2L^2 & (1+0.7L)(1-0.4L) & 0.2L-0.2L^2 \\ 0.2L-0.2L^2 & 0.2L-0.2L^2 & (1+0.7L)(1-0.4L) \end{bmatrix} \begin{bmatrix} y_t \\ x_t \\ z_t \end{bmatrix} = \begin{bmatrix} \epsilon_t \\ \mu_t \\ \omega_t \end{bmatrix} \quad (3.4.4)$$

où  $\epsilon_t \sim N(0,1)$  ,  $\mu_t \sim N(0,1)$  et  $\omega_t \sim N(0,1)$

$$\begin{bmatrix} (1+0.7L)(1-0.4L) & 0.2L-0.2L^2 & 0.2L-0.2L^2 \\ 0.2L-0.2L^2 & (1+0.7L)(1-0.5L) & 0.2L-0.2L^2 \\ 0.2L-0.2L^2 & 0.2L-0.2L^2 & (1+0.7L)(1-0.6L) \end{bmatrix} \begin{bmatrix} y_t \\ x_t \\ z_t \end{bmatrix} = \begin{bmatrix} \epsilon_t \\ \mu_t \\ \omega_t \end{bmatrix} \quad (3.4.5)$$

où  $\epsilon_t \sim N(0,1)$  ,  $\mu_t \sim N(0,1)$  et  $\omega_t \sim N(0,1)$

$$\begin{bmatrix} (1+0.7L)(1-0.4L) & 0.2L-0.2L^2 & 0.2L-0.2L^2 \\ 0.2L-0.2L^2 & (1+0.6L)(1-0.5L) & 0.2L-0.2L^2 \\ 0.2L-0.2L^2 & 0.2L-0.2L^2 & (1+0.5L)(1-0.6L) \end{bmatrix} \begin{bmatrix} y_t \\ x_t \\ z_t \end{bmatrix} = \begin{bmatrix} \epsilon_t \\ \mu_t \\ \omega_t \end{bmatrix} \quad (3.4.6)$$

où  $\epsilon_t \sim N(0,1)$  ,  $\mu_t \sim N(0,1)$  et  $\omega_t \sim N(0,1)$  .

Les résultats des simulations de ces processus générateurs de données sont présenté dans le tableau 3.4.

Tableau 3.4

Comparaison des erreurs de prévisions pour des VAR(2) à trois variables

	Erreurs quadratiques de prévisions moyennes
<b>Si le pgd* est le modèle eq. 3.4.4 – avec 2 racines communes</b>	
$q_1^a$ : prévision avec un VAR(2)	3.132
$Q_1^a$ : prévision avec un AR(p <sub>AIC**</sub> ) agrégé.	3.107
$\Delta Q_1^a = Q_1^a - q_1^a$	-0.025
<b>Si le pgd* est le modèle eq. 3.4.5 – avec 1 racine commune</b>	
$q_1^a$ : prévision avec un VAR(2)	3.104
$Q_1^a$ : prévision avec un AR(p <sub>AIC**</sub> ) agrégé.	3.109
$\Delta Q_1^a = Q_1^a - q_1^a$	0.005
<b>Si le pgd* est le modèle eq. 3.4.6 – sans racine commune</b>	
$q_1^a$ : prévision avec un VAR(2)	3.107
$Q_1^a$ : prévision avec un AR(p <sub>AIC**</sub> ) agrégé.	3.174
$\Delta Q_1^a = Q_1^a - q_1^a$	0.067

\*Processus générateur de données

\*\*Les retards sont choisis avec le critère d'Akaike pour le modèle agrégé  
et avec un maximum de 9 retards.

L'erreur moyenne de prévision de la représentation agrégée pour les processus générateurs de données avec une racine commune est légèrement plus élevée que l'erreur moyenne de prévision du modèle désagrégé. Par conséquent, nos simulations démontrent que remplir la condition d'agrégation développée au chapitre suivant n'est pas suffisante pour que le modèle agrégé soit le meilleur indicateur. Nous remarquons à nouveau que la présence de racine commune permet de diminuer l'écart entre les erreurs de prévisions de la représentation agrégée et désagrégée. Ces deux résultats confirment notre modèle théorique.



