

CHAPITRE 2

Modèles VAR, modèles VAR structurels

Michel LUBRANO

March 17, 2008

Contents

1	Introduction	2
2	Introduction aux modèles VAR stationnaires	2
2.1	Ecrire un VAR(p) comme un VAR(1)	3
2.2	Autocovariance	4
2.3	Représentation MA	5
2.4	Moments dans un modèle VAR stationnaire	6
2.5	Prévision dans un VAR(1)	7
2.6	Moments de l'erreur de prévision et représentation MA(∞)	8
2.7	Inférence	8
2.8	Choix de modèles	9
2.9	Tests de mauvaise spécification	10
3	Analyse structurelle	10
3.1	Non-causalité au sens de Granger	11
3.2	Tests de causalité	12
3.3	Non causalité instantanée	12
3.4	Analyse impulsionnelle	12
3.5	Décomposition de la variance	13
4	Modèles VAR structurels	15
4.1	Le modèle théorique	15
4.2	Dynamique des chocs d'offre et de demande aux USA	16
5	Conclusion	18

1	INTRODUCTION	2
6	Lectures additionnelles	19
7	Exercices	19
7.1	Développement d'un VAR(1)	19
7.2	Equation caractéristique	19
7.3	Prévision	19
7.4	Examen 2007	19

1 Introduction

Dans ce chapitre, nous allons nous intéresser à la modélisation jointe de séries temporelles. Dans les modèles de régression que nous avons examinés, on expliquait le comportement d'une variable endogène par ses retards et par une ou plusieurs variables exogènes et leurs retards. Ici, nous allons maintenant nous intéresser à la modélisation jointe d'un groupe de variables sans distinguer dans un premier temps variables endogènes et variables exogènes. Les modèles mis en oeuvre sont une généralisation directe des modèles autorégressifs. Il s'agit des modèles VAR (Vector AutoRegressive) qui ont été popularisées par Sims (1980) et puis développés dans Sims (1981). Plus que d'estimer des coefficients individuels, il s'agira d'estimer la dynamique générale d'un système et d'arriver à décrire son comportement par rapport à un choc sur les termes d'erreurs. Les modèles VAR initiaux sont intrinsèquement des formes réduites dont les paramètres ne sont pas interprétables. L'analyse impulsionnelle repose sur une décomposition de la matrice de variance covariance des résidus qui est arbitraire. Aussi, Sims (1986) et d'autres auteurs ont proposé une classe plus générale de modèles VAR, les modèles VAR structurels. L'identification des chocs repose alors sur un choix raisonné de restrictions d'identification.

2 Introduction aux modèles VAR stationnaires

Un modèle autoregressif simple modélise une série univariée sous la forme

$$y_t = c + \alpha y_{t-1} + \epsilon_t$$

Ce modèle est stationnaire quand $|\alpha| < 1$. Un modèle AR(p) sera stationnaire si les racines en z de l'équation caractéristique

$$\alpha_1 z + \alpha_2 z^2 + \dots + \alpha_p z^p = 0$$

sont toutes en dehors du cercle unité.

On va maintenant considérer une modélisation similaire pour une série multivariée X_t de \mathbb{R}^n . Un modèle VAR à l'ordre 1 se notera

$$X_t = m + AX_{t-1} + \epsilon_t$$

avec $E(\epsilon_t) = 0$ et $E(\epsilon_t \epsilon_t') = \Sigma$. La matrice A est une matrice carrée $n \times n$ de coefficients. Le modèle sera stationnaire si les valeurs propres de cette matrices sont toutes plus petites

que 1 en valeur absolue, les valeurs propres étant définies comme les racines de l'équation

$$|I_n \lambda - A| = 0.$$

On peut développer cette équation sous la forme d'un polynôme de degré n en λ . Par exemple pour une matrice 2×2 :

$$\begin{aligned} \left| \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} - \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \right| &= \left| \begin{pmatrix} \lambda - a_{11} & -a_{12} \\ -a_{21} & \lambda - a_{22} \end{pmatrix} \right| \\ &= \lambda^2 - \lambda(a_{11} + a_{22}) + (a_{11} + a_{22}) - a_{12}a_{21}. \end{aligned}$$

Il y a donc deux racines réelles ou complexes. La dynamique a dépendre du signe et de la valeur de ces racines. De manière équivalente, on aurait pu écrire l'équation caractéristique du polynôme matriciel de retards

$$|I - Az| = 0.$$

Les racines de cette équation sont simplement l'inverse des valeurs propres de la matrice A . La stationnarité implique que ces racines soient toutes plus grandes que 1 en valeur absolue.

2.1 Ecrire un VAR(p) comme un VAR(1)

On peut généraliser le modèle initial à l'ordre p en considérant

$$X_t = m + A_1 X_{t-1} + \dots + A_p X_{t-p} + \epsilon_t.$$

Il est alors utile de définir le polynôme matriciel de retards $A(L)$:

$$A(L) = I_n - A_1 L - A_2 L^2 - \dots - A_p L^p,$$

et l'on va noter le modèle sous la forme

$$A(L)X_t = m + \epsilon_t.$$

On peut toujours étudier les conditions de stationnarité à partir de ce polynôme matriciel, mais on préfère considérer une forme plus simple. Il est en effet possible par une simple reparamétrisation de réécrire un VAR(p) sous la forme d'un VAR(1) plus grand. Pour cela, on va définir une nouvelle variable par empilement de la variable initiale et de ses retards (en laissant tomber m pour l'instant). Par exemple si $p = 3$, cela donne

$$Y_t = \begin{pmatrix} X_t \\ X_{t-1} \\ X_{t-2} \end{pmatrix}.$$

Ce vecteur est donc à $n \times p$ lignes. On définit ensuite une matrice carrée $np \times np$ de paramètres F comme

$$F = \begin{pmatrix} A_1 & A_2 & A_3 \\ I_n & 0 & 0 \\ 0 & I_n & 0 \end{pmatrix}$$

et un vecteur de résidus

$$v_t = \begin{pmatrix} \epsilon_t \\ 0 \\ 0 \end{pmatrix}.$$

On peut alors écrire le VAR(p) initial comme le VAR(1) suivant

$$Y_t = FY_{t-1} + v_t.$$

Cette écriture est commode car elle justifie la simplification qui consiste à se contenter d'étudier les conditions de stationnarité d'un VAR(1).

