La contagion financière

Introduction :

Le problème majeur dans la gestion des risques est l'accroissement de l'interdépendance des marchés financiers en période de crise. Ce qui pousse les chercheurs à s'intéresser de plus en plus à l'étude des périodes de tension économique et à la notion de contagion. Dans ce chapitre, nous donnons un bref aperçu sur la littérature empirique et théorique concernant la crise et la contagion en se référant à l'article de Costinot *et al*. (2000).

La multiplicité des travaux théoriques et empiriques qui ont été menées sur le thème de la contagion suggère une diversité des définitions théoriques de la contagion et des modèles explicatifs de la diffusion géographique des chocs.

Section 1 : Définition de la contagion

La notion de contagion ne fait pas encore l'objet d'une définition unique. Certains économistes considèrent une contagion dès qu'il y a une influence mutuelle et que les voies de la contagion sont celles de l'interdépendance. Dans cette veine, Masson (1998) distingue deux types de contagion, ce qui lui permet de faire ressortir une forme de contagion non mécanique la «pure» contagion. D'autres auteurs rejettent cette définition et n'acceptent l'idée de contagion que si elle est associée à une augmentation exceptionnelle et transitoire de l'intensité des dépendances économiques pendant une crise. Introduite par King et Nadhwani (1990), elle a été appelée *shift contagion* par Forbes et Rigobon (2000).

Forbes et Rigobon (2001) distinguent deux grandes catégories d'approches de la contagion :

L'approche contingente aux crises d'une part, connu dans la littérature anglo-saxonne sous le nom *Crisis Contingent Theories,* qui traite l'effet du stress sur les interdépendances structurelles. Dans ce cadre le mécanisme de transmission durant la crise ou juste après est fondamentalement différent de celui qui prévalait avant la crise. Cette dernière provoque un changement structurel de telle sorte que les chocs se propagent par l'intermédiaire d'un canal qui n'existe pas durant les périodes de stabilité financière (Marais et Bates (2004). L'approche non contingente aux crises d'autre part, connu dans la littérature anglo-saxonne sous le nom *Non Crisis Contingent Theories,* traite les transmissions de chocs par les canaux des interdépendances stables. Marais et Bates (2004) expliquent que dans le cas de la contagion non contingente les mécanismes de transmission après un choc ne sont pas significativement différents de ceux avant la crise[1].

Section 2 : Les théories de transmission de crises

2. 1.Les approches de non contingence des crises

Cette approche est la résultante de l'interdépendance économique de plusieurs pays. L'émergence d'une crise dans une économie, génère la modification des fondamentaux macroéconomiques des autres économies. Cet effet systématique de la simultanéité des chocs est la conséquence des canaux existants avant le choc, tels les liens économiques qui entretiennent les pays.Calvo et Reinhart (1996) et Kaminsky et Reinhart (2000) appellent ce type de propagation de Crise, contagion basée sur les fondamentaux.

[1] BRANA Sophie, Mars 2004, Les facteurs de propagation des crises financières : La crise asiatique a t'elle été discriminante ?, Larfi, UNIV Montesquieu bordeaux IV, pp 4-7.

2.1.1. Chocs idiosyncratiques :

1er canal : liens réels : liens commerciaux et dévaluation compétitive :

Les liens réels permettent aux pays d'entretenir des relations appuyées sur des fondamentaux économiques similaires. Ainsi, une crise affecte un groupe d'économies étroitement liées par des liens externes tel que le commerce international ou les relations financières qui assurent un rôle prépondérant dans la propagation des crises et l'étude de la contagion.

Masson (1999) qualifie la contagion, par le moyen des liens commerciaux, de l'interdépendance normale entre les pays. L'auteur considère que le canal commercial est un facteur de dévaluation lors d'une crise dans un pays, suite à une dévaluation dans sa monnaie locale. Il bénéficie de sa compétitivité et pénalise ces partenaires commerciaux internationaux. En conséquence, la dévaluation entraîne deux effets nocifs. Un effet direct, relatif à la détérioration de la balance commerciale des partenaires suite à la chute de leurs exportations vers le pays en crise. Le deuxième effet est indirect, associé à la dévaluation compétitive des exportations des produits concurrents, ce facteur engendre des pressions monétaires sur les économies en question qui se verront obliger de dévaluer leurs monnaie et ainsi maintenir leurs niveaux d'exportations et conserver leurs devises ; (Gerlach et Smets (1994), Corsetti *et al.* (2000) et Kaminsky *et al.* (2003)), Connu dans la littérature anglo-saxonne sous le nom *fundamentals based contagion.*

2ème canal : les liens financiers

Dire qu'un certain nombre de pays sont interconnectés à partir des liens commerciaux, laisse supposer l'existence d'interactions financières, notamment, à partir des conventions financières mises en place entre les pays dans le but de faciliter l'échange des différents biens et services, ils sont donc reliés par un même système financier et international qui fait intervenir différents investisseurs internationaux.

Masson (1999) explique: en cas de survenance d'une crise dans un pays, les investisseurs sont conduits à *débalancés* leurs portefeuilles dans l'espoir de compenser leurs pertes et réduire leur exposition au risque. D'où l'importance des interdépendances financières dans la transmission des chocs entre les pays.

Par ailleurs, en cas de crise, les appels de marges poussent les investisseurs à liquider leurs positions dans d'autres économies n'ayant pas encore subies de chocs pour gérer le

risque ou pour des raisons de liquidité (Kaminsky et Reinhart (1999), Kaminsky et Reinhart (2000),Van Rijckeghem et Weder (2001) et Broner et Gelos (2003)).

2.1.2. Chocs communs et *Monsoonal effect :*

Les chocs communs (globaux, aléatoires, exogènes), qui sont aussi spécifiés à l'effet de Mousson, affectent d'une manière simultanée les fondamentaux d'un bon nombre de pays (Dornbush *et al.* (2000)). En conséquence, des séries des perturbations sont conjointement déclenchées et affectent l'intégrité du système financier international. D'après Masson (1999), la crise est généralisée et pour ne pas toucher de premier pays. Il justifie que la contagion des crises financières est assurée par la corrélation des variables macroéconomiques fondamentales.

Suite à un choc commun, les crises concomitantes dans les marchés d'un bon nombre de pays, inhibent la croissance de ces derniers. On assiste donc à une forte corrélation des marchés et des fluctuations des prix en groupe.

Les causes communes peuvent être dues à des modifications dans les politiques des pays industrialisés, générant des dégâts à l'économie mondiale. Selon Marais (2004), ces chocs aléatoires peuvent prendre différentes formes : une augmentation des taux d'intérêts internationaux, une hausse du prix du pétrole, une diminution de l'offre mondiale ou de la demande globale.

Moreno et Trehan (2000) étudient la simultanéité des chocs externes sur la base des taux d'intérêt et le taux d'inflation américain. Ils justifient la simultanéité des crises monétaires par une variation supérieure à leurs niveaux de base.

2.1.3. Learning :

La manifestation d'une crise dans un pays, peut mener à ce qu'on l'appel *Wake up call* (Goldstein (1998)). Dès lors, les investisseurs réévaluent leur aversion pour le risque et leur perception pour des pays ayant les mêmes structures macroéconomiques. Pour exemple, en cas de défaillance du système bancaire, les investisseurs peuvent changer leurs opinions concernant le système bancaire ou financier dans d'autre pays. Ainsi ils ajustent leurs espérances de crise en fonction de leurs jugements. Il en résulte l'observation d'un changement de comportement des investisseurs qui paraissaient-être irrationnel avant la survenance de la crise puisqu'ils sous estimaient la faiblesse de l'économie. Mais après le choc, leur comportement semble être rationnel (Kodres et Pritsker (2002)).

En conclusion, plus un pays est intégré dans le système financiers international et donc dans l'économie mondiale, plus élevé est le risque de contagion. Tous ces canaux jouent un rôle significatif dans le mécanisme de propagation des crises et l'accélération de l'effet de contagion. Cependant, ces fondamentaux peuvent être insuffisants pour expliquer la transmission des chocs. Donc l'influence transfrontière trouve son explication dans d'autres facteurs, soit, l'approche de contingence des crises.

2.2. Les approches de contingence des crises :

Elle est définie comme la propagation internationale des chocs tout en supposant que le mécanisme de transmission à changer après le choc initial. Contrairement à l'approche de non contingence des crises, ces théories se réfèrent aux cas où la transmission n'est pas justifié par des bases financières ou économiques ni par des liens réels entre les marchés. L'influence transfrontière est, via des liens qui n'existaient pas durant la période tranquille, ils sont plutôt contingents et spécifiques à la crise.

Cette théorie propose trois explications majeures : comportement mimétique, équilibres multiples et les chocs de liquidité endogène.

2.2.1. Comportement mimétique :

Calvo (1996), a développé un modèle dans lequel la majorité des investisseurs admettent un comportement mimétique dû à un manque de l'information précise. Etant donnée le coût élevé du contrôle et de l'évaluation de chaque marché, il sera plus optimal pour les investisseurs non informés d'agir simultanément quand un problème apparaît dans l'un de ces marchés.

2.2.2. Equilibres multiples :

Pour être capable d'élucider le phénomène de contagion, nous introduisons les modèles macroéconomiques auto réalisatrices avec équilibres multiples. A travers ces modèles, Masson (1999), démontre que le phénomène apparaît lorsque l'économie d'un pays, particulièrement son marché financier, passe d'un bon équilibre à un mauvais équilibre suite à une crise dans un premier pays touché.

Le saut entre les équilibres suppose l'existence d'une partie invisible. Celle-ci est fortement rattachée à la psychologie des investisseurs, tel, que le manque de confiance vis-à-vis du marché et les modifications des anticipations, plutôt que l'existence de liens réels entre les pays ou des changements des fondamentaux macroéconomiques.

Cette théorie, permet d'expliquer la concentration des crises et les raisons d'attaques spéculatives; (Radelet et Sachs (1998)).

2.2.3. Les chocs de liquidité endogène :

En réponse à un choc sur un marché, les investisseurs agissent par des ajustements automatiques de leur portefeuille à partir des mouvements de liquidation et de réallocation.

Ces mouvements provoquent un Co-mouvement excessif des prix des actifs en excès par rapport aux fondamentaux ; aussi, ils sont réalisés dans le cadre d'asymétrie informationnelle qui affecte d'autant plus le comportement des investisseurs ;(Masson (1998), Calvo (1999) et Kodres et Pritsker (2002)).

D'après ce qui précède, nous voyons que les chocs de liquidité endogène et l'asymétrie informationnelle disposent d'un rôle capital dans la transmission des chocs. A cet égard, Forbes & Rigobon (2000) ont constaté une augmentation significative de la corrélation entre les actifs financiers. Un tel mécanisme ne se produit pas durant les périodes de stabilité financière mais uniquement en l'occurrence d'une crise.

Section 3 : Mise en évidence empirique de la contagion

La littérature qui traite le problème de crise et sa propagation est abondante. Quatre méthodes de repérage de la contagion sont les plus utilisées.

1/ La premier approche s'intéresse aux effets d'un choc dans un pays sur d'autres. Les estimations économétriques par les modèles probit/logit permettent de tester la significativité de différents canaux réels de transmission tels que les liens commerciaux, les interdépendances financières ou les similitudes macroéconomiques (Marais et Bates (2004)).

Eichengreen, Rose et Wyplosz (1995), développent la définition de la contagion comme une augmentation de la probabilité d'un choc conditionnellement à l'occurrence d'une crise de change dans un autre pays, ce qui justifie l'utilisation des modèles dichotomiques (probit ou logit) pour définir la propagation des crises de change.

Kaminsky et Reinhart (1998), vérifient une telle interaction dans le cas des pays appartenant à la même région géographique.

2/ La deuxième approche traite les modèles ARCH et GARCH. Chou *et al.* (1994) traitent la propagation de la volatilité *Spillovers* par les modèles GARCH pour identifier un

effet de contagion.Engle,Ito et Lin (1990)[1] ont utilisé un modèle GARCH pour étudier en premier temps les effets de transmission de la contagion intra-quotidienne du taux de change entre les marchés Japonais et Américaines second temps sur les liens entre les marchés des actions Américaines et Japonaises L'effet de transmission de la contagion est significatif dans les deux études. En effet ils ont abouti à la conclusion que les rendements et les volatilités de jour d'un marché sont corrélés avec les rendements et volatilités de nuit de l'autre marché.