Si on compare les résultats des tableaux 3.3 et 3.4, on remarque que pour les VAR(2), il existe un gain à estimé un modèle agrégé en présence de racine commune alors que ce n'est pas le cas pour les AR(2) indépendants. Par conséquent, on confirme à nouveau l'intuition de Grunfeld et Griliches (1960) : il y a une certaine corrélation entre les variables désagrégées pour que l'agrégation soit utile.

### 3.5 Simulations de processus agrégés identiques et l'impact des racines communes

Selon la théorie de l'agrégation parfaite de Khon (1980) et de Clark (2000), utiliser des racines autorégressives différentes sans altérer les relations entre les variables désagrégées modifie la représentation agrégée et les conditions d'agrégation. Le travail de ces deux auteurs implique que pour vérifier si les racines communes diminuent l'écart des erreurs de prévision entre la représentation agrégée et désagrégée nous devons nous assurer que  $Ac(L)$ , où  $c(L)$  est le processus désagrégé et  $A$  la matrice d'agrégation, demeure constante pour les processus comparés.

Nous allons donc créer deux processus qui satisfont cette condition dans l'objectif de vérifier que la présence de racine commune diminue l'écart des erreurs de prévisions de la représentation agrégée et désagrégée. Nous vérifions ainsi les implications de notre modèle théorique.

Considérons donc les deux processus suivants :

$$\begin{bmatrix} (1+0.7L)(1-0.3L) & 0.1L-0.1L^2 & 0.1L-0.1L^2 & 0.1L-0.1L^2 \\ 0.1L-0.1L^2 & (1+0.7L)(1-0.4L) & 0.1L-0.1L^2 & 0.1L-0.1L^2 \\ 0.1L-0.1L^2 & 0.1L-0.1L^2 & (1+0.7L)(1-0.5L) & 0.1L-0.1L^2 \\ 0.1L-0.1L^2 & 0.1L-0.1L^2 & 0.1L-0.1L^2 & (1+0.7L)(1-0.6L) \end{bmatrix} \begin{bmatrix} y_t \\ x_t \\ z_t \\ k_t \end{bmatrix} = \begin{bmatrix} \epsilon_t \\ \mu_t \\ \omega_t \\ \nu_t \end{bmatrix} \quad (3.5.1)$$

et

$$\begin{bmatrix} (1+0.7L)(1-0.2L) & 0.2L-0.1L^2 & -0.1L-0.1L^2 & 0.1L-0.18L^2 \\ -0.1L-0.17L^2 & (1+0.8L)(1-0.4L) & 0.1L-0.1L^2 & -0.1L-0.18L^2 \\ 0.1L-0.1L^2 & -0.1L-0.1L^2 & (1+0.9L)(1-0.5L) & 0.1L-0.18L^2 \\ 0.2L-0.1L^2 & 0.1L-0.06L^2 & 0.1L-0.05L^2 & (1+0.6L)(1-0.3L) \end{bmatrix} \begin{bmatrix} y_t \\ x_t \\ z_t \\ k_t \end{bmatrix} = \begin{bmatrix} \epsilon_t \\ \mu_t \\ \omega_t \\ \nu_t \end{bmatrix} \quad (3.5.2)$$