Ce modèle sera stationnaire si toutes les valeurs propres λ de la matrice F sont plus petites que 1, ce qui signifie que l'on doit chercher en λ les solutions de l'équation

$$|\lambda I - F| = 0.$$

On peut développer cette équation selon les composantes de la matrice F . Les valeurs propres de la matrice F sont obtenues comme solutions de l'équation:

$$|I_n \lambda^p - A_1 \lambda^{p-1} - \dots - A_p| = 0.$$

On obtiendra la stationnarité si toutes les valeurs propres sont en module inférieures à 1.

Il est aussi d'usage d'étudier les racines de l'équation caractéristique suivante:

$$|I_n - A_1 z - A_2 z^2 - \dots - A_p z^p| = 0.$$

Il s'agit d'une équation qui ressemble fort à la précédente, mais cette fois-ci les puissances de la variable z vont par ordre croissant au lieu d'aller par ordre décroissant. On aura la stationnarité quand les racines en z de cette équation seront toutes en module supérieures à 1.

2.2 Autocovariance

On va noter μ le vecteur des espérances de X_t . L'autocovariance de la série $\Gamma(h)$, h donné est une matrice

$$\Gamma(h) = E((X_t - \mu)(X_{t-h} - \mu)') \quad (1)$$

Pour $h = 0$, on a la matrice de variance-covariance de la série. Sur la diagonale de cette matrice, se trouve la variance de chaque série, et sur les éléments hors diagonaux, les covariances entre deux séries. On peut alors généraliser au cas multivarié la définition de la stationnarité au second ordre que l'on a donné pour les processus univariés:

Définition 1 *un processus stochastique multivarié X_t de \mathbb{R}^n est stationnaire au second ordre si sa moyenne μ et ses autocovariances $\Gamma(h)$ sont bornées et indépendantes de t .*

Nous montrerons que l'on ne peut bien sûr calculer ces deux moments que si les conditions de stationnarité exprimées sur la forme VAR au moyen des valeurs propres sont remplies.

2.3 Représentation MA

On peut mettre un modèle VAR sous une forme MA(∞) en écrivant

$$X_t = \mu + \sum_{j=0}^{\infty} C_j \epsilon_{t-j}$$

où C_j est une suite de matrices carrées de taille $n \times n$ avec $C_0 = I_n$. Il est d'usage d'écrire cette représentation sous une forme matricielle $C(L)$ avec

$$C(L) = \sum_{j=0}^{\infty} C_j L^j \quad |C(1)| \neq 0. \quad (2)$$

Quelle est la relation avec la forme VAR(p)? On a bien évidemment que $C(L) = A^{-1}(L)$ que l'on va résoudre par récurrence en posant

$$C(L) A(L) = I. \quad (3)$$

La récurrence est la suivante:

$$\begin{aligned} C_0 &= I \\ C_1 C_0 A_1 &= 0 \\ C_2 - C_1 A_1 - C_0 A_2 &= 0 \\ \dots &\dots \\ C_i - \sum_{j=1}^i C_{i-j} A_j &= 0 \quad \text{avec} \quad A_j = 0 \text{ pour } j > p \end{aligned}$$

Si $p = 1$, on a bien évidemment $C_i = A^i$. Car on peut généraliser le cas connu dans le cadre univarié. Dans le VAR(1) $X_t = AX_{t-1} + \epsilon_t$, on a la relation:

$$\sum_{j=0}^{\infty} A^j = (I_n - A)^{-1}.$$

On peut partir de la représentation VAR en $t = 1$, $X_1 = AX_0 + \epsilon_1 + \mu$ et développer la suite récurrente:

$$\begin{aligned} X_1 &= AX_0 + \epsilon_1 + \mu \\ X_2 &= AX_1 + \epsilon_2 + \mu = A^2 X_0 + A\epsilon_1 + A\mu + \epsilon_2 \\ X_3 &= AX_2 + \epsilon_3 + \mu = A^3 X_0 + A^2 \epsilon_1 + A^2 \mu A\mu + A\epsilon_2 + \epsilon_3 + \mu \\ \dots & \\ X_t &= A^t X_0 + \sum_{j=0}^{t-1} A^j (\epsilon_{t-j} + \mu) \end{aligned}$$

2.4 Moments dans un modèle VAR stationnaire

Le modèle VAR est particulièrement populaire pour modéliser les séries temporelles multivariées à cause de sa simplicité et de sa richesse. Il permet de modéliser les interactions existantes entre les variables. Détaillons maintenant l'expression des moments et des prédicteurs.

Considérons le modèle VAR suivant avec comme terme déterministe, un terme constant qui sera la moyenne du processus:

$$A(L)(X_t - \mu) = \epsilon_t \quad \text{avec } \epsilon_t \sim N(0, \Sigma) \quad (4)$$

que l'on peut encore noter :

$$A(L)X_t = m + \epsilon_t \quad (5)$$

avec $m = A(1)\mu$. Si le processus est stationnaire, on peut passer à la forme MA(∞) en multipliant chaque terme par $A^{-1}(L) = C(L) = I + C_1L + C_2L^2 + \dots$:

$$X_t = \mu + C(L)\epsilon_t \quad (6)$$

Pour ce qui concerne le terme constant, on a utilisé cette fois $C(1)m = \mu$. Cette forme moyenne mobile est particulièrement utile pour définir les moments théoriques du processus, car on voit tout de suite que :

$$\begin{aligned} E(X_t) &= \mu \\ \text{Var}(X_t) &= \sum_{i=0}^{\infty} C_i \Sigma C_i' \end{aligned} \quad (7)$$

L'auto-covariance de X_t se définit comme

$$\Gamma(h) = E[(X_t - \mu)(X_{t-h} - \mu)] = \sum_{i=0}^{\infty} C_{i+h} \Sigma C_i' \quad (8)$$

L'expression de ces moments fait intervenir la suite infinie des coefficients de la représentation MA. Mais comme c'est la forme VAR que l'on va estimer, on peut chercher à exprimer l'auto-covariance en fonction de ces paramètres là et l'on obtiendra alors une formule par récurrence pour $\Gamma(h)$. Prenons l'exemple d'un VAR(1) que l'on va écrire

$$X_t - \mu = A(X_{t-1} - \mu) + \epsilon_t$$

Pour calculer $\Gamma(h)$, on va remplacer dans sa définition (8) $X_t - \mu$ par son expression donnée par ce modèle, c'est à dire $A(X_{t-1} - \mu) + \epsilon_t$. On a donc

$$\Gamma(h) = AE[(X_{t-1} - \mu)(X_{t-h} - \mu)'] + E[\epsilon_t(X_{t-h} - \mu)'].$$