Hamao et al (1990)[2], ont analysé les relations entre les marchés boursiers de New york,Londres et Tokyo à partir d'un modèle ARCH. Ils ont étudié la volatilité des prix des actions dans chacun de ses marchés et leur éventuelle transmission d'un marché à l'autre. Les résultats des estimations ont montrés des effets de transmission de la volatilité des prix de New York vers Tokyo et de Londres vers Tokyo. mais pas de Tokyo vers New York ou Londres.Koutmous et Bouth (1995)[3] ont observé des effets de transmission asymétriques de la volatilité à travers les mêmes marchés boursiers que ceux étudiés par Hamao et al (1990), le model exponentiel GARCH est estimé en utilisant les données journalières recueillies aux heures d'ouverture et de fermeture des indices boursiers les plus représentatifs de ces marchés, le S&P 500 pour les Etats-Unis, le FTSE 100 pour le Royaume-Uni et le NIKKEI 225 pour le Japon. Ils ont obtenu des résultats similaires à ceux de Hamao et *al*. Cependant, contrairement à ces derniers, ils ont pris en considération les possibilités de l'existence d'effets asymétriques dans les mécanismes de transmission de la volatilité, c'est-à-dire la possibilité que la mauvaise nouvelle sur un marché donné ait un plus grand impact sur la volatilité des rendements de l'autre marché. Bekaert et Harvey (1997)[4] ont étudié l'intégration des marchés de vingt pays émergents à l'économie mondiale pour voir l'effet de transmission de l volatilité des actions des pays émergents. Pour le faire, ont appliqué un modèle GARCH multivarié dont la moyenne et la variance conditionnelle sont composées de variable lesquelles permet de capter des chocs provenant du marché mondial et des chocs propres à chaque pays. En intégrant aussi dans leur modèle des variables macroéconomiques permettant de mesurer le degré d'intégration de chaque pays à l'économie mondiale, comme par exemple

[1]ENGLE.R.F& lTO.T&, LIN .W, Meteor shower or Heat Waves ?Heteroskedastic Intra-Daily volatility in the foreign Exchange Market, Econometra, vol 58, pp525-542.

[2]HAMAO.Y.R&MASULIS.R.W, Correlations in price changes and volatility Across international stock markets, Review of Financial studies, vol 3, pp 281-307.

[3]KOUTMOS.G&BOOTH. G, 1995, Asymetric volatility transmission i international stock markets, journal of international money and finance, vol 14, pp747-762.

[4] BEKAERT Geert& HARVEY.C,1997,Emerging Aquity Market volatility,journal of financial Economics,vol 43,pp29-77.

la part des échanges internationaux sur le PIB. Les résultats montrent que plus un pays est intégré, plus il est exposé à subir un choc provenant de l'extérieur par les canaux de transmission.Yew-Kwang Ng (2000)[1] a également constaté u effet de transmission de la volatilité des marchés américains et japonais vers les marchés du bassin du pacifique. Il a utilisé un modèle GARCH multivarié le même que Bakaert et Harvey (1997). Sauf que Ng explique clairement dans son modèle les chocs étrangers provenant de sources régionales (le Japon) et globales (les Etats-Unis) En utilisant des données hebdomadaires des rendements des indices des marchés boursiers des pays de bassin du pacifique (thailande,HongKong, Corée du sud, Taiwan…) et les indices américains et japonais respectivement S&P 500 et le Tokyo Stock price Index. Afin de distingué l'influence relative des Etats-Unis et celle du Japon sur les marchés du bassin du pacifique, il a utilisé un modèle de transmission de volatilité uni-variée et a effectué pour chaque modèle un test de l'existence de l'effet de transmission Les résultats ont révélé un effet de transmission de la volatilité significatif des marchés japonais et américain en Malaisie, Singapour, Taiwan et en Thaïlandeet contrairement aux attentes, aucun effet de transmission n'est observé des Etats-Unis vers Hong Kong qui sont pourtant deux marchés étroitement liés. Baele (2002)[2], se base sur la même étude, mais en considérant les marchés des actions américains et européens. Il a cherché à mesurer l'importance et la nature des variations dans le temps de la transmission de la volatilité prévenant des marchés des Etats-Unis et des marchés agrégés de l4union européenne(UE) vers chacun des marchés des 13 pays d'Europe occidentale étudiés. Le modèle qu'il a proposé est une extension du modèle de Bekaert et Harvey(1997) dans le sens où il ya deux sources de chocs régionaux (US et EU) au lieu d'un choc mondial. C'est également un prolongement du modèle de Ng(2000) en prenant en compte les changements de régimes dans les paramètres de transmission. Les données utilisées dans cette étude sont composées des rendements boursiers des pays membres et non membres de l'union monétaire européenne (UME),et des marchés régionaux (US et EU).Il en découle de cette étude que la transmission des chocs de l'US et de l'UE s'est intensifiée au courant des décennies 1980 et 1990.L'augmentation de la transmission de choc de l'UE état remarquée vers la fin de ces années, mais cette hausse n'a pas empêché aux marchés américains d'avoir à prépondérance

[1]NG A, 2000, Volatility spillover Effects from Japan and the US to the Pacific-Basin, journal of international money and finance,vol 19,pp 207-233.

[2]BAELE.L, 2002, Volatility Spillover Effects in European Equity Markets : Evidence from a regime switching model mimico,university of Ghent ,p255.

sur les marché européens est aussi mis en évidence en période de forte volatilité des marchés des actions.

Naoui et al (2010)[1], ont testé comportement de marché boursiers développés et 10 marchés boursier émergents, qui ont été touchés par la crise des subprime déclenchée aux Etats-Unis d'Amérique ; via la contagion. En appliquant la méthode DCC-GARCH dynamique corrélation conditionnelle. Ils relèvent un effet de contagion entre les Etats Unis et les pays émergents et développés au cours de la crise des subprimes.

Hwang e al.(2010)[2], ont étudié le phénomène de contagion lors de la crise des subprime 2007 sur les marchés boursiers internationaux en utilisant un modèle DCC-GARCH sur 38 données par pays. En conclusion, ils ont trouvé des preuves de la contagion financière, non seulement dans les marchés émergents, mais aussi dans les pays développés au cours de la crise des subprimes.

Bouaziz et al.(2012)[3] ont examiné empiriquement le phénomène de contagion lors de la crise de subprime via la technique de DCC GARCH sur les marchés boursiers des pays développés avec le marché américain pendant la crise des subprimes,ce qui relève une preuve évidente de la contagion.

3/ La troisième approche fait appel aux tests de co-intégration pour repérer les relations de long terme. Longin et Solnik (1995) recherchent des changements dans les relations structurelles entre les marchés financiers des différents pays.

4/ La quatrième approche calcule les coefficients de corrélations entre les rendements des titres des différents marchés. C'est la plus utilisée et la plus directe de toutes les méthodes. Appliquée au *shift contagion*, elle consiste à évaluer la variation des comouvements entre les périodes de crise et la période de stabilité. Nous concluons d'un effet de contagion si les corrélations croisées augmentent significativement en période de crise. King et Wadhani (1990) calculent le coefficient de corrélation entre les Etats-Unis, le Royaumes Unis et le Japon pour le Krach boursier américain d'octobre 1987. Calvo et

[1]NAOUI. K& LIOUANE. N& BRAHIM. S, 2010, A dynamic conditional correlation analysis of financial contagion : the case of the subprime crédit crisis international,journal of Economics and Finance ,pp 85-96.

[2]HWANG,I, & AEUCKIN.F.&KIM.T.S,2010,Contagion effects of the U.S subprime crisis on international stock markets.Finance and Corporate Governance,Conference Paper.

[3]BOUAZIZ.,M.C& SELMI.N & BOUJELBENE.Y,2012,Contagion effect of the subprime financial crisis :evidence of DCC multivariate GARCH models,European journal of economics,finance and Administrative Sciences n°44,pp66-76.

Reinhart (1996) utilisent cette approche pendant la crise mexicaine de 1994. Ils étudient la contagion sur les marchés des actions et des obligations pour pays émergents d'Asie et d'Amérique Latine.

Baig et Goldfajn (1998), traitent une gamme de marchés très large : indices boursiers, taux de change, taux d'intérêt et différentiel de rendements des dettes des pays émergents durant la période 1997-1998.

Toutes ces études montrent une nette augmentation de la corrélation entre les différents marchés. Néanmoins, l'utilisation de la corrélation linéaire pour détecter la contagiosité, peut mener à des résultats fallacieux ou bien à des différentes interprétations selon la manière dont on l'utilise. Nous avons donc besoin d'une mesure de dépendance plus robuste que le coefficient de corrélation linéaire.

Section 4 : Méthode de détection et mesure de contagion

Les travaux empiriques sur la contagion sont de plus en plus fréquents. En fait, on distingue deux axes de travaux. Le premier concerne les recherches dont l'objectif est de déterminer les causes de la contagion ainsi que les canaux de sa transmission. Le deuxième axe comprend les travaux qui ont essayé de démonter l'existence de la contagion pure, dans les récentes crises.

La plupart de ces derniers travaux sont basés explicitement ou implicitement sur les définitions de la contagion déjà discutées notamment celle de Forbes et Rigobon (2001). Ces auteurs ont cherché à comparer le degré des liens financiers avant et après la crise indépendamment des fondamentaux. En terme technique, les liens entre les marchés peuvent être mesurés par plusieurs statistiques à savoir la corrélation, la probabilité des attaques spéculatives ainsi que la transmission des chocs ou de volatilité. Ils ont aussi utilisé diverses approches très développées comme l'étude des corrélations (Bain et Goldfajn, 1998, Park et Song, 1999 ; Forbes et Rigobon, 2001), les processus ARCH et GARCH (Edwards, 1998), le modèle VAR et les tests de causalité (Masih et Masih, 1999 ; Khalid et Kawai, 2003 ; Sander et Kleimeir, 2003).

4.1. Les processus ARCH et GARCH :

4.1.1. Processus ARCH :

Engle (1982) a développé les modèles ARCH (Auto Regressive Conditional Engle (1982) a développé les modèles ARCH (*Auto Regressive Conditional Heteroskedasticity*) afin de permettre à la variance d'une série de dépendre de l'ensemble des informations disponibles, et du temps. La raison principale de ce phénomène est la leptokurticité des séries. Une série est leptokurtique si son coefficient de Kurtosis est élevé, c'est à dire supérieur à trois lorsque le coefficient n'est pas centré. Les modèles ARCH (*Autoregressive Conditional Heteroskedasticity*) peuvent modéliser la leptokurticité présente dans les données financières Cette classe de modèles a pour objet de remédier aux insuffisances des représentations ARMA, en réalité peu adaptées aux problématiques financières. Les séries financières sont en effet caractérisées par une volatilité variable et par des phénomènes d'asymétrie qui ne peuvent être pris en compte par ce type de modélisations.

En outre, si le processus suivi par la volatilité est correctement spécifié, celui-ci peut fournir des informations utiles à la détermination du processus générant les rentabilités. Les modèles ARCH sont fondés sur une par métrisation endogène de la variance conditionnelle.

Dans un modèles de types ARCH, ε_t est processus tel que :

$$\begin{cases} E(\varepsilon_t | \varepsilon_{t-1}) = 0 \\ V(\varepsilon_t | \varepsilon_{t-1}) = \sigma_t^2 \end{cases}$$

$\varepsilon_t \sim N(0, \sigma^2)$

$\varepsilon_{t-1} = (\varepsilon_{t-1}\varepsilon_{t-2}.....)\sigma_t^2$ Représente la variance conditionnelle du processus ε_t On voit donc que la variance conditionnelle qui est un indicateur de la volatilité d'un titre-peut varier au cours du temps dans ce type de processus, à la différence des processus ARMA ou elle est supposée constante.