Si on multiplie les différentes racines on trouve pour ces deux modèles:

$$\begin{bmatrix} y_t \\ x_t \\ z_t \\ k_t \end{bmatrix} = \begin{bmatrix} 0.4 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.3 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.2 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \\ z_{t-1} \\ k_{t-1} \end{bmatrix} + \begin{bmatrix} -0.21 & -0.1 & -0.1 & -0.1 \\ -0.1 & -0.28 & -0.1 & -0.1 \\ -0.1 & -0.1 & -0.35 & -0.1 \\ -0.1 & -0.1 & -0.1 & -0.42 \end{bmatrix} \begin{bmatrix} y_{t-2} \\ x_{t-2} \\ z_{t-2} \\ k_{t-2} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ \mu_t \\ \omega_t \\ \nu_t \end{bmatrix} \quad (3.5.1a)$$

et

$$\begin{bmatrix} y_t \\ x_t \\ z_t \\ k_t \end{bmatrix} = \begin{bmatrix} 0.5 & 0.2 & -0.1 & 0.1 \\ -0.1 & 0.4 & 0.1 & -0.1 \\ 0.1 & -0.1 & 0.4 & 0.1 \\ 0.2 & 0.1 & 0.1 & 0.3 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \\ z_{t-1} \\ k_{t-1} \end{bmatrix} + \begin{bmatrix} -0.14 & -0.1 & -0.1 & -0.18 \\ -0.17 & -0.32 & -0.1 & -0.18 \\ -0.1 & -0.1 & -0.40 & -0.18 \\ -0.1 & -0.06 & -0.05 & -0.18 \end{bmatrix} \begin{bmatrix} y_{t-2} \\ x_{t-2} \\ z_{t-2} \\ k_{t-2} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ \mu_t \\ \omega_t \\ \nu_t \end{bmatrix} \quad (3.5.2a)$$

Le processus agrégé est déterminé par la somme de ces processus, si on multiplie ces deux équations par le vecteur d'agrégation A on obtient:

$$Y_t = [0.7 \quad 0.6 \quad 0.5 \quad 0.4] \begin{bmatrix} y_{t-1} \\ x_{t-1} \\ z_{t-1} \\ k_{t-1} \end{bmatrix} + [-0.51 \quad -0.58 \quad -0.65 \quad -0.72] \begin{bmatrix} y_{t-2} \\ x_{t-2} \\ z_{t-2} \\ k_{t-2} \end{bmatrix} + e_t \quad (3.5.1b)$$

et

$$Y_t = [0.7 \quad 0.6 \quad 0.5 \quad 0.4] \begin{bmatrix} y_{t-1} \\ x_{t-1} \\ z_{t-1} \\ k_{t-1} \end{bmatrix} + [-0.51 \quad -0.58 \quad -0.65 \quad -0.72] \begin{bmatrix} y_{t-2} \\ x_{t-2} \\ z_{t-2} \\ k_{t-2} \end{bmatrix} + e_t \quad (3.5.2b)$$

Ce qui démontre que les processus agrégés seront les mêmes. La comparaison des résultats de simulation de ces deux modèles est présenté dans le tableau 3.5.

Tableau 3.5

Comparaison des erreurs de prévisions pour des VAR(2) à quatre variables

	Erreurs quadratiques de prévisions moyennes
<b>Si le pgd* est le modèle eq. 3.5.1 – avec 1 racine commune</b>	
$q_1^a$ : prévision avec un VAR(2)	4.157
$Q_1^a$ : prévision avec un AR( $p_{AIC^{**}}$ ) agrégé.	4.174
$\Delta Q_1^a = Q_1^a - q_1^a$	<b>-0.025</b>
<b>Si le pgd* est le modèle eq. 3.5.2 – sans racine commune</b>	
$q_1^a$ : prévision avec un VAR(2)	4.191
$Q_1^a$ : prévision avec un AR( $p_{AIC^{**}}$ ) agrégé.	4.248
$\Delta Q_1^a = Q_1^a - q_1^a$	<b>0.005</b>

\*Processus générateur de données

\*\*Les retards sont choisis avec le critère d'Akaike pour le modèle agrégé et avec un maximum de 9 retards.