En posant $h = 0$, on trouve immédiatement

$$\Gamma(0) = A\Gamma(-1) + \Sigma = A\Gamma(1)' + \Sigma,$$

car

$$E[\epsilon_t(X_t - \mu)'] = E[\epsilon_t(A X_{t-1} - \mu)'] + E(\epsilon_t \epsilon_t') = \Sigma.$$

Passons maintenant aux ordres supérieurs pour dégager une relation de récurrence. Pour $h = 1$, on a :

$$\Gamma(1) = A\Gamma(0) + 0,$$

et pour les ordres supérieurs on obtient la récurrence

$$\Gamma(h) = A\Gamma(h - 1) = A^h\Gamma(0).$$

Pour un VAR(p), cette récurrence se généralise à

$$\Gamma(h) = A_1\Gamma(h - 1) + \dots + A_p\Gamma(h - p).$$

Rappelons que l'on obtient le même type de relation dans un modèle AR(p).

2.5 Prévision dans un VAR(1)

La prévision de X à l'horizon h est traditionnellement définie comme l'espérance conditionnelle de X_{t+h} calculée en utilisant l'information disponible au temps t :

$$E_t(X_{t+h}) = E(X_{t+h}|X_1^t) \quad (9)$$

où X_1^t représente tout le passé de X entre 1 et t . Considérons un modèle VAR d'ordre 1 :

$$X_t = m + A X_{t-1} + \epsilon_t \quad (10)$$

Pour $h = 1$, on va écrire le modèle en $t + 1$, ce qui fait

$$X_{t+1} = m + A X_t + \epsilon_{t+1}, \quad (11)$$

et prendre l'espérance conditionnelle. La prévision s'écrit :

$$E_t(X_{t+1}) = m + A X_t$$

étant donné que X_t est observé. Pour $h = 2$, on écrit le modèle en $t + 2$

$$X_{t+2} = m + A X_{t+1} + \epsilon_{t+2}, \quad (12)$$

et on prend l'espérance conditionnelle, ce qui donne cette fois-ci :

$$\begin{aligned} E_t(X_{t+2}) &= m + A E_t(X_{t+1}) \\ &= (I_n + A) m + A^2 X_t \end{aligned}$$

On en déduit par récurrence que :

$$E_t(X_{t+h}) = (I_n + A + \dots + A^{h-1})m + A^h X_t. \quad (13)$$

Il est intéressant de calculer la limite de cette espérance pour $h \rightarrow \infty$. Si le polynôme $A(L)$ est inversible, A^h converge vers zéro pour $h \rightarrow \infty$ et l'on a :

$$\lim_{h \rightarrow \infty} E_t(X_{t+h}) = C(1)m = \mu \quad (14)$$

avec $C(1) = \sum_{h=1}^{\infty} A^{h-1}$. Quand h tend vers l'infini, la prévision de X_{t+h} tend vers la moyenne de X_t . Ce résultat reste valide quel que soit le degré du polynôme $A(L)$.

2.6 Moments de l'erreur de prévision et représentation MA(∞)

L'erreur de prévision est définie par $X_{t+h} - E_t(X_{t+h})$. On va calculer ses moments en utilisant la représentation MA du processus:

$$X_t = \mu + \sum_{i=0}^{\infty} C_i \epsilon_{t-i}.$$

Écrivons maintenant le modèle théorique en $t + h$:

$$X_{t+h} = \mu + \sum_{i=0}^{\infty} C_i \epsilon_{t+h-i}. \quad (15)$$

Si maintenant on veut prédire X_{t+h} à partir de la connaissance que nous avons du processus jusqu'en t , les erreurs $\epsilon_{t+1} \cdots \epsilon_{t+h}$ seront inconnues, ce qui fait que le prédicteur linéaire optimal s'écrit:

$$E_t(X_{t+h}) = \mu + \sum_{i=h}^{\infty} C_i \epsilon_{t+h-i}. \quad (16)$$

Calculons maintenant l'erreur de de prévision :

$$X_{t+h} - E_t(X_{t+h}) = \sum_{i=0}^{\infty} C_i \epsilon_{t+h-i} - \sum_{i=h}^{\infty} C_i \epsilon_{t+h-i} = \sum_{i=0}^{h-1} C_i \epsilon_{t+h-i}.$$

Son espérance est nulle. Sa variance s'écrit:

$$\text{Var}(X_{t+h} - E_t(X_{t+h})) = \sum_{i=0}^{h-1} C_i \Sigma C_i' \quad (17)$$

Quand $h \rightarrow \infty$ la matrice de variance-covariance de l'erreur de prévision se rapproche de la matrice de variance covariance de X_t . La prévision optimale de long terme de X_t est simplement sa moyenne μ (en l'absence d'autre composante déterministe) et la variance de l'erreur de prévision, la variance de la série.

2.7 Inférence

On suppose que l'on dispose de T observations pour la variable X_t et que X_t suit un VAR(p). On va regrouper ces observations de manière à définir un modèle de régression multivariée de Y sur Z :

$$Y = X_{p+1}^T \quad Z = [1, X_p^{T-1}, \dots, X_1^{T-p}] \quad B = [\alpha_1, \dots, \alpha_n] \quad U = \epsilon_{p+1}^T$$

où les vecteurs α_i représentent les vecteurs de paramètres associés à chaque équation du VAR, y compris le terme constant. Alors le modèle VAR(p) peut s'écrire sous la forme

$$Y_{[T,n]} = Z_{[T,1+np]} B_{[1+np,n]} + U_{[T,n]}$$

où les dimensions sont indiquées sous les matrices et où l'on a remplacé $T - p$ par T pour simplifier les notations. Si le modèle ne comporte aucune restriction, il peut s'estimer par moindres carrés ordinaires avec :

$$\hat{B} = (Z'Z)^{-1}Z'Y.$$

On peut montrer que cet estimateur est équivalent à l'estimateur des moindres carrés appliqué sur chacune des équations séparément. Ceci vient du fait que les équations ont toutes la même structure et les mêmes variables explicatives. Il n'y a aucune restriction d'exclusion provenant par exemple du fait que l'on n'aurait pas mis le même nombre de retards dans chaque équation.