Dans le cadre de ce modèle, on postule que la variance des erreurs est constante (homoscedasticité). Dans le cas où la variance des erreurs n'est pas constante, le processus est heteroscedastique. Dans le cadre des modèles ARCH, la variance conditionnelle des erreurs σ^2 dépend du carré de la dernière erreur :

$$\sigma_t^2 = a + b_1 \varepsilon_{t-1}^2 .....(1)$$

Le modèle ci-dessus est connu sous le nom d ARCH(1) car il y a une seule erreur carrée retardée. En revanche, l'équation de moyenne conditionnelle, c'est-à-dire l'équation qui montre comment la variable dépendante x_t évolue dans le temps, peut prendre quasiment toutes les formes estimées adéquates par le chercheur. Le modèle précédent peut donc être étendu à un modèle général avec q erreurs carrées retardés connu sous le nom d'ARCH (q)

Dans la littérature financière, la variance conditionnelle (σ^2_t) est appelée. Le modèle peut ht ainsi s'écrire de la manière suivante :

$$X_t = \eta + X_{t-1} + \varepsilon_t \text{.....(2)} \qquad Ou \quad \varepsilon_t \approx N(0, \sigma^2)$$

$$h_t = a + b_1\varepsilon^2_{t-1} + b_2\varepsilon^2_{t-2} + \text{.....} + b_q\varepsilon^2_{t-q} \text{.....(3)}$$

La relation peut s'écrire aussi :

$$h_t = a + \sum_{i=1}^{q} b_i\varepsilon^2_{t-i} \text{.....(5)} \qquad a_0 \succ 0 b_i \succ 0$$

Le modèle ARCH(q) permet notamment de prendre en compte les groupements de volatilité, c'est à dire le fait que les fortes (faibles) variations de prix sont suivies par d'autres fortes (faible) variations dont le signe n'est pas prévisible. Pour appliquer une description plus parcimonieuse de la structure des retards, la littérature a généralisé le modèle ARCH et établit le modèle GARCH (p, q).

4.1.2. Processus GARCH :

Analyser et comprendre comment le modèle GARCH univarié fonctionne est fondamental pour l'étude de la corrélation conditionnelle GARCH multivariée dynamique modèle d'Engle et Sheppard (2001). Il est important, non seulement en raison du fait que le modèle est une combinaison linéaire des modèles GARCH univariés, mais aussi parce que la matrice de corrélation conditionnelle dynamique est basée sur la façon dont le processus GARCH (1.1) univariée fonctionne.

La volatilité du rendement de l'actif désigne l'écart type des variations de la valeur sur un horizon de temps donné. Sur le long terme, les rendements ont la tendance à se déplacer vers une valeur moyenne (retour à la moyenne). Les variations de la valeur qui apparaissent durant cette période sont à la fois positif et négatif (asymétrique), le plus souvent sont proche de la valeur moyenne, mais certains changements obtiennent des valeurs extrêmes (leptokurtique).

La volatilité des rendements d'aujourd'hui est conditionnelle à la volatilité passée et elle a la tendance à se regrouper dans la volatilité.

Supposons que le processus stochastique $(r_t)_t^T$ décrit le retour à une époque spécifique d'horizon où r_t le rendement observé à l'instant t de model des rendements est comme suivant :

$$r_t = \mu_t + \varepsilon_t \text{.....(6)}$$

Ou $\mu_t = E(r_t|\psi_{t-1})$ désigne les espérances conditionnelles de la série de rendements, ε_t erreurs conditionnelles et $\psi_{t-1} = \sigma(\{r_s : s \le t-1\})$ représente l'ensemble d'information au :

Temps t-1.

Supposons que les erreurs conditionnelles sont les écarts-types conditionnelles des rendements $h_t^{1/2} = Var(r_t|\psi_{t-1})$ fois l'unité de la variable stochastique de la variance moyenne nulle indépendante $h_t \ et \ Z_t$ qui sont distribuée normalement.

On note que $h_t \ et \ Z_t$ Independent de t.

$$\varepsilon_t = \sqrt{h_t z_t} \approx N(0, ht) \text{.....(7)}$$

On suppose que l'espérance conditionnelle μ_t=0 ce qui implique que :

$$r_t = \sqrt{h_t z_t} \quad Ou \quad r_t|\psi_t \text{.....(8)}$$

Un modèle GARCH peut être considéré comme un modèle ARMA sur le carré des innovations. Lorsque μ_t = 0 la variance du rendement coïncide avec la variance des erreurs et leur espérance conditionnelle est nulle, donc l'erreur processus est un processus d'innovation.

Souvent dans un modèle de Bollerslev (1986), qui a généralisé le modèle initial d'Engle en établissant le modèle GARCH (p, q) en l'introduisant des valeurs retardées de la variance dans son équation (la variance conditionnelle (h_t) dépend aussi de celles retardées). Ainsi, un processus GARCH (p, q) est défini par :

$$h_t = a + \sum_{i=1}^{q} b_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} c_j h_{t=j}^2 \text{.....(9)}$$

Ou $a \succ 0, b_i \ge 0, c_j \ge 0 \forall i, \forall_j$

En d'autres termes le GARCH (p, q) est composé de trois termes :

a : La variance de long terme pondéré

$\sum_{i=1}^{q} b_i \varepsilon_{t-i}^2$ Les carrées des résidus retardés

$\sum_{j=1}^{p} c_j h_{t=j}^2$ La variance conditionnelle retardée

Notez que les innovations dans le terme de la moyenne mobile sont les carrés et le GARCH (p, q) ne comprend pas l'asymétrie des erreurs, ce qui est un inconvénient. Le modèle EGACH de Nelson (1991) et le modèle GJR-GARCH de Glosten, Jagannathan et Runkle (1993) sont deux exemples de modèles GARCH étendues qui prennent en compte l'asymétrie des rendements. Par définition la variance est non-négatif, le processus doit être positive ou nulle. Nelson and Cao (1992).

4.1.2.1 GARCH (1.1) :

Le modèle GARCH (1.1) est le plus utilisé. Il est donné par l'équation (10) :

$$h_t = a + b\varepsilon_{t=1}^2 + ch_{t-1} a \succ 0, b \geq 0, c \geq 0 \text{.....(10)}$$

Les trois termes peuvent être interprétés comme pour le GARCH (p, q) mais avec seulement un décalage pour chacun, carré des 'innovations et la variance.

On remplace t par (t-j) dans (10) on obtient l'équation suivante :

$$h_t = a(1 + c + c^2 + ...c^{j=1}) + \sum_{k=1}^{j} c^{k-1} \varepsilon_{j=l}^2 + c^j h_{t-j} = a\frac{1-c^j}{1-c} + a\sum_{t-k}^{j} c^{k-1} \varepsilon_{t=k}^2 + h_{t-j} \text{....(11)}$$

Si « j » tend vers l'infini alors $\lim_{\infty} h_t = \frac{a}{1-c} + a\sum_{k=1}^{\infty} c^{j-} 1\varepsilon_{t-k}^2$

Cela implique que la volatilité actuelle est une moyenne mobile exponentielle pondérée des derniers carrés d'innovations. Bien qu'il existe des différences fondamentales entre le modèle ARCH (1,1) et EMWA (moyenne mobile exponentielle pondérée) les paramètres du GARCH doivent être estimés, et le retour à la moyenne doit être intégré dans le modèle.

Pour produire le nombre de paramètres de trois à deux dans le GARCH (1.1), ainsi pour qu'on puisse rendre le calcul de la « variance de ciblage » d'Engle et Mezrich (1996) plus facile à utiliser. On ajoute la variance inconditionnelle h à l'équation (10) :

$$h_t - \dot{h} = a + \dot{h} + b(\varepsilon_{t=1}^2 - \dot{h}) + c(h_{t-1} - h) + c\dot{h} + b\dot{h} \text{.....(12)}$$

$$h_t = a - (1-b-c)h + (1-b-c)h + b\varepsilon_{t=1}^2 + c\dot{h}_{t-1} \text{.....(13)}$$

Si a= (1-b-c)h alors l'équation (10) peut être comme suite :

$$h_t = (1-b-c)h + b\varepsilon_{t=1}^2 + c\dot{h}_{t-1} \text{.....(14)}$$

Le modèle est non seulement plus facile à calculer, mais il implique également que la variance inconditionnel $h = \frac{a_0}{(1-b-c)}$ fonctionne tout simplement sous l'hypothèse que $(b+c) \prec 1 \; et \;\; b \succ 0 \;\; c \succ 0 \;\; a \succ 0$

4.1.2.2. Les contraintes du GARCH (1,1) :

Comme on la déjà mentionné au-dessus, la variance conditionnelle doit être positive ou nul, pour cela il faut que $a \succ 0, b \geq 0, c \geq 0$ Une autre caractéristique qui maintient la variance conditionnelle non négative est que le processus soit stationnaire. Le processus GARCH (1.1) est faiblement stationnaire avec la valeur attendue de la covariance inconditionnelle $E(r_t) = 0$

$$COV(r_t, r_{t-s}) = a_0 /(1-b-c)$$

Si et seulement si b + c < 1

Rappel, lors de l'utilisation de la variance de ciblage dans le GARCH (1,1) la variance inconditionnelle $\dot{h} = \frac{a}{(1-b-c)}$ qui coïncide avec la variance inconditionnelle lorsque le processus est faiblement stationnaire.

En résumé, les contraintes liées au modèle GARCH (1.1), lors de l'utilisation de la variance de ciblage sont :

$$a \succ 0, b \geq 0, c \geq 0 \;\; et \;\; b + c \prec 1$$

Si ces contraintes se réalisent, le procédé est considéré comme covariance stationnaire.

4.1.2.3. Le DCC (*The Dynamic Conditional Correlation*) :

Pour prolonger les hypothèses citées précédemment au cas multivarié, supposons que nous ayons N actifs dans un portefeuille et le vecteur de rendements est le vecteur de côlon.

$$r = (r_{1,t}, r_{2,t} .....r_{n,t})$$

En outre, on suppose que les rendements conditionnels sont distribués normalement et la matrice de covariance conditionnelle.

$$H_t = E(r_t r'_t | \psi_{t-1})$$

$$r = \gamma_0 + \gamma_1 r_{t-1} + \gamma_2 r_{t-1} + \varepsilon_t$$
$$r = (r_{1,t}, r_{2,t} .....r_{n,t}). \qquad n = 11, \qquad \varepsilon_t = (\varepsilon_{1,t}, \varepsilon_{2,t}, ...\varepsilon_{n,t})$$
$$et \ \ \varepsilon_t | \psi_{t-1} \approx N(0, H_t)$$

Ht : matrice de variance-covariance conditionnelle à la date t, ψ_{t-1} l'ensemble d'information en t-1. La matrice de variance-covariance conditionnelle peut s'écrire :

$$H_t = D_t R_t D_t$$

Ou le D_t matrice diagonale des écarts type temporelle conditionnelle du rendement, obtenu à partir de l'estimation d'un modèle GARCH (q, p) avec

$$\sqrt{h_{i,t}} i = 1,2,.....K$$

$$\sqrt{h_{i,t}} i = a + \sum_{q=1}^{qi} b_{iq} \varepsilon_{it-q}^2 + \sum_{p=1}^{pi} c_{ip} h_{it-p}$$

R_t : Matrice des corrélations conditionnelles

Le modèle de DCC proposé par Engle (2002) implique deux étapes d'estimation de la matrice de covariance conditionnelle H_t, dans la première étape le modèle de la volatilité variable consiste à estimer les rendements des indices et la variance conditionnelle $\sqrt{h_{i,t}}$ obtenue.

Dans la deuxième étape, les résidus des rendements sont transformés par leurs écarts-types estimés à partir de la première étape c'est-à-dire $\delta_{i,t} = \varepsilon_{i,t} / \sqrt{h_{i,t}}$

Est ensuite utilisée pour estimer les paramètres de la corrélation conditionnelle

$$Q_t = (1 - \alpha - \beta)\overline{Q} + \alpha \delta_t - 1\delta'_{t-1} + \beta Q_{t-1}$$

$$R_t = Q_t^{*-1} Q_t Q_t^{*-1}$$

$R_t=(\text{diag}(Q_t))^{-1/2}Qt(\text{diag}(Qt))^{-1/2}$

$$(diag(Q_t))^{-1/2} = diag(\frac{1}{\sqrt{q_{11,t}}}.....\frac{1}{\sqrt{q_{kk,t}}})$$

$Q_t = \left|q_{ij,t}\right|$ Matrice de variance –covariance des résidus standardisés $\delta_{i,t}$

$\overline{Q}_t = E\left[\delta_t \delta'_t\right]$ La matrice des variances –covariances inconditionnelles de δ_t

(α), (β) : sont des paramètres non négative (α + β < 1) sont intercepté, respectivement, les effets des chocs et des corrélations dynamiques retardées sur le niveau contemporain de ces dernières.