Les résultats confirment que la présence d'une racine commune améliore l'erreur de prévision du modèle agrégé et réduit l'écart des erreurs de prévision entre la représentation agrégée et désagrégée. Ce résultat est d'autant plus valable que nous avons respecté la condition d'agrégation de Khon (1980) et de Clark(1980) qui impliquerait que les deux modèles aient une erreur moyenne de prévision identique.

### 3.6 Les résultats des simulations appuient-ils notre modèle théorique?

Les résultats des simulations que nous avons effectuées sont en ligne avec les conclusions du modèle théorique que nous avons développé. Les simulations nous ont permis de vérifier trois éléments importants de notre modèle théorique :

- La représentation agrégée peut avoir une erreur de prévision plus faible que la représentation désagrégée si le nombre de racines autorégressives communes total est plus grand que le nombre de retards d'une représentation désagrégée.

- Remplir cette condition n'assure en rien que le modèle agrégé ait une erreur de prévision plus faible que le modèle désagrégé.
- Que la présence de racine commune fait diminuer l'écart des erreurs de prévisions entre le modèle agrégé et désagrégé.

De plus, nous avons déterminé que la théorie de Grunfeld et Griliches (1960) peut expliquer la disparité observée entre les résultats des VAR(p) et des AR(p) indépendants et que la présence de corrélation positive entre les termes d'erreurs diminue l'écart entre les erreurs de prévision de la représentation agrégée et désagrégée. Par conséquent, l'existence de corrélation ou de lien causal entre les variables désagrégées est importante pour obtenir une agrégation efficiente.

## CONCLUSION

Les avantages et les inconvénients de l'agrégation contemporaine dans le cadre de la prévision d'un agrégat sont méconnus des économistes, et ce, malgré le grand nombre de modèles de prévision agrégés et désagrégés qui sont utilisés en macroéconomie. Le but de ce mémoire était d'unir les différentes théories sur l'agrégation et la désagrégation contemporaine pour aider les économistes dans le choix d'un modèle de prévision.

Pour atteindre notre objectif, nous avons effectué un survol de la littérature en nous attaquant particulièrement à sept auteurs : Grunfeld, Griliches, Lutkepohl, Hendry, Hubric, Granger et Newbold. Grunfeld et Griliches (1960) ont bien cerné l'importance des corrélations entre les variables désagrégées comme une condition essentielle à l'agrégation efficiente. De plus, Lutkepohl (1984) a identifié un coût à l'agrégation d'un nombre plus élevé de paramètres pour le modèle désagrégé alors qu'Hendry et Hubric (2006) ont déterminé que la représentation désagrégée possédait cependant plus d'information comparativement à la représentation agrégée. Finalement, Granger et Newbold (1977) ont révélé que l'agrégation de processus AR(p) et la création de processus ARMA(p,q) qui en résulte a un impact plus complexe que ce que Lutkepohl (1984), Hendry et Hubric (2006) ont trouvé.

Nous avons ensuite repris les démonstrations de Rose (1977) et de Reinsel (1980) pour plus aisément incorporer les idées de ces différents auteurs dans notre modèle théorique que nous avons ensuite validé à travers plusieurs simulations.

La théorie que nous avons élaborée nous a permis de déterminer que la présence de racine autorégressive commune dans les séries désagrégées est nécessaire, mais non suffisante pour que la représentation agrégée soit un meilleur indicateur que la représentation désagrégée. De plus, nous avons déterminé que la présence de racine commune diminuait l'écart des erreurs de prévisions entre la représentation agrégée et désagrégée.

Les simulations effectuées suggèrent que l'intuition de Grunfeld et Griliches (1960) pourrait aussi expliquer la disparité entre les erreurs de prévision des modèles agrégés et

désagrégés. Nous avons observé que s'il existe de la corrélation positive entre le terme d'erreur des variables désagrégées l'écart entre les erreurs de prévision de la représentation agrégé et désagrégé diminue alors qu'elle augmente si la corrélation est négative.