Un estimateur pour Σ est donné par

$$\hat{\Sigma} = \frac{1}{T} \sum U_t U_t' = \frac{1}{T} Y'(I_T - Z(Z'Z)^{-1}Z')Y.$$

On peut montrer la consistance et la normalité asymptotique de cet estimateur sous deux conditions. Il faut que

$$\text{plim} \frac{1}{T} Z'Z = Q \quad \text{avec } Q \text{ finie et non singulière}$$

Cette condition est en général vérifiée si le processus est stationnaire. Il faut ensuite que les termes d'erreur ϵ_t soient des bruits blancs. On aura alors

$$\text{plim} \hat{B} = B$$

et

$$\sqrt{T} \text{vec}(\hat{B} - B) \xrightarrow{d} N(0, \Sigma \otimes Q^{-1}).$$

On choisira comme estimateur de Q la matrice $Z'Z/T$ et comme estimateur de Σ $\hat{\Sigma}$.

Si l'on fait une hypothèse de normalité sur les termes d'erreurs, on peut écrire la fonction de vraisemblance du modèle. Sa maximisation donne exactement les mêmes résultats que l'estimation par moindres carrés dans le modèle standard, tant que l'on n'impose pas de restrictions particulières. On préférera donc en général utiliser ces derniers. Mais quand le modèle comporte des restrictions, la méthode du maximum de vraisemblance devient la seule méthode disponible. Pour plus de détails, on pourra consulter le chapitre 3 de Lütkepohl (1991).

2.8 Choix de modèles

Pour choisir la taille d'un modèle VAR, c'est à dire le nombre global de retards, on se servira usuellement d'un critère d'information. Les trois critères traditionnels (Akaike, Hannan et Quinn et Schwarz) s'écrivent, si $X_t \in R^n$ et p est le nombre de retards et $\hat{\Sigma}$

l'estimateur de la variance des résidus dans le cadre des moindres carrés:

$$\begin{aligned} AIC(p) &= \log |\hat{\Sigma}| + \frac{2}{T}n^2p, \\ HQ(p) &= \log |\hat{\Sigma}| + \frac{2 \log \log T}{T}n^2p, \\ SC(p) &= \log |\hat{\Sigma}| + \frac{\log T}{T}n^2p. \end{aligned}$$

On choisira alors le nombre de retards qui minimise l'un de ces critères d'information. On sait que le critère de Akaike est asymptotiquement biaisé, alors que les deux autres ne le sont pas. Mais en petit échantillon, le critère de Akaike a de meilleures propriétés.

On peut aussi utiliser un test de rapport de vraisemblance construit à partir de l'estimation de la variance des résidus. Si $\hat{\Sigma}(p)$ est la variance des résidus du modèle non contraint et $\hat{\Sigma}(p-1)$, celle du modèle contraint avec $p-1$ retards, alors le test du rapport de vraisemblance est:

$$T(\log |\hat{\Sigma}(p-1)| - \log |\hat{\Sigma}(p)|).$$

Ce test a une distribution du χ^2 avec un nombre de degrés de libertés égal au nombre de contraintes. Quand on élimine un retard, on élimine en fait une matrice de n^2 éléments. Le nombre de degrés de liberté de la χ^2 sera donc n^2 . Il est connu que ce test conduit à une paramétrisation moins parcimonieuse que celle obtenue avec les critères d'information.

Enfin, on peut décider également d'examiner les statistiques de Student associées à chacun des coefficients individuels du modèle et éliminer les paramètres dont la valeur du test de Student est inférieure à la valeur critique. Cette procédure conduit en général à avoir des équations comportant des nombres différents de retards. Il faut alors estimer le modèle équation par équation si l'on veut continuer à utiliser les moindres carrés. Mais l'estimateur ne sera pas efficace. La solution consiste à tenir compte de la variance estimée des résidus et donc à employer des moindres carrés généralisés ou utiliser l'estimateur du maximum de vraisemblance. On vérifiera l'adéquation du modèle final au moyen d'un test du rapport de vraisemblance qui cette fois-ci s'écrit

$$\lambda_{LR} = 2(\log L - \log L_c)$$

où L représente la vraisemblance du modèle initial et L_c la vraisemblance du modèle contraint. Ce test est distribué selon une loi du χ^2 ayant comme nombre de degrés de libertés le nombre de contraintes imposées.

2.9 Tests de mauvaise spécification : à compléter

Voir la généralisation au cas multivarié des tests usuels de mauvaise spécification. Sinon, on peut faire les tests équation par équation. Le dernier livre de Lutkepohl et al.

3 Analyse structurelle

Un modèle VAR indique comment le passé d'un ensemble de variables agit sur le présent de ces mêmes variables et comment des chocs sur une variable se transmettent au reste du système. Dans cette section, nous allons détailler ces deux aspects. Il s'agira tout d'abord de voir si un sous ensemble de variables a un effet différent des autres. Ce sera l'analyse de la causalité. L'autre question concerne l'analyse impulsionnelle.

3.1 Non-causalité au sens de Granger

Granger (1969) a introduit une notion de non-causalité qui repose sur les propriétés de prévision des modèles VAR. L'idée est que la cause ne peut pas venir après l'effet. Si une variable Z affecte une variable Y , Z sera utile pour améliorer la prévision de Y . Plus formellement on a :

Définition 2 Z ne cause pas Y au sens de Granger si quelque soit h positif :

$$\text{Var}(Y_{t+h} - E(Y_{t+h}|Y_1^t, Z_1^t)) = \text{Var}(Y_{t+h} - E(Y_{t+h}|Y_1^t))$$

Pour caractériser un peu mieux cette définition, considérons un processus multivarié X_t dans sa représentation $MA(\infty)$:

$$X_t = \mu + C(L)\epsilon_t \quad (18)$$

que l'on partitionne en :

$$X_t = \begin{bmatrix} Y_t \\ Z_t \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} C_{11}(L) & C_{12}(L) \\ C_{21}(L) & C_{22}(L) \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}. \quad (19)$$

Si l'on se sert de ce modèle pour prédire Y_{t+1} , on aura

$$E_t(Y_{t+1}) = \mu_1 + \sum_{i=1}^{\infty} C_{11,i} \epsilon_{1,t+1-i} + C_{12,i} \epsilon_{2,t+1-i}.$$

Par contre, si l'on n'utilise que le modèle marginal en Y_t , le prédicteur sera simplement

$$E_t(Y_{t+1}) = \mu_1 + \sum_{i=1}^{\infty} C_{11,i} \epsilon_{1,t+1-i}.$$

La prévision par les deux méthodes aura la même variance si et seulement si $C_{12}(L) = 0$. Alors Z ne causera pas Y au sens de Granger. La représentation MA est commode pour calculer la variance de l'erreur de prévision. Mais pour tester la non causalité, on aura besoin de la forme VAR estimable. On peut par inversion trouver la représentation AR de ce processus en posant :

$$\begin{bmatrix} C_{11}(L) & C_{12}(L) \\ C_{21}(L) & C_{22}(L) \end{bmatrix}^{-1} = \begin{bmatrix} A_{11}(L) & A_{12}(L) \\ A_{21}(L) & A_{22}(L) \end{bmatrix} \quad (20)$$

avec $A(L)(X_t - \mu) = \epsilon_t$. Alors, la condition $C_{12}(L) = 0$ se traduit immédiatement en $A_{12}(L) = 0$. C'est dans cette représentation AR que l'on pourra facilement tester la non-causalité au sens de Granger.