Q_t^* : C'est une matrice diagonale contenant la racine carrée des éléments de la diagonale principale de Q_t ($Q_t^* = \left[q_{iit}^*\right] = \sqrt{q_{ii,t}}$)

Les corrélations conditionnelles pour une paire de marchés (i) and (j) à l'instant (t) peuvent être défini comme suite :

$$p_{ij,t} = q_{ij,t} / \sqrt{q_{ii,t} q_{jj,t}} i, j = 1,2,.....n, i \neq j$$

$$p_{ij,t} = \frac{(1-\alpha-\beta)\bar{q}_{i,j} + \alpha\delta_{t-1}\delta'_{t-1} + \beta q_{ij,t-1}}{\sqrt{(1-\alpha-\beta)\bar{q}_{ii} + \alpha\delta_{t-1}\delta'_{t-1} + \beta q_{ii,t-1}}\sqrt{(1-\alpha-\beta)\bar{q}_{jj} + \alpha\delta_{t-1}\delta'_{t-1} + \beta_{qjj,t-1}}}$$

$q_{ij,t}$: est un élément de la matrice Q_t ou i^{th} c'est la ligne et j^{th} c'est la colonne

Les paramètres du modèle DCC d'Engle (2002) sont estimés par la méthode de l'og-vraisemblance introduite par Ballerslev et al. (1992).

$$L(\Phi) = 1/2\sum_{t=1}^{t}\left[(n\log(2\Pi) + \log\left|D_t\right|^2 + \varepsilon_t' D_t^{-1}\varepsilon_t) + (\log\left|R_t\right| + \delta_t' R_t^{-1}\delta_t - \delta'_t\,\delta_t)\right]_{ij,t}$$

N : Le nombre d'équations

T : Le nombre d'observation

Φ : Le vecteur des paramètres à estimer

La première partie de la fonction de vraisemblance de l'équation représente la volatilité, qui est la somme des probabilités individuelles du modèle GARCH. Dans la première partie La fonction de log-vraisemblance peut être maximisée sur les paramètres D_t Compte tenu des paramètres estimés dans la première étape, la fonction de probabilité dans la deuxième étape, qui contient les paramètres de corrélation peut être maximisée pour estimer les coefficients de corrélation.

4.2. Le Processus VAR :[1]

Les processus VAR(p) (*Vactor Auto Regressive*) est une généralisation des processus AR au cas multivarié. Ils ont été introduits par Sims (1980), ces modèles macro-économétriques keynésiennes comportent des manques, telles que :

- Restrictions a priori trop fortes sur les paramètres par rapport à ce que prédit la théorie ;

- Absence de tests sérieux sur la structure causale

- Traitement inadéquat des anticipations.

En outre, d'un point de vue, ces modèles ont été mis à mal par les évènements au cours des années 70 (chocs pétroliers, récession mondiale, etc.) conduisant à de très importantes erreurs de prévision. Pour ces raisons, Sims (1980 a proposé une modélisation multivarié sans autre restriction a priori que le choix des variables sélectionnées et du nombre des retards.

La modélisation VAR repose sur l'hypothèse selon laquelle l'évolution de l'économie est bien approchée par la description du comportement dynamique d'un vecteur de N variables dépendant linéairement du passé. Depuis les travaux de Sims (1980), les techniques basées sur les modèles VAR ont connu de nombreuses évolutions.

4.2.1. Définition :

Considérons deux variables stationnaires x_{1t}, x_{2t}. Chaque variable est en fonction de ses propres valeurs passées mais aussi des valeurs passées et présentes des autres variables.

Supposons que l'on ait p=4 le modèle VAR (4) décrivant ces deux variables s'écrit : (1)

[1] BENABED Rachid, Econométrie théorie et applications, édition et publication OPU, Algérie, 2001, p71.

$$\begin{cases} x_{1t} = a_1 + \sum_{i=1}^{4} b_{1i} x_{1t-i} + \sum_{j=1}^{4} c_{1j} x_{2t-j} - d_1 x_{2t} + \varepsilon_{1t} \\ x_{2t} = a_2 + \sum_{i=1}^{4} b_{2i} x_{1t-i} + \sum_{j=1}^{4} c_{2j} x_{2t-j} - d_2 x_{1t} + \varepsilon_{2t} \end{cases}$$

$\varepsilon_{1t}, \varepsilon_{2t}$: sont deux bruits blancs non corrélés.

Ce modèle implique l'estimation de 20 coefficients. Le nombre de paramètres à est augmenté rapidement avec le nombre de retards, PN^2 ou P est le nombre de retards et N le nombre de variables du modèle.

Sous forme matricielle, le processus Var (4) s'écrit :

$$BX_t = \Phi_0 + \sum_{i=1}^{4} \Phi_i X_{t-i} + \varepsilon_t \text{.....(2)}$$

$$B = \begin{pmatrix} 1 & d_1 \\ d_2 & 1 \end{pmatrix} \Phi_0 = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} X_t = \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} \Phi_i = \begin{pmatrix} d_{1i} & c_{1i} \\ d_{2i} & c_{2i} \end{pmatrix} \varepsilon_t = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \text{.....(3)}$$

Il suffit de multiplier chaque (2) par B^{-1} en supposant B inversible, afin d'obtenir l'écriture usuelle du modèle VAR

Dans le cas général *Xt* contient N variables et pour un ordre de retard p quelconque. Un processus VAR (p) à N variables s'écrit sous forme matricielle :

$$X_t = \Phi_0 + \Phi_1 x_{t-1} + ... + \Phi_p X_{t-p} + \varepsilon_t \text{.....(4)}$$

$$X_t = \begin{pmatrix} x_{1t} \\ . \\ x_{Nt} \end{pmatrix} \varepsilon_t = \begin{pmatrix} \varepsilon_{it} \\ . \\ .\varepsilon_{Nt} \end{pmatrix} \Phi_0 = \begin{pmatrix} a_{1p}^N \\ . \\ a_N^0 \end{pmatrix} \Phi p = \begin{pmatrix} a_{1p}^1 & a_{1P}^2 & a_{1p}^N \\ a_{NP}^1 & a_{NP}^2 & a_{NP}^N \end{pmatrix} \text{.....(5)}$$

ε_t : est bruit blanc de la matrice de variance covariance $\sum_{\varepsilon}$

On peut écrire :

$$(I - \Phi_1 L - \Phi_2 L^2 - .....\Phi_p L^p) X_t = \Phi_0 + \varepsilon_t \text{.....(6)}$$

Soit :

$$\Phi(L) X_t = \Phi_0 + \varepsilon_t \text{.....(7)}$$

Avec $\Phi(L) = I - \sum_{i=1}^{p} \Phi_i L^i$

Plus formellement on retiendra la définition suivante.

On écrit que $X_t \approx VAR(p)$ si et seulement si il existe un bruit blanc $\varepsilon_t (\varepsilon_t \approx BB(0, \sum_{\varepsilon})).\Phi 0\varepsilon$

R^N et p matrices $\Phi_1, ..., \Phi_p$ que :

$$X_t - \sum_{i=1}^{p} \Phi_i X_{t=i} = \Phi_0 + \varepsilon_t \text{.....(8)}$$

Soit encore : $\Phi(L)X_t = \Phi_0 + \varepsilon_t$.....(9)

Ou Φ_0 est la matrice identitéI[1] et

$$\Phi(L) = I - \sum_{i=1}^{p} \Phi_i L^i \text{.....(10)}$$

Remarque :

1. VAR peut être généralisé afin de tenir compte d'une auto corrélation des erreurs d'ordre q.

$$X_t = \Phi_0 + \Phi_1 X_{t=1} + ... + \Phi_p X_{t=p} + \varepsilon_t + \Theta_1 \varepsilon_{t=1} + ... + \Theta_q \varepsilon_{t=q} \text{.....(11)}$$

Soit encore :

$$\Phi(L)X_t = \Theta(L)\varepsilon_t + \Theta_0 \text{.....(12)}$$

Φ : Est un polynôme matriciel d'ordre p et Θ un polynôme matriciel d'ordre

2. Dans un processus Ma multivarié (VMA) chaque composante suit un M.A mais dans un VAR chaque composante ne suit pas forcément un AR.

3. Tout processus VAR(p) peut s'écrit sous la forme d'un VAR (1), mais de dimension supérieure (Np au lieu de N)

Soit $\Phi(L)X_t = \Phi_0 + \varepsilon_t$.....(13)

$$Y_t = \begin{pmatrix} x_t \\ X_{t-1} \\ . \\ . \\ X_{t-(p-1)} \end{pmatrix} \text{.....(14)}$$

On peut alors écrire :

$$Y_t = \Phi Y_{t-1} + \Phi_0 + \varepsilon_t \text{.....(15)}$$

$$\Phi_0 = \begin{bmatrix} \Phi_0 \\ 0 \\ 0 \end{bmatrix} \quad \varepsilon_t = \begin{bmatrix} \varepsilon_t \\ 0 \\ 0 \end{bmatrix} \qquad \Phi = \begin{pmatrix} \Phi_1 & \Phi_2 & .. & \Phi_p \\ I_N & 0 & & \\ 0 & I_N & . & . \\ 0 & . & . & . \end{pmatrix} \text{.....(16)}$$

Ou I_N la matrice d'identité :

$$Y_t = \begin{pmatrix} xt \\ Xt-1 \\ . \\ . \\ Xt-(p-1) \end{pmatrix} = \begin{pmatrix} \Phi_1 & \Phi_2 & .. & \Phi_p \\ I_N & 0 & & \\ 0 & I_N & & \\ . & & & \\ 0 & . & . & . \end{pmatrix} * \begin{bmatrix} x_{t-1} \\ X_{t-2} \\ . \\ . \\ X_{t-(t-p)} \end{bmatrix} + \begin{pmatrix} \Phi_0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \\ 0 \end{pmatrix} \text{.....(17)}$$

4.2.2. Représentation canonique et processus des innovations :

Le processus VAR(p) centré, c'est-à-dire $\Phi_0 = 0$

$$\Phi(L) X_t = \varepsilon_t \text{.....(18)}$$

On peut écrire :

$$X_t = \Phi^{-1}(L)\varepsilon_t = \frac{\Phi(L)}{\det \Phi(L)} \varepsilon_t \text{.....(19)}$$

Remarque :

- Si toutes les racines des déterminants de $\Phi(L)$ son supérieur à 1 alors l'équation $\Phi(L)\ X_t = \varepsilon_t$ définit un unique processus VAR(p) stationnaire. On dit que X_t est en représentation canonique et ε_t est appelé l'innovation du processus[1].

[1]. Les innovations canoniques sont associées au processus VAR non contraints. Elles représentent des chocs ou impulsion dont propagation de traduit par fluctuation du système dynamique

- Si les racines de $det\,\Phi(L)$ sont inférieur à un, on peut changer les racines en leur inverse et modifier le bruit blanc associé afin de se ramener à la représentation canonique.

- Si au moins nue racine de $det\,\Phi(L)$ est égale à 1, le processus n'est plus stationnaire et on ne peut pas se ramener à une représentation canonique.

En représentation canonique, la prévision s'écrit :

$$E\left[X_{t+1}\middle|X_t\right]=\sum_{i=1}^{p}\Phi_i X_{t+1-i}\text{.....(20)}$$

X_t désigne le passé de X jusqu'à la date t incluse

4.2.3. Fonction d'auto covariance, fonction d'auto corrélation et densité spectrale :

Etudions les principales caractéristiques des processus VAR.