Ces résultats sont un pas en avant vers une meilleure compréhension de la théorie derrière l'agrégation de séries désagrégées. En effet, les recherches antérieures n'avaient trouvé que des conditions relativement simples d'agrégation parfaite (Rose, 1977 ; Pesaran, 2003) alors que certains modèles théoriques avaient déçu lors des simulations (Lutkepohl, 1984).

Il reste cependant beaucoup de travail à effectuer dans ce domaine. En effet, il existe un test de racine autorégressive commune développé par Gouriéroux, Montford et Renault (1989) mais son efficacité en petit échantillon n'a pas encore été démontrée. De plus, il serait intéressant d'étudier les facteurs qui peuvent affecter la taille du gain d'information du modèle désagrégé en commençant avec les résultats de Grunfeld et Griliches (1960). Finalement, il serait aussi intéressant d'étudier l'impact de l'intégration fractionnaire sur l'agrégation de séries désagrégées.

## RÉFÉRENCES

- Benalal, N., et J.L.D. del Hoyo, B. Landau, M. Roman et F. Skudelny. 2004. « To Aggregate or not to Aggregate? Euro Area Inflation Forecasting ». Working Paper no 374, European Central Bank, 65 p.
- Clements, M.P. et D.F. Hendry. 1995. « Forecasting in Cointegrated Systems ». *Journal of Applied Econometrics*, vol.10, no 2 (avril-juin), p. 127-146.
- Clark, Todd E. 2000. « Forecasting an Aggregate of Cointegrated disaggregates ». *Journal of Forecasting*, vol.19, no 1 (janvier), p. 1-21.
- Demers, F., et A. De Champlain. 2005. « Forecasting Core Inflation: Should we forecast the Aggregate or the Components? Some Empirical Evidence from Canada ». Banque du Canada, document de travail 2005-44, 58 p.
- Demers, F., et D. Dupuis. 2005. « Forecasting Canadian GDP : Region-Specific versus Countrywide Information ». Banque du Canada, document de travail 2005-31, 33 p.
- Garderen, K.J.V., K. Lee, et M.H. Pesaran. 2000. « Cross-sectional aggregation of non-linear models ». *Journal of Econometrics*, vol. 95, no 2 (avril), p. 285-331.
- Gouriéroux, C., A. Montford, et E. Renault. 1989. « Testing For Common Roots ». *Econometrica*, vol.57, no 1 (Janvier), p. 171-185.
- Granger, C.W.J. 1990. « Aggregation of Time-Series Variables: A Survey ». *Disaggregation in Econometric Modelling*, sous la dir. de T. Barker et M.H. Pesaran, p. 17-34. London: Routledge.
- Granger, C.W.J. et P. Newbold. 1977. *Forecasting Economic Time Series*. New York: Academic Press.
- Grunfeld, Y. et Griliches Z. 1960. « Is Aggregation Necessaliry Bad ». *The Review of Economics and Statistics*, vol. 42, no 1 (février), p. 1-13.
- Hamilton, James D., 1994, *Time Series Analysis*, Princeton, New Jersey: Princeton University Press.
- Hendry, David F. et Hubrich K. 2006. « Forecasting Economic Aggregates by Disaggregates ». Centre for Economic Policy Research, Discussion Paper Series No. 5485, février 2006, 42 p.