3.2 Tests de causalité

Considérons comme cas particulier un modèle VAR(p) de dimension 2 expliquant y_t et z_t . On suppose que l'on a déjà sélectionné le nombre optimal de retards. La première équation de ce modèle s'écrit

$$y_t = \mu_1 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \beta_1 z_{t-1} + \beta_2 z_{t-2} + \epsilon_{1t}$$

Pour tester la non-causalité au sens de Granger de z sur y , il suffit de tester la nullité jointe de β_1 et β_2 . Cela va se faire par un test en F dont l'expression générale est

$$S_1 = \frac{(RSS_0 - RSS)/p}{RSS/(T - 2p - 1)} \sim F(p, T - 2p - 1).$$

Toutefois, ce test n'est pas un test exact car il y a des endogènes retardées dans la régression OLS. On emploie donc un test asymptotique du $\chi^2(p)$ basé sur

$$S_2 = \frac{T(RSS_0 - RSS)}{RSS} \sim \chi^2(p).$$

3.3 Non causalité instantanée

La non causalité instantanée est un concept différent de la non causalité de Granger. Elle regarde si pour un instant donné du temps, c'est à dire en t , deux ou plusieurs variables évoluent de manière indépendante. C'est à dire si un choc sur une variable n'a pas de répercussion instantanée sur les autres variables. On devine sans qu'il soit nécessaire de faire de développements mathématiques que ce sera le cas si les innovations du processus sont indépendantes. Plus précisément, il n'y aura pas de causalité instantanée entre x_{1t} et x_{2t} si ϵ_{1t} et ϵ_{2t} sont non corrélés, c'est à dire si $E(\epsilon_{1t}\epsilon_{2t}) = 0$.

3.4 Analyse impulsionnelle

Nous arrivons maintenant au cœur de l'analyse des modèles VAR. Un modèle VAR modélise par essence les relations dynamiques entre un groupe de variables choisies pour caractériser un phénomène économique particulier. L'analyse impulsionnelle va permettre de déterminer l'influence d'un choc relié à l'évolution d'une des variables sur les autres variables du système. La non-causalité de Granger va jouer un rôle important si elle est introduite dans le système car elle spécifie justement que les chocs associés à une variable n'ont possiblement pas d'influence sur d'autres variables.

Considérons un modèle VAR d'ordre un sans constante

$$X_t = A X_{t-1} + \epsilon_t \tag{21}$$

à partir duquel on va analyser l'effet d'un choc en $t = 0$. On aura donc ϵ_0 qui portera un choc et puis la suite des ϵ_i sera nulle pour $i = 1, \dots, t$. Développons le modèle selon ces hypothèses. On a la suite $X_0 = \epsilon_0, X_1 = A X_0, X_2 = A X_1 = A^2 X_0$ et donc $X_t = A^t X_0$.

Si on suppose que le choc sur ϵ_0 est égal à 1 et porte sur la première variable, on définit $\epsilon'_0 = (1, 0, \dots, 0)$. Alors la première colonne de A^t représentera l'effet sur le système d'un choc sur la première variable après t périodes. Comme:

$$(I + AL + A^2L^2 + \dots) = (I - AL)^{-1} = C(L) \quad (22)$$

les coefficients de la forme $MA(\infty)$ vont donner la suite des réponses impulsionnelles du système à un choc unitaire sur les innovations du processus, car

$$\frac{\partial X_{t+h}}{\partial \epsilon'_t} = C_h.$$

étant donné que $X_t = C(L)\epsilon_t$. L'effet total sera donné par $C(1)$. Il est parfois utile de normaliser les colonnes des matrices C par l'écart type des innovations ϵ_t .

Cette première analyse avoue sa limitation dans la mesure où un choc ne peut logiquement affecter une innovation de manière isolée que si la matrice de variance covariance Σ est diagonale. Si ϵ_{1t} et ϵ_{2t} sont fortement corrélés, un choc sur ϵ_{1t} sera forcément accompagné par un choc sur ϵ_{2t} . Considérons maintenant la décomposition de Σ en:

$$\Sigma = PP' \quad (23)$$

Il s'agit de la décomposition de Choleski d'une matrice PDS où P est une matrice triangulaire supérieure avec ses éléments diagonaux positifs. On peut alors réécrire la forme $MA(\infty)$ en:

$$X_t = \mu + \sum_{i=0}^{\infty} C_i PP^{-1} \epsilon_{t-i} = \mu + \sum_{i=0}^{\infty} \Phi_i u_{t-i} \quad (24)$$

avec:

$$\Phi_i = C_i P \text{ et } u_t = P^{-1} \epsilon_t \quad (25)$$

Il est facile de vérifier que les u_t ont une matrice de variance covariance égale à la matrice identité. Les colonnes de Φ_i représenteront la réponse du système par rapport à un choc indépendant et normalisé sur l'innovation d'une variable après i périodes. Ce type d'analyse été popularisé par Sims (1980), Sims (1981).