Considérons un processus VAR (1) :

$$X_t=\Phi_0+\Phi_1X_t-1+\varepsilon_t\text{.....(21)}\qquad\qquad \varepsilon\approx BB(0,\sum_{\varepsilon})$$

4.2.3.1. Espérance :

$$E\left[X_t\right]=E\left[\Phi_0+\Phi_1X_{t-1}+\varepsilon_t\right]\text{.....(22)}$$

Le processus étant stationnaire, on a : $E(X_t)=E(X_{t-1})$ on peut écrire sachant que $E(\varepsilon_t)=0$

$$E\left(X_t\right)=\Phi_0+\Phi_1(X_t)\text{.....(23)}$$

$$E\left(X_t\right)=(I-\Phi_1)^{-1}\Phi_0\text{.....(24)}$$

4.2.3.2. Fonction d'auto covariance :

Considérons le processus centré $Y_t=X_t-E(X_t)$ soit :

$$Y_t=\Phi_1Y_{t-1}+\varepsilon_t\text{.....(25)}$$

La fonction d'auto covariance I est donnée par :

étudié.Cependant, l'analyse statistique correspondante n'est facile à mettre en œuvre que dans le seul cas ou les impulsions sont corrélées instantanément.

$$I'(0) = E(Y_t Y_t^{'}) = E(\Phi_1 Y_{t-1} Y_t^{'} + \varepsilon_t Y_t^{'}).....(26)$$

$$E(\varepsilon_t Y'_t) = E(\varepsilon_t(\Phi_t Y_{t-1} + \varepsilon_t) = \Phi_1 E(\varepsilon_t Y'_{t-1}) + E(\varepsilon_t \varepsilon'_t).....(27)$$

Or comme $\varepsilon_t \approx BB(0, \sum_{\varepsilon})$. On a :

$$E(\varepsilon_t Y'_{t-1}) = 0.....(28)$$

On a donc

$$E(\varepsilon_t Y'_{t-1}) = E(\varepsilon_t \varepsilon_t^{'}) = \sum_{\varepsilon} .....(29)$$

En regardant l'équation (26) :

$$I'(0) = \Phi_1 E\left[Y_{t-1} Y_t^{'}\right] + \sum_{\varepsilon} .....(30)$$

Enfin, en remarquant que $E\left[Y_{t-1} Y_t^{'}\right] = I'(1)$ on déduit

$$I'(0) = \Phi_1 I'(1)' + \sum_{\varepsilon} .....(31)$$

On calcule la matrice d'auto covariance d'ordre 1 :

$$I'(0) = E\left[Y_t Y_{t-1}^{'}\right] = E\left[\Phi_1 Y_{t-1} + \varepsilon t\right] = \Phi_1 E(Y_{t-1} Y_{t-1}^{'}) = \Phi_1 I'(0).....(32)$$

On déduit la formule de récurrence suivante pour la matrice d'autocovariance d'ordre h d'un processus Var (1) :

$$I'(h) = \Phi_1 I'(h-1) \forall h \geq 1.....(33)$$

4.2.3.3. Fonction d'auto corrélation partielle : identification des processus VAR(p) :

Dans le cas univarié pour identifier le nombre de retard p d'un processus VAR on utilise la fonction d'auto corrélation partielle. En multivarié on dispose de la matrice d'auto corrélations partielles. Cependant celles-ci sont très difficiles à calculer, en général on n'utilise pas. En pratique, on impose des ordres de retard p suffisamment grands puis d'en réduire la taille à l'aide de tests. On choisit ainsi en général un VAR (4) pour données trimestrielles, un VAR (12) et VAR (8) pour données mensuelles, il existe des outils comme les critères d'information qui évitent de fixer la valeur de p afin que le nombre de paramètres à estimer pN^2 ne doivent pas trop grand. Même les tests de causalité vont permettre de

hiérarchiser les variables et réduire le nombre de paramètre à estimer voir notamment Caines et al (1981). Densité spectrale

Dans le cas des processus VAR, la densité spectrale est donnée par :

$$F(w) = \frac{1}{2II}\Phi^{-1}(e^{-iw})\sum_{\varepsilon}\Phi^{-1}(e^{iw})' \text{.....(34)}$$

Cette fonction donne la même information que l'auto covariance.

4.2.4. Estimation des paramètres d'un Var(p) :

4.2.4.1. Estimation par les moindres carrés ordinaire des VAR non contraints :

Considérons le processus VAR (p) :

$$\Phi(L)X_t = \varepsilon_t \text{.....(35)}$$

Ou $\varepsilon_t \approx BB(0, \sum_{\varepsilon})$

Déterminons tout d'abord le nombre de paramètres à estimer :

$\frac{N(N+1)}{2}$ Paramètres à estimer dans $\sum_{\varepsilon}$

N^2p paramètres à estimer dans Φ

Au total, on a donc $N^2 p + \frac{N(N+1)}{2}$ paramètres à estimer pour un VAR (p).

Décomposons l'écriture du VAR (p) La jième équation du VAR (p) s'écrit :

$$X_j = \begin{pmatrix} X_{j1} \\ X_{j2} \\ . \\ . \\ X_{jT} \end{pmatrix} = \begin{pmatrix} X'_0 & \ldots & .. & X'_{1-p} \\ X'_1 & .. & .. & X'_{2-p} \\ X'_{t-1} & \ldots & .. & X'_{t-p} \\ . & \ldots & .. & \\ X'_{T-1} & \ldots & .. & X'_{T-p} \end{pmatrix} \psi_j + \varepsilon_{j.} \text{.....(36)}$$

Soit encore :

$$X_j = X\psi_j + \varepsilon_j \text{.....(37)}$$

$$X = \begin{pmatrix} X'_0 & \dots & .. & X'_{1-p} \\ X'_1 & .. & .. & X'_{2-p} \\ X'_{t-1} & \dots & .. & X'_{t-p} \\ . & \dots & .. & \\ X'_{T-1} & \dots & .. & X'_{T-p} \end{pmatrix} \text{.....(38)} \qquad \varepsilon_{j.} = \begin{pmatrix} \varepsilon_{j1} \\ \varepsilon_{j2} \\ . \\ . \\ \varepsilon_{jT} \end{pmatrix} \text{.....(39)}$$

La variable contient T observations. La matrice X est de format (T, *Np*) soit une ligne X_t de cette matrice :

$$X_t^{'} = (X_{1t-1}X_{2t-1}X_{Nt-1}X_{1t-2}X_{Nt-2}X_{1t-p}X_{Nt-p}) \text{.....(40)}$$

Le modèle, est un processus VAR (p) à N composante indicées par le temps ψ_j est de dimension (Np,1). On a :

$$\psi_j = \begin{pmatrix} \Phi_{1j}^1 \\ \Phi_{1j}^2 \\ . \\ . \\ \Phi_{2j}^N \\ \Phi_{2j}^N \\ . \\ \Phi_{pj}^N \end{pmatrix} \qquad \varepsilon_j = \begin{pmatrix} \varepsilon_{j1} \\ \varepsilon_{j2} \\ . \\ . \\ \varepsilon_{jT} \end{pmatrix} \text{.....(41)}$$

La matrice X ne dépend pas de j :

$$X_j = X\psi_j + \varepsilon_j \text{.....(42)}$$

On empile les N équations pour retrouver le Var :

$$X = \begin{pmatrix} X_1 \\ X_2 \\ . \\ X_N \end{pmatrix} = \begin{pmatrix} X_{11} \\ X_{12} \\ . \\ X_{1T} \\ X_{2T} \\ . \\ X_{NT} \end{pmatrix} = \begin{pmatrix} X & 0 & 0 \\ 0 & X & 0 \\ . & . & X \end{pmatrix} \begin{pmatrix} \Psi_1 \\ \Psi_2 \\ . \\ \Psi_N \end{pmatrix} . + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ . \\ \varepsilon_{1T} \\ \varepsilon_{21} \\ . \\ \varepsilon_{NT} \end{pmatrix} \text{.....(43)}$$

On cherche à estimer $(\psi_1\psi_2....\psi_N)$

La matrice de variance – covariance des erreurs devient un peu compliquées et s'écrit :

$$\begin{pmatrix} \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{11} \end{pmatrix} \begin{pmatrix} \sigma_{12} & 0 \\ 0 & \sigma_{12} \end{pmatrix} & .. & \\ \begin{pmatrix} \sigma_{21} & 0 \\ 0 & \sigma_{21} \end{pmatrix} \begin{pmatrix} \sigma_{22} & 0 \\ 0 & \sigma_{22} \end{pmatrix} & .. & \\ . & . & \begin{pmatrix} \sigma_{NN} & 0 \\ 0 & \sigma_{NN} \end{pmatrix} \end{pmatrix} \quad \text{.....(44)}$$

L'observation de cette matrice indique la présence d'hétéroscédasticité (il n'y a, en effet aucune raison pour que $\sigma_{11}=\sigma_{22}=\ldots=\sigma_{NN}$ et d'auto corrélation. Il se pose en conséquence un problème pour l'application des MCO. Rappelons en effet que les estimateurs sont sans biais, mais ne sont plus de variance minimale. Il convient dès lors d'utiliser la technique des moindres carrées généralisés (MCG) qui fournit un estimateur Blue (*Best lineaur unbiaised Estimator*).

On peut réécrire la matrice de variance covariance comme :

$$V(\varepsilon) = \sum_{\varepsilon} * I = \Omega \text{.....(45)}$$

Ou $\sum_{\varepsilon} =$ $et*$ désigne le produit de Kronecker rappelons que :

$$A * B = a_{ij} B$$

Nous venons de voir que la matrice de variance des résidus est-elle que l'on devrait théoriquement appliquer les MCG. Cependant, puisque la matrice des variables explicatives est bloc diagonale, on peut appliquer les MCO bloc par bloc le théorème de Zellner nous montre ainsi qu'estimer chacune des I équations par les MCO est équivalent à estimer le modèle par les MCG. Afin de le prouver, considérant le modèle suivant :

$$Y = Xa + \varepsilon \text{.....(48)}$$

ε :Est bruit blanc.

Rappelons que l'estimateur des MCO est donnée par :

$$\hat{a}_{MCO} = (X'X)^{-1} X'Y \text{.....(49)}$$

Et que l'estimateur des MCG s'écrit :

$$\hat{a}_{MCO} = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}Y \text{.....(50)}$$

Ou Ω désigne la matrice de variance covariance de ε

Dans notre cas, on a :

$$X = \begin{pmatrix} X & 0 & 0 \\ 0 & X & 0 \\ . & . & X \end{pmatrix} + 1 * X \quad \text{.....(51)}$$

Ou I est la matrice identité.

Remarque : avant d'appliquer les MCG, rappelons que l'on a les égalités concernant le produit de Kronecker :

- $(A * B)(C * D) = AC * BD$

- $(A * B)' = A' * B'$

- $(A * B)^{-1} = A^{-1} * B^{-1}$

Afin de calculer l'estimateur des MCG, commençons par étudier la matrice :

$$X'\Omega^{-1}X = (I * X')(\sum_{\varepsilon}^{-1} *)(I * X) \quad \text{.....(52)}$$

Avec $\Omega^{-1} = (\sum_{\varepsilon}^{-1} *)I$

On déduite :

$$(X'\Omega^{-1}X)^{-1} = \sum_{\varepsilon} *(X'X)^{-1} \quad \text{.....(53)}$$

Pour la matrice $X'\Omega^{-1}X$, il vient :

$$X'\Omega^{-1}X = (I * X')(\sum_{\varepsilon}^{-1} * I)Y \quad \text{.....(54)}$$

$$= (\sum_{\varepsilon}^{-1} * X')Y \quad \text{.....(55)}$$

On a donc :

$$\hat{a}_{MCG} = \sum_{\varepsilon} *(X'X)^{-1}(\sum_{\varepsilon}^{-1} * X')Y \quad \text{.....(56)}$$

$$= (I * X'X)^{-1}X')Y \quad \text{.....(57)}$$

$$\hat{a}_{MCG} = \begin{pmatrix} (X'X)^{-1}X' & ..... & ..... \\ 0 & 0 & . \\ . & . & (X'X)^{-1}X' \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ . \\ . \\ Y_N \end{pmatrix} .....(57)$$

$$= \begin{pmatrix} (X'X)^{-1}X'Y_1 \\ (X'X)^{-1}X'Y_2 \\ . \\ (X'X)^{-1}X'Y_N \end{pmatrix}$$

On retrouve l'estimateur des MCO équation. Cependant, cette technique d'estimation des VAR n'est plus valable dès lors qu'il existe des contraintes sur les paramètres. Il convient alors d'utiliser la technique du maximum de vraisemblance.