- Hubrich, K. 2005. « Forecasting euro area inflation: Does aggregating forecasts by HICP component improve forecast accuracy? ». *International Journal of Forecasting*, vol. 21, no.1 (janvier-mars), p.119-136
- Khon, Robert. 1982. « When is an aggregate of a time series efficiently forecast by its past? ». *Journal of Econometrics*, vol. 18, no 3 (avril), p. 337-349.
- Lütkepohl, H. 1984. « Forecasting Contemporaneously Aggregated Vector ARMA Processes ». *Journal of Business and Economic Statistics*, vol. 2, no 3 (juillet), p. 201-214.
- Lütkepohl, H. 1986. « Forecasting Temporally Aggregated Vector Arma Processes ». *Journal of Forecasting*, vol. 5, no 2 (avril-juin), p. 85-95.
- Marcellino, M. 1999. « Some Consequences of Temporal Aggregation in Empirical Analysis ». *Journal of Business and Economic Statistics*, vol. 17, no 1 (janvier), p. 129-136.
- Marcellino, M. 2000. « Linear aggregation with common trends and cycles ». *Research in Economics*, vol. 54, no 1 (juin), p. 117-131.
- Marcellino, M., J. H. Stock, M.W. Watson 2003. « Macroeconomic forecasting in the Euro area : Country specific versus area-wide information ». *European Economic Review*, vol. 47, no 1 (Février), p. 1-18.
- Pesaran, M.H., R.G. Pierse et M.S. Kumar. 1989. « Econometric Analysis of Aggregation in the Context of Linear Prediction Models ». *Econometrica*, vol. 57, no 4 (Juillet), p. 861-888.
- Pesaran, M.H., K. Lee et R.G. Pierse. 1993. « Persistence, Cointegration and aggregation. A disaggregated analysis of output fluctuations in the U.S. economy ». *Journal of Econometrics*, vol. 56, no 1 et 2 (Mars), p. 57-88.
- Pesaran, M.H., K. Lee et R.G. Pierse. 1994. « Choice Between Disaggregate and Aggregate Specifications Estimated by Instrumental Variables Methods ». *Journal of Business and Economic Statistics* vol. 12, no. 1 (Janvier), p. 11-21.
- Pesaran, M.H. 2003. « Aggregation of linear dynamic models: an application to life-cycle consumption models under habit formation ». *Economic Modelling*, vol. 20, no 2 (Mars), p. 383-415.



- Reinsel, G. 1980. « Asymptotic Properties of Prediction Errors for Multivariate Autoregressive Model Using Estimated Parameters ». *Journal of the Royal Statistical Society, Ser. B*, vol. 42, no 3, p. 328-333
- Reinsel, G. et Richard Lewis 1985. « Asymptotic Properties of Prediction Errors for Multivariate Autoregressive Model Using Estimated Parameters ». *Journal of Multivariate Analysis*, vol. 16, no 3, p. 393-411
- Rose, D.E. 1977 « Forecasting Aggregates of Independent Arima Processes ». *Journal of Econometrics*, vol. 5, p. 323-345
- Theil, H., 1954, *Linear aggregation of economic relations*, Amsterdam: North-Holland.
- Tiao, G.C., et I. Guttman. 1980. « Forecasting Contemporaneous aggregates ». *Journal of Econometrics*, vol. 12, no 2 (février), p. 219-230.
- Tobias, J., et A. Zellner. 2000. « A Note on Aggregation, Disaggregation and Forecasting Performance ». *Journal of Forecasting*, vol. 19, no 5 (septembre), p. 457-469.
- Yamamoto, Taku. 1976. « Asymptotic Mean Square Prediction Error for an Autoregressive Model with Estimated Coefficients ». *Applied Statistics*, vol. 25, no 2, p. 123-127.
- Yamamoto, Taku. 1980. « On the treatment of Autocorrelated Errors in the Multiperiodic Prediction of Dynamic Simultaneous Equation Models ». *International Economic Review*, vol. 21, no 3 (octobre), p. 735-748.
- Yamamoto, Taku. 1981. « Predictions of Multivariate Autoregressive-moving Average Models ». *Biometrika*, vol. 68, no 2 (août), p. 485-492.
- Zaffaroni, Paolo. 2004. « Contemporaneous aggregation of linear dynamic models in large economies ». *Journal of Econometrics*, vol. 120, no 1 (juin), p. 75-102.
- Zotteri, G., M. Kalchschmidt et F. Caniato. 2005. « The impact of aggregation level on forecasting performance ». *International Journal of Production Economics*, vol. 93-94, no 5 (septembre), p. 479-491.