3.5 Décomposition de la variance

Quand on a normalisé et orthogonalisé les chocs, l'analyse impulsionnelle se fait donc au moyen du modèle

$$X_t = \mu + \sum_{i=0}^{\infty} \Phi_i u_{t-i} \quad (26)$$

L'erreur de prévision à l'horizon h se note

$$X_{t+h} - E_t(X_{t+h}) = \sum_{i=0}^{h-1} \Phi_i u_{t+h-i}$$

On va décomposer cette erreur de prévision pour chaque composante de X_t que l'on va noter $x_{j,t}$. On a :

$$x_{j,t+h} - \mathbf{E}_t(x_{j,t+h}) = \sum_{i=0}^{h-1} (\phi_{j1,i}u_{1,t+h-i} + \phi_{j2,i}u_{2,t+h-i} + \dots + \phi_{jn,i}u_{n,t+h-i})$$

où $\phi_{jk,i}$ est l'élément (j, k) de la matrice Φ_i et $u_{1,t}$ l'élément 1 du vecteur u_t . On peut exprimer différemment la somme de gauche en intervertissant les deux signes sommes implicites

$$x_{j,t+h} - \mathbf{E}_t(x_{j,t+h}) = \sum_{k=1}^n (\phi_{jk,1}u_{k,t+h} + \dots + \phi_{jk,h-1}u_{k,t+1}).$$

Comme les u sont non-corrélés et de variance 1, il est facile de calculer la variance de l'erreur de prévision

$$\mathbf{E}(x_{j,t+h} - \mathbf{E}_t(x_{j,t+h}))^2 = \sum_{k=1}^n \phi_{jk,1}^2 + \dots + \phi_{jk,h-1}^2$$

On interprète alors la somme

$$\phi_{jk,1}^2 + \dots + \phi_{jk,h-1}^2 = \sum_{i=0}^{h-1} (e_j' \Phi_i e_k)^2$$

où e_i est la i^{ieme} colonne de la matrice identité I_n comme la contribution de l'innovation de la variable k à la variance de l'erreur de prévision à l'horizon h de la variable j . Pour pouvoir ces chiffres à des proportions, on peut normaliser l'expression trouver par

$$\sum_{k=1}^n \sum_{i=0}^{h-1} \phi_{jk,i}^2$$

On va donc pouvoir décomposer la variance de prévision pour chaque variable d'un VAR. Dans un VAR sur trois variables par exemple, la variance de l'erreur de prévision de la variable 1 pourra être due pour 30% à ses propres innovations, pour 20% aux innovations de la variable 2 et pour 50% aux innovations de la variable 3 pour une prévision à l'ordre 1. Ces chiffres vont changer quand on les calculera pour une prévision à l'ordre 2, 3 ou à long terme.

Il existe une relation non triviale entre la décomposition de la variance de l'erreur de prévision et la non causalité au sens de Granger. On s'attendrait à ce que si z ne cause pas y dans un système bivarié, alors la variance de l'erreur de prévision de y soit entièrement due aux innovations de y et aucunement à celles de z . Cette propriété n'est vérifiée que si la matrice de variance-covariance des innovations est diagonale, c'est à dire s'il n'y a pas de causalité instantanée entre les deux variables y et z . En effet comme $\Phi_i = C_i P$, les contraintes sous forme de zéros portant sur les C_i au titre de la non causalité disparaîtront dans la matrice Φ_i car la matrice P ne sera diagonale que si la matrice Σ l'est.

4 Modèles VAR structurels

Pour pouvoir effectuer une analyse impulsionnelle, on a orthogonalisé les chocs. La méthode a un certain côté arbitraire dans la mesure où la matrice P n'est pas indépendante de l'ordre dans lequel les variables sont indicées dans X_t . Aussi, au lieu de se reposer sur une procédure mathématique automatique, on va proposer une nouvelle approche qui consiste à introduire un minimum de raisonnement économique dans la construction d'un modèle VAR. C'est l'approche des VAR structurels que l'on va maintenant présenter.

4.1 Le modèle théorique

Watson (1994) introduit le modèle VAR structurel sous sa forme $MA(\infty)$

$$X_t = C_0\epsilon_t + C_1\epsilon_{t-1} + C_2\epsilon_{t-2} + \dots$$

Elle exprime l'évolution de X_t comme une réponse à une somme infinie de chocs structurels, les ϵ_t . Les chocs contemporains peuvent avoir des effets croisés sur X_t à cause de la présence de la matrice C_0 qui n'est pas contrainte à l'identité comme dans les modèles VAR que l'on a vu jusqu'ici. Si ce processus est inversible, on peut trouver sa représentation VAR que l'on va supposer d'ordre p . L'inversion de $C(L)$ avec $C_0 \neq I_n$ donne un polynôme $A(L)$ qui a un terme constant qui lui aussi n'est pas contraint à la matrice identité avec $A(L) = A_0 - A_1L - A_2L^2 - \dots - A_pL^p$. Le modèle sous sa forme VAR s'écrit donc

$$A_0X_t = A_1X_{t-1} + \dots + A_pX_{t-p} + \epsilon_t \quad \text{Var}(\epsilon_t) = \Sigma$$

Par rapport à un modèle VAR traditionnel, la matrice supplémentaire de paramètres A_0 vient multiplier le vecteur des observations contemporaines X_t et autorise donc des relations contemporaines entre les variables. On est de ce fait en présence d'un modèle à équations simultanées sans variables exogènes, mais seulement avec des variables prédéterminées. On peut en calculer la forme réduite en pré-multipliant tout par A_0^{-1} pour obtenir

$$X_t = B_1X_{t-1} + \dots + B_pX_{t-p} + u_t \quad \text{Var}(u_t) = \Omega = A_0^{-1}\Sigma(A_0^{-1})'$$

La distribution conditionnelle des X_t est entièrement caractérisée par cette forme réduite. Ce qui permet de voir que le modèle a un problème d'identification. La forme structurelle a n^2 paramètres supplémentaires par rapport à la forme réduite et ils viennent de la matrice A_0 . Il va donc falloir introduire n^2 restrictions. Il existe plusieurs façons d'introduire ces restrictions. Notons immédiatement que la littérature n'a pas considéré la possibilité de restrictions sur les matrices associées aux retards, car cela signifierait contraindre la dynamique, ce qui est à l'opposé du but recherché. On va donc considérer les trois possibilités suivantes:

- On peut contraindre la matrice A_0 qui lie les chocs contemporains, donc la partie court terme du modèle
- On peut contraindre la matrice Σ des erreurs structurelles

- On peut contraindre la matrice des multiplicateurs de long terme $A(1)^{-1}$.

Bien souvent il va falloir faire un mélange de ces trois possibilités pour arriver au nombre requis de restrictions. Toutefois, on évitera en général d'imposer plus de restrictions que nécessaires à la juste identification, procédant ainsi à l'inverse de la pratique des modèles à équations simultanées. Dans sa fameuse critique Sims (1980) juge que les restrictions de sur-identifications sont totalement irréalistes.