4.2.4.2. La technique du maximum de vraisemblance :

Considérons un processus VAR (p) :

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + ... + \Phi_p X_{t-p} + \varepsilon_t .....(58)$$

ε_t est un bruit blanc de matrice de variance covariance $\sum_{\varepsilon}$

On écrit la vraisemblance conditionnellement à toutes les valeurs passées du processus :

$$L(X_1.....X_T) = \Pi_{t=1}^{T} L(X_t | X_{t-1}) .....(59)$$

Ou X_{t-1} désigne tout le passé de X_t jusqu'à la date (t-1) incluse.

$$L(X_1.....X_T) = \Pi_{t=1}^{T} \frac{1}{(\sqrt{2\pi})^N \sqrt{\det \sum_{\varepsilon}}} .....(60)$$

$$\exp\left[-\frac{1}{2} \sum_{t=1}^{T} (X_t - \Phi_t X_{t-1} - .. - \Phi_p X_{t-p})' \sum_{\varepsilon}^{-1} (X_t - \Phi_1 X_{t-1} - ... - \Phi_p X_{t-p}) \right]$$

On déduit l'expression de la log-vraisemblance :

$$LogL(X_1.....X_T) - \frac{NT}{2} \log 2\pi - \frac{T}{2} \log \det \sum_{\varepsilon} \quad - \frac{1}{2} \sum_{t=1}^{T} \varepsilon'_t \sum_{\varepsilon}^{-1} \varepsilon_{t.} .....(61)$$

On maximise ensuite cette expression afin d'obtenir les estimations de $\Phi_1 \ldots \Phi_p$ de $\sum_{\varepsilon}$

4.2.5. Validation : tests de spécification :

4.2.5.1. Test du rapport de maximum de vraisemblance :

On peut effectuer des tests sur l'ordre p du VAR. considérons le test suivant :

$H_0 : \Phi_{p+1} = 0$: Prosseces VAR (p)

$H_1 : \Phi_{p+1} \neq$: Processus VAR (p+1)

La matrice d'information de Fisher est difficile à calculer, ce qui explique que l'on peut utiliser un test du rapport du maximum de vraisemblance. La technique consiste à estimer un modèle (VAR(p) et un modèle non contraint (VAR (p+1)) et à effectuer le rapport des log-vraisemblances. Rappelons que la log-vraisemblance d'un processus VAR s'écrit :

$$LogL(X_1,.....X_T) = -\frac{NT}{2}\log 2\pi - \frac{T}{2}\log\det\sum_{\varepsilon} \quad -\frac{1}{2}\sum_{t=1}^{T}\varepsilon_t^{'}\sum_{t=1}^{-1}\varepsilon_t \text{.....(62)}$$

$\sum_{t=1}^{T}\varepsilon_t^{'}\sum_{t=1}^{-1}\varepsilon_t$ Est un scalaire, on a donc, en notant *Tr* la trace

$$\sum_{t=1}^{T}\varepsilon_t^{'}\sum_{t=1}^{-1}\varepsilon_t = Tr\left(\sum_{t=1}^{T}\varepsilon_t\sum_{\varepsilon}^{-1}\varepsilon_t\right) = Tr\left(\sum_{\varepsilon}^{-1}\sum_{t=1}^{T}\varepsilon_t\varepsilon_t^{'}\right)$$

$$= Tr\left(T\sum_{\varepsilon}^{-1}\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t\varepsilon_t^{'}\right)$$

$$= Tr\left(T\sum_{\varepsilon}^{-1}.\sum_{\varepsilon}\right) = T_r(TI_N)NT$$

Soient log L^C la log-vraisemblance estimé du modèle contraint :

$$LogL^C = \frac{NT}{2}\log 2\pi - \frac{T}{2}\log\det\sum_{\varepsilon}^{C} - \frac{1}{2}NT \text{.....(64)}$$

Et $logL^{NC}$ la log-vraisemblance estimé du modèle non contraint :

$$LogL^{NC} = \frac{NT}{2}\log 2\pi - \frac{T}{2}\log\det\sum_{\varepsilon}^{NC} - \frac{1}{2}NT \text{.....(65)}$$

Ou $\sum_{\varepsilon}^{NC}$ désigne l'estimateur de la matrice de variance covariance des résidus du modèle (respectivement non contraint)

On calcule la statistique de test ε=TT*RMV ou RMV désigne le rapport du maximum de vraisemblance :

$$\varepsilon = T \log \frac{\det \sum_{\varepsilon}^{C}}{\det \sum_{\varepsilon}^{NC}} \text{.....(66)}$$

Sous les l'hypothèse nulle, cette statistique suit une loi de Khi-deux à r degrés de liberté ou r désigne le nombre de contraintes.

Si l'on accepte l'hypothèse nulle, on peut effectuer un deuxième test :

H_0 :$Ø_p$=0 : prosseces VAR (p-1)

H_1 :$Ø_p$≠0 : processus VAR (p)

Ce test s'effectue de même façon que précédemment. On a ainsi une séquence de tests emboités le but est de déterminer l'ordre du processus VAR.

Remarque : dans le cas des processus AR en plus des tests sur les paramètres, on effectue des tests sur les résidus afin de valider le processus. Dans le cadre des processus VAR, ces tests ne sont pas très puissants et l'on préfère réaliser un graphe de résidus.Notons cependant qu'il convient d'examiner attentivement les résidus surtout lors d'utilisation des modèles VAR pour l'analyse de réponse impulsionnelle ou l'absence de corrélation des innovations est cruciale pour l'interprétation.

4.2.5.2. Critères d'information :

Afin de déterminer le nombre de retard p du processus VAR, on peut également utiliser les critères d'information. Ainsi, on estime un certain nombre de modèle VAR pour un ordre p allons de 0 à h ou h est le retard maximum. On retient le retard p qui minimise les critères AIC, SIC et Hannan-Quinn (HQ)[1] sont comme suites :

$$AIC = \log \det \sum_{\varepsilon} + \frac{2N^2 p}{T}$$

[1] Il existe d'autre critère d'information que ceux présentés ici- on pourra consulter sur ce point la revue de la littérature de Goojer et al (1985), que l'article de Deniau et Mathis (1992) pour une application VAR.

$$SIC = \log\det\sum_{\varepsilon} + \frac{\log T}{T}$$

$$HQ = \log\det\sum_{\varepsilon} + \frac{2N^2 p_2(\log T)}{T}$$

N est le nombre de variables du système T est le nombre d'observation t $\sum_{\varepsilon}$ Est un estimateur de la matrice de variance covariance des résidus. Les critères SIC et HQ à des estimateurs convergents de p le critère AIC donne des estimateurs efficaces de p.

4.2.6 Prévision des processus VAR :

Considérons un processus VAR (p) :

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + ... + \Phi_p X_{t-p} + \varepsilon_t \text{.....(70)}$$

On suppose que p a été choisi que Φ_i ont été estimés et que la matrice de variance covariance associée à ε_t à été estimée. Afin de réaliser des prévisions, il est nécessaire de vérifier que le modèle est bien en représentation canonique. Pour cela on calcule le déterminant du polynôme Φ (L) et l'on regarde si les racines sont bien à l'extérieur du disque unité. Si tel est le cas alors la prévision en (T+1) du processus est :

$$E(X_{T+1} \mid X_T) = \Phi_1 X_T + ... + \Phi_p X_{T-p+1.} \text{.....(71)}$$

X_T désigne le passé de x à la date T.

Afin de limiter le nombre de paramètres à estimer dans un processus VAR on peut effectuer des tests de causalité.

4.2.7. Les tests de causalité :

Soient Y_t,X_t deux processus aléatoires et I_t un univers, c'est-à-dire un ensemble de processus contenant notamment Y_t,X_t.

Désignons par It l'information relative au passé :

$$\underline{I_t} = (I_s \mid s \prec t) \text{.....(72)}$$

Et soit :

$$\underline{\underline{I_t}} = (I_s \mid s \leq) \text{.....(73)}$$

De la même façon, on introduit les notations :

$$\underline{Y}_t = (Y_s \mid s \leq) \quad et \quad \underline{\underline{Y_t}} = (Y_s \mid s \leq t) \text{.....(74)}$$

$$\underline{X}_t = (X_s \mid s \prec t) \quad et \quad \underline{\underline{X_t}} = (X_s \mid s \leq t) \text{.....(75)}$$

Si l'on suppose que tous les processus Y_t, $X_t$sont stationnaires, aux définitions générales suivantes :

(i) X cause Y si l'erreur de prévision de Y est telle que $\sigma^2 = (Y_t \mid I_t) \prec \sigma^2(Y_t \mid I_t - X_t)$ ou

$\underline{I_t} - \underline{X_t}$ est l'information obtenue en retirant de I les valeurs passées de X.

(ii) X cause instantanément Y si $\sigma^2 = (Y_t \mid I, X_t) \prec \sigma^2(Y_t \mid I_t)$.

Ou σ² désigne la variance de l'erreur de prévision[1]

La condition (i) signifié que pour prévoir Y_t, le passé de X apporte une information supplémentaire par rapport à la seule prise en compte des autres variables figurant dans I.

La condition (ii) signifie que la valeur présente de X apporte une information supplémentaire par rapport à la connaissance du passé des variables figurant dans I.

Les inégalités sont toujours satisfaites. Elles se transforment en égalités et si seulement si :

- i) X ne cause pas Y si $E(Y_t \mid I_t) = E(Y_t \mid I_t - X_t)$
- ii) X ne cause pas instantanément Y si $E(Y_t \mid I_t, X_t)$

La définition de la causalité est ici relative à univers. On distingue trois grands types de causalité, dont le plus utilisé la causalité au sens de Granger (1969)[2]

[1] Un processus réel $V = (V_t, t \in Z)$ et une information noté I_t on peut déterminer la prévision optimale de V_t sachant $I_t$cette dernière est notée $E = (V_t \mid I_t)$ l'erreur de prévision correspondante est $V_t - E = (V_t \mid I_t)$ et la variance de l'erreur de prévision est donnée par $\delta^2 = (V_t \mid I_t) = E(V_t - E(V_t \mid I_t))^2$

[2]Ainsi que le rappelle de Bruneau l'analyse de causalité apparait dans des travaux antérieurs à ceux de Granger (1969), on peut notamment citer Simon (1953) Stortz wold (1960) et Basmann (1963). La causalité y est introduite dans l'écriture du modèle de référence constitué d'un ensemble de relations fonctionnelles entre variables. Par ailleurs le lecteur intéressé pourra consulter avec profil le même article de Bruneau (1996) une représentation reposant sur des références structurelle que ceux

4.2.7.1. Causalité au sens de Pierce et Haugh (1977) :

Si I= (X, Y) alors la causalité de *X* vers *Y* peut être caractérisée par les corrélations des innovations des deux processus *X*, *Y*.

Soient μ_t le processus des innovations de $Y_t$etη_t le processus des innovations de X_t.

La fonction d'autocorrélation est donnée par :

$$P_{\eta\mu}(h) = \frac{Cov(\eta_t \mu_t)}{\sigma_\eta \sigma_\eta} \text{.....(76)}$$

Dans ce cas Y ne cause pas X si $P_{\eta\mu}(h) = 0\ h \succ 0$ (ou $h \geq 0$ si la causalité instantanée est exclue).

L'innovation de X_t doit être non corrélée avec toutes les innovations passées associées au processus Y_t. On peut donner une autre caractérisation de cette non causalité en termes la régression par le bais de notion de causalité au sens de Granger.

4.2.7.2. Causalité au sens de Granger (1969) :

De façon heuristique, on dira que X cause Y si la prévision de Y fondée sur la connaissance des passés conjoints de X et de Y est meilleurs que la prévision fondée sur la seule connaissance du passé de Y. Plus formellement, n reteindra la définition suivante :

Définition :

Causalité au sens de Granger :

X cause *Y* à la date *y* si :

$$E(Y_t \left| \underline{Y_{t-1}}, \underline{X_{t-1}} \right.) \neq E(Y_t \left| \underline{Y_{t-1}} \right.)$$

- *X* cause instantanément *Y* à la date t si :

$$E(Y_t \left| \underline{Y_{t-1}}, \underline{X_t} \right.) \neq E(Y_t \left| \underline{Y_{t-1}}, \underline{X_{t-1}} \right.)$$

- *X* ne cause pas *Y* à la date *t* si :

---

présentés dans le cadres de cet ouvrage la causalité reposant sur des références structurelles (basée sur des travaux de Feigl (1953) et différence entre causalité persistante et transitoire)

$$V_{\varepsilon}(Y_t \left| \underline{Y_{t-1}}, \underline{X_{t-1}} \right.) = V\varepsilon(Y_t \left| \underline{Y_{t-1}} \right.)$$

V_ε Désigne la matrice de variance covariance de l'erreur de prévision.

$$\underline{X_t} = \{X_{t-i}, i \geq 0\}, \underline{X_{t-1}} = \{X_{t-i}, i \geq 1\}, \underline{Y_{t-1}} = \{Y_{t-i} \geq 1\}$$

Rappelons que dans le cas linéaire, l'opérateur d'espérance conditionnelle désigne la meilleure prévision linéaire d'une variable fondée sur certain ensemble d'information. Si l'on suppose que les prévisions sont effectuées par régression linéaire. L'opérateur d'Espérance conditionnelle représente une fonction de régression. Ainsi $E(Y_t \left| \underline{Y_{t-1}}, \underline{X_{t-1}} \right.)$ désigne la régression linéaire de Y_t sur son passé Y_{t-1}et sur de X_t jusqu'à la date t-1 incluse (X_{t-1}). De même la notion $V_\varepsilon(Y_t \left| \underline{Y_{t-1}}, \underline{X_{t-1}} \right.)$ désigne la matrice de variance-covariance de l'erreur de prévision associé à la régression linéaire de Y_t sur son passé Y_{t-1}et sur le passé de X_t jusqu'à la date t-1 incluse ($\underline{X_{t-1}}$).

De même la notation $V\varepsilon(Y_t/Y_{t-1},X_{t-1})$ désigne la matrice de variance-covariance de l'erreur de prévision associée à la régression linéaire de Y_t sur son passé Y_{t-1} et sur le passé de X_t jusqu'à la date t-1 icluse (X_{t-1}).

Mesure de causalité : A partir de la définition, il est claire de définir les mesures de la causalité.

- Mesure de causalité de X vers Y :

$$C_{x,y} = \log \frac{\det V_\varepsilon \left[Y_t \left| Y_{t-1} \right. \right]}{\det V_\varepsilon \left[Y_t \left| Y_{t-1}, X_{t-1} \right. \right]} \text{.....(77)}$$

Bien évidemment si X ne cause pas Y,$C_{x,y}=0$ dans le cas contraire on $C_{x,y}>0$

- Mesure de causalité instantanée de X vers Y :

$$C_{x,y} = \log \frac{\det V_\varepsilon \left[Y_t \left| \underline{Y_{t-1}} \right. \right]}{\det V_\varepsilon \left[Y_t \left| \underline{Y_{t-1}}, X_t \right. \right]} \text{.....(78)}$$

4.2.7.2.1. Tests de non causalité :

Il est possible de tester l'hypothèse nulle de non causalité au moyen de statistique du rapport du maximum de vraisemblance :

$$E = TC_{x,y} \text{.....(79)}$$

Sous les hypothèses nulle cette statistique suit une loi de Khi-deux à r (T-r) p le degré de liberté ou r est le nombre de contraintes imposées. La règle de décision est la suivante :

Si $\varepsilon < X^2_{(r(T-r)p)}$ on accepte l'hypothèse nulle d'absence de causalité.

Si $\varepsilon \geq X^2_{(r(T-r)p)}$ on rejette l'hypothèse nulle en faveur de l'hypothèse alternative de présence de causalité.

4.2.7.3. Causalité au sens de Sims (1980) :

Sims (1980), a introduit un concept de causalité légèrement diffèrent. Il propose de considérer les valeurs futures de Y_t. Si les valeurs futures de Y peuvent permettre d'expliquer les valeurs présentes de X, alors que X est la cause de Y, de façon similaire on dira que X cause Y si les innovations de X contribuent à la variance de l'erreur de prévision de Y.

Considérons un processus VAR(p) à deux variables :

$$\begin{cases} Y_t = a_1^0 + \sum_{i=1}^{p} a_{1i}^1 Y_{t-i} + \sum_{i=1}^{p} a_{1i}^2 X_{t-i} + \sum_{i=1}^{p} b_i^2 X_{t+1} + \varepsilon_{1t} \\ Y_t = a_2^0 + \sum_{i=1}^{p} a_{2i}^1 X_{t-i} + \sum_{i=1}^{p} a_{2i}^2 Y_{t-i} + \sum_{i=1}^{p} b_i^2 Y_{t+1} + \varepsilon_{2t} \end{cases}$$

Dans ce cas :

Y ne cause pas *X* si l'hypothèse nulle suivante est vérifiée $H_0 : b_1^2 = b_2^2 = ... = b_p^2 = 0$.

X ne cause pas *Y* si l'hypothèse nulle suivante est vérifiée $H_0 : b_1^1 = b_2^1 = ... = b_p^1 = 0$.

Il s'agit là encore de tests de Fisher de nullité des coefficients.

4.2.8. Exogéniété :

Considérons un processus VAR (p) écrit sous forme matricielle :

$$\begin{bmatrix} Y_t \\ X_t \end{bmatrix} = \begin{bmatrix} a0 \\ b0 \end{bmatrix} + \begin{bmatrix} a_1^1 & b_1^2 \\ a_1^2 & b_1^2 \end{bmatrix} \begin{bmatrix} Y_{t-1} \\ X_{t-1} \end{bmatrix} + ... + \begin{bmatrix} a_p^1 & b_p^1 \\ a_p^2 & b_p^2 \end{bmatrix} \begin{bmatrix} Y_{t-p} \\ X_{t-1p} \end{bmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon 2t \end{pmatrix} \text{.....(82)}$$

Le bloc de variables $(X_{t-1}, X_{t-2},....X_{t-p})$ est exogène par rapport au bloc de variables $(Y_{t-1}, Y_{t-2},....Y_{t-p})$ Si le fait de rajouter le bloc de X_t n'améliore pas significativement la détermination des variables Y_t. On effectue un test de restrictions sur les coefficients de variables X_t de la représentation VAR plus précisément en accord avec Engle, Hendry et Richard (1983), une variable est dite strictement exogène si ces valeurs à chaque période[1] sont indépendantes des valeurs des perturbations aléatoires de toutes les périodes. Le concept de stricte exogéniété est ainsi lie au concept de causalité au sens de Granger (1969).

4.2.9. Analyse de réponse impulsionnelle :

La fonction repense impulsionnelle représente l'effet d'un choc d'une innovation sur les valeurs courantes et futures des variables endogènes. Un choc sur1ième variable peut affecter directement cette ième variable, mais il se transmet également à l'ensemble des autres variables au travers de la structure dynamique du VAR.

Nous considérons deux réalisations différentes de notre processus VAR X_t en t+T soit X_{t+T}. supposons que la première réalisation soit telle qu'entre t et t+T, le système connaisse un seul choc (cet intervenant en t). La deuxième réalisation suppose que le système ne subit pas de choc entre t et t+T. la fonction de réponse impulsionnelle est alors définie comme ses deux réalisation qui sont identiques jusqu'en t-1.

4.2.9.1. Représentation VMA d'un processus VAR :

Nous avons vu qu'un VAR (centré) en représentation canonique était donné par :

$$\Phi(L)X_t = \varepsilon_t \text{.....(83)}$$

Ou ε_t est un bruit blanc, soit encore :

$$X_t=\sum^{p}_{i=1}\phi_i X_{t-i}+\varepsilon_t \text{ .....(84)}$$

Dans ce cas, selon le théorème de Wold, ce processus peut être écrit comme suite :

[1] Notons que le concept de prédétermination est lié à la notion d'exogéniété. Une variable est dite prédéterminée en t si toutes ses valeurs présentes et passés sont indépendantes des erreurs présentes et si celles-ci ne sont pas auto-corrélées.

$$X_t = \sum_{i=0}^{\infty} \theta_j \varepsilon_{t-i} = \theta(L)\varepsilon_t \text{.....(85)}$$

$$\theta(L) = \sum_{j\geq 0}^{\infty} \theta_j L^j, \theta_0 = I \text{.....(86)}$$

Sur cette forme, ε_t représente le vecteur des innovations canonique du processus. Les innovations canoniques représentent la plus petite partie non révisable de chacune des variables composant le système VAR.

Cette représentation VMA est utile dans l'analyse de repense impulsionnelle. En effet, des innovations canoniques peuvent être interprétées comme chocs dont la propagation est caractérisée par la dynamique VAR ou d'une autre façon par les multiplicateurs dynamiques $\theta_j, j \geq 0$. C'est au travers des matrices qu'un choc s'introduit tout de long du processus.

On caractérise les réponses de différentes séries $X_{it}(i = 1,.....N)$ aux différentes innovations $\varepsilon_{is}(s \leq t)$ à partir des multiplicateurs dynamiques :

$$\Theta_{i,j,t-s} = \partial X_{it} / \partial \varepsilon_{js} \text{.....} (87)$$

Le multiplicateur $\theta_{i,j,t-s}$ représente l'effet du choc j sur la variable i, h périodes après le choc.

4.2.9.2. VAR structurel :

Le processus VAR en représentation canonique pouvait prendre la forme d'un processus VMA (∞). Et peut même prendre une représentation d'un Var structurel.

Soit W_t le vecteur des chocs structurels. Il s'agit de chocs interprétables économiquement. On suppose que l'économie, représentée par un vecteur de séries observables $X_t = (X_{1t}.....X_{Nt})'$.

A chaque date t résulte de la combinaison dynamique de N chocs structurel passés

$$(w_{1s}, ..., w_{Ns}), s \leq t$$

La représentation VAR structurel se déduit de la représentation VAR canonique en supposant que le vecteur des innovations canoniques ε_t est une combinaison linéaire des innovations structurelles w_t de la même date :

$$\varepsilon_t = Pw_t, \text{.....(88)}$$

P est une matrice de passage (inversible et de dimension N*N) qui doit être estimée.

On considère la représentation canonique suivante :

$$X_t = \sum_{i=1}^{p} \Phi_i X_{t-i} + \varepsilon_t \text{.....(89)}$$

Et que l'on pré multiple les deux membres par la matrice $\hat{P}^{-1}$ ($\hat{P}$ étant un estimateur de p)

$$\hat{P}^{-1} X_t = \hat{P}^{-1} \sum_{i=1}^{p} \Phi_i X_{t-i} + \hat{P}^{-1} \varepsilon_t \text{.....(99)}$$

Soit encore :

$$X_t = X_t - \hat{P}^{-1} X_t + \sum_{i=1}^{p} \hat{P}^{-1} \Phi_i X_{t-i} + \hat{P}^{-1} \varepsilon_t \text{.....(91)}$$

On déduit l'expression du processus Var structurel :

$$X_t = \sum_{i=0}^{p} \psi_i X_{t-i} + \psi_t \text{ .....(92)}$$

Avec $w_t = \hat{P}^{-1} \varepsilon_t \psi_0 = I - \hat{P}^{-1}$ et $w_i = \hat{P}^{-1} \Phi_t$, pour $I \leq i \leq p$

On constate que l'estimation du modèle Var structurel est estimée, l'identification des chocs est réalisée puisqu'il est possible de passer des chocs estimes aux chocs structurels (interprétable économiquement) par :

$$\hat{w}_t = \hat{P}^{-1} \varepsilon_t \text{.....(93)}$$

Si les chocs ont été correctement identifiés et si les effets significatifs et conformes à la théorie, alors l'intérêt économique de l'analyse impulsionnelle est qu'est permet de mesurer et d'anticiper les effets d'une politique économique.

4.2.9.2. Orthogonalisation des chocs :

Exemple introductif

Considérons par exemple le processus VAR(1) suivant :

$$\begin{cases} Ip_t = a_{11} IP_{t-1} + a_{12} M_{t-1} + \varepsilon_{1t} \\ M_t = a_{21} IP_{t-1} + a_{22} M_{t-1} + \varepsilon_{2t} \end{cases}$$

Ou IP représente la production industrielle et m l'offre de monnaie (masse monétaire). On voit qu'un choc sur ε_1 affectera immédiatement la valeur présente de IP. Il affectera cependant

également les valeurs futures d'IP et M puisque les valeurs passées d'IP interviennent dans les équations.

Si les innovations $\varepsilon_{1,}$ ε_2 ne sont pas corrélés, l'interprétation de la fonction de réponse impulsionnelle est très simple. En effet ε_{1t} est l'innovation de : IP et ε_{2t} est l'innovation de M. la fonction de réponse impulsionnelle pour ε_{2t} mesure l'effet d'un choc monétaire sur les valeurs présentes et passées de la production industrielle et de l'offre de monnaie.

Cependant, en pratique, les innovations sont généralement corrélées. Elles ont donc une composante commune qui ne peut pas être associée à une variable spécifique. Une méthode quelque peu arbitraire mais très fréquemment utilisée consiste à attribuer la totalité de l'effet de la composante commune à la variable qui intervient en premier dans le système VAE (ici c'est IP). Dans l'exemple la composante commune de ε_{1t}, ε_{2t} est totalement attribué à ε_{1t} car ε_{1t} précède ε_{2t}. ε_{1t} est l'innovation d'IP et ε_{2t} est transformée de façon à éliminer la composante commune.

De façon plus technique, les erreurs peuvent être orthogonalisées en utilisant la décomposition de Cholesky: la matrice de variance covariance des innovations qui en résulte est diagonale.

Position du problème

Afin d'interpréter une analyse de réponse impulsionnelle, il faut que les chocs (les innovations canoniques) ne soient pas corrélés entre eux. Si tel n'est pas le cas, alors l'analyse de la propagation des chocs est rendue délicate, même impossible. Il faut orthogonaliser les chocs à l'aide d'une transformation linéaire. En effet en multipliant le vecteur des innovations canoniques par une matrice P préalablement définie, on obtient des innovations interprétables car non corrélées instantanément.

On peut retenir divers matrices P. certaines ne font références à aucune théorie économique. C'est le cas de la matrice issue de la décomposition de Cholesky. Même si cette technique est fréquemment utilisée, les résultats obtenus dépendent forcément de l'ordre dans lequel on range les séries, puisque la matrice obtenue est triangulaire inférieure[1]. Mais l'inconvénient de cette méthode est qu'on ne peut pas réduire d'interprétation économique des

[1] Cette décomposions est la méthode préconisée par Sim (1980) dans ses premiers travaux.

impulsions obtenues puisque l'ordre établi des variables n'est justifié que par des méthodes purement statistiques.

L'approche des modèles VAR structurels répond à cette critique en permettant d'identifier les chocs interprétables économiquement, puisque les matrices utilisées font explicitement référence à la théorie économique. Ainsi depuis les travaux de Shapiro et Waston (1988) et Blanchard et Quanh (1989), la matrice P d'Orthogonalisation est choisie de manière à pouvoir interpréter économiquement les chocs transformés comme des chocs d'offre, de demande, de politique monétaire ou budgétaire, etc.…dont on connait a priori l'effet économique. C'est l'identification des chocs par introduction de contraintes identifiâtes structurelles, c'est-à-dire déduite par la théorie économique.

4.2.9.3. Méthodes d'identification des chocs :

La matrice de passage P comprend N² paramètres inconnus. En général, pour faciliter l'identification de ces paramètres on suppose que :

$$V(w_t) = I \text{.....(95)}$$

Ceci signifie que les différents chocs structurels à une date ne sont pas corrélés entre eux et ont une variance unitaire.

Soit $\sum_{\varepsilon}$ la matrice de variance covariance des innovations canoniques ε_t. On a alors :

$$V(\varepsilon_t) = P^{-1}V(w_1)P^{-1'} = P\ \ P' = \sum_{\varepsilon} \text{.....(96)}$$

Puisque la matrice$\sum_{\varepsilon}$ est systématique, on impose avec l'hypothèse (95) seulement N (N+1)/2 contraintes sur les éléments de la matrice P. Ces contraintes sont appelées contraintes d'Orthogonalisation. Et pour identifier les N² éléments de la matrice P, il reste à imposer N (N-1)/2 contraintes supplémentaire pour estimer le modèle VAR structurel. Ces contraintes supplémentaires sont appelées contraintes identifiâtes structurelles.

Remarque :

Les contraintes d'Orthogonalisation ont une justification plus technique qu'économique. On peut en effet se demander par exemple pourquoi un choc d'offre ne doit pas être lié à chaque instant avec un choc de demande. C'est un point faible pour les VAR structurel.

Les N(N-1)/2 contraintes supplémentaires portent toujours sur les réponses du système au différentes impulsions structurelles. Ces contraintes présentent des faiblaisses de la méthode VAR structurel : le nombre N(N-1)/2 de contraintes utilisé par l'économètre pour analyser la situation)économique augmente rapidement avec le nombre de variable introduites.

4.2.9.3.1. La décomposition de Cholesky :

Cette méthode est celle préconisée par Sims (1980). Il s'agit d'une méthode statistique d'imposer les N(N-1)2 contraintes supplémentaires. Pour imposer ces contraintes, Sims propose d'utiliser une matrice de passage P la décomposition de Cholesky de la matrice de variance covariance des innovations canoniques. La décomposition de Cholesky fournit l'unique matrice covariance triangulaire inférieur P telle que PP'=$\sum_{\varepsilon}$.

Cette méthode ne montre aucune a priori économique seulement le choix de l'ordre des séries ; elles doivent être rangées de la plus exogène à la plus endogène. La matrice P correspondant à la décomposition de Cholesky est alors définie de manière unique pour un ordre donné des composantes du VAR.

Par exemple un VAR de dimension 2 avec $X_t = (X_{1t}, X_{2t})$. En représentation canonique, ce processus s'écrit :

$$\begin{pmatrix} \Phi_{11}(L) & \Phi_{12}(L) \\ \Phi_{21}(L) & \Phi_{22}(L) \end{pmatrix} \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \text{.....(97)}$$

Ecrivons à présent la représentation Var structurelle. P est la matrice de passage issue de la décomposition de Cholesky de $\sum_{\varepsilon}$ et telle que $\varepsilon_t = Pw_t$

On a donc :

$$\begin{pmatrix} \Phi_{11}(L) & \Phi_{12}(L) \\ \Phi_{21}(L) & \Phi_{22}(L) \end{pmatrix} \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{pmatrix} \begin{pmatrix} w_{1t} \\ w_{2t} \end{pmatrix} \text{.....(98)}$$

Ainsi, la seconde innovation structurelle n'a aucun impact courant sur la première. Ce choix se justifie par le fait que les séries sont influencées par un nombre croissant de variables contemporaines et sont donc de plus en plus « endogènes ».

Cette méthode d'Orthogonalisation a été remise en cause par les partisans « durs » de l'approche VAR structurel pour lesquels les contraintes identifiâtes doivent être issues de la

théorie économique. Dans ce cas, les N(N-1)2 contraintes supplémentaires identifiâtes portent sur l'effet de court et de long termes des chocs structurels sur les différentes composantes du système.

4.2.9.3.2. Les contraintes de court terme :

Lorsque la dynamique est stationnaire[1], ce sont toujours des contraintes de court terme qui sont imposées. Celles-ci expriment l'absence de réponse instantanée de certaine série à certaines impulsions structurelles. D'un point de vue pratique, les contraintes de court terme se traduisent simplement par la nullité d'un certain nombre de coefficients dans la matrice P.

Pour le voir, il suffit d'écrire le processus VAR sous forme moyenne mobile structurelles.

$$X_t = \sum_{i=0}^{\infty} \theta_j \varepsilon_{t-j} = \theta(L)\varepsilon_t \text{.....(99)}$$

On peut encore écrire en introduisant la matrice de passage P :

$$X_t = \sum_{j=0}^{\infty} \theta_j P P^{-1} \varepsilon_{t-j} \text{.....(100)}$$

Soit finalement la forme VMA structurelle :

$$X_t = \sum_{j=0}^{\infty} \Omega_j w_{t-j} \text{.....(101)}$$

Avec $w_t = P^{-1}\varepsilon_t \ et \ \Omega_j = \theta_j P \forall J$

Ω_j Représente la matrice des multiplicateurs dynamiques structurels :

$$\Omega_{i,j,t-s} = \frac{\partial X_{it}}{\partial w_{is}} \text{.....(102)}$$

[1] Supposons l'absence de tendance déterministe. Dans ce cas si Xt est stationnaire on peut estimer un modèle Var sur cette série. On parle de modèle Var en niveau. Si Xt est non stationnaire, on la stationnaire en la différenciation $\Delta X_t = X_t - X_{t-1}$ on peut dès lors estimer un modèle Var sur ΔX_t on parle dans ce cas d'un modèle Var en différence.

Ainsi le multiplicateur $\Omega_{i,j,h}$ représente l'effet du choc structurel j sur la variable i, h périodes après le choc. En prenant l'écriture VMA structurelle, on constate que les réponses instantanées sont données par :

$$\frac{\partial X_{it}}{\partial w_{js}} = \Omega_{i,j,0} = P_{i,j} \text{.....(103)}$$

$$\Omega_{i,j,0} = \theta_{ij,0P} \text{.....(104)}$$

$$\Omega_{ij,0} = I \text{.....(105)}$$

Les réponses instantanées aux chocs sont fournies par les éléments de la matrices p. les contraintes identifiâtes de court terme se traduisent donc par la nullité de certains éléments de cette matrice.

4.2.9.3.3. Les contraintes de long terme :

Lorsque la dynamique du VAR n'est pas statiquement pas stationnaire, on peut introduire des contraintes de long terme. Ces contraintes expriment le fait que certaines impulsions structurelles n'ont pas d'effet de long terme sur certaines composantes du système.

Les effets de long terme sont caractérisés par le multiplicateurs dynamiques de long terme définis à partir de l'écriture VMA (ou décomposition de Wold) du VAR structurel en différence première.

$$\Delta X_t = \sum_{h=0}^{\infty} \Omega_h w_{t-h} \text{.....(106)}$$

On a de plus

$$X_{it} = \sum_{h=0}^{t-1} \Delta X_{it-h} + X_{i0} \text{.....(107)}$$

La réponse de X_{it} ou choc w_{js} soit $\frac{\partial X_{it}}{\partial w_{js}}$ est égale au cumul des réponses des différences premières $\frac{\partial \Delta X_{it-h}}{\partial w_{js}}, h \le t-s$ à ce même choc. Puisque $\frac{\partial \Delta X_{it-h}}{\partial w_{js}} = \Omega_{i,j,h}$, la réponse de X_{it} au choc w_{js} est donc égale à $\sum_{h=0}^{t-s} \Omega_{i,j,h}$

La réponse de long terme, notée $\Omega_{i,j}(1)$ est finalement en faisant tendre t vers l'infini :

$$\Omega_{i,j}(1) = \lim_{\infty} \sum_{h=0}^{t-s} \Omega_{i,j,h} \text{.....(108)}$$

Cette équation est le multiplicateur dynamique de long terme.

Remarque : Un effet de long terme d'une impulsion w_j sur une série Xi est aussi un effet persistant, c'est-à-dire un effet sur la composante permanente de la série telle qu'elle définit à partir d'une décomposition tendance cycle suivant les principes de Beveridge et Nelson (1981).

Une contrainte de long terme exprime l'absence de réponse à long terme d'une composante Xi à une impulsion w_j et se traduit par la nullité du multiplicateur dynamique de long terme $\Omega_{ij}(1)$ correspondant. $\Omega_{ij}(1) = 0$ représente une contrainte linéaire sur les éléments de la matrice P puisque $\Omega(1) = \theta(1)P.$ la contrainte de long terme s'écrit comme suit :

$$\Omega_{ij}(1) = (\theta(1)P)_{i,j} = 0 \text{.....(109).}$$

Ce qui est équivalent

$$\sum_{K=1}^{N} \theta_{iK}(1) P_{Kj} = 0 \text{.....(110).}$$

Il est important de noter qu'une contrainte de long terme ne peut porter que sur les réponses d'une série stationnaire en différence et en aucun cas sur les réponses d'une série stationnaire.

Dans ce dernier cas, la décomposition de Wold caractérise la dynamique de la série en niveau et non différenciée, de sorte que le cumul des réponses ne correspond à aucune mesure pertinente.

Conclusion :

D'une manière générale, les phénomènes de contagion apparaissent désormais comme un facteur explicatif incontournable du déclanchement des crises financières, plusieurs études théoriques et empiriques ont été élaborées dans le but d'expliquer les déterminants, les causes et les canaux de transmission d'un choc d'un pays à un autre.

On oppose deux types de contagion ; une contagion mécanique induite par les interdépendances réelles et financières entre pays et une contagion psychologique qui mettent en jeu le comportement des investisseurs. Cette contagion psychologique ou contagion pure met en lumière le fait que la transmission d'une crise peut être relative davantage au comportement des investisseurs qu'a l'évolution des fondamentaux macroéconomiques des pays concernés.

La modélisation de ce consensus a fait l'œuvre d'innombrables études empiriques ayant pour cible d'avoir une meilleure mesure afin de mieux cerner ces effets.

Nous allons ainsi recourir, dans le chapitre suivant, à une investigation empirique du phénomène de contagion en choisissant le cas de la crise de « subprime » des Etats-Unis de 2007 étant qu'un champ d'application, on se demande alors de savoir s'il existe une contagion entre les économies faisant l'objet de l'étude tout en se basant sur des modèles et des tests économétriques.