La pratique la plus courante, utilisée par exemple dans Sims (1980) consiste à contraindre la matrice Σ à être diagonale, à normaliser la diagonale de A_0 à 1 et à imposer une structure récursive sur A_0 (triangulaire inférieure ou supérieure). On se retrouve alors dans le cas similaire de l'orthogonalisation par Choleski, à ceci près que c'est maintenant l'économètre qui doit choisir l'ordre des variables. Il faut noter qu'il est équivalent de contraindre $\Sigma = I_n$ et de lever alors la normalisation de la diagonale de A_0 .

On peut imaginer d'autres types de restrictions sur A_0 . Il suffit que le nombre y soit, c'est à dire qu'il faut en plus des n restrictions de normalisation trouver $n(n-1)/2$.

Blanchard and Quah (1989) ont choisi de contraindre un aspect de la dynamique de long terme. Ils étudient la propagation des chocs de demande et d'offre. Ils veulent imposer que les chocs de demande n'ont pas d'influence à long terme et que seuls les chocs d'offre en ont une. Le modèle VAR structurel peut s'écrire en utilisant un polynôme de retard matriciel $A(L) = A_0 - A_1L - \dots - A_pL^p$. Comme $C(1) = A(1)^{-1}$, introduire une restriction sur $A(1)$ revient à introduire une restriction sur la somme des multiplicateurs. Si l'on pose par exemple que $A(1)$ est triangulaire inférieure, on obtient bien nos $n(n-1)/2$ restrictions nécessaires à l'identification. On a alors imposé une chaîne causale sur la façon dont les chocs se propagent dans le long terme.

Ce type de modèle n'est en général pas trop difficile à estimer s'ils sont juste identifiés. Il suffit alors de procéder par moindres carrés indirects. On va commencer par estimer les paramètres de la forme réduite Ω et les B_i . Il suffira ensuite de résoudre les deux systèmes

$$\begin{aligned} A_0^{-1}A_i &= \hat{B}_i \\ A_0\Sigma A_0' &= \hat{\Omega} \end{aligned}$$

par une méthode itérative. **Indiquer la méthode pour le cas général. Signaler JMulti de Luthkephol.**

4.2 Dynamique des chocs d'offre et de demande aux USA

Blanchard and Quah (1989) veulent étudier l'influence macroéconomique des chocs d'offre et de demande en utilisant pour cela un modèle VAR structurel. Sans pour cela procéder à des tests de racine unitaire, ils partent du principe que le PIB est caractérisé par une tendance stochastique où les chocs s'accumulent. L'analyse serait simple s'il n'y avait qu'une seule sorte de choc, mais il est probable qu'il y en ait plus. Afin de les identifier, il faut faire appel à au moins une variable supplémentaire. Ils supposent qu'il y a deux types de chocs, non corrélés et qui n'ont aucune influence de long terme sur le chômage. De plus

on suppose que le premier type de choc a une influence de long terme sur le PIB tandis que l'autre non. On voit donc immédiatement que la nature des chocs sera identifiée par les restrictions d'identification du modèle. On interprétera les chocs qui ont une influence durable comme des chocs d'offre et ceux qui n'ont qu'une influence transitoire comme des chocs de demande.

Le modèle se caractérise par deux variables Y et U , le logarithme du PIB en volume et le niveau du taux de chômage. Comme un VAR s'écrit sur des variables stationnaires, on va considérer

$$X_t = [\Delta Y_t, U] \quad \epsilon = [\epsilon_d, \epsilon_s]$$

Leur modèle est un VAR structurel écrit en forme MA(∞)

$$X_t = C_0 \epsilon_t + C_1 \epsilon_{t-1} + \dots$$

Ils posent que la matrice Σ de variance-covariance des erreurs structurelles ϵ est égale à l'identité de manière à rendre les deux chocs structurels indépendants. Comme on a supposé que X_t était stationnaire, les chocs n'ont aucune influence de long terme sur le taux de croissance du PIB ΔY_t et sur le niveau du taux de chômage U . La restriction $C_{11}(1) = 0$, impose également que le choc de demande ϵ_d n'a pas d'influence à long terme sur le niveau du PIB Y_t . Le fait que les chocs de demande puissent avoir une influence à court terme est dû à la présence de rigidités nominales. Cette influence disparaît dans le long terme.

Si l'on considère maintenant la forme MA(∞) réduite obtenue en prémultipliant par C_0^{-1}

$$X_t = u_t + \bar{C}_1 u_{t-1} + \dots \quad \text{Var}(u_t) = \Omega$$

on peut se rendre compte que C_0 est identifié par les restrictions introduites. Le fait que $\Sigma = I_2$ implique que $\Omega = C_0 C_0'$. Nous avons donc que les 3 éléments de Ω doivent déterminer les 4 éléments de C_0 . La 4^{ième} restriction est fournie par $[\bar{C}(1)C_0]_{11} = 0$.

Blanchard et Quah utilisent des données trimestrielles US couvrant la période 1950:2-1987:4. Ces données présentent une légère augmentation tendancielle du taux de chômage et une baisse dans le taux de croissance du PIB depuis 1974:1. Au lieu d'introduire explicitement ces caractéristiques dans le modèle au moyen d'un trend et d'une rupture dans le terme constant, les auteurs préfèrent traiter les données à l'avance et ne retenir donc que les résidus d'une régression du taux chômage sur un trend et une constante et soustraire leurs moyennes calculées sur deux sous périodes pour ΔY_t .

Les chocs positifs de demande ont un effet qui s'amortit au bout de 5 ans et qui est maximum au bout de 3 trimestres. L'effet positif sur Y est le miroir de l'effet négatif sur U . Les effets d'un choc d'offre par contre sont positifs sur Y et ne s'amortissent pas. Sur U l'effet commence à être positif, puis devient négatif et s'amortit en 5 ans.

Ceci permet de revisiter la courbe d'Okun. On voit sur le graphique des chocs de demande qu'il existe une relation constante entre la variation du niveau de l'emploi (taux de chômage) et la variation du PIB avec un coefficient de réaction de 2 au plus fort de

la dynamique. Par contre dans le cas d'un choc d'offre, il ne semble pas exister de telle relation.

Enfin, Blanchard and Quah (1989) décomposent la variance de erreur de prévision entre les deux types de choc. Ceci permet de constater l'importance des chocs de demande dans

Table 1: Décomposition de la variance de prévision

Pourcentage de la variance due à la demande		
Horizon	Production	Chômage
1	99.0	51.9
2	99.6	63.9
3	99.0	73.8
4	97.0	80.2
8	81.7	87.3
12	67.6	86.2
40	39.3	85.6

le court terme sur l'emploi et la production.

L'article de Blanchard and Quah (1989) est remarquable à plus d'un titre. Il accepte tout d'abord de se situer dans un cadre où le PIB présente une racine unitaire (même s'ils ne le teste pas). Dans ce cadre, les théorie classiques tentent de ramener l'explication des fluctuations à un cadre traditionnel où seules comptent les politiques de l'offre. Ils parviennent à montrer empiriquement que l'on peut retrouver une relation d'Okun réaliste, que les chocs de demande ont une influence transitoire sur Y et donc que les politiques de demande peuvent être utilisées à des fins conjoncturelles. Enfin, dans l'explication de la conjoncture l'influence des chocs de demande est primordiale par rapport aux chocs d'offre. Tout ceci a été obtenu dans un cadre économétrique finalement assez simple. La présence d'une racine unitaire joue un rôle important pour expliquer la persistance des chocs d'offre. Mais elle ne joue pas au niveau de la modélisation dans la mesure où il n'y a qu'une seule variable $I(1)$. Il ne peut donc y avoir de cointégration dans ce modèle.

Il faudrait détailler plus à partir du livre de Lutkepohl.

Autre papier Evans (1989) avec autre règle d'identification.

5 Conclusion

Ce chapitre s'est centré sur les problèmes de modélisation dans un cadre multivarié. Nous avons rappelé certaines notions sur les modèles VAR et en particulier les notions de non-causalité et d'exogénéité faible. En suivant Florens and Moucahrt (1985), il est possible

de construire un modèle à équations simultanée à partir d'un VAR par conditionnement et marginalisation. Le modèle conditionnel, de nature plus "économétrique", est analysable séparément pourvu qu'une hypothèse d'exogénéité soit satisfaite. Dans le cas des VAR simples, elle l'est toujours.

6 Lectures additionnelles

On consultera avec profit les chapitres 10 et 11 de Hamilton (1994) qui traitent spécifiquement des modèles VAR. Le chapitre 12 de cet ouvrage est également intéressant pour notre propos car il introduit l'analyse Bayésienne des modèles VAR avec les a priori de Litterman qui ont eu une grande influence pour l'analyse des modèles VAR. Les chapitres 2, 3 et 4 de Lütkepohl (1991) traitent spécifiquement des modèles VAR stationnaires ainsi que des réponses impulsionnelles. Mais il ne traite pas les modèles VAR structurels. Le chapitre 4 de Lütkepohl and Krätzig (2004) est lui entièrement dédié aux modèles VAR structurels. Mais un des exemples traités fait appel à la notion de cointégration, qui n'est pas abordé dans ce chapitre.

7 Exercices

7.1 Développement d'un VAR(1)

Considérons un VAR(1) que l'on fait partir en 0. Ecrire ensuite pour $t=1$, puis $t = 2$, etc et donner la formule générale. Que veut dire l'expression A^t et sous quelle condition cette expression est finie?

7.2 Equation caractéristique

Exercice sur les valeurs propres et l'équation caractéristique. Passage entre les deux. Racines = 1/valeur propres. Il doit y avoir un exemple dans Johansen (1995).

7.3 Prévision

La prévision à l'ordre h dans un modèle VAR(1) implique

$$E_t(X_{t+h}) = (I_n + A + \dots + A^{h-1})m + A^h X_t$$

Montrez que sous certaines conditions que l'on devra préciser la matrice A^h tend vers zéro quand h tend vers l'infini. (On utilisera pour cela la décomposition en valeurs propres vecteurs propres de la matrice A).

7.4 Examen 2007

Question 1 : Considérez le modèle VAR à l'ordre 1

$$A(L)(X_t - \mu) = \epsilon_t$$

Quelle est la moyenne de X_t ? Si l'on écrit maintenant

$$A(L)X_t = m + \epsilon_t$$

Quelle est la relation entre μ et m ?

Question 2 : Soit le modèle VAR structurel bivarié suivant avec $X_t \in R^2$ ($n = 2$ et $p = 1$);

$$A_0 X_t = A_1 X_{t-1} + \epsilon_t \quad \text{Var}(\epsilon_t) = \Sigma.$$

La variable X_t se décompose donc en $X_t' = [X_{1t}, X_{2t}]$. Les matrices A_0 , A_1 sont des matrices de taille 2×2 ainsi que la matrice Σ .

- Quelles hypothèses fait-on habituellement sur une matrice de variance-covariance?
- Combien de paramètres comporte cette forme structurelle?
- Quelles sont les hypothèses que l'on peut faire sur la matrice A_0 en terme de restrictions?
- Écrivez la forme réduite de ce modèle. Combien de paramètres comporte la forme réduite?
- Quelles sont les types de contraintes d'identification que l'on peut introduire dans la forme structurelle et combien en faut-il pour que le modèle soit juste identifié?

References

- BLANCHARD, O. J., AND D. QUAH (1989): "The dynamic effects of aggregate demand and supply disturbances," *The American Economic Review*, September, 655–673.
- EVANS, G. W. (1989): "Output and Unemployment Dynamics in the United States: 1950–1985," *Journal of Applied Econometrics*, 4, 213–237.
- FLORENS, J.-P., AND M. MOUCAHRT (1985): "Conditioning in Dynamic Models," *Journal of Time Series Analysis*, 6, 15–34.
- GRANGER, C. W. (1969): "Investigating Causal Relations by Econometric Models and Cross Spectral Methods," *Econometrica*, 37, 424–438.
- HAMILTON, J. D. (1994): *Time Series Analysis*. Princeton University Press, Princeton.
- LÜTKEPOHL, H. (1991): *Introduction to Multiple Time Series Analysis*. Springer Verlag, Berlin.
- LÜTKEPOHL, H., AND M. KRÄTZIG (eds.) (2004): *Applied Time Series Econometrics, Themes in Modern Econometrics*. Cambridge University Press, Cambridge.

SIMS, C. (1980): "Macroeconomics and Reality," *Econometrica*, 48, 1–48.

——— (1981): *An Autoregressive Index Model for the US 1948-1975* chap. Large Scale Macro-Econometric Models, pp. 283–327. North Holland, Amsterdam.

SIMS, C. (1986): "Are forecasting models usable for policy analysis," *Quarterly Review, Federal Reserve Bank of Minneapolis*, 10, 2–16.

WATSON, M. (1994): "Vector autoregression and cointegration," in *Handbook of Econometrics*, ed. by R. Engle, and D. McFadden, vol. IV, chap. 47, pp. 2843–2915. Elsevier, New